Check for updates

American Sociological Association

Creative Destruction: The Structural Consequences of Scientific Curation

American Sociological Review
1–36
© American Sociological Association 2021

DOI: 10.1177/0003122421996323
journals.sagepub.com/home/asr

Peter McMahan[a] and Daniel A. McFarland[b]

Abstract

Communication of scientific findings is fundamental to scholarly discourse. In this article, we show that academic review articles, a quintessential form of interpretive scholarly output, perform curatorial work that substantially transforms the research communities they aim to summarize. Using a corpus of millions of journal articles, we analyze the consequences of review articles for the publications they cite, focusing on citation and co-citation as indicators of scholarly attention. Our analysis shows that, on the one hand, papers cited by formal review articles generally experience a dramatic loss in future citations. Typically, the review gets cited instead of the specific articles mentioned in the review. On the other hand, reviews curate, synthesize, and simplify the literature concerning a research topic. Most reviews identify distinct clusters of work and highlight exemplary bridges that integrate the topic as a whole. These bridging works, in addition to the review, become a shorthand characterization of the topic going forward and receive disproportionate attention. In this manner, formal reviews perform creative destruction so as to render increasingly expansive and redundant bodies of knowledge distinct and comprehensible.

Keywords

reviews, disciplines, scientific communication, networks, citations

The communication of scientific findings is fundamental to scholarly discourse. Isolated findings are understood only when they are viewed in relation to other scholarly output. Any particular claim is without substance unless it is interpreted contextually. What theories does it support? Which research agendas does it contradict? How does it fit into the overarching structures of scientific knowledge? Scholarly discourse relies on the relational interpretation of research findings to codify claims. Such interpretive work is present in virtually every piece of scientific writing, but formal academic review articles represent a quintessential means through which scientific findings are brought into context with one another and sense is made of collected research. Reviews aim to gather the relevant published findings in a domain of inquiry and to synthesize those findings into a coherent body. If we want to know how scientific discourse progresses, how ideas move from tentative propositions to accepted knowledge, formal reviews offer us a window into the mechanisms involved.

[a]McGill University
[b]Stanford University

Corresponding Author:
Peter McMahan, McGill University, Room 717, Leacock Building, 855 Sherbrooke Street West, Montreal, Quebec H3A 2T7
Email: peter.mcmahan@mcgill.ca

Much of the work in social studies of science that investigates relational synthesis focuses on novel juxtapositions of scientific ideas (Fleming 2001; Foster, Rzhetsky, and Evans 2015; Latour 1987; Leahey and Moody 2014; Uzzi et al. 2013). This research consistently finds that scientific projects that posit unexpected relationships between domains—as long as they have positive findings—receive greater attention and are more richly rewarded than projects that explore more commonplace connections. The implication is that scientists who are able to bring disparate domains into conversation are more likely to generate significant innovation by defining a newly relevant area of research. Scientific research that synthesizes divergent knowledge moves away from Kuhn's (1970) *normal science*, introducing unconventional ideas and potentially defining new domains of research.

But most of the curatorial work that takes place in scientific discourse is not oriented toward finding novel or surprising combinations. The bulk of synthetic work takes the opposite approach: it is concerned with assessing and interrogating the immediate relationships among bodies of research. Such acts of curation have long been recognized as playing an important role in scientific discourse, addressing, among other things, the need for scientists to handle an ever-increasing deluge of new research findings. Scholars faced with the reality that they will be unable to consume everything published in their own specialization, much less in subfields outside of their expertise, will turn to curated expositions of relevant research communities to help contextualize and expand their own work.

This phenomenon is not new. Price (1986) describes the emergence of the scientific paper (as opposed to published book) in the seventeenth century as a remedy for the perceived overabundance of research available for scientists to absorb. Such papers, published in periodical format that today is the norm for scientific journals, "had the stated function of digesting the books and doings of the learned all over Europe. Through them the casual reader might inform himself without the network of personal correspondence, private rumor, and browsing in Europe's bookstores, formerly essential" (Price 1986:57). Scientific output outpaces researchers' ability to consume and make sense of it. Synthetic work addresses this problem not only by sorting through a relevant literature and summarizing its findings, but, as we will argue, by contextualizing those findings among themselves and within a broader scholarly field.

Scholarly review articles are perhaps the most explicit form of synthetic production in contemporary scientific publication.[1] The curatorial work they do is not incidental: it is the expressed purpose of their existence. Review articles target a range of audiences, attempting to be valuable not only to communities actively engaged with their topic but also those that may be wholly unfamiliar with the relevant scientific domain. They aim to "inform interested readers who have limited knowledge of [a] topic, whether students new to the field or seasoned researchers from other domains" (Freeman and Jeanloz 2015). Beyond simply identifying a relevant body of literature, authors of review articles are expected to tell their readers how, exactly, the works they cite relate to one another. The authoritative account of research findings provided by formal reviews is intended to help insiders and outsiders alike make sense of those findings.

A good review article tells a *story* about what ideas are important and why, and synthesis of this type, although necessary, need not be neutral. In presenting an "official" account of a complex set of research to readers from diverse disciplinary backgrounds,[2] reviews must translate a specialized discourse into a more accessible description, and that translation, like any, is apt to add an interpretive aspect to the exposition. A review "selects from [the relevant] papers, juxtaposes them, and puts them in a narrative that holds them together, a narrative with actors and events but still without an ending. It draws the reader into the writer's view of what has happened, and by ordering the recent past, suggests

what can be done next" (Myers 1991:46). By synthesizing the findings of an emerging research area into a coherent narrative, reviews can affect the future direction of research in that area. The unique discursive orientation of reviews, focused on clarity of synthesis for newcomers from a position deeply embedded within existing research, suggests a generative role in the production of scientific output.

The question of what review articles "do" thus becomes: how do review articles affect the knowledge they digest and the research areas they interrogate? In this article, we address this question by studying, at large scale, the histories of publications cited by review articles, examining the attention papers receive both before and after they are included in a review's synthesis. We argue that reviews are much more than mere summaries of relevant findings: they induce novel structure on a research domain and define relevant actors, alliances, divisions, and omissions within the literature. The structure they impose comes to represent the domain and becomes a scaffolding around which future discourse is formed.

We will show that academic review articles perform curatorial work that substantially transforms the emerging research areas they aim to summarize. Using an exhaustive corpus of millions of journal articles, we analyze the consequences of review articles for the publications they cite, focusing on citation and co-citation as indicators of scholarly attention. Our analysis shows that being included in a review article is, on average, detrimental for scientific publications and leads to a dramatic overall decrease in future citations. However, reviews bring increased attention to a research area as a coherent whole, and they impose a novel structure on that area's discourse moving forward.

By analyzing the co-citation networks of both reviewed and unreviewed areas of knowledge, we show that reviews dramatically simplify a specialized domain of knowledge, focusing future scholarly attention onto a few key publications and the relations between them at the expense of the broad majority of the research in a domain. Reviews help establish and relate a set of exemplars in a specialized domain, structuring subsequent conversations within and outside of the existing research area. In short, reviews perform a type of creative destruction:[3] in identifying a coherent subdomain centered on a set of exemplars, they diminish the effect of the non-exemplars going forward. The articulation of a topic as a legitimate scientific research area comes at the expense of a pared-down conception of what that research area entails. We discuss the implications of these findings, arguing that review articles' work of synthesis suppresses many of the core discussions and conflicts in a specialized subdomain of knowledge, and that in doing so they help to constitute that subdomain as a coherent scholarly field.

THEORETICAL FRAME

The assumed goal of a review article is to create a bite-sized synthesis of work from a distinct research area. A glance at the table of contents for a recent volume of the *Annual Review of Sociology* is illustrative: "Wealth Inequality and Accumulation" (Killewald, Pfeffer, and Schachner 2017); "The Development of Transgender Studies in Sociology" (Schilt and Lagos 2017); "The Second Demographic Transition Theory: A Review and Appraisal" (Zaidi and Morgan 2017). Each of these review articles is a window into a specific, relatively well-defined subdomain of research.

The separation of scientific research into diverse research areas is well established in the sociology of science and knowledge. These implicit areas, which researchers have variously termed *invisible colleges* (Price 1986), *epistemic communities* (Holzner 1972), and *scientific collectivities* (Woolgar 1976), among many other terms, arise from the need for some level of insulation from the torrent of scientific output. Specialized domains emerge as scholars form relationships, reputations, and regularized modes of communication among their peers (Chubin

1976; Collins 2009; Kuhn 1970; Lievrouw 1992). Tight-knit research areas can act as a hotbed for knowledge production and scientific development, but they can also promote research segregation, becoming inaccessible to the uninitiated (Collins 1998; cf. Callon et al. 1983). Many areas of research would remain invisible to the broader community without the aid of expert summary.

Consequentially, the common (and often implicit) theoretical view of review papers sees them simply as tools to identify the most important results emerging from a specialized domain. Indeed, in the rare cases where review articles are discussed in scholarly literature, they are often treated as mere synopses: "Review papers . . . can be considered simply as summary reports of research results in the specialty" (Morris and Van der Veer Martens 2008:260). In this view, authors of review papers sort through the scientific output in an area, discerning the significant findings and omitting the false starts and intermediary research that would be of less use to an outsider. Review articles, the argument goes, are little more than a "packing down" (Price 1986:178) of the contributions of a research area—they act as an undiscriminating spotlight, illuminating relevant work without interpretation.

But scholars concerned with the dissemination and use of scientific knowledge argue that the act of transporting ideas from one specialized domain to another, as review articles do, is not neutral. Synthesizing knowledge makes a body of work accessible to those not familiar with it and entails active translation. Translators of specialized knowledge do not merely expose that knowledge, they respond to their audiences' perceived expectations, transforming and presenting a body of work in a way they hope will be palatable (Latour 1987; Star and Griesemer 1989).

A considerable literature examines the types and prevalence of scientific translation, focusing in part on the phenomenon of knowledge encapsulation often referred to as *black-boxing* (Whitley 1970). For a collection of research to be useful outside of its original, narrow domain—the argument goes—it must be perceived as an unproblematic "black box" so that outside researchers can use the specialized knowledge without concern about the details of the provenance of that knowledge. For science to be truly cumulative, its outcomes need to be encapsulated as uncontentious knowledge (Fuchs and Spear 1999), building an increasing corpus of consensus-based facts (Shwed and Bearman 2010). Black-boxed knowledge may be utilized by scientists without consideration of the roadblocks, conflicts, or methodological innovations (the inner workings of the box) that went into its creation (Latour 1987).

Black-boxing presents a plausible hypothesis for the impact of review articles on knowledge translation and consumption. As a domain of specialized research resolves internal disputes and reaches consensus on its core questions, a review article may distill the findings, crafting portable language for those findings to be utilized by a wider community (see, e.g., Oreskes's [2004] review of anthropogenic climate change). In terms of the attention that research in the specialized domain receives, a review article may "poach" citations from the work it references. Faced with a limited budget for citations, scholars interested in utilizing the findings from one corner of a research area could instead cite the review itself as a black-boxed representation of the area as a whole.

This idea of black-boxing, however, may oversimplify the process of scientific discovery and dissemination. Although there are clear instances of scientific consensus that have reached the point of undissected fact (see Latour 1987; Shwed and Bearman 2010), a purely cumulative vision of science in which questions are posed, hypotheses tested, and answers eventually added to the body of scientific fact is incomplete. As many scholars argue, scientific research subsists as much on conflict and dialogical contention as it does on agreement and consensus-seeking (Abbott 2001; Bourdieu 1975). A review article that treats a research specialty as a unified whole—as a black box—may be less useful

to scholars trying to engage with and understand a topic area than one that packages that specialty as a conversation with diverse viewpoints and internal divisions. Black-box representations are appropriate for translation of science for lawmakers and the public (Callon et al. 1983; Star and Griesemer 1989), but research scientists, who constitute the primary audience for the academic review article, may seek the discursive structure of a field as much as they do its established findings.

For scholars and students trying to understand a new research domain, there is value in knowing something of the internal workings, at least in schematic form, of a black-box machine. The relevance of a specialized domain's work can be understood through the processes by which particular researchers and their findings came to dominate the discourse within that domain (Bourdieu 1990; Kim 2009). Kuhn's (1970) conception of a scientific paradigm emphasizes the role of *exemplars* for structuring disciplines, but many scholars argue that the same structuring occurs at a smaller scale, constituting the less extreme evolution of scholarly knowledge that builds on previous work without subverting its paradigmatic core (Frickel and Gross 2005; Hedgecoe 2002; Price 1986). In this view, reviews should not be understood as external or peripheral to a field of study, but as active participants in that field. Scientific discourse is distinctly reflexive: exemplary publications can act as signposts for a specialized body of knowledge, both describing and defining the connections among a set of researchers. As scholarly subdomains develop, the shape of their discourse is based on a shared understanding of this relational structure and its history—the canonical citations, theoretical divisions, agreed-upon terminology, and myriad other features that allow researchers to feel they are on the same page (Myers 1985). In making a particular claim about the relations that define a research area, review articles are engaging in a specific *intervention* in the evolution of that area (Ketcham and Crawford 2007; Myers 1991; Sinding 1996).

This suggests a third possibility for the structuring role of review articles in the digestion of an emerging research area. Reviews may define an opinionated representation of a subdomain, singling out a set of key exemplars and the relationships between them to tell a specific story of the past, present, and future of their area. Like a black box, this type of review translates a research specialty by omitting details of its history, but by focusing on a representative set of exemplary objects, such a review would be selective in its omissions. Like a subway map, it would present its audience with a caricatured version of its subject, highlighting important landmarks and, by tracing the relationships between them, describing in broad strokes the shape of the area they inhabit.

ANALYTIC STRATEGY

Through rigorous empirical analysis of a large corpus of scholarly output, we will show that formal review articles have significant and consequential influence on scientific discourse at multiple levels. The transformations we describe occur at several distinct analytic scales, which we address with a series of statistical models. As a first description, we will examine how reviews affect the trajectories of the individual articles they cite. A naive perspective would regard reviews as little more than digests of a subdomain, simply bringing the most relevant research to the attention of a larger audience. In contrast, we will show that reviews heavily curate a subdomain, drawing attention *away* from the majority of the articles they cite. The bulk of the existing work in a research area experiences loss of attention as a result of its inclusion in a formal review.

Having established the strong curatorial influence of reviews, the subsequent analyses focus on their structural consequences for the specialized domains they address. Do reviews universally suppress attention for the publications they cover, or do they selectively promote a subset of that work? To address these structural questions, we examine the

relational networks of individual research areas before and after they become the subject of review. We show that reviews do draw an area into tighter relation with itself, but they also erase much of the structure that defined it. Reviews shift the focus of a research area from a collection of small specialized clusters onto a relative few research exemplars. After review, a handful of previously noncentral articles become the hubs around which a subdomain revolves, and the subdomain's focus shifts to relations with these hubs at the expense of relations between nonhubs. We further demonstrate that such hubs are unlikely to be central to any specific specialized cluster in a research area but tend to hold bridging positions between those clusters. These analyses show that review articles dramatically restructure the patterns of attention that specialized domains receive by constructing a simplified narrative of the major discourses in those domains.

A key step in studying the effects of review articles is their identification among a sea of published works. The category of "review article" is far from clear-cut.[4] It is often not obvious whether a particular publication is intended as a review, and a great many articles do at least some review work in citing previous literature. However, the analysis presented here is not concerned with literature review as a characteristic of research publications in general, but with formal review as a specific form of academic publication. Articles that perform the "summing-up" discussed in the previous section have a particular authoritativeness among published articles. Approaches using citation patterns or keyword matching can identify reviews as a *style* of publication (see, e.g., Ketcham and Crawford 2007), but it is the *role* of formal reviews in the ecosystem of journals that is most relevant to our analysis.

