

The Temporal Structure of Scientific Consensus Formation

Uri Shwed^a and Peter S. Bearman^b

Abstract

This article engages with problems that are usually opaque: What trajectories do scientific debates assume, when does a scientific community consider a proposition to be a fact, and how can we know that? We develop a strategy for evaluating the state of scientific contestation on issues. The analysis builds from Latour's black box imagery, which we observe in scientific citation networks. We show that as consensus forms, the importance of internal divisions to the overall network structure declines. We consider substantive cases that are now considered facts, such as the carcinogenicity of smoking and the non-carcinogenicity of coffee. We then employ the same analysis to currently contested cases: the suspected carcinogenicity of cellular phones, and the relationship between vaccines and autism. Extracting meaning from the internal structure of scientific knowledge carves a niche for renewed sociological commentary on science, revealing a typology of trajectories that scientific propositions may experience en route to consensus.

Keywords

sociology of science, consensus, black boxing, network analysis, citations

When and how did we become certain that smoking causes cancer, coffee does not, and human activity is producing global climate change? Since the coining of scientific consensus as closure (Pinch and Bijker 1984), various branches of the sociology of science have made great strides in exposing its mechanisms (e.g., Collins 2004; Frickel and Moore 2006; Fujimura 1996; Gieryn 1996; Latour 1987; Shapin 1996; Star and Griesemer 1989). Yet such studies fall short of providing a tool to monitor the formation of closure as consensus among relevant scientists. Existing work in this area is case specific and limited with respect to comparative research. As such, it has yet to develop an analytic typology of possible patterns of consensus formation.

This article offers a quantitative strategy to measure scientific consensus/contestation levels, which enables comparative research and thus extends the generalizability of the sociology of science. Such a measure may prove useful to scholars of innovation, and it provides a new tool for anyone interested in network structures and their outcomes. Applying this strategy to several different cases, we identify three trajectories scientific

^aBen Gurion University

^bColumbia University

Corresponding Author:

Uri Shwed, Department of Sociology and Anthropology, Ben Gurion University of the Negev, PO Box 653, Beer-Sheva 84105 Israel
E-mail: shwed@bgu.ac.il

propositions assume on their way through contestation to consensus among practicing scientists: (1) spiral, in which substantive questions are answered and revisited at a higher level; (2) cyclical, in which similar questions are revisited without stable closure (Abbott 2001); and (3) flat, in which there is no real scientific contestation. This typology emerges from empirical analysis of controversies and offers terms for rapprochement between qualitative and quantitative analyses of science.

Detecting consensus is not trivial, as some level of contestation is always present in science. Scholars dispute previous findings or point to literature gaps to establish their own footholds in the field (Bourdieu 1975; Merton 1973). This everyday, normative level of contestation is benign, particularly compared with debates of epistemic rivalry, in which strongly entrenched camps disagree on core issues.¹ In contexts that harbor extra-scientific interests, such as the hazards of smoking, interested groups may exaggerate normative contestation levels to claim that the extant scientific knowledge is inconclusive (McCright and Dunlap 2000; Proctor and Schiebinger 2008). An important contribution of science and technology studies (STS) is to make such political manipulation more transparent and harder to exercise (Oreskes 2004a) by delivering science to many publics (Collins and Evans 2002; Latour 2004; Moore 1996) and delivering lay voices to science (Epstein 1998; Jasanoff 2004; Rowe and Frewer 2005). However, qualitative engagement with truth, rather than consensus, can only pass judgment after the fact. A third wave in the sociology of science seeks ways of assessing science and promoting informed policy discussions (Collins and Evans 2007; Latour 2009; Weinel 2008). This is the task we embark on here.

The logic of the strategy we propose is rooted in Collins's (1975) "bottled ships" metaphor, which Latour's conception of black boxing elaborates: When a proposition is stable (i.e., a consensual scientific proposition or

a functioning machine), its internal elements are concealed. While a proposition is in the making and still contested, however, the interactions between its internal elements are visible. We export this insight to the macrostructure of scientific citation networks and employ a network community-detection algorithm (Leicht and Newman 2008) to evaluate the degree of internal divisions in scientific literatures. The analysis allows us to distinguish epistemic rivalries from benign contestation: When different factions debate a scientific issue, they create distinct regions in a citation network. In epistemic rivalries, such network regions are a defining characteristic of a network's structure. When consensus over an issue emerges (and only benign contestation remains), the salience of these regions to the overall network structure diminishes. To account for temporality, we develop a novel approach to answer Pickering's (1993) critique that quantitative analyses are inherently *ex-post* in their relation to science. We unfold the emergent temporality of scientific debates by identifying a new property of such disputes: the relevant temporal length of scholarly interaction. By capturing the temporality of consensus formation, we discover that scientific literatures can follow three trajectories—spiral, cyclical, or flat.