In light of this, we consider review articles to be anything published in an Annual Reviews (AR) journal.[5] Annual Review articles do not follow the standard peer-review process of most academic journal publications—they are written by authors considered experts in a field who are invited by the journal's editorial board to submit a review. These publications are overtly situated as authoritative sources of the state of the fields they review. Although there are many overt and authoritative review articles that do not appear in AR journals (including discipline-specific sources), the deliberately conservative approach we take minimizes the number of falsely identified review articles at the expense of an increased likelihood that a "genuine" review will not be categorized as such. Erring on the side of "false negatives" implies that the effects shown in our analysis will be biased toward the null; the chances that we conclude reviews have no effect on knowledge production when in fact they do is significantly higher than the converse.

In line with previous research (Price 1986; Small 1973; Small and Sweeney 1985), we focus on scholarly attention as measured by citations. Our analyses are based on data from Clarivate Analytics *Web of Science* (WoS).[6] The WoS has extensive coverage of journal publications across a wide array of scientific disciplines. Although it includes information on over 100 years of publications, the coverage is far from complete for most of that time. To ensure the data we use are representative, we limit our analysis to work published from 1990 through 2016. Relatively few scientific publications are ever included in a review article, especially articles published in less widely read journals. We therefore restrict our analyses to a subset, albeit a large subset, of the full WoS corpus. We generated our sample by calculating the 50 academic journals most-cited by each of the 52 Annual Reviews we consider, and collecting every article published in those journals. Review citations are heavily skewed toward top journals, so this sampling covers the large majority of articles likely to be reviewed—retaining 80 percent of all cited articles in reviews.[7] The final sample yields approximately 5.9 million articles published in 1,155 journals across the 27 years of the sample. Table 1 displays summary statistics for the sample, broken down by subject area.

Table 1. Summary Statistics for the Full Sample, by Journal Subject Area and Total

	Journals	Articles	Articles reviewed > 0 times	Articles reviewed > 1 time	Mean citations/ year (median, IQR)	Citations referenced (median, IQR)	Prop. citations to WoS (median, IQR)
Nat. Sci.	636	3,938,866	365,226	105,538	.85 (.33, 1.91)	29 (18, 44)	.56 (.29, .76)
Eng. & Tech.	173	1,005,023	34,489	5,552	.58 (.25, 1.32)	23 (13, 34)	.5 (.22, .72)
Health Sci.	264	1,727,804	133,197	30,751	1 (.33, 2.33)	30 (15, 44)	.62 (.31, .82)
Agr. Sci.	64	258,336	8,741	972	.62 (.29, 1.29)	27 (18, 38)	.44 (.19, .66)
Soc. Sci.	277	310,736	33,061	8,054	.62 (.25, 1.50)	34 (19, 53)	.27 (.10, .49)
Humanities	39	45,759	2,437	375	.44 (.17, 1.00)	31 (14, 52)	.27 (.07, .62)
Total	1,155	5,901,565	509,325	140,044	.85 (.33, 1.96)	29 (17, 44)	.55 (.26, .77)

Note: Many journals are listed in multiple subjects, so total counts will be less than the sum of the subject counts. In addition to the number of journals and number of articles, the number of articles cited at least once and at least twice by a review article are shown. Median and inter-quartile range across articles are shown for the mean annual number of citations received, the number of works cited, and the proportion of those citations that reference an article in the WoS database.

PROMOTE OR POACH?

The discussion of theoretical perspectives on the translation of scholarly knowledge presented above suggests several distinct ways reviews might affect the research specialties they cover. As an important first pass at disentangling these ideas, we ask a relatively simple question: what are the implications of inclusion in a review article for future citations? Can an article that is discussed in a review expect to enjoy newfound awareness in a wider scholarly circle, or will it fall into relative obscurity as attention is diverted toward a smaller set of literature? More succinctly, *do review articles promote or poach citations*?

The question is simple, but it is vital to untangling the different theoretical processes discussed above (and the mechanisms they imply). If reviews simply shine uncritical spotlights on a literature, bringing wider visibility to research that was previously obscure, then we should expect the publications highlighted to enjoy increased attention in the form of citations. In contrast, if reviews are taking a more active role in the interpretation or creation of scientific knowledge, as suggested by the theories of black-boxing and exemplars, then their effect on the articles they cite will be more complex. If reviews indeed construct unproblematic black boxes for scientific topics, then future work will be more likely to ignore the individual publications cited, either citing the review as a placeholder for the entire area or not bothering to cite the ideas at all (Garfield 1977). Similarly, if reviews identify a small subset of exemplars as the core of a field, successfully reconfiguring discourse on the subject along specific, simplified lines, the average effect on future citations received by the subjects of review should also be negative or ambiguous.

Answering this question is not as straightforward as it might appear. Myriad factors confound and complicate the effect under consideration. The number of citations an article receives is highly time sensitive, and much of its year-to-year variation is linked to

waxing and waning attention (Redner 2004). This suggests the identification of a "review effect" on future citations needs to take into account the natural cycle of attention for any given publication, seeking out discontinuities in the pattern of citations. Furthermore, being cited by a review article can hardly be considered an independent event in the lifecycle of a publication, and care must be taken to account for confounding factors that may be related to an article's citation by reviews and non-reviews alike.

We account for these complications using a hierarchical negative-binomial model of citations over time for a large set of articles in the dataset. The model predicts the number of citations an article will receive in a given year, based on covariates specific to the year, the article, and the journal of publication. The model, specified in Equation 1, has three nested levels: years, articles, and journals. Each journal contains many articles, and each article exists for multiple years. The dependent variable, denoted $\mathbf{C}_{ijt}$, is the number of non-review citations received by article i, published in journal j, t years after its publication. Because citation by a review article is our primary explanatory variable, we exclude those citations from the measurement of scholarly attention in the outcome. We represent the cumulative number of review citations an article has received $t + 1$ years[8] after publication (on a log scale) with $\mathbf{Reviews}_{ijt}$.[9] Using a cumulative measure allows the model to account for lasting effects of review and for the effects of multiple reviews (see Table 1 for incidence of multiple review).

The remaining covariates in the model are included primarily as controls for the main effect of interest, although we will show they indicate some interesting patterns on their own. At the coarsest level, the model has variables associated with each article's publication journal, indexed with subscript j. We model each journal's status using its impact factor,[10] represented in Equation 1 with $\mathbf{IF}_j$. We also account for different disciplinary publication cultures by including dummy variables for six subject areas: natural sciences ($\mathbf{Sub1}_j$), engineering and technology ($\mathbf{Sub2}_j$), medical and health sciences ($\mathbf{Sub3}_j$), agricultural sciences ($\mathbf{Sub4}_j$), social sciences ($\mathbf{Sub5}_j$), and humanities ($\mathbf{Sub6}_j$). Subject areas are assigned to each journal according to its WoS subject tags, mapped to OECD major subject categories (OECD 2015). These subjects are nonexclusive, allowing many of the journals to be included in multiple categories.

The next level of the model, indexed with subscript i, includes variables specific to each article that help explain the number of citations it receives. Because citation practices vary considerably even within subject areas, these covariates capture details of the publication style of each article. $\mathbf{NumCites}_i$ indicates the number of items each article cites, allowing the model to account for specialties in which citations tend to be numerous or sparse. Similarly, $\mathbf{Pages}_i$ measures the number of pages the article uses. Number of pages can vary according to journal format and style, but it is a good indicator of overall article length. Together with the other covariates, article length can be associated with both the number of citations made and the number of citations received. We also measure the proportion of an article's citations that reference items in the Web of Science database ($\mathbf{WoS}_i$). This variable acts as a rough measure of the degree to which a publication is engaged with journal publications (well represented in WoS) as opposed to books or conference proceedings (not included in WoS).

The model includes several covariates at the lowest level, representing relevant features of an article that vary over time. In addition to the outcome variable and the number of reviews (described above), temporal patterns are accounted for by a quadratic function on the number of years since publication ($\mathbf{t}$ and $\mathbf{t}^2$). In the negative-binomial framework with random effects, this form allows considerable variation in underlying patterns of waxing and waning attention.

Finally, to be sure we are capturing the effect of being included in a published review rather than some feature of the types of articles that are *likely* to be reviewed, we include

a covariate intended to proxy the cumulative *reviewability* (**Reviewability**$_{ijt}$) of an article. This control is calculated from the frequency of co-citation with reviewed texts[11] and can be thought of as the aggregate number of times an article could have been reviewed but was not.[12] To interpret the main temporal intervention in the model (being included in a review article) as a causal factor, it is important to control for the possibility that a discontinuity in the outcome is associated with review articles specifically, and not simply a characteristic of the types of articles that tend to be cited by reviews. The variables for time, journal impact factor, average pages, and reviewability are standardized to have zero mean and unit standard deviation.

$$
\begin{aligned}
&\mathbf{C}_{ijt} \sim \text{NBinom}\left(\lambda_{ijt}, d\right) \\
&\log\left(\lambda_{ijt}\right) = \beta_{0ij} + \beta_{1i}\mathbf{t} + \beta_{2i}\mathbf{t}^2 + \beta_{3ij}\,\mathbf{Reviews}_{ij(t\text{-}1)} \\
&\qquad + \beta_{4j}\,\mathbf{Reviews}_{ij(t\text{-}1)} \\
&\qquad \beta_{0ij} = \gamma_{00\,j} + \gamma_{01}\mathbf{NumCites}_i + \gamma_{02}\mathbf{WoS}_i \\
&\qquad + \gamma_{02}\mathbf{Pages}_i + \eta_{0i} \\
&\qquad \beta_{1i} = \gamma_{10} + \eta_{1i} \\
&\qquad \beta_{2i} = \gamma_{20} + \eta_{2i} \\
&\qquad \beta_{3ij} = \gamma_{30\,j} + \gamma_{31}\mathbf{NumCites}_i + \gamma_{32}\mathbf{WoS}_i \\
&\qquad + \gamma_{32}\mathbf{Pages}_i + \eta_{3i} \\
&\qquad \beta_{4\,j} = \gamma_{40\,j} \\
&\qquad \gamma_{00\,j} = \alpha_{00} + \nu_{0\,j} \\
&\qquad \gamma_{30\,j} = \alpha_{30} + \alpha_{31}\mathbf{IF}_j + \alpha_{32}\mathbf{Sub1}_j \\
&\qquad + \cdots + \alpha_{37}\mathbf{Sub6}_j \\
&\qquad \gamma_{40\,j} = \alpha_{40} + \nu_{4j}
\end{aligned}
\tag{1}
$$

The model described in Equation 1 allows considerable structured flexibility through the random-effects vectors η_i and ν_j. β_{0ij}, β_{1i}, and β_{2i} describe the average evolution of citations over time, but each article's underlying citation curve is accounted for through random variations in the values of η_{0i}, η_{1i}, and η_{2i}. The main explanatory coefficient of interest is β_{3ij}, which measures the impact of citation by a formal review article on future citations. We are especially concerned with the ways this effect may be heterogeneous, so we take particular care to account for variation in this parameter between articles. In addition to allowing article- and journal-level characteristics to moderate this effect, the random-effect term η_{3i} models variation in β_{3ij} that is unexplained by covariates.[13]

Finally, recognizing that the number of citations an article receives may depend strongly on the venue of publication for a given article, each journal similarly has a random vector ν_j to allow otherwise unmodeled variation. Estimation of such a model poses significant computational difficulties, due both to the overall size of the sample (88,590,720 observations across 5,901,566 articles), and the need to estimate random effects for each article in the sample.[14] We therefore restrict the estimation to five-percent subsamples of the articles, weighted to over-sample less prolific subjects and years. To avoid bias related to the heavy right skew of the time variable, we consider only articles with at least 10 years of citation information. The resulting sample has 2,931,604 year-level observations nested within 144,097 articles in 1,069 journals. Repeated estimation on independent subsamples yields virtually identical results, indicating the analysis is robust to the choice of subsample.

Table 2 presents the model estimates. The large negative estimates for the intercept and time-related coefficients tell a dismal but predictable story about the evolution of citations over time (see Figure 1): most publications are rarely cited at all, and their chances of receiving citations tend to peak at about the fifth year after publication. The quite large standard deviations (10.03, 22.70, 5.34) for these article-level random effects, however, indicate there is considerable variability among articles' citation curves, as indicated by the thin lines in Figure 1. More important for the current analysis are the estimates affecting β_{3ij}, which measures the effect of inclusion in review articles on future citations for each article. The strong negative estimate for α_{30} (labeled "Reviews" in Table 2) suggests the first time a given paper is cited by a review, that paper can expect approximately 11 percent fewer citations in *every* subsequent year.[15] Additional citations from other Annual Review articles further

Table 2. Multilevel Negative-Binomial Model Estimates of Model 1

Population-Level Effects	Estimate	Std Error	(*Z* value)
(Intercept)	–1.48926	.05147	(–28.94)
t	–3.47962	.01326	(–262.48)
t^2	–2.18744	.00645	(–339.02)
Reviews	–.11920	.03639	(–3.28)
Reviewability	.28620	.01189	(24.07)
Impact Factor	.28647	.01747	(16.4)
Pages	–.08607	.0125	(–6.89)
Citations	.38025	.00394	(96.53)
Prop. WoS	1.22897	.01415	(86.85)
Subj: Nat. Sci.	.05144	.04757	(1.08)
Subj: Eng. & Tech	–.02725	.05142	(–.53)
Subj: Med. & Health	.18860	.05101	(3.7)
Subj: Agr. Sci.	–.19620	.07352	(–2.67)
Subj: Soc. Sci.	–.10433	.05527	(–1.89)
Subj: Humanities	–.60872	.10442	(–5.83)
Reviews × Impact Factor	.02643	.00633	(4.18)
Reviews × Pages	–.01779	.00945	(–1.88)
Reviews × Citations	.04250	.00726	(5.86)
Reviews × Prop. WoS	–.52196	.04075	(–12.81)
Reviews × Subj: Nat. Sci.	–.00867	.02895	(–.3)
Reviews × Subj: Eng. & Tech.	–.06128	.03219	(–1.9)
Reviews × Subj: Med. & Health	.11775	.02759	(4.27)
Reviews × Subj: Arg. Sci.	–.10006	.05467	(–1.83)
Reviews × Subj: Soc. Sci.	.10427	.0354	(2.95)
Reviews × Subj: Humanities	–.05329	.08834	(–.6)
Group-Level Effects		**Level**	**Std Dev**
(Intercept)		Article	10.0341
t		Article	22.7016
t^2		Article	5.3351
Reviews		Article	2.1927
(Intercept)		Journal	.2743
Reviewability		Journal	.1265

Note: Coefficients reflect expected effects on total yearly citations received. All covariates except *reviews* standardized to have zero mean and unit standard deviation.

reduce expected future citations—an article reviewed for the third time will receive about 30 percent fewer citations over the remainder of its lifetime.