We apply our analysis strategy to four cases in which experts' reports identify the timing of consensus accomplishment: the carcinogenicity of smoking, solar radiation, and coffee, as well as anthropogenic climate change.² A fifth validating case is the controversy regarding gravitational waves, which Harry Collins (e.g., 2004) studied extensively. Our analysis reveals consensus earlier than do expert evaluations. Having validated our measure compared with existing authoritative mechanisms, we then apply the method to two additional cases of current relevance: Do cellular phones cause cancer, and do vaccinations cause autism? We find that there is no real contestation within the scientific community about these questions.

EXISTING LITERATURE

Studying scientific consensus and closure is a major focus of recent work in the sociology of science. Different streams within the sociology of science expose how scientific consensus is not a direct consequence of new findings but is shaped by extra-scientific factors such as culture (Hess 1995), power and funding (Martin and Richards 1995), politics (Epstein 1998), and personal credibility (Leahey 2005; Shapin 1994). In addition, consensus results from social processes within the core set of practicing scientists who negotiate results (Collins 1974, 1992, 2004), demarcate knowledge claims (Gieryn 1999; Wynne 1996), construct boundary objects to conceal conflict (Star and Griesemer 1989), employ micro-politics of translation (Latour 1987), and fortify bandwagon practices (Fujimura 1996). A next step is to develop a strategy that allows for easy comparison across cases.

Generally speaking, scholars identify consensus by immersing themselves in a cognitive scientific domain and then report their conclusions on the status of the field. Studies of scientific consensus thus leave its detection entirely in the hands of experts, be they practitioners of the issue under scrutiny or expert sociologists of science. Oreskes's (2004b) report of consensus on climate change offers one influential example. However, when deep understanding of each case is required merely to assess its consensus level, comparative research that sorts and qualifies the plethora of consensus forming mechanisms becomes difficult. The sociology of expertise suggests moving beyond expert filtration of knowledge by developing general rules to assess different expert claims (Collins and Evans 2007; Weinel 2008). In this article, we provide such a strategy using scientific products as data.

Sociologists of science used to quantify consensus either by asking practitioners about it (Biglan 1973; Hargens and Kellywilson 1994), by journal rejection rates (Hargens 1988; Hargens and Kellywilson 1994), by

(inverse) lengths of published abstracts, by cohesiveness of graduate training (Ashar and Shapiro 1990), or by agreement between different reviewers of grant proposals (Cole 1983). These studies examine consensus levels in disciplines. Disciplines, however, are not the arenas of scientific progress (Cole 1983; Kuhn 1970). While detecting consensus in disciplines answers an interesting question, it is not the central concern here. Science advances around sub- and multi-disciplinary puzzles, as evident by the breadth of contributors to the report of the Intergovernmental Panel on Climate Change (IPCC 2007).

Quantitative attempts at measuring scientific consensus on issues, rather than disciplines, are rare (Cole and Zuckerman 1975). Following their lead, Evans (2007a) recently analyzed discursive-consensus formation on subdebates (which might proxy the substantive consensus we investigate). Overall, however, Cole and Zuckerman's effort ended prematurely (Wray 2005). Before finding a meaningful measure of scientific consensus, sociologists stopped searching.

This abandonment of the search for a consensus measure followed the emergence of the sociology of scientific knowledge (SSK), which diverted attention from classic, Mertonian sociology of science (Zuckerman 1988). SSK mastered a qualitative gaze into scientific knowledge. Classic sociologists of science sought a consensus measure because they were reluctant to engage with scientific content. SSK and later scholars abandoned the search for a consensus measure because they had no reason to short circuit their engagement with domain knowledge. In doing so, SSK gave up on the manipulation of scale that makes comparative studies important and influential.