This effect is highly heterogeneous across articles, due to predicted differences in article characteristics (e.g., article length, journal of publication) and the large standard deviation in the random-effect term η_{3i} (2.20). Figure 2 summarizes this variation, showing the distribution of the net multiplicative "review effect" across all articles that received at least one review citation. The large majority of publications that are included in Annual Review articles can expect a substantial dampening effect on their future citations, with a median value of $\exp(\beta_{3ij}) = .617$—an expected decrease of nearly 40 percent for each review citation received. Figure 2 also shows that the distribution has a long right tail. This indicates that a substantial minority of articles (around 30 percent) are predicted to have at least a small increase in their future citations; of these, only around 12 percent

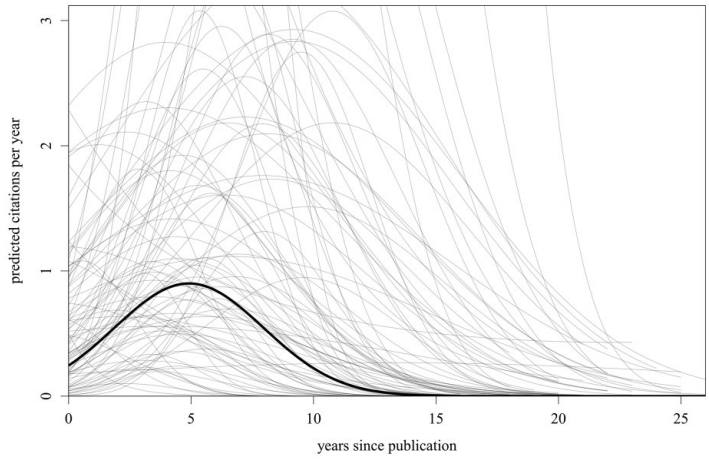

Figure 1. Predicted Citations per Year (setting all other covariates to zero)
Note: The thick line represents the average (fixed) effect estimate; thin lines represent a random subsample of articles and demonstrate the flexibility of the model.

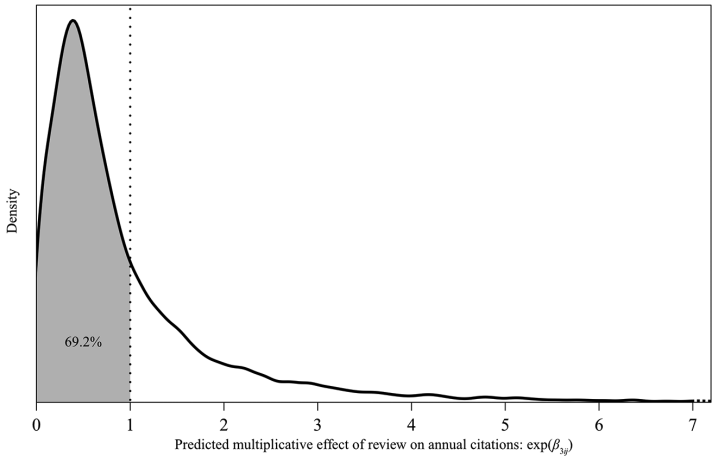

Figure 2. Net Multiplicative Effect of Annual Review Citation on Subsequent Citations Received, Corresponding to the Predicted Value of $\exp(\beta_{0ij})$ for All Articles Receiving at Least One Annual Review Citation
Note: The figure shows that the large majority of cited articles have a value of $\exp(\beta_{0ij}) < 1$ and can expect a substantial negative impact on their future citations. The horizontal axis is truncated at the 99th percentile of the empirical distribution.

are predicted to have their future citations increase more than twofold.

Figure 3 illustrates this pattern with six sample articles. In most cases, reviews lead to significantly diminished predicted citations (panels a through e). But for some articles, like the one depicted in panel f, citation by review predicts a small *increase* in future citation. The variability in the main effect of interest is an important outcome of the analysis so far: although the effect of inclusion in a review article is detrimental for the large bulk of scholarly publications, a small minority "rise to the top" and enjoy considerably expanded attention. Indeed, the remainder of our quantitative analysis will focus on uncovering the structural characteristics that differentiate these fortunate

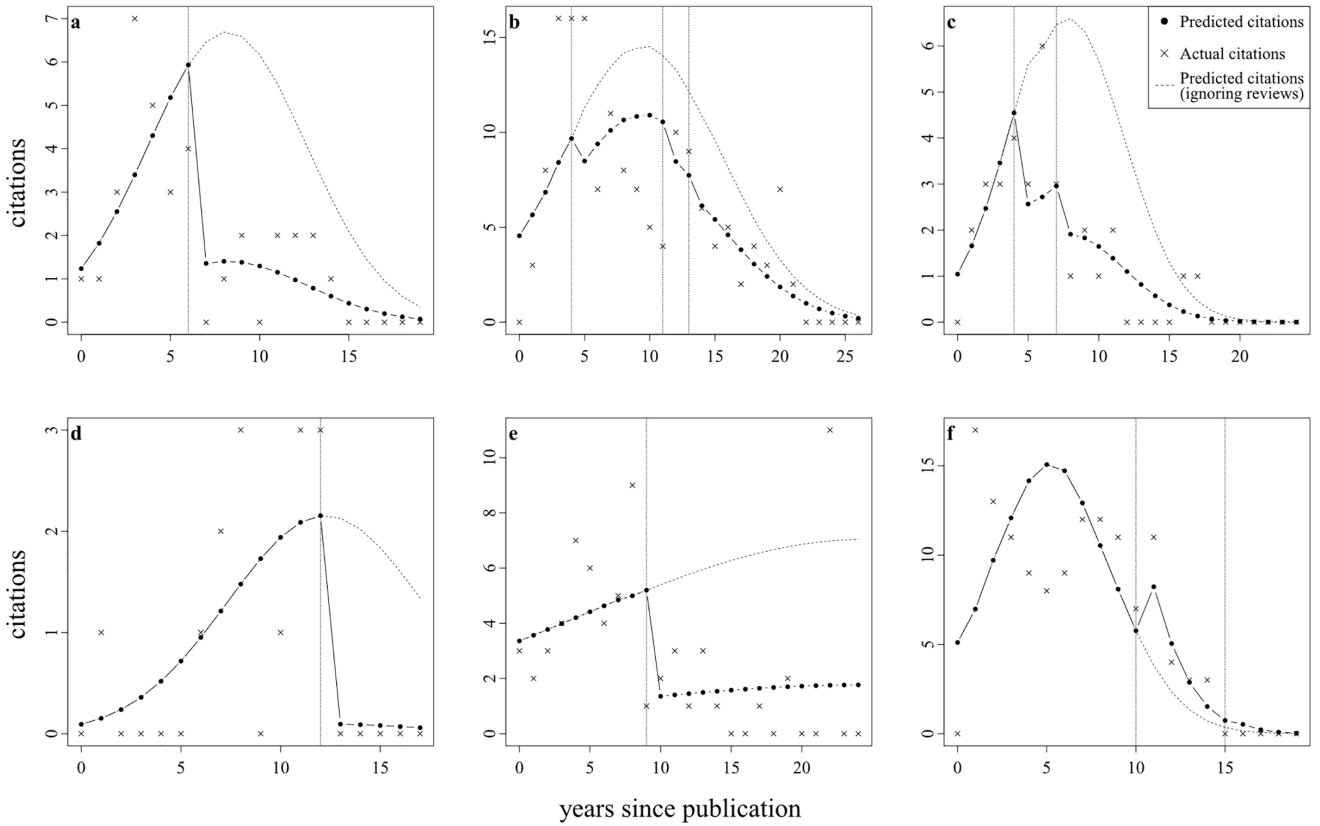

Figure 3. Actual versus Predicted Number of Citations Received by Six Sample Articles
Note: Vertical dashed lines represent years an article was included in a formal review. The model predicts a smooth expected–citation curve (dashed curve) for each publication, interrupted by discontinuities associated with reviews.

few articles from the rest of the published literature.

In relation to the theoretical perspectives outlined above, the implications of these results are significant. In contrast with the "spotlight" theory, review articles tend to *diminish* the attention their citations can expect to receive. As these results suggest, and as we will investigate, reviews have this diminishing effect on the majority of articles they cite, with only a handful of reviewed publications gaining increased attention in the future. The results of the regression make it clear that reviews do not act as neutral observers, merely raising the awareness of a body of literature. They are not simple spotlights that bring attention to a scholarly domain. In fact, reviews execute curatorial power. By representing a research specialty for general scrutiny, reviews perform a selection on existing research. Through their curation, reviews turn eyes away from (most of) what they highlight as important in a research area.

RESTRUCTURING DISCOURSE

Having shown that review articles have a heterogeneous and largely stifling effect on the individual articles they cite, we now turn to an analysis of their effects on the research area they aim to summarize. Formal review undercuts the attention given to published work, on average, but considerable variation exists in the magnitude of this effect. Indeed, some of the articles cited by a particular review may bear the brunt of the obscuring effect, while others could experience minimal impact or even a boost to their future citations.

The remainder of this article will focus on the relationship between the structure of scholarly subdomains and the positions of individual works within those structures. Dissecting such structural effects is crucial to differentiating between the different theoretical mechanisms summarized above. If reviews package knowledge into concise, unproblematic units (as the black-box theory suggests), then scholarly attention should shift away from the individual publications in a field more or less uniformly, focusing instead on the apparently settled concepts the field has produced. If instead, as we argue, formal reviews actively *remake* domains of knowledge, constructing new perspectives and new understandings of that knowledge, then the structure of scholarly attention in a domain should change in form and not just magnitude. As we will show, areas of knowledge that are recast in novel ways by review articles undergo a transformation in their topology—which findings are central, which are peripheral, and how they relate to one another shift in observable ways. We begin to uncover these differential effects by examining the structural changes that occur within scholarly discourse when a research domain finds itself the subject of a formal academic review.

The Structure of Reviewed Work

To interrogate the structural features of reviewed research domains, it is necessary to identify a representative collection of publications within the domain. Reviews generally concern emerging domains of scientific activity that are considered to speak to one another. The emerging domain can represent different scholarly moments: an emerging topic that is not yet widely recognized, the recasting of a developed subfield, or even articulation of an existing school of thought. All reflect areas of intellectual activity that the reviewer posits as related in some way. As such, a review defines a specialized domain and its subdomains through a body of related published work.

The reliable identification of specialized scholarly domains of this sort is a long-standing problem in research on scientific processes (Morris and Van der Veer Martens 2008). Although the idea of a scientific research specialty can seem straightforward from the perspective of an individual researcher, the appropriate definition of a specialized domain starkly differs depending on the research question being asked. A dominant thread in the identification of specialized domains focuses on the text of scientific output, using either domain-specific terminology (Foster et al. 2015; Rzhetsky et al. 2015) or statistically

determined lexicons (Anderson, McFarland, and Jurafsky 2012; Munoz-Najar Galvez, Heiberger, and McFarland 2020) to locate researchers and publications that use similar language in similar ways.

Such methods have been put to productive use, but their emphasis on lexical similarity has certain shortcomings in relation to the questions we pose here. Reviews are often concerned with emerging research areas that may not yet have an established lexicon. The research they describe may be the work of a small subcommunity in an established field that has not differentiated its terminology from its parent domain. Moreover (as we will illustrate in our analyses), the establishment of specialized research areas is often the result of a conceptual merging of existing groups of scholars who might use different language to talk about the central ideas they have in common.[16]

In light of this, we use a comparatively simple approach to identifying sets of related articles by utilizing the reference lists from published articles in the corpus. For each publication in the data (review or non-review), we define its *reference set* as the subset of its cited works that are contained within the WoS corpus. These reference sets exploit the expert knowledge within scholarly domains to find groups of related publications. The reference sets we define are necessarily incomplete collections of work in a field, but they capture the most important work being done in a particular domain. References have long been utilized as an effective way to identify interrelated research (Chen, Ibekwe-SanJuan, and Hou 2010; Gmür 2003; Mullins et al. 1977), and these reference sets provide a practical scaffolding for analyzing the structural dynamics of the research domain targeted by a review article.

Using reference sets as representative subsets of a specialized domain, we construct co-citation networks to describe the *structure* of the scholarly conversation within that domain. A co-citation network represents each publication with a node and creates an edge between a pair of nodes for each other publication that cites them both (Ennis 1992; Moody 2004; Price 1986; Small 1977, 1986; Stokes and Hartley 1989).[17] By focusing on co-citation between the set of articles that authors consider core to a topical area, we are able to discern structural characteristics that the simple citation counts from the previous analysis cannot. Co-citation reveals the relationships between scientific research as it is perceived by the scholars who are most engaged with and invested in that research (McCain 1986; White and Griffith 1981). Authors cite work for any number of contrasting reasons (Garfield 1979; Jurgens et al. 2018; Krampen et al. 2007), and high rates of co-citation between publications do not necessarily indicate agreement between those publications' claims. But citation provides a signal that the referenced work bears direct relevance toward a publication's arguments, be it supportive or argumentative. Frequent co-citation is therefore a strong indication that a pair of publications are in conversation with the same texts.

When aggregated across a set of articles, co-citation relations enable a host of network-analytic measures to be used, revealing the structure of similarity-in-use of a group of articles. The core of the published research in an area has an internal structure that is revealed by these patterns of co-citation among the individual works, and co-citation is especially well suited to identify the key scholarly roles and communities we will discuss. Figure 4 illustrates the structure revealed by the co-citation network for one such review (to be discussed in more detail below). If reviews have substantial structural effects on the reference set of research represented in a topic—fragmentation or unification, centralization or democratization—those changes will be revealed by comparing the set's co-citation network before and after the review is published.[18]

Identifying Structural Transformation

Structural change in networks can be difficult to consistently measure. Many measures, such as those relating to automatically

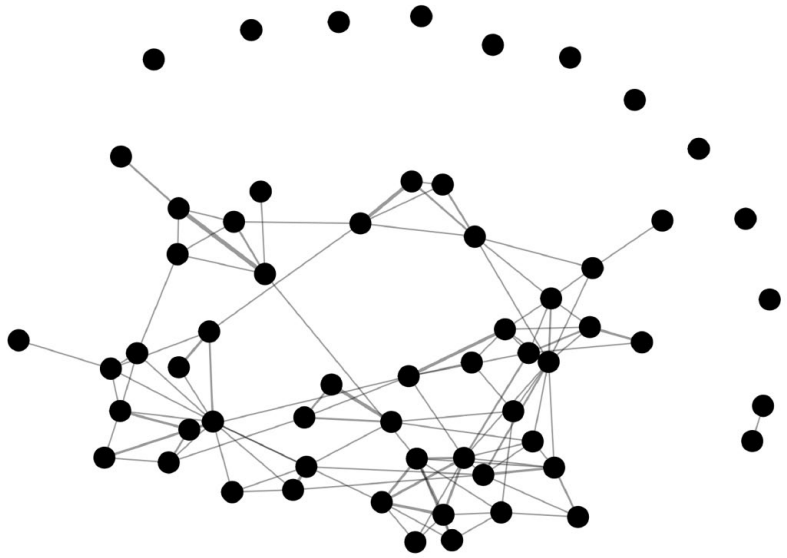

Figure 4. The Reference Set and Co-citation Network Associated with the *Annual Review of Entomology* Review "Geographic Structure of Insect Populations: Gene Flow, Phylogeography, and Their Uses" (Roderick 1996)
Note: The nodes in the network represent the set of references from the review; the edges designate co-citations between those publications in the seven preceding years. Thicker edges indicate co-citation by more publications.

identified communities, are descriptively rich but not robust to comparisons between different networks. Other measures (e.g., assortativity) offer consistent comparisons across time and communities but do not afford clear interpretations in the context of the co-citation networks under consideration here. We restrict our attention to two features of network structure, *closeness* and *transitivity*. Together, these features yield rich structural descriptions of how sets of references get co-used, and they provide robust measures of within-network change and between-network difference.