Recent works from organizational theory and social movement research offer a systematic examination of the dynamics of scientific claims and fields, as well as their effects on society (see Frickel and Gross 2005; Frickel and Moore 2006; Owen-Smith and Powell 2008). Such studies abandon the distinction

between external and internal influences on science, revealing both as crucial. Science operates in a co-constitutive environment of organizations and networks that is shaped by social, technical, and economic changes (Powell et al. 2005; Smith-Doerr, Manev, and Rizova 2004), as well as by funding (Evans 2007b; Hess 2006), geographical embeddedness (Whittington, Owen-Smith, and Powell 2009), and a host of micro strategies (Powell and Colyvas 2008). Therefore, this tradition may benefit from our conceptualization of consensus.

In this article, we supplement such investigations of “why we know” with a comparative framework to examine “when we know.” We show that it is possible to time consensus and to compare it across cases without demanding domain expertise of the analyst. The focus on timing opens new questions for organizational theory and the sociology of science. For science, the simplest of these—and perhaps the most important—is what do the trajectories of scientific propositions look like?

BLACK BOXING FACTS

How do we know when something is a fact? Oreskes (2004b) asserts that science should be evaluated by its inscriptions—that is, when an entire scientific literature agrees on something, we can treat it as a fact. Making such a claim requires domain-specific expertise. This demand hinders comparative research and introduces potential biases. Oreskes, for example, codes climate change publications into groups characterized as supportive, skeptical, and indifferent. It is not surprising that skeptics challenged her selection and coding schemes. By identifying a structural measure of consensus, we minimize experts’ discretion. Actor Network Theory (ANT) shares this goal.

Our theoretical model can be traced back to Collins’s (1974) idea that facts are like ships in bottles and we should study these ships/facts as they are being built; that is, as

a core set of practicing scientists transform several possible answers to a question into one correct and several erroneous answers (Collins 2004), where correct and erroneous reflect only an agreement in the scientific community at a given time, not some transcendental truth. Latour (1999) elaborated this notion into the broad ANT concept of black boxing and detailed what we may find in its formation process: A black box is an apparatus that conceals its internal elements, which are viewed only through inputs and outputs (Latour 1987, 1999). A working computer is a black box with a keyboard (for inputs) and a screen (for output). Only when the computer malfunctions, or as it is assembled, can we see that it is really a network, tying together chips, magnets, service providers, and so on. Similarly, the proposition “smoking causes cancer,” stated today, needs no proof; it works and is tied into a vast network of persons, substances, studies, and policies. The entire epoch (of statistical inference, tragic deaths, chemical processes, and genetics) that once showed that smoking causes cancer is concealed. Its elaboration is no longer required because the proposition is connected to every cigarette carton and life insurance application. Its internal elements (e.g., chemicals and statistics) already work together, so their connections do not require explication. Consensus formation is a black-boxing process: the weaving together of multiple elements of scientific propositions until their internal divisions are well hidden.

We can observe black boxing in citation networks, or more precisely, in representations of scientific papers connected by citations. Empirical and theoretical work suggests that citations most often signal agreement.³ This probabilistic property induces identifiable areas in the citation network characterized by denser interactions—what can be identified as network communities—even when one is blind to citation type (e.g., favorable, opposing, or ceremonial). Consequently, the network structure that emerges from citation networks of contentious

literatures is characterized by relatively segregated communities. Of course, communities are not completely segregated, as not all citations are favorable. Yet both Merton's and Latour's theories, as well as Hanney and colleagues' (2005) findings, suggest that contentious literatures should exhibit a salient community structure. It follows that the salience of communities to a network's typology is measurable; a reduction in community salience of citation networks over time should point to consensus formation in the literature.

In this article, our empirical task is to consider this proposition. After empirically showing that community salience measures consensus, we can identify patterns of consensus formation. The many insights on micro processes of consensus formation reviewed earlier provide little help for identifying the macro patterns of its formation over time. Mulkay, Gilbert, and Woolgar (1975) argue that there are three stages in the lives of problem areas: exploration, in which unconnected scientists explore a new problem independently; followed by unification, in which the explorers become the leaders of a unified, exponentially growing field; and finally decline/displacement, caused by the institutionalization of the field, which restricts new discovery. In passing, Mulkay and colleagues argue that the unification stage may lead to a redefinition of the field, rather than to its decline or displacement. Our findings show that in Khunian normal science controversies, such as the carcinogenicity of solar radiation, Mulkay and colleagues' model fits well. We call this a spiral pattern, in which many new questions pop up following the unification stage. A different pattern is evident for public controversies that are not really controversial among scientists, such as the carcinogenicity of coffee or the debate over autism and vaccines. We call these controversies flat because they exhibit the same exponential growth of papers but with flat (and low) contestation levels. Finally, scientific controversies such as the carcinogenicity of tobacco exhibit the redefinition of the field mentioned by Mulkay

and colleagues in passing. We call these controversies cyclical because they reveal how consensus forms, is destroyed, and is rebuilt again around different formulations of similar substantive questions.