Our measure of *network closeness* is based on the average, weighted path length within a network; it is a description of how "narrow" a group of publications is. Does the set of publications expand over a wide range, with certain articles virtually unrelated to others? Or do the works address very similar ideas, touching on different aspects of the same topic? Weighted path length captures this idea succinctly: if two publications are topically similar then they are likely to be either co-cited often or both be co-cited frequently with an intermediary work. Starkly dissimilar works, in contrast, will likely require many "jumps" along co-citation edges in the network to reach one another.

We measure the length of a path in the network as the sum of the inverted[19] co-citation counts of the edges included in the path $d_P = \sum_{(i,j)\in P} e_{ij}^{-1}$, and the *minimal* path length between two vertices i and j as the minimum of this value across all possible paths connecting them in the graph. The *average, minimal path length* is simply the average of this value across all pairs of connected vertices in the graph: $\bar{\pi}_G = \sum_{(i,j)\in G} \pi_{ij} / Z$, where π_{ij} is the minimal path length between i and j, and Z is the number of connected pairs of vertices in G. The average path length of a network is descriptive, but it is quite sensitive to the number of vertices and edges within the network—random networks with more

or higher-weighted edges will tend to have shorter paths, and those with more vertices or lower weights will tend to have longer paths. We are not interested in the absolute average path length, but whether a given network has uncharacteristically long or short paths. Following Leskovec and colleagues (2008), we therefore measure the closeness of a network by regularizing the average path length in relation to the average path length of a random graph with the same number of nodes and degree sequence (where degree is measured as the sum of the weights of a node's edges). In Equation 2, $E\left(\bar{\pi}_{DS}\right)$ is the expected average path length for such a random graph.[20] Values of P_G less than zero indicate a graph with uncharacteristically short paths between vertices.

$$\textbf{AvgPathLength}_G = \log\left(\frac{\bar{\pi}_G}{E\left(\bar{\pi}_{DS}\right)}\right) \quad (2)$$

$$\textbf{Clustering}_G = \log\left(\frac{\gamma_G}{E\left(\gamma_{DS}\right)}\right) \quad (3)$$

The second network characteristic we investigate is *clustering*, the degree to which the papers listed in a review come to be co-cited in clusters. In a network with high clustering, the individual articles are embedded in tight-knit groups, with members of those groups having a high probability of being connected to one another through co-citation. A network with low levels of clustering, in contrast, is more heterogeneous, allowing certain publications to enjoy privileged positions in the global structure.

We measure network clustering (also referred to as a network's *transitivity*) using an extension to the standard equation of the clustering coefficient for weighted graphs. If publication A is frequently co-cited with publications B and C, the clustering coefficient summarizes how frequently publications B and C are co-cited with each other—it is a straightforward measure of triadic closure. It is calculated by dividing the number of closed triplets (triangles) in the network by the number of connected triplets (triangles and open two-paths): $\gamma_G = \#\textbf{closed} / \left(\#\textbf{closed} + \#\textbf{open}\right)$.[21] As with average path length, it is important to measure the clustering coefficient relative to a random graph with the same degree sequence, converting it to a measure of how uncharacteristically clustered or unclustered a network is and allowing for meaningful comparison across time and reference sets. Equation 3 shows the calculation used in our analysis.

Together, these two measures do a remarkably good job of summarizing much of the overall structure of a co-citation network. Although a simple two-dimensional space obviously cannot capture the nuances of an entire network of relationships, average path length and the clustering coefficient can indicate a wide array of network topographies of interest. Figure 5 illustrates some common network configurations associated with high and low values of these measures. Networks with low clustering coefficients and short path lengths can be characterized by star structures, with central hubs connecting groups of vertices that are not connected to one another. Networks with many long, sprawling paths that do not connect back on themselves will have a low clustering coefficient and a large average path length. A high clustering coefficient with short average paths between nodes is the measure of a "small world" network (Watts and Strogatz 1998), in which tight, insular clusters are connected by bridging nodes that span subgroups. Finally, a network with long average paths but a relatively high clustering coefficient will have dense clusters of vertices separated by longer sequences of connected, intransitive vertices.

Modeling change in networks is complicated by the interdependencies between their various structural features. A network's size tends to be closely linked to the density of network relations, which itself is highly correlated with transitivity, and so on. Certain robust statistical models designed to address such interdependencies exist, including dynamic exponential random graph models (Krivitsky and Handcock 2013) and

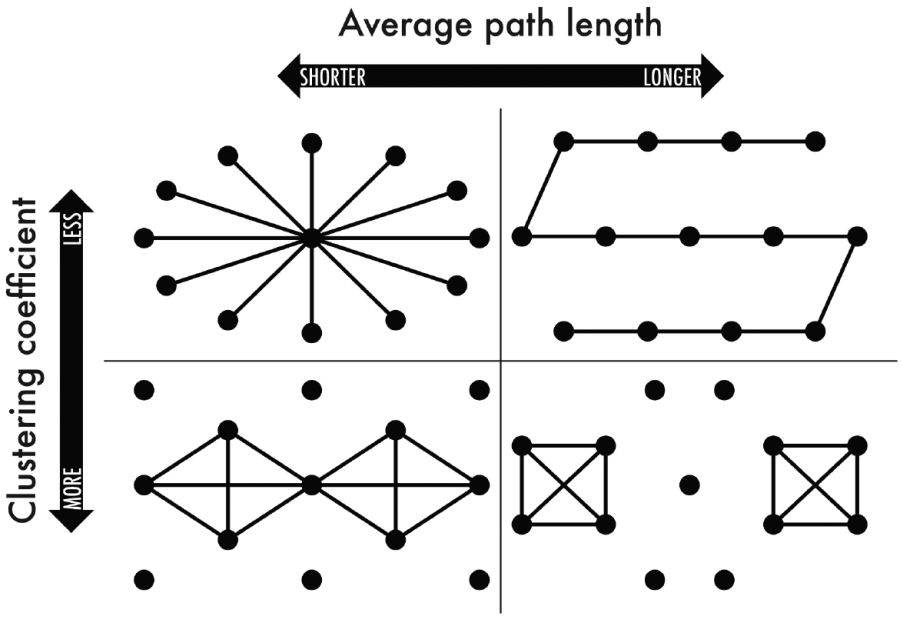

Figure 5. Typical Structures Found in Networks with High and Low Clustering Coefficients and Long and Short Average Path Lengths

stochastic actor-oriented models (Snijders 2005). These models are excellent for uncovering the dynamics within a particular community, but they are ill-suited for comparing dynamics across communities. Moreover, the complexity of most network-evolution models does not scale to the order of millions of distinct networks containing tens of millions of vertices. We therefore take a simplified approach to the estimation of structural changes associated with scholarly review.

We created a dataset with each observation representing the reference set of one article in the corpus (reviews and non-reviews), measuring for each co-citation network its structural characteristics in the seven years leading up to publication of the focal article, and the change in those structural characteristics observed over the seven subsequent years. We then used a multivariate regression (see Equation 4) to predict the changes to the structure that can be explained by review.[22] This approach is robust to the pitfalls of network prediction in several ways. This modeling approach will allow us to compare the evolution of Annual Review reference sets to general non-review articles and to matched review-like articles. As described earlier, our primary measures of interest, change in mean path length and change in clustering coefficient, are calculated with respect to deviations from the average over random rewritings of the network. This is a widely used approach in the network literature to construct variables that are not sensitive to the particularities of an individual community (e.g., Kolaczyk and Csárdi 2014:5.3). In addition, our measures of change focus on the difference between the before- and after-publication structures, so by including covariates for the initial structure of each network, we are effectively controlling for that network's idiosyncrasies.

$$\begin{pmatrix} \mathbf{AvgPathLenth}_i^{\Delta} \\ \mathbf{Clustering}_i^{\Delta} \end{pmatrix} \sim \text{MVNorm}\left(\begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix}, \Sigma \right)$$
$$\begin{aligned} \mu_{ji} = {} & \beta_{j0} + \beta_{j1}\mathbf{Review}_i \\ & + \beta_{j2}\left(\#\mathbf{Vertices}_i\right) \\ & + \beta_{j3}\left(\#\mathbf{Vertices}_i\right)^2 \\ & + \beta_{j4}\left(\mathbf{Density}_i^0\right) \\ & + \beta_{j5}\left(\mathbf{Density}_i^0\right)^2 \\ & + \beta_{j6}\mathbf{AvgPathLenth}_i^0 \\ & + \beta_{j7}\mathbf{Clustering}_i^0 \end{aligned} \quad (4)$$

Equation 4 specifies the multivariate linear model in detail, with reference sets indexed by i and $j \in \{1,2\}$ indicating the response variable. To describe change in network structure, we adopt notation for the average path length (Equation 2) and clustering coefficient (Equation 3) to indicate their values for each network in the time leading up to review, and the difference in those values after review. $\mathbf{AvgPathLength}_i^0$ is the average path length of co-citation network i before being cited by the focal article, and $\mathbf{AvgPathLength}_i^{\Delta} = \mathbf{AvgPathLength}_i^1 - \mathbf{AvgPathLength}_i^0$ is the change in that network's average path length in the period after being cited together. The primary explanatory covariate for expected change for each outcome variable is $\mathbf{Review}_i$, an indicator variable for whether the focal article for reference set i was published in an Annual Review journal. We include several control variables to account for potentially confounding structural features: network size ($\#\mathbf{Vertices}_i$) and its square, network density before review ($\mathbf{Density}_i^0$) and its square, and the structures of interest in the time period before being cited by the focal article ($\mathbf{AvgPathLength}_i^0$ and $\mathbf{Clustering}_i^0$).

The model described in Equation 4 is a straightforward way to identify changes in network structure that coincide with the publication of Annual Review articles, but care must be taken to justify the broader argument we make about the effect of formal review on research domains. Academic discourse is a distinctly reflexive process—every stage of research is carried out with an awareness of the context in which it will be viewed by others. No part of the creation or dissemination of scholarly work is done in a vacuum, an observation that is especially true for review articles. Annual Review publications are often targeted directly at emerging fields that are likely experiencing characteristic structural transformations on their own. Disentangling the types of changes that are the result of formal review from those that would have taken place independently in a field requires careful consideration of both the theorized mechanisms in play and the analytic methods used. We therefore estimate the model in Equation 4 on two versions of the data, each emphasizing a different aspect of the transformations under consideration. The first version uses a representative subsample of the full corpus—reviews and non-reviews alike—to describe the overall features of structural change we observe. The second version uses a more restrictive sample of articles, matching Annual Review articles to Annual Review-like articles published elsewhere, to capture the narrow effect of Annual Review journals.

The initial estimation of the model parameters compares Annual Reviews to all other articles in the corpus using a weighted 5 percent sample from the corpus. A subsample is necessary for computational efficiency—the calculation of network statistics like average path length as well as estimation of the multivariate regression are impractical on the full set of citation data.[23] The sample is taken at the level of the referencing article for each reference set and is weighted to ensure (a) representation of less prolific subject areas and (b) retention of all Annual Review articles. In addition, degenerate networks (very sparse or having fewer than 20 referenced articles in the Web of Science) for which the structural statistics are not calculable were removed from the sample.

The first two columns of Table 3 present estimates[24] for the model using the 606,979 observations in this sample. The control variables suggest some interesting patterns, but the primary covariate of interest for our

Table 3. Results Describing Structural Changes to Reference Clusters

	5% Sample		Matched Sample	
	AvgPathLength$^{\Delta}$	**Clustering$^{\Delta}$**	**AvgPathLength$^{\Delta}$**	**Clustering$^{\Delta}$**
(Intercept)	−.008 (−.010, −.005)	−.032 (−.035, −.029)	−.002 (−.027, .023)	.012 (−.008, .031)
Review	−.145 (−.164, −.125)	−.208 (−.228, −.188)	−.109 (−.136, −.082)	−.07 (−.090, −.049)
#Vertices	−.135 (−.138, −.132)	−.026 (−.029, −.023)	−.207 (−.222, −.192)	.01 (−.001, .021)
#Vertices^2	.008 (.007, .009)	.006 (.005, .008)	.044 (.031, .058)	.005 (−.006, .016)
Density^0	−.105 (−.107, −.102)	−.083 (−.085, −.081)	−.107 (−.125, −.089)	−.034 (−.048, −.020)
(Density^0)^2	.002 (−.000, .003)	.027 (.025, .028)	.012 (.001, .023)	.017 (.009, .026)
AvgPathLength^0	−.638 (−.640, −.636)	.001 (−.001, .003)	−.625 (−.639, −.610)	.012 (.001, .022)
Clustering^0	.033 (.031, .035)	−.62 (−.622, −.617)	.026 (.010, .043)	−.817 (−.829, −.804)
Res. Std. Dev.	.778 (.776, .779)	.803 (.802, .804)	.789 (.780, .799)	.597 (.590, .604)
Res. Cor.	.07 (.062, .072)		.186 (.169, .202)	

Note: The left two columns of results show estimates using a weighted, 5 percent sample of the full data; the remaining columns show the same results estimated on the propensity score–matched sample. The final two columns use co-citation networks that exclude articles citing the root article. Values in parentheses represent 95 percent credible intervals on all estimates.

analysis is the indicator for sets of publications referenced by review articles. For both outcome variables, the coefficient on **Review** (β_{j1}) is significant and negative. This means groups of articles cited by a formal review are expected to have an average path length about .145 standard deviations shorter, and a clustering coefficient about .208 standard deviations smaller, than groups cited by other published papers. These results suggest formal reviews might be inducing distinctive structural changes for the sets of articles they cite, making them both narrower in scope and less cohesive as specialties.

However, as mentioned earlier, we must be careful about the specific comparison this initial analysis makes. The question we hope to address is this: how does publication of a formal review affect the scientific discourse in a field? But the comparison implicit in the representative sample contrasts formal reviews with all other scholarly articles. To interrogate the effects of formal review specifically, we ought to compare "officially sanctioned" reviews to review-like articles that are not endorsed by a publisher like Annual Reviews. This is important because the estimates produced on a representative sample (like those just described) may confound the effect we are seeking with a more general pattern of the evolution of research specialties. It is plausible that specialized subfields develop in predictable ways and that the estimates in the first two columns of Table 3 are an artifact of the tendency for reviews to target fields at specific moments in their development. In this situation, the **Review** variable would simply be an indicator of a newly emerging specialization, marking the structural dynamics typical of that stage of development without having an independent effect on those dynamics.