CONSENSUS AND COMMUNITY

The strategy we develop in this article diverges from previous work by focusing on papers without disciplinary boundaries. Most previous studies that attempt to measure consensus deal with whole disciplines, and the few exceptions focus on a qualitative selection of a core set of scientists (i.e., authors) that compose a sub-disciplinarian cognitive domain (e.g., Cole and Zuckerman 1975; Collins 1974). By using papers as our focal units (see elaboration below and in the online supplement [<http://asr.sagepub.com/supplemental>]), we avoid reducing pieces of knowledge into their authors' dynamics or institutions⁴ as we observe how the products of science—peer-reviewed papers—obtain verisimilitude (Latour 1999) and become building blocks of a single proposition. Measures focusing on author degree or expert opinions are problematic because they select on authors, leading to loss of important information. If, for example, a crucial step in the black boxing of “smoking causes cancer” was a statistical innovation, simply following cancer scholars would never reveal it.

Although it seems anti-intuitive, brief reflection indicates that when all papers about a subject are black-boxed together, their network structure is not defined by the presence of disjoint communities. When papers promote the same views⁵ and cite the same sources, the science behind them is conclusive. It may turn out that this science was wrong, or that published consensus was a result of fiat, but regardless of the reasons behind consensus, the community structure of research literatures' citation networks can reveal if the

scientific questions that produced them are black boxed or contentious. Approaching the problem from this perspective allows us to develop a strategy for observing consensus formation in scientific literatures, without reliance on mediating experts' interpretations.

METHODS

Since Price (1965) suggested that the degree distribution of citations could point to important papers and journals, network analysis has become prominent in evaluating journals' importance (Garfield 1972) and inducing mappings of science (Moody 2004). Generally speaking, the strategies reflected in the network literature on citations take for granted predetermined categories (e.g., disciplines or journals) and restrict analysis to a predetermined subset of the literature. This is unfortunate.

As Figure 1 shows, citation networks are too complex to reveal anything by simple observation. The tools scholars have used to extract meaning from such graphs demand reduction and simplification, for example by removing infrequently cited papers (Small 2006) or predetermining the sample of authors (Collins 2004). Such automatic deletion distorts network measures and gives citation indexes a critical level of importance. This assumption is not self-evident. If, for example, all papers citing a specific paper in a network were never cited themselves, deleting them would hide this important finding. Our strategy also involves data reduction, but one that is data driven and analytic. We model the internal structure of citation networks to reveal consensus without classifying papers or authors into membership of different camps. Our measure enables agnosticism toward papers' content because we extract meaning—that is, the contestation level of scientific debates—from the structure of the networks indexed by their organic community structure.

Modularity as a Measure of Scientific Consensus

In network terms, a community is a subset of a larger population where internal ties are more prevalent than ties to other subsets. In a network of asphalt roads, for example, communities are villages, cities, and states. In a network of scientific papers linked by citations, communities are groups of papers that deal with the same issues and cite each other. Papers that agree are likely to cite each other much more than their protagonists (Hanney et al. 2005), giving rise to communities of agreement. The simple intuition underlying our strategy is that when different communities are salient to the global structure, the field is contentious.

It follows that changes in a citation network's community structure represent changes in consensus levels on an issue: Contentious networks are well defined by communities, and consensual networks are not. Consensus formation exhibits a decline in community salience; the literature produces a common, core community and many miniscule communities (e.g., in the case of smoking and cancer, miniscule communities are populated by studies that retain smoking as a control variable when studying, say, the hazards of solar radiation). A consensual literature set is black boxed, and its internal divisions carry little structural meaning. Statistically, we measure this as the amount of information that communities carry regarding a network's structure. In the current context, as a domain gains consensus, its citation network's community salience declines.⁶

What does community salience mean, and how does one find communities? Different methods are suitable for different cases (Reichardt 2009). Recently, Newman (2006) introduced a method for partitioning a network into communities by maximizing modularity. For a given network division, modularity compares the odds of within-community ties with these odds after a random rewiring of the network. If a division does not include