To account for this, we estimate the model described in Equation 4 again, this time on an unweighted, propensity score–matched subsample of the data.[25] The propensity score used for matching is built from a logistic regression predicting formal review articles using the structural characteristics of the articles' reference co-citation sets (Equation 5). In addition to the structural features used as controls in Equation 4, we add predictors that are likely to be associated with reference lists of review versus non-review articles. Because reviews may tend to focus on novel subfields, we account for structural features of the co-citation network that are indicative of newly emerging areas of research: the mean age of the cited references (**AvgCiteAge**) and the standard deviation of those ages (**StdDevCiteAge**). We also incorporate the total number of citations received by any member of the set at the time the reference article cites them (**TotalCites**) to control (along with the network's size) for the relative popularity of the references themselves. Finally, formal reviews might be more likely to cite work by one of the review's own authors, so we include a covariate for the number of publications in the set that share at least one author with the reference article (**SelfCites**).

$$\begin{aligned}\mathbf{Review}_i &\sim \text{Bernoulli}\left(\text{logit}^{-1}(\mu_i)\right)\\ \mu_i &= \beta_0 + \beta_1\left(\#\mathbf{Vertices}_i\right)\\ &\quad + \beta_2\left(\#\mathbf{Vertices}_i\right)^2 + \beta_3\left(\mathbf{Density}_i^0\right)\\ &\quad + \beta_4\left(\mathbf{Density}_i^0\right)^2 + \beta_5\mathbf{AvgPathLenth}_i^0\\ &\quad + \beta_6\mathbf{Clustering}_i^0 + \beta_7\mathbf{AvgCiteAge}_i\\ &\quad + \beta_8\mathbf{StdDevCiteAge}_i + \beta_9\mathbf{TotalCites}_i\\ &\quad + \beta_{10}\mathbf{SelfCites}_i\end{aligned} \quad (5)$$

Before discussing the results of the propensity score–matched analysis, it is worth considering the results of the propensity score model itself (presented in Table 4). At the point of publication, the set of works that review articles reference are distinct from a regular citation community in important ways. Unsurprisingly, reviews are strongly associated with larger reference lists than are standard articles (**#Vertices**).[26] The networks also have significantly fewer co-citation edges than do non-reviews, and the articles within them have received somewhat fewer citations overall, suggesting a sparser research domain consistent with a newer field. Interestingly, review articles cite works that are slightly older than those of a standard article, on average, but that were published in a much narrower period of time. Reviews are also disproportionately likely to self-cite, reflecting the policy of Annual Review journals to invite reviewers who are considered pivotal in a field. Reviews are more likely to cite communities with high clustering coefficients (**Clustering**^{0}) and very short path lengths (**AvgPathLength**^{0}) in their co-citation networks, providing important context for the initial state of the changes described in the multivariate model. This combination of high clustering and short paths is characteristic of a small-world network structure—communities in which most interaction occurs between members of the same clusters of individuals, but longer-range influence is accommodated through inter-cluster bridges (Newman 2001b, 2003; Uzzi and Spiro 2005).

These results suggest the literature cited by reviews forms a constellation of distinct, active communities; small research clusters, each engaged in vigorous conversation within itself, are linked by more sporadic connections that tie them together into a topically coherent whole (as summarized in the top panel of Figure 6). As a whole, estimates from the propensity model describe what many would expect of a newly emerging scientific specialty: clusters of vibrant research activity spanning a short time frame and just beginning to realize their connections with one another. This suggests the propensity score we use to produce the matched sample is identifying the types of research communities that *could* be selected for an Annual Review article but were not.[27] Analysis of this matched sample therefore allows us to look for divergence in the histories of research communities that were the subject of such formal review and those that were not.

The model estimates presented in the last two columns of Table 3 are based on the matched sample of about 6,500 pairs of

Table 4. Logistic Regression Estimates for the Propensity Score Model (Equation 5)

Coefficient	Estimate	95% Cred. Int
(Intercept)	–5.293	(–5.336, –5.250)
#Vertices	1.230	(1.186, 1.274)
#Vertices^2	–.097	(–.108, –.086)
Density^0	–.198	(–.231, –.165)
(Density^0)^2	.043	(.022, .064)
AvgPathLength^0	–.149	(–.178, –.119)
Clustering^0	.100	(.082, .118)
AvgCiteAge	.087	(.046, .128)
StdDevCiteAge	–.475	(–.516, –.434)
TotalCites	–.326	(–.357, –.296)
SelfCites	.213	(.196, .230)

reviews and non-reviews ($N = 12{,}990$). Strikingly, these results display the same overall effect of reviews for the matched pairs as we see in the 5 percent sample. Formally reviewed research areas experience a significant negative change in both their average path lengths (about 11 percent of a standard deviation shorter) and clustering coefficients (about 7 percent of a standard deviation lower) when compared to similar reference sets that were not the subject of formal review. However, note that the estimated effect of review on clustering coefficient is significantly smaller in magnitude when estimated on the matched sample (–.070 versus –.208), suggesting that at least some of the dramatic "declustering" we observe in these communities would have taken place even in the absence of a formal review. Nonetheless, when taken together with the propensity-score estimates themselves, these results describe a striking restructuring of the discourse in a field. Although reviewed networks were already very narrow, the further decrease in path lengths suggests these networks shrink even further. The negative estimate for the effect of review on change in clustering coefficient shows that the individual clusters characteristic of the small-world sets that are formally reviewed in an Annual Review journal become considerably less cohesive.[28]

The estimates presented in Tables 3 and 4 describe a network that becomes much more centralized, with more of the literature in their specialty relating exclusively to a small subset of publications. After a review is published, more of the co-citation relations are centered on fewer of the cited works (a structure Gondal [2011] suggests is typical of a newly-emerging research domain). Communities of literature that are the subject of a review article collapse into a hub-and-spoke structure, as illustrated in Figure 6. Edges from the hubs to any one of the peripheral works become more common, and edges between those peripheral works diminish. This transformation suggests reviews are performing an act of selection: certain works are singled out as exemplary in a scientific subject, so much so that the remaining work already published in that area is cited only in relation to the newly anointed exemplars. The story that a review tells about an emerging field—a narrative of its past, present, and future as a coherent specialty (Sinding 1996)—shapes that field in consequential ways. The legitimacy of a formal review in an Annual Review journal grants its authors considerable influence over the development of a research domain.

Identifying Exemplars

Still, if review articles are indeed performing curatorial work on the articles they consider, centering a few as exemplars of a field while sidelining others, what criteria does that

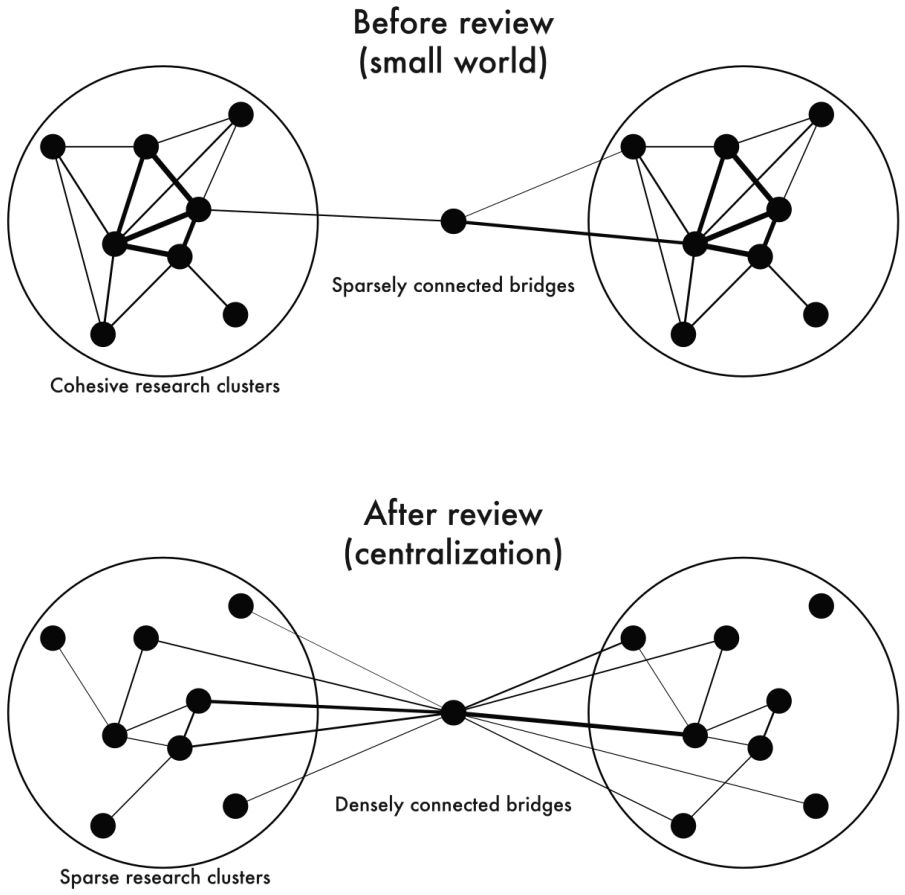

Figure 6. Simplified Illustration of the Structural Changes to Co-citation Networks Associated with Review Articles
Note: The network in the top panel typifies a small-world network, with a small number of links bridging tightly connected clusters. The bottom panel shows a highly centralized network in which intra-cluster edges are eschewed in favor of ties to a central hub.

curation use? Do reviews simply amplify the attention received by articles that are canonical to a topic, drawing further accolades to the already celebrated? Or do reviews perform a more dramatic form of synthesis, drawing previously marginal work into the spotlight to compose a novel portrait of a specialized domain of knowledge? We address this question by examining the structural features of individual articles that lead to changes in the attention they receive.

Many of the articles a review cites do not become significantly more or less central to their subject area as a result, but some migrate from the periphery to the center of the domain's focus. Our aim is to identify the articles that gain significantly in this regard, and to identify which features of their initial position help them achieve increased attention. We focus on two facets of initial network position associated with the importance of a vertex in its community: central nodes and bridging nodes. Structurally, a central publication is one that is co-cited frequently within a relatively tight cluster of other publications that are themselves co-cited frequently to one another. This recursive definition means a central article is one that is cited alongside many other publications in the network and is in the core of a cohesive cluster of

publications that are all frequently cited by the same body of published work. *Eigenvector centrality* (Bonacich 1987) is ubiquitous in the literature, as it succinctly captures this recursive notion of structural importance—an article is central if it is tightly connected to other central articles.

In contrast to central nodes, network bridges exist in the spaces between clusters in a network. In the literature, bridges are discussed in terms of brokerage (Fleming, Mingo, and Chen 2007) and structural holes (Ahuja 2000). To capture the degree to which a specific article acts as a bridge in a reference set, we measure the local *transitivity* of each vertex.[29] It is important to note that transitivity has an inverted relation to bridging: a vertex with *low* transitivity holds a *highly* bridging position in its network and vice versa. The transitivity of a given vertex is the proportion of the pairs of its neighbors that are themselves connected to one another (we use the generalization of this measure for weighted networks from Barrat and colleagues [2004]). Co-citation networks, as a class of affiliation networks, tend to have high numbers of closed (transitive) triads, making local transitivity an especially powerful measure of bridging in the communities we are studying.

To discern which publications are boosted by review articles relative to the reference set, we use a publication-level regression. Equation 6 specifies a multilevel representation of the model, where *i* indexes articles embedded in the reference set indexed with *j*. For each referencing article, we calculate centrality and bridging statistics for every article in its co-citation network before and after publication of the reference article, and we use these statistics to predict those articles' change in citations received. The dependent variable is computed as the difference in the total number of citations each article receives in the seven-year windows before and after publication of the review article.

To allay spurious patterns that may emerge from variability between reference set networks, we standardize the measures of citation change, centrality, and transitivity to ensure they describe each publication's evolution relative to the reference set in which it is embedded. Thus, each value is centered at the set's mean and divided by the set's standard deviation. Such group-mean centering is vital for interpretation of the model—we are interested in the change for each article *relative* to the other articles in its reference set, and group-mean centering allows us to measure that change while minimizing the confounding effects of set-specific characteristics. Articles with high values of the dependent variable are those that gained especially greatly after being cited by the focal article, for instance, moving from a position of relative obscurity to one of high visibility in the field.

$$
\begin{aligned}
\mathbf{EVCentrality}_{ij}^{\Delta} &\sim \text{Normal}(\mu_{ij}, \sigma_2) \\
\mu_{ij} &= \beta_{1j}\mathbf{Citations}_{ij}^{0} + \beta_{2j}\mathbf{EVCentrality}_{ij}^{0} \\
&\quad + \beta_{3j}\mathbf{Transitivity}_{ij}^{0} \\
&\quad + \beta_{4j}\left(\mathbf{EVCentrality}_{ij}^{0} \times \mathbf{Transitivity}_{ij}^{0}\right) \\
&\quad + \beta_{5j}\mathbf{SelfCite}_{ij} + \beta_{6j}\mathbf{ReviewCited}_{ij} \\
\beta_{kj} &= \gamma_{k0} + \gamma_{k1}\mathbf{Review}_{j}
\end{aligned}
\tag{6}
$$

As independent variables, we use measures of each work's eigenvector centrality and transitivity in the co-citation network that preceded the focal article's publication (as well as their interaction). We also include the total number of citations each article received in the seven years preceding its citation by the focal article. Each of these article-level variables is group-mean centered—a value of $\mathbf{Citations}_{ij}^{0} = 1.0$, for example, would indicate that article *i* has received one standard deviation more citations than the average for its reference set. Together, these covariates predict which articles will experience a boost in attention based on their initial attention and position in the network.

Finally, we include two non-structural covariates to account for potentially confounding characteristics of each referenced article. $\mathbf{SelfCite}_{ij}$ is an indicator variable with a value of 1 if the referencing article (*j*) shares any authors with the referenced article

(i), and **ReviewCited**$_{ij}$ is an indicator variable that takes a value of 1 if the referenced article (i) is itself a formal review. Self-citations are an important consideration because, as shown earlier, reviewers are very likely to cite their own work, and might therefore present their own work as especially relevant to the topic they are reviewing. A control for a referenced article being a review itself mitigates the potential that review articles are both structurally distinct (occupying, e.g., bridging positions) and independently more likely to gain scholarly attention.

Because we aim to measure the difference in these predictions for networks referenced by formal reviews, we interact all the covariates with a dummy indicating whether the focal article was published in an Annual Review journal. In addition, we restrict our sample to the matched pairs of referencing articles described earlier (see Table 3). Although the current model (Equation 6) is specified at the level of the individual cited publication, the causal effect of interest remains at the level of the referencing article. Using the reference–set level matched pairs allows us to compare the changes in individual article positions within review and non-review sets, and the estimates indicate the degree of divergence between scientific specialties that were chosen for review and those that were not. The full model is estimated for 589,735 articles across about 13,000 reference sets using OLS with cluster-robust standard errors to mitigate the within-network interdependencies of observations.[30]

Table 5 reports the coefficient estimates (along with 95 percent credible intervals) for the model. The first six variables, those not interacted with **Review**, are controls indicating the expected change in centrality for non-review articles. Although these estimates do not address our primary questions, they are nonetheless interesting. These networks are matched to formal reviews on structural and temporal criteria, and they largely represent emerging specialties. The first six variables therefore represent general trends for these types of specialties, whether or not they are the subject of an Annual Review. The highest-magnitude of these coefficients (**EVCentrality**^{0}) is directly related to the dependent variable and predictable in its sign; the outcome is change in relative centrality, so simple regression to the mean suggests articles with a high relative centrality would tend to become less central and vice versa. The coefficient on (**Citations**^{0}) is significant in magnitude and negative, indicating highly cited articles are likely to move somewhat toward the periphery of their networks. This could be the result of highly visible research losing its relevance as it ages or being supplanted from the core of the community by newer work.

Local transitivity is similarly negatively associated with increased centrality. For articles with low transitivity, the articles they

Table 5. Coefficient Estimates and 95 Percent Credible Intervals from a Linear Regression Predicting Relative Change in Article Eigenvector Centrality

	EVCentrality$^\Delta$
Citations^0	−.0612 (−.068, −.054)
EVCentrality^0	−.504 (−.515, −.493)
Transitivity^0	−.083 (−.089, −.077)
EVCentrality^0 × Transitivity^0	−.101 (−.108, −.094)
SelfCite	.037 (.022, .051)
ReviewCited	.107 (.066, .147)
Citations^0	−.001 (−.011, .009)
Review × EVCentrality^0	.02 (.005, .036)
Review × Transitivity^0	−.02 (−.028, −.011)
Review × EVCentrality^0 × Transitivity^0	−.014 (−.024, −.004)
Review × SelfCite	.012 (−.007, .030)
Review × ReviewCited	−.024 (−.073, .025)
Observations	589,735
Adjusted R^2	.260

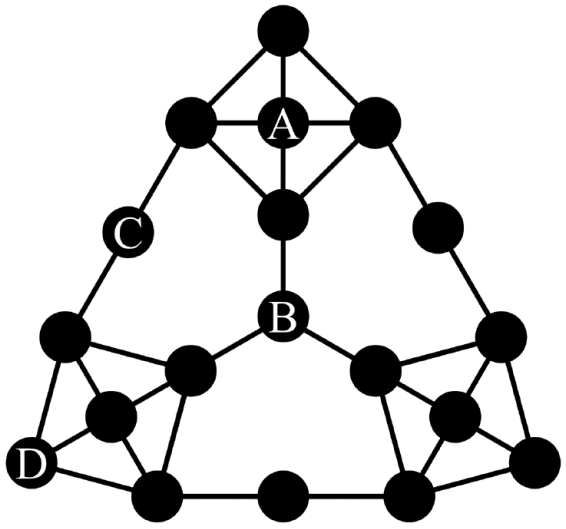

Figure 7. Illustration of the Interaction of Eigenvector Centrality and Local Transitivity in a Network

Note: In this figure, vertex A has high (eigenvector) centrality and high (local) transitivity, vertex B has high centrality and low transitivity, vertex C has low centrality and low transitivity, and vertex D has low centrality and high transitivity. Formal reviews reward articles in positions like B over those in positions like A, C, or (especially) D.

are co-cited alongside are not likely to be co-cited alongside one another, suggesting these low-transitivity articles act as bridges in a network. The results of this model indicate that even when not cited in a review, bridging articles are likely to become more central over time. Articles that share an author with the referencing article are also more likely to have a positive change in their eigenvector centrality, whether or not the referencing article is a formal review. Finally, articles that are themselves reviews are predicted to become more central after being cited by non-review and review articles alike.

These patterns tell us something about the general structural evolution of scientific specialties, but we are most interested in the comparison between research areas that receive formal reviews and those that do not. Once we account for the types of research fields that are prone to review, what is the residual difference in fields that are chosen for review by an Annual Review journal? The remainder of results in Table 5—those that include an interaction with **Review**—describe exactly these differences.

First, we see no evidence that reviews have an independent effect on the centrality of self-cited articles or of articles that are themselves reviews. Moreover, reviews do not seem to change the trajectory of highly cited publications, at least in comparison with similarly structured communities that were not cited by a review. However, the coefficients measuring the structural positions of articles before being referenced tell a different story. The estimated effect of eigenvector centrality, local transitivity, and their interaction are all significantly different than zero, and they suggest review articles will tend to disproportionately reward publications that hold a central position or act as bridges in their co-citation network. The negative estimate on the interaction term can be interpreted to mean articles that are both central *and* occupy bridging positions—those like vertex B in Figure 7—are especially likely to become more central after being cited by

a formal review. To get a sense of the magnitude of these effects, one can think of an article that is in the top fifth percentile of eigenvector centrality and bottom fifth percentile of transitivity in its network before being cited by a review. Such an article will be expected to experience an increase of more than 10 percent of a standard deviation in eigenvector centrality in the seven years following review, being co-cited much more frequently alongside the other important work in the domain.[31]

Together, the estimates from the two models (Tables 3 and 5) suggest review articles redefine how cited works are interrelated going forward by weakening the existing, tight-knit clusters of research, and by recentering the conversations that relate these clusters to one another. An important caveat to these findings lies in the small magnitude of the coefficient estimates in Table 5. It is clear the model is not describing the majority of the variation in centrality change among these communities (the R^2 for the model is around .26). Nevertheless, the findings are significant for understanding the effects of formal review articles. The dynamics of citation are immensely complex, determined by a multitude of scholarly, social, institutional, and structural factors. The above analysis examines *only* the effects of the network's structure, but it still uncovers significant regularities in the way those networks change. The contrasting predictions for review and non-reviewed communities reveal the acute, atypical restructuring that results from formal scholarly review, even when accounting for the types of changes typical for emerging specializations. Referring to Figure 6, it is exactly the bridging articles that are most likely to become the central hubs in the reshaped network that reviews create.

LOOKING CLOSER

These statistical analyses paint a vivid picture of how formal scholarly review contributes to the structuring of emergent research domains. Research output is treated much differently by the larger academic community after it has been reviewed. Specific findings are drawn into a broader conversation with one another, with fewer stark divides between different approaches to the same topic. Research projects that are relevant to multiple conversations become exemplars that relate disparate threads into a single cohesive discourse. To interrogate this process in more detail, and to cement the ideas in a practical framework, it is useful to dissect specific cases of domain transformation associated with a formal review.

We selected three such cases for dissection: "Integrated Assessment Models of Global Climate Change" from the *Annual Review of Energy and the Environment* (Parson and Fisher-Vanden 1997); "Geographic Structure of Insect Populations: Gene Flow, Phylogeography, and Their Uses" from the *Annual Review of Entomology* (Roderick 1996); and "Surface Treatments of Polymers for Biocompatibility" from the *Annual Review of Material Science* (Elbert and Hubbell 1996) (see Figure 8). The close examination of individual cases serves two purposes. First, in a large-scale quantitative analysis such as this, looking at the outcomes as they play out in real scenarios gives substance to the abstract results revealed by the models. Clear examples of the structural changes described in this article can aid in understanding how they are realized within actual publication communities. These cases are not intended to validate the mechanisms of change we describe—a group of three exemplary samples from such a large corpus would be ill suited for that purpose—rather, they allow a partial elaboration of a dynamic structural evolution. Any case, including those shown in Figure 8, will afford multiple, potentially contradictory explanations of the mechanisms at play. We chose these cases not to be *representative* of the specific, diverse circumstances of academic fields, but to be *illustrative* of the transformations that the quantitative analyses uncovered at scale.

The second purpose for showing a selection of real-world examples is to underscore some of the limitations of the story of

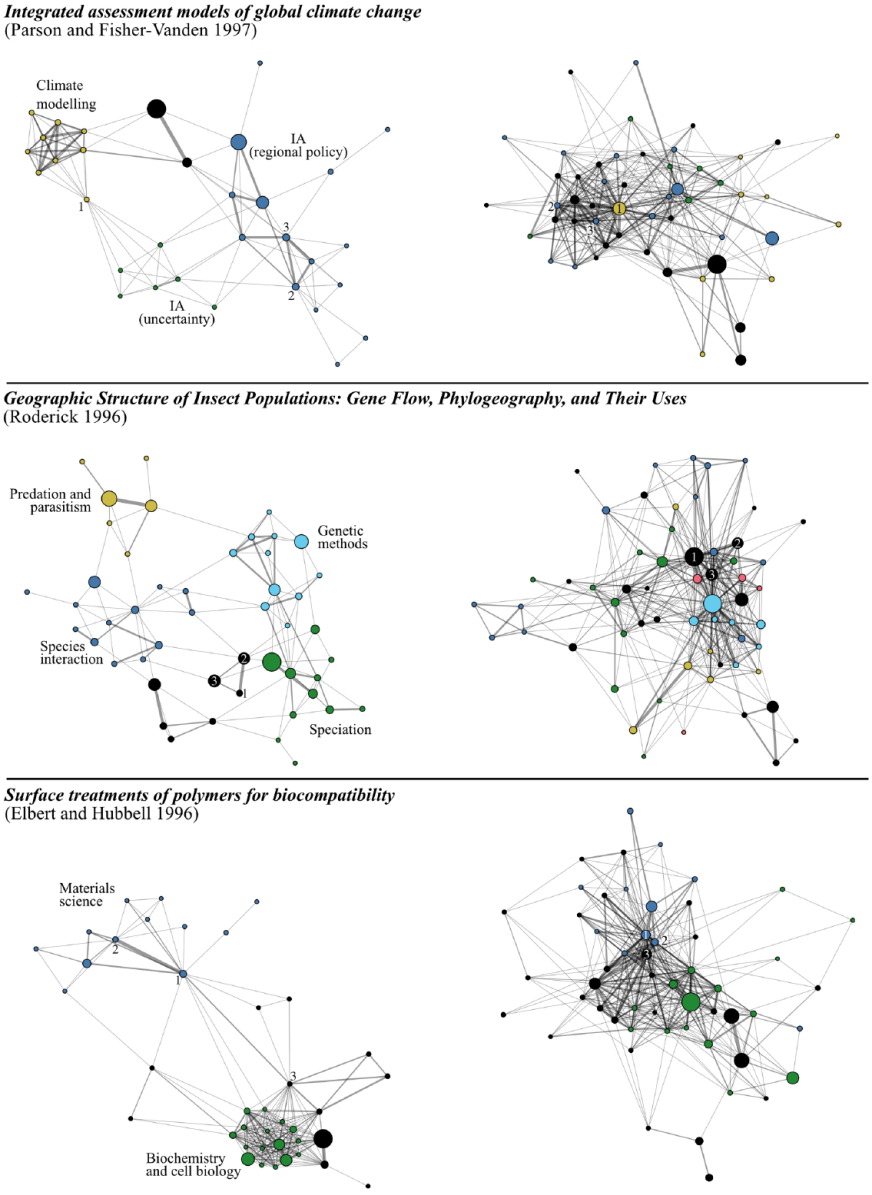

Figure 8. Illustration of Network Change in Three Cases
Note: Each panel contains visualization for a reference set's co-citation network in the seven years leading up to (left) and following (right) the publication of the Annual Review article. Vertex size reflects the total citations received, and edge widths reflect co-citation count. In each case, the three articles with the highest post-review eigenvector centrality are numbered. Unconnected isolates and dyads are omitted from the networks for visual clarity. Clusters in the networks before review (left-hand diagrams) are colored (see the online version of the article for a colored figure) and indicated with descriptive labels.

structural change we describe. The evolution of any social network is an inherently complex process. The reflexivity and multifaceted nature of scholarly publication mean co-citation networks display especially complicated dynamics. The mechanisms we describe

here—those of scholarly centralization and the promotion of bridging exemplars—form just one component of the process of disciplinary evolution. However, as the examples in Figure 8 illustrate, myriad other factors influence the type and magnitude of reshaping these fields undergo. The success of a research project and its impact on a field are influenced by forces related to institutional affiliations, personal relationships between scientists, geographic location, political and cultural climate, technological innovations, and any number of other particularities. In spite of the complexity and heterogeneity of scholarly fields' evolution, formal review appears to exert consistent influence over the shape of that evolution.

Each of the three panels in Figure 8 represents one Annual Review publication. The cases were chosen to demonstrate the types of transformations suggested by the quantitative analysis from the previous section and to span a variety of scientific disciplines. In each panel, the network on the left represents the co-citation structure of the reference set collected in the seven years immediately preceding publication of the review, and the network on the right represents that same set over the seven years after the review was published. Perhaps most striking when comparing the networks before and after review is the consistent change in structure with clustered, small-world networks transforming into centralized, core–periphery networks. The size of the vertices in the figure represents the number of citations received by the article, demonstrating the high level of centralization in the post-review reference sets. Although the networks become more dense overall after review, the co-citation relations accumulate predominantly among a small core of publications. We argue that this centralization is a result of reviews breaking down the boundaries between insulated research clusters and lifting a smaller number of publications up as a "hub" holding the emerging scientific subfield together.

The examples in Figure 8 are consistent with our explanation, but they also illustrate the domain-specific particularities that underlie the pattern of centralization—the processes by which reviews reconfigure the field are by no means uniform. In each case, the network diagram on the left maps the contours of a burgeoning subfield defined by the institutions, disciplinary norms, and existing research agendas that constitute it, and the diagram on the right describes a unified, centralized structure that might represent a more established area of research. But each of these three transformations is contingent on the specific context of its own domain.

Edward A. Parson and Karen Fisher-Vanden's 1997 review (top panel) surveys three distinct approaches to integrated assessment (IA) models of climate change, all in the context of the more traditional atmospheric climate modeling that does not focus on political, economic, or social factors. George K. Roderick's 1996 review (middle panel) describes several distinct research clusters[32] concerned with the interaction between geography and the evolution of insect species, linking them together with recent work on genetic analyses. And Donald L. Elbert and Jeffrey A. Hubbell's 1996 article (bottom panel) summarizes a class of biomaterial surface treatments that exists at the boundary between the otherwise often distinct biological and material sciences. Each of these situations is embedded in and responsive to a different scholarly setting, but they all share a common feature of illustrating connections between research domains that in other respects might have little in common. These three cases are typical of Annual Review articles in that they appear to meld a disparate archipelago of research clusters into a singular island of the targeted subfield.

The merging process is elucidated by examining the roles of key publications in the co-citation networks. The numbered vertices in each panel of Figure 8 mark the publications that are most central to the post-review co-citation network.[33] With few exceptions and across the three examples, the articles that end up near the core of the network seven years after the review occupied central

bridging positions when the review was published. In practical terms, this means that as reviews selected specific publications for their relevance to more than one of the disparate communities those reviews were linking, those same publications became key citations in the academic literature that followed. However, the drift toward the core that these boundary articles experience coincides with an unraveling of the dense research clusters they bridge. The most dramatic examples of this process in the cases listed here are the "climate modeling" (top), "species interaction" (middle), and "biochemistry and cell biology" (bottom) clusters in Figure 8. Each of these clusters represents a distinct and highly cohesive research domain that is cited by the review. However, after the review is published, these clusters become sparse, being characterized more by their articles' connections to the new hub than to each other.

Although these cases are just single examples, they demonstrate the processes that underlie the transformations characteristic of reviews. Each review in Figure 8 is tied to a major reshaping of the conversation surrounding its topic. In every case, the literature became more integrated in the eyes of those who cited it. The climatological research investigating atmospheric processes became more solidly engaged with integrated assessment models of economic and political change. Publications discussing insect evolution and geography began to reference the then-new work on genetic "microsatellites" in insect populations. However, this integration came at the expense of the cohesion that had been inherent in the distinct communities. By shifting the focus of discourse from the narrow scope of the highly specialized subdomains to the integrated whole, a larger portion of the conversation shifted to relationships between the distinct literatures rather than within them. The consequences of this shift were significant. The cost of greater integration was the minimization of the highly specialized work in smaller communities, and the complete marginalization of research that did not fit in to the new, centralized narrative.

SYNTHESIZING SCIENTIFIC SPECIALTIES

Our analyses describe the significant discursive transformations, both dramatic and subtle, that accompany curated academic review. Articles in Annual Review journals affect the future of the works they cite. We show that, contrary to conventional wisdom, the majority of reviewed publications are cited less than if they had been omitted from the review. Paradoxically, reviews tend to draw attention away from the specific articles they cite, an outcome that holds across disciplines and publication cultures.

One might explain this outcome by suggesting a process of knowledge encapsulation: reviews describe scientific specialties that have reached their conclusion and can be incorporated as resolved scientific fact. In this view, reviewed articles represent a step on the path to a conclusive finding and need not be cited once that finding has been achieved. This sort of black-boxing of knowledge may be present in scholarly discourse, but we argue that there is a more complex, and in many ways more significant, process taking place. Our examination of the structural changes that occur in reviewed scientific specialties suggests reviews shift discourse in a manner that simultaneously simplifies and collapses a knowledge community. A reviewed specialty may receive more attention overall, but that attention is directed toward a small group of exemplars. Peripheral research is included primarily through its relation to a central hub. The metamorphoses specialties undergo make them more closely connected while limiting the kinds of discourse they use.