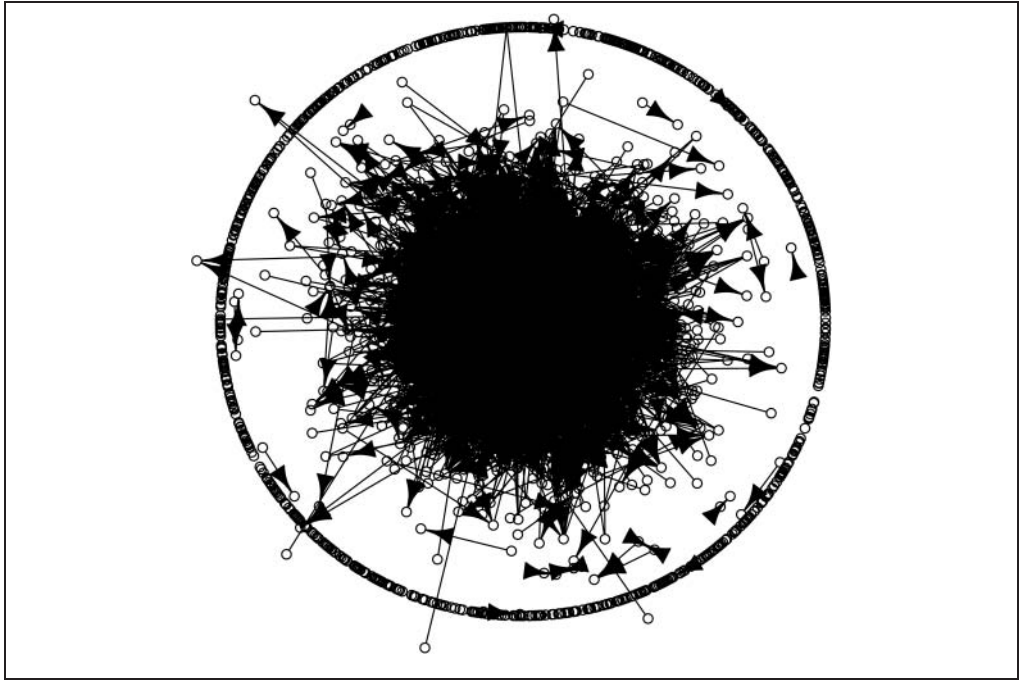


Figure 1. Citation Network of 4,276 Papers about Smoking and Cancer, 1920 to 1995

Note: The outside ring is populated by 906 isolated papers that do not cite other papers and were not cited by 1995. Most of the network is connected in a large crowded component. Different graphic algorithms may draw this network differently, but it remains hard to extract meaning from such representations.

more within-community ties than it would with random ties, it is an artifact of individual-level properties (i.e., degree distribution) and harbors no further information about the structure. The division's modularity in such a case is 0. Modularity, then, measures the salience of communities for a given division. Maximizing this property is one way to get a division.

Our focus is on the dynamics of community salience, so the partition that maximizes this property is appropriate for our purposes. Figure 2 presents some simple networks and their modularity scores and shows how meaningful internal groupings—groupings that are not defined by a node's properties—increase modularity.

Modularity maximization algorithms identify an important network property: the

maximal amount of information that groups carry about a network. We argue that maximal modularity—that is, modularity of the division obtained by modularity maximization—indexes community salience. If no partition of a network reveals much about it, communities are not salient. Black boxing suggests that community salience—the importance of communities to the macro structure—is highest when a proposition is combating objection and lowest when it is consensual fact. When consensus on an issue arises and contestation levels decrease, modularity scores decrease too—which is what we observe. The contestation we wish to reveal, however, is historically-specific epistemic rivalries. Modularity does not distinguish between these and benign contestation. The course of professionalization in science, regardless of consensus, also