Review articles mark a particular moment in the evolution of an academic specialty—a moment when a domain of research that is still relatively young has become established enough to warrant attention beyond that of its pioneering scholars. Annual Review articles are written about areas that have a sizeable literature, a range of committed scholarship, and often consist of multiple complementary or competing internal specializations.

The transformations they describe are not as grand as the Kuhnian paradigm shifts or black-boxing of facts studied at length by scholars of science. Moments like those are marked not by reviews but by textbooks and encyclopedias. Rather, formal reviews indicate the meso-scale transformation of knowledge—from marginal collections of research to legitimate scientific specialties—that is missing from the macro-historical accounts focused on crisis, paradigmatic conflict, and revolution. Our work concentrates on review and synthesis as a site of negotiation of the definition and organization of a research area that, along with other forms of discourse like published replies and responses, may be integral to the meso-level process by which fields and their contents evolve and change rather than revolutionize.

Still, legitimation of knowledge requires translation of its principal claims and accepted theoretical stances into a more general framework. Novel domains are initially messy, full of contradictions, confusion, and exploration. Not until an area of knowledge has been schematized into a coherent and simplified framework will it enjoy acceptance in the wider community. Research specialties must make sense within existing disciplinary logics to be recognized within a wider scholarly context. This type of sense-making is the aim of all review articles, but our claim is that formal, invited Annual Review articles provide a distinctly authoritative source for this type of schematization. The quantitative analyses we describe compare Annual Review publications to non-AR articles that take the same summarizing form. The results indicate there is something distinctive about the authority conveyed by an Annual Review. The specific mechanism of this authority is not entirely clear—it could be the simple notoriety of AR journals, or it could be the high prestige of the authors they recruit.[34] But it is clear from our analysis that this authority allows the formal reviews to exercise curatorial control over the future direction a scholarly field takes.

It is not enough to simply publish an article that aims to summarize the important work in an emerging specialization. Reviews published in an Annual Review journal present a distinct form of legitimized scholarly knowledge that allows them not just to observe an evolving subfield, but to alter its course. They essentialize the small-scale conversations that constitute a specialty, ignoring the minutia that have little relevance to discipline-wide discourse and painting with broad strokes the different camps engaged with the reviewed topic. This erasure of detail is necessary to draw those camps into relation with one another as a singular whole. These types of restructuring, essential to the constitution of a legitimate research domain, are precisely those observed in the analyses we present here. The changes to scholarly discourse that accompany AR reviews create a constellation of simplified topics, tracing connections between them through exemplary publications and casting distinct bodies of work as a holistic gestalt.

But the sense-making entailed in scholarly reviews is more than simple translation. Novel areas of research are characterized by a lack of consensus not only on formal findings, but on their identity as a research specialty (Hill and Carley 1999). Much of the work of establishing a specialty lies in negotiating the definition of the underlying norms of discourse—a *story* with which to frame the conversation (Goffman 1974; Morrill and Owen-Smith 2002; Sinding 1996). Because of this, Annual Review articles occupy a privileged position in a newly formed area of research. By externalizing one particular definition of a scholarly situation and consecrating it in a published review, their particular form of academic discourse has disproportionate social influence over ensuing works that relate to the specialty. Review articles' interpretations set an agenda for the future of scholarly domains (Myers 1991). In so doing, reviews do not just define what was, but what could be and will be.

Our findings offer a counterpoint to the view of scientific development as a process of conflict, challenge, and supplanting of paradigmatic frameworks. Sense-making

efforts are ubiquitous in published scientific discourse. Researchers are constantly engaging in synthesis, attempting to make sense of observed scholarly developments by positioning them relative to other work and thus (re)defining some part of the field. In this way, science engages in a process of continuous self-reflection.

Formal review articles, although they occupy a niche in the domain of academic publishing, are ideal illustrations of such synthetic processes of knowledge creation. We have shown that formal reviews are much more than simple summaries of scientific subfields. By curating the published research in an area, reviews highlight certain connections between publications while obscuring others, dramatically simplifying a domain of knowledge. They focus scholarly attention around a few key publications and the relations between them at the expense of the broad majority of the research in a domain. Upon inclusion in a review article, the seminal research in a domain is apt to become forgotten, replaced by work that drew connections between existing ideas rather than generating new ones. This suggests the substance of scientific progression may be located somewhere between revolutionary shifts in paradigms of thought at one extreme and the ordinary science of cumulative advancement at the other. The synthetic work that is foundational to scientific discovery is an ongoing process of redefined frames imposed through micro-erasure. This continual, destructive restructuring of discourse constitutes the churning substrate on which significant swaths of knowledge are created.

APPENDIX

Table A1. Replications of Estimates in Table 3

	5% replication 1		5% replication 2	
	AvgPathLength$^{\Delta}$	**Clustering$^{\Delta}$**	**AvgPathLength$^{\Delta}$**	**Clustering$^{\Delta}$**
(Intercept)	−.013 (−.016, −.010)	−.032 (−.035, −.029)	−.014 (−.016, −.011)	−.029 (−.032, −.026)
Review	−.148 (−.168, −.127)	−.209 (−.229, −.188)	−.148 (−.167, −.128)	−.207 (−.228, −.185)
#Vertices	−.135 (−.138, −.132)	−.026 (−.029, −.023)	−.134 (−.137, −.131)	−.023 (−.026, −.020)
#Vertices^2	.008 (.007, .009)	.006 (.005, .008)	.008 (.007, .009)	.005 (.003, .006)
Density^0	−.106 (−.109, −.104)	−.084 (−.087, −.082)	−.107 (−.109, −.105)	−.082 (−.085, −.080)
(Density^0)^2	.003 (.002, .005)	.028 (.026, .030)	.003 (.002, .005)	.027 (.025, .028)
AvgPathLength^0	−.657 (−.659, −.655)	.001 (−.001, .003)	−.66 (−.662, −.658)	0 (−.002, .002)
Clustering^0	.032 (.030, .034)	−.621 (−.623, −.619)	.031 (.029, .034)	−.622 (−.624, −.620)
Res. Std. Dev.	.78 (.778, .781)	.803 (.802, .805)	.779 (.777, .780)	.801 (.800, .803)

Note: Each replication represents an independent, weighted, 5 percent subsample of the complete dataset.

Acknowledgments

The authors would like to thank Bas Hofstra, Vamshi Krishna, Hjalmar Carlsen, Sebastian Munoz-Najar Galvez, Tamara Gilkes, Taylor LiCausi, Zack Almquist, Daniel Reese, Sanne Smith, Elizabeth Reddy, and James Murphy for their invaluable feedback and assistance.

Funding

This research was funded in part by NSF award #1633036 and SMA-1829240.

ORCID iDs

Peter McMahan https://orcid.org/0000-0003-0996-0465
Daniel A. McFarland https://orcid.org/0000-0002-6805-0798

Notes

1. Meta-analyses represent another potential site of scientific syntheses, but they are distinct from reviews in key ways. Whereas reviews integrate and map out an account for an area of research, meta-analyses frequently seek generalization. These two processes—integration versus generalization—may be quite distinct, in that the former realigns the focal concerns for an area of research, whereas the latter generalizes existing findings within it. Nevertheless, generalization as a discourse act and genre of scientific argumentation is a potentially fruitful place to focus additional studies of these meso-level practices.
2. "Readers of Annual Reviews articles include researchers who want to keep abreast of their field and integrate this information with their own activities; researchers who want an introduction to new fields, with a view to developing an interface between different areas of research; students at all levels who want to gain a thorough understanding of a topic; and business people, journalists, policy makers, practitioners, patients and patient advocates, and other who wish to be informed about developments in research" (Annual Reviews n.d.).
3. The term "creative destruction" mirrors Schumpeter's ([1943] 2010) discussion of recombinative innovation in capitalist economies.
4. Our primary data source, Web of Science, categorizes some articles as reviews, but the methodology is not reliable for rigorous analysis (Harzing 2013).
5. There are 54 total Annual Review journals with 6,495 individual reviews represented in our corpus, spanning disciplines in the biological, physical, and social sciences.
6. © Copyright Clarivate Analytics 2018. All rights reserved.
7. The reviews cite 1,515,737 publications in 10,685 distinct journals. Our sample includes 1,242,599 publications in 1,163 journals, since most journals are cited only once and never reoccur.
8. In light of the mechanics of academic publishing, it is unrealistic that reviews have an immediate effect on citation patterns. We therefore estimated the model in Equation 1 with lags of zero through five years on the main explanatory variable. Each model gave substantively similar results, but we report here the model with a lag of one year because it had a dramatically lower value of the AIC when estimated on a fixed sample.
9. The variable is calculated by adding 1 to the count and then taking its logarithm with base 2. This transformation aids considerably in interpretation of coefficient estimates.
10. Impact factor is calculated with reference to the year 2000 as a fixed measure of a journal's overall influence. Impact factors change over time, but they do not do so dramatically. The impact factor of the journals used in this analysis range from .03 to 72.40 between 1997 and 2016, but the standard deviation of each journal's impact factor across this time has a mean of just .187.
11. More precisely, reviewability is calculated for each review article by enumerating all publications that cite at least one of the same articles the review cites. We then count the number of such co-citations for all articles not cited in the review, discarding counts for articles with fewer than 20 co-citations. An article's cumulative reviewability is the log (base 2) of the sum of all such co-citation counts. This yields results analogous to certain effective citation recommendation systems (e.g., Küçüktunç et al. 2012; Zhou et al. 2008). The variable is standardized to make its effect interpretable.
12. We include reviewability to ensure the appropriateness of the model; however, variations of the model that omit the covariate yield estimates of the same sign and comparable magnitude for the remaining variables.
13. Incorporating article-level random effects for the coefficient on reviews complicates the model significantly but allows for a robust analysis of inter-article variation in the effect of review. One consequence is that the large proportion of articles never cited by an Annual Review can yield no information about the effect of review on those articles, and the restricted maximum-likelihood estimation we use may therefore lead to an *under*-estimate of the effect of reviews overall as the never-reviewed articles pull the average toward zero. Alternative formulations of the model that allow only journal-level variation in β_{3ij} yield an estimate for the average effect of review of approximately –.40, which is similar to the effect in the current model averaged across the subset of articles that received at least one Annual Review citation.
14. We estimate this model in R using maximum likelihood and the glmmTMB package (Brooks et al. 2017) as Monte Carlo methods were computationally prohibitive.

15. This estimate describes an "average" article, before taking any subject-area effects into account.
16. In pilot work on this project, both purely lexical and hybrid lexical–structural approaches endemically identified domains that were either too broad to be associated with a single review article or were associated with only a small subset of the work included in the review. Still, utilization of a more general framework of subfield identification would be a beneficial extension for future research.
17. The network analyses presented here represent edges as unique among pairs of publications rather than multiplex. Each edge is weighted with the count of co-citations between its termini.
18. A potential shortcoming of the reference sets we use is their restriction to articles that were published earlier than the referring article. If new research after a review is published comes to dominate a field, then co-citations within the reference set may diminish as the field focuses on the newer work. However, as we will show, the co-citation networks overwhelmingly become much *more* dense after publication of a review, suggesting the structural changes we observe are not simply artifacts of this exclusion.
19. We follow Newman's (2001a) interpretation of weighted edges in scholarly networks.
20. To obtain an approximation of the expected value of the statistic, we generated a sample of 100 random graphs using Milo and colleagues' (2003) Monte Carlo method, and we calculated the mean of the statistic across every graph in the sample.
21. To account for edge weights, we use the arithmetic mean method to calculate the clustering coefficient, as described in Opsahl and Panzarasa (2009).
22. The use of multivariate regression allows simultaneous estimation of linear models for each of the dependent variables (change in average path length and change in clustering coefficient), while allowing for correlation among error terms between the two models. The results can be interpreted as if each model was run independently, but with more complete error structure.
23. To ensure the robustness of the sampling procedure, we completed the analysis on two additional 5 percent samples. The results, available in Appendix Table A1, are nearly identical to those in Table 3.
24. We estimate multivariate models using Hamiltonian Monte Carlo (No-U-Turn) samplers using the R package brms (Bürkner 2017) for modeling and Stan (Carpenter et al. 2017) for estimation. Both versions of the model are estimated with Norm (0,2) priors for all coefficients, a Cholesky decomposition of the covariance matrix with an LKJ(2) prior (Lewandowski, Kurowicka, and Joe 2009) for correlation, and a half student's t distribution t (3,0,10) for standard deviations. Data and R code to replicate portions of the analysis are available at https://github.com/mcmahanp/reviews_replication.
25. Propensity score–matching creates a quasi-experimental framework in which each "treatment" observation (in this case Annual Review articles) is matched with a comparison case that did not receive the treatment, but whose propensity score ($\text{logit}^{-1}(\mu_i)$) is as close as possible to that treatment observation (Ho et al. 2007).
26. In fact, the number of citations is the primary criteria used in the WoS categorization criteria for reviews (Clarivate Analytics n.d.).
27. Manual inspection of the matched samples confirms that the non-Annual Review articles used in the analysis are either overt reviews or more general interventions in a specific field (e.g., "Childhood Predictors of Adult Obesity: A Systematic Review" [Parsons et al. 1999]; "Game Theory and Neural Basis of Social Decision Making" [Lee 2008]; "Social Capital and Economic Development: Toward a Theoretical Synthesis and Policy Framework" [Woolcock 1998]).
28. To determine the scope of the effect described—whether formal reviews truly change the course of a field rather than just the attention of researchers that cite them—we performed a further robustness check. We recalculated the co-citation networks and all associated statistics while ignoring the citations of any article that itself cited the focal article. This change limits the co-citation networks to the perspective of articles that are not directly influenced by the root article itself and therefore measures the non-proximal effect of Annual Reviews. The results are nearly identical to those listed in Table 3, suggesting that formal reviews shape subfields even outside of their direct influence.
29. We use transitivity (local bridging) rather than betweenness (global bridging) for its conceptual appropriateness for co-citation relations and its sensitivity to the prevalence of cliques in co-citation networks. Still, local transitivity and betweenness are strongly correlated in our sample ($\rho = -.36$).
30. Article-level variables are centered and standardized around the mean of their reference set (group-mean centered). The intercept and non-interaction review indicator are thus analytically zero and are omitted from the model.
31. As a further robustness check, we estimated this model using the restricted version of the co-citation networks described earlier, with virtually indistinguishable results.
32. Colored research clusters in Figure 8 are identified using "fast, greedy" approximate modularity maximization (Clauset, Newman, and Moore 2004).
33. As measured with eigenvector centrality: EVCentrality[1].
34. An examination of the distinctive ways Annual Review articles are read and utilized is outside the scope of this article, but it would be an important topic for future research.

References

Abbott, Andrew Delano. 2001. *Chaos of Disciplines*. Chicago: University of Chicago Press.

Ahuja, Gautam. 2000. "Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study." *Administrative Science Quarterly* 45(3):425–55.

Anderson, Ashton, Daniel A. McFarland, and Dan Jurafsky. 2012. "Towards a Computational History of the ACL: 1980–2008." Pp. 13–21 in *Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries*. Association for Computational Linguistics.