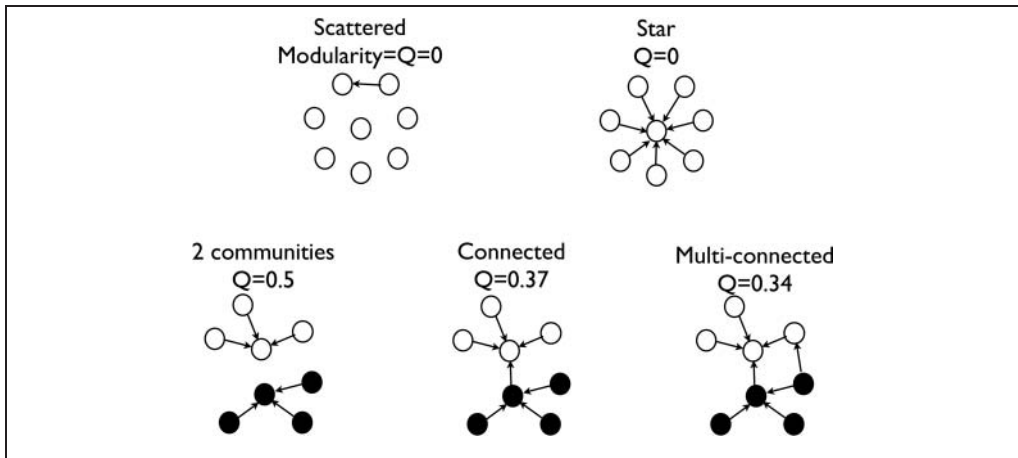


Figure 2. Modularity of Five Artificial Networks of Eight Nodes

Note: Modularity is 0 for the two cases in the top panel. While the networks are very different, they have in common the fact that ties are completely dependent on individual properties. In the top panel, random rewiring would only reproduce the same network, and thus communities contribute no information about the network structure. When all ties in a network are within communities, as in the bottom-left network, modularity is high—random rewiring would allocate half of the ties in this case between communities, and the original state has all of them within communities, so modularity is .5. As more between-community ties are introduced in the original state (in the two remaining examples) modularity declines.

creates salient network communities that modularity detects. Such benign contestation is a product of network size and reflects scientists' struggles to establish their own niches in growing literatures. To discuss epistemic rivalries, we therefore scale raw modularity scores with a literature's size (see the online supplement).

DATA

So far, we have outlined a theoretical concept for thinking about consensus and a way to express it quantitatively. To show that our concept actually measures consensus, we consider five different cases and compare them with expert reports made in real time. Two of the cases we selected because they pertain to scientific issues that were once contentious but became consensual: smoking and cancer, and anthropogenic climate change. We supplement these with two cases that were historically less contentious—the

carcinogenicity of solar radiation and coffee (this was a case of consensus on a null finding, as the scientific community quickly exonerated coffee from suspicions of carcinogenicity). Our fifth case provides an iconic example: the claim of gravitational waves, which has had an impressive history in SSK. These five cases validate our strategy across different periods, contexts, and scales. We then apply our strategy to the literatures about the possible carcinogenicity of cellular phones and the possible relationship between vaccinations and autism, both of which lack an authoritative expert report. This analysis suggests that scientists agree that vaccinations do not cause autism and that there is consensus on the inconclusiveness of science on cellular radiation. Media reports of these issues overrepresent minority views (Boykoff and Boykoff 2004).

Using keywords, we define our cases by their cognitive domain rather than select on authors. For each case, we use specific

keywords that define a cognitive domain to extract a comprehensive set of all papers indexed by ISI Web of Science. We selected keywords with the aim of including every relevant paper. For example, the keywords used to construct the “smoking and cancer” dataset are (Smoking OR nicotine OR cigar* OR Tobacco) AND (cancer OR carci*).⁷ Any paper about cancer *and* smoking is included. We limit the data to articles and reviews, forming a comprehensive set of peer-reviewed scientific works on a subject. Table 1 presents general properties of these datasets.

As noted earlier, defining a cognitive domain by sampling core-set authors (e.g., Cole and Zukerman 1975; Collins 2004) or a broader set of contributing scientists (Weinell 2008) may omit important parts of the domain. Our approach, on the other hand, is exposed to the danger of being over inclusive. Selecting only on keywords, we let in some irrelevant papers (e.g., Wallace [1994] discusses carcinogenicity risk evaluation in pesticides, noting in the abstract that it is insignificant compared with tobacco). Our sense is that inclusion is superior to the risk of exclusion because our method is robust to noise. Theoretically, no other criteria can define a cognitive domain with more accuracy than the terms used by the papers (indeed, Wallace’s paper takes for granted tobacco’s carcinogenicity, maintaining the consensus). If a paper is irrelevant, it will not connect to other relevant papers and will have no effect on the modularity score. Our approach is not immune to deliberate manipulation (e.g., constructing a literature with the keywords “baby” and “murder” and reporting it as a set about abortions), but no data collection method is. Our approach, however, is easy to assess through evaluation of keywords. Furthermore, a sensitivity analysis shows that (honest) changes in keyword selection do not change the results (see the online supplement).

Using Garfield’s *HistCite*[®] software, we generated a graph representation of the data and further modified it in R (Csardi and

Nepusz 2006; R Development Core Team 2008) to account for temporality, as described below. We then evaluated the salience of community structure over time using the Leicht and Newman (2008) algorithm, which adds directionality to Newman’s (2006) algorithm.

MODELING TIME

Our effort is inherently historical; we try to answer Pickering’s (1993) call to understand science in its temporal unfolding. We seek the critical years in which, and the dynamic patterns by which, propositions were black boxed from contentious to consensual literatures. Our modeling of time is critical; we need to be sensitive to new developments without neglecting old papers that remain relevant. Our method relies on published papers, which produces a latency period from the moment a discovery is made to its journal publication.

How can we define scientific knowledge at a given moment? Observing only the latest research severs ties to old papers, while observing all previous research greatly extends the latency period. Common strategies are a cumulative approach (e.g., Leicht et al. 2007), a cross-sectional method (e.g., Cole and Zuckerman 1975), or a moving window strategy (e.g., Small 2006) that uses sliding, fixed-width observation windows. The latter two methods ignore citations to older papers and require an analyst to predetermine either discrete periods or a uniform observation width. By predetermining these properties, an analyst imposes an ex-post view of the field (Pickering 1993). Moreover, these strategies ignore the fact that some papers are more reachable than others, and that this difference in accessibility is itself time variant (Evans 2008). To properly account for temporal unfolding, we need a mechanism to model window width for each point in time.

A meaningful observation period improves the moving window approach by determining

Table 1. Propositions and Datasets

Case	Validation Cases				Current Controversies		
	Coffee is Cancerous	Solar Radiation is Cancerous	Gravitational Waves	Smoking is Cancerous	Anthropogenic Climate Change	Vaccinations Cause Autism	Mobile Phones are Cancerous
Articles (nodes)	1,544	4,009	5,113	8,872	9,432	245	622
Citations (edges)	5,569	17,190	26,413	38,920	30,478	1,358	4,505
Time Span	'39-'08	'14-'05	'13-'09	'20-'00	'75-'08	'76-'09	'93-'09
Search String	(Coffee OR caffeine) AND (Cancer OR carci*)	(Sun OR solar OR photo OR "UV radiation" OR tanning) AND (cancer OR carci* OR melanoma OR sarcoma OR basal)	"Gravitational waves" OR "gravitational radiation" (see Collins 2004:1826)	(Tobacco OR cigar* OR nicotine OR smoking) AND (cancer or carci*)	(Global Warming OR climat*) AND (Greenhouse effect OR Greenhouse gas* OR anthropogenic)	Autis* AND Vaccin*	(Cellular OR mobile) AND (Phone* or telephone*) AND (cancer OR Carci* OR tumor OR brain)
Top Cited; Year	3,348; 1991	970; 2001	1,256; 1962	2,336; 1991	1,246; 1999	549; 1998	182; 2001
Median Citations	15	14	7	21	6	8	7

window width from the changing structure of citations. We call this the dynamic window approach. For each year Y , we note a distribution of citation-ages, defined as the difference between Y and the year of publication of each paper cited in Y . The median of this distribution serves as window width for year Y .⁸ We then define every paper published within the width years from Y as a focal paper, relevant for the year at the end of the window. We include older papers that are still cited by any focal paper to keep influential papers in our analysis, regardless of their age; we do not include the papers cited by these older papers. For example, of all citations made in papers about smoking hazards published in 1987, the median citation-age is 4. The dynamic network for 1987, then, contains all papers published from 1984 to 1987, and all the papers they cited, regardless of year. This procedure is superior to cross-sectional and fixed-width window approaches because its observation window is theoretically justified and sensitive to the varying time frames of scientific activity.

To demonstrate the advantage of dynamic windows over the cumulative approach, we mimic Leicht and colleagues (2007). They calculated authority scores⁹ for each court ruling (paper), pointing to its importance in the network (Kleinberg 1999), and plotted the mean age of top authorities over time. For each year Y , we calculate the mean difference between Y and the publication years of the top 15 authorities, derived from the cumulative