Annual Reviews. n.d. "Editorial Principles and Policies." Retrieved November 20, 2020 (http://www.annualreviews.org/page/authors/editorial-policies).

Barrat, A., M. Barthelemy, R. Pastor-Satorras, and A. Vespignani. 2004. "The Architecture of Complex Weighted Networks." *Proceedings of the National Academy of Sciences* 101(11):3747–52.

Bonacich, Phillip. 1987. "Power and Centrality: A Family of Measures." *American Journal of Sociology* 92(5):1170–82.

Bourdieu, Pierre. 1975. "The Specificity of the Scientific Field and the Social Conditions of the Progress of Reason." *Information (International Social Science Council)* 14(6):19–47.

Bourdieu, Pierre. 1990. *In Other Words: Essays Towards a Reflexive Sociology*. Stanford, CA: Stanford University Press.

Brooks, Mollie E., Kasper Kristensen, Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Mächler, and Benjamin M. Bolker. 2017. "GlmmTMB Balances Speed and Flexibility among Packages for Zero-Inflated Generalized Linear Mixed Modeling." *The R Journal* 9(2):378–400.

Bürkner, Paul-Christian. 2017. "Brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80(1):1–28.

Callon, Michel, Jean-Pierre Courtial, William A. Turner, and Serge Bauin. 1983. "From Translations to Problematic Networks: An Introduction to Co-word Analysis." *Information (International Social Science Council)* 22(2):191–235.

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76(1):1–32.

Chen, Chaomei, Fidelia Ibekwe-SanJuan, and Jianhua Hou. 2010. "The Structure and Dynamics of Co-citation Clusters: A Multiple-Perspective Co-citation Analysis." *Journal of the American Society for Information Science and Technology* 61(7):1386–1409.

Chubin, Daryl E. 1976. "The Conceptualization of Scientific Specialties." *The Sociological Quarterly* 17(4):448–76.

Clarivate Analytics. n.d. "Impact Factor." *Clarivate*. Retrieved July 22, 2019 (https://clarivate.com/essays/impact-factor/).

Clauset, Aaron, M. E. J. Newman, and Cristopher Moore. 2004. "Finding Community Structure in Very Large Networks." ArXiv:Cond-Mat/0408187 (https://doi.org/10.1103/PhysRevE.70.066111).

Collins, H. M. 1998. "The Meaning of Data: Open and Closed Evidential Cultures in the Search for Gravitational Waves." *American Journal of Sociology* 104(2):293–338.

Collins, Randall. 2009. *The Sociology of Philosophies*. Cambridge, MA: Harvard University Press.

Elbert, Donald L., and Jeffrey A. Hubbell. 1996. "Surface Treatments of Polymers for Biocompatibility." *Annual Review of Material Science* 26:365–94.

Ennis, James G. 1992. "The Social Organization of Sociological Knowledge: Modeling the Intersection of Specialties." *American Sociological Review* 57(2):259–65.

Fleming, Lee. 2001. "Recombinant Uncertainty in Technological Search." *Management Science* 47(1):117–32.

Fleming, Lee, Santiago Mingo, and David Chen. 2007. "Collaborative Brokerage, Generative Creativity, and Creative Success." *Administrative Science Quarterly* 52(3):443–75.

Foster, Jacob G., Andrey Rzhetsky, and James A. Evans. 2015. "Tradition and Innovation in Scientists' Research Strategies." *American Sociological Review* 80(5):875–908.

Freeman, Katherine H., and Raymond Jeanloz. 2015. "Mission Statement." *Annual Review of Earth and Planetary Sciences* 43(1).

Frickel, Scott, and Neil Gross. 2005. "A General Theory of Scientific/Intellectual Movements." *American Sociological Review* 70(2):204–32.

Fuchs, Stephan, and Joseph H. Spear. 1999. "The Social Conditions of Cumulation." *The American Sociologist* 30(2):21–40.

Garfield, Eugene. 1977. "The 'Obliteration Phenomenon' in Science—and the Advantage of Being Obliterated!" In *Essays of an Information Scientist*, Vol. 2. Philadelphia: ISI Press.

Garfield, Eugene. 1979. "Is Citation Analysis a Legitimate Evaluation Tool?" *Scientometrics* 1(4):359–75.

Gmür, Markus. 2003. "Co-citation Analysis and the Search for Invisible Colleges: A Methodological Evaluation." *Scientometrics* 57(1):27–57.

Goffman, Erving. 1974. *Frame Analysis; an Essay on the Organization of Experience*. Cambridge, MA: Harvard University Press.

Gondal, Neha. 2011. "The Local and Global Structure of Knowledge Production in an Emergent Research Field: An Exponential Random Graph Analysis." *Social Networks* 33(1):20–30.

Harzing, Anne-Wil. 2013. "Document Categories in the ISI Web of Knowledge: Misunderstanding the Social Sciences?" *Scientometrics* 94(1):23–34.

Hedgecoe, Adam M. 2002. "Reinventing Diabetes: Classification, Division and the Geneticization of Disease." *New Genetics and Society* 21(1):7–27.

Hill, Vanessa, and Kathleen M. Carley. 1999. "An Approach to Identifying Consensus in a Subfield: The Case of Organizational Culture." *Poetics* 27(1):1–30.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15(3):199–236.

Holzner, Burkart. 1972. *Reality Construction in Society*. Rochester, VT: Schenkman Publishing Company.

Jurgens, David, Srijan Kumar, Raine Hoover, Dan McFarland, and Dan Jurafsky. 2018. "Measuring the Evolution of a Scientific Field through Citation Frames." *Transactions of the Association for Computational Linguistics* 6:391–406 (https://doi.org/10.1162/tacl_a_00028).

Ketcham, Catherine M., and James M. Crawford. 2007. "The Impact of Review Articles." *Laboratory Investigation* 87(12):1174–85.

Killewald, Alexandra, Fabian T. Pfeffer, and Jared N. Schachner. 2017. "Wealth Inequality and Accumulation." *Annual Review of Sociology* 43(1):379–404.

Kim, Kyung-Man. 2009. "What Would a Bourdieuan Sociology of Scientific Truth Look Like?" *Social Science Information* 48(1):57–79.

Kolaczyk, Eric D., and Gábor Csárdi. 2014. *Statistical Analysis of Network Data with R*. New York: Springer.

Krampen, Günter, Ralf Becker, Ute Wahner, and Leo Montada. 2007. "On the Validity of Citation Counting in Science Evaluation: Content Analyses of References and Citations in Psychological Publications." *Scientometrics* 71(2):191–202.

Krivitsky, Pavel N., and Mark S. Handcock. 2013. "A Separable Model for Dynamic Networks." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76(1):29–46.

Küçüktunç, Onur, Kamer Kaya, Erik Saule, and Ümit V. Çatalyürek. 2012. "Fast Recommendation on Bibliographic Networks." Pp. 480–87 in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (https://doi.org/10.1109/ASONAM.2012.82).

Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Leahey, Erin, and James Moody. 2014. "Sociological Innovation through Subfield Integration." *Social Currents* 1(3):228–56.

Lee, Daeyeol. 2008. "Game Theory and Neural Basis of Social Decision Making." *Nature Neuroscience* 11:404–9.

Leskovec, Jure, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. 2008. "Statistical Properties of Community Structure in Large Social and Information Networks." Pp. 695–704 in *Proceedings of the 17th International Conference on World Wide Web, WWW '08*. Beijing, China: Association for Computing Machinery.

Lewandowski, Daniel, Dorota Kurowicka, and Harry Joe. 2009. "Generating Random Correlation Matrices Based on Vines and Extended Onion Method." *Journal of Multivariate Analysis* 100(9):1989–2001.

Lievrouw, Leah A. 1992. "Communication, Representation, and Scientific Knowledge: A Conceptual Framework and Case Study." *Knowledge and Policy* 5(1):6–28.

McCain, Katherine W. 1986. "Co-cited Author Mapping as a Valid Representation of Intellectual Structure." *Journal of the American Society for Information Science* 37(3):111–22.

Milo, R., N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon. 2003. "On the Uniform Generation of Random Graphs with Prescribed Degree Sequences." ArXv:Cond-Mat/0312028.

Moody, James. 2004. "The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999." *American Sociological Review* 69(2):213–38.

Morrill, Calvin, and Jason Owen-Smith. 2002. "The Emergence of Environmental Conflict Resolution: Subversive Stories and the Construction of Collective Action Frames and Organizational Fields." Pp. 90–118 in *Organizations, Policy, and the Natural Environment: Institutional and Strategic Perspectives*, edited by A. J. Hoffman and M. Ventresca. Stanford, CA: Stanford University Press.

Morris, Steven A., and Betsy Van der Veer Martens. 2008. "Mapping Research Specialties." *Annual Review of Information Science and Technology* 42(1):213–95.

Mullins, Nicholas C., Lowell L. Hargens, Pamela K. Hecht, and Edward L. Kick. 1977. "The Group Structure of Co-citation Clusters: A Comparative Study." *American Sociological Review* 42(4):552–62.

Munoz-Najar Galvez, Sebastian, Raphael Heiberger, and Daniel A. McFarland. 2020. "Paradigm Wars Revisited: A Cartography of Graduate Research in the Field of Education (1980–2010)." *American Educational Research Journal* 57(2):612–52.

Myers, Greg. 1985. "Texts as Knowledge Claims: The Social Construction of Two Biology Articles." *Social Studies of Science* 15(4):593–630.

Myers, Greg. 1991. "Stories and Styles in Two Molecular Biology Review Articles." In *Textual Dynamics of the Professions: Historical and Contemporary Studies of Writing in Professional Communities*, edited by C. Bazerman and J. G. Paradis. Madison: University of Wisconsin Press.

Newman, M. E. J. 2001a. "Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality." *Physical Review E* 64(1):016132.

Newman, M. E. J. 2001b. "The Structure of Scientific Collaboration Networks." *Proceedings of the National Academy of Sciences* 98(2):404–9.

Newman, M. E. J. 2003. "The Structure and Function of Complex Networks." *SIAM Review* 45(2):167–256.

OECD. 2015. *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*. Paris: OECD Publishing.

Opsahl, Tore, and Pietro Panzarasa. 2009. "Clustering in Weighted Networks." *Social Networks* 31(2):155–63.

Oreskes, Naomi. 2004. "The Scientific Consensus on Climate Change." *Science* 306(5702):1686–6.

Parson, Edward A., and Karen Fisher-Vanden. 1997. "Integrated Assessment Models of Global Climate Change." *Annual Review of Energy and the Environment* 22(1):589–628.

Parsons, T. J., C. Power, S. Logan, and C. D. Summerbell. 1999. "Childhood Predictors of Adult Obesity: A Systematic Review." *International Journal of Obesity* Supplement 8:S1–107.

Price, Derek J. de Solla. 1986. *Little Science, Big Science—and Beyond*. New York: Columbia University Press.

Redner, S. 2004. "Citation Statistics from More Than a Century of Physical Review." ArXiv:Physics/0407137.

Roderick, George K. 1996. "Geographic Structure of Insect Populations: Gene Flow, Phylogeography, and Their Uses." *Annual Review of Entomology* 41(1):325–52.

Rzhetsky, Andrey, Jacob G. Foster, Ian T. Foster, and James A. Evans. 2015. "Choosing Experiments to Accelerate Collective Discovery." *Proceedings of the National Academy of Sciences* 112(47):14569–74.

Schilt, Kristen, and Danya Lagos. 2017. "The Development of Transgender Studies in Sociology." *Annual Review of Sociology* 43(1):425–43.

Schumpeter, Joseph A. [1943] 2010. *Capitalism, Socialism and Democracy*. London, UK: Routledge.

Shwed, Uri, and Peter S. Bearman. 2010. "The Temporal Structure of Scientific Consensus Formation." *American Sociological Review* 75(6):817–40.

Sinding, Christiane. 1996. "Literary Genres and the Construction of Knowledge in Biology: Semantic Shifts and Scientific Change." *Social Studies of Science* 26(1):43–70.

Small, Henry. 1973. "Co-citation in the Scientific Literature: A New Measure of the Relationship between Two Documents." *Journal of the American Society for Information Science* 24(4):265–69.

Small, Henry. 1977. "A Co-citation Model of a Scientific Specialty: A Longitudinal Study of Collagen Research." *Social Studies of Science* 7(2):139–66.

Small, Henry. 1986. "The Synthesis of Specialty Narratives from Co-citation Clusters." *Journal of the American Society for Information Science* 37(3):97–110.

Small, Henry, and Ernest Sweeney. 1985. "Clustering the Science Citation Index® Using Co-citations: I. A Comparison of Methods." *Scientometrics* 7(3–6):391–409.

Snijders, Tom A. B. 2005. "Models for Longitudinal Network Data." Pp. 215–47 in *Models and Methods in Social Network Analysis*, edited by P. J. Carrington, J. Scott, and S. Wasserman. Cambridge, UK: Cambridge University Press.

Star, Susan Leigh, and James R. Griesemer. 1989. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39." *Social Studies of Science* 19(3):387–420.

Stokes, T. D., and J. A. Hartley. 1989. "Coauthorship, Social Structure and Influence within Specialties." *Social Studies of Science* 19(1):101–25.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. "Atypical Combinations and Scientific Impact." *Science* 342(6157):468–72.

Uzzi, Brian, and Jarrett Spiro. 2005. "Collaboration and Creativity: The Small World Problem." *American Journal of Sociology* 111(2):447–504.

Watts, Duncan J., and Steven H. Strogatz. 1998. "Collective Dynamics of 'Small-World' Networks." *Nature* 393(6684):440–42.

White, Howard D., and Belver C. Griffith. 1981. "Author Co-citation: A Literature Measure of Intellectual Structure." *Journal of the American Society for Information Science* 32(3):163–71.

Whitley, Richard D. 1970. "Black Boxism and the Sociology of Science: A Discussion of the Major Developments in the Field." *The Sociological Review* 18(1_suppl):61–92.

Woolcock, Michael. 1998. "Social Capital and Economic Development: Toward a Theoretical Synthesis and Policy Framework." *Theory and Society* 27(2):151–208.

Woolgar, Steve W. 1976. "The Identification and Definition of Scientific Collectivities." Pp. 233–46 in *Perspectives on the Emergence of Scientific Disciplines*, edited by G. Lemaine, R. Macleod, M. Mulkay, and P. Weingart (https://doi.org/10.1515/9783110819038).

Zaidi, Batool, and S. Philip Morgan. 2017. "The Second Demographic Transition Theory: A Review and Appraisal." *Annual Review of Sociology* 43(1):473–92.

Zhou, Ding, Shenghuo Zhu, Kai Yu, Xiaodan Song, Belle L. Tseng, Hongyuan Zha, and C. Lee Giles. 2008. "Learning Multiple Graphs for Document Recommendations." Pp. 141–50 in *Proceedings of the 17th International Conference on World Wide Web, WWW '08*. New York: ACM.

Peter McMahan is an Assistant Professor of Sociology at McGill University. He completed his PhD at the University of Chicago. His work centers on the ways that communication informs and reinforces knowledge and cultural domains. He specializes in statistical modeling and computational methodologies, with particular focus on network analysis and methods for social inference on large, unstructured datasets.

Daniel A. McFarland is a Professor of Education, and by courtesy, Sociology and Business, at Stanford University. He completed his PhD at the University of Chicago. His work broadly concerns social and cultural dynamics, and the structuring of educational systems and their practice. Dan teaches courses in the sociology of science, social network methods, relational sociology, computational sociology, and the sociology of play and games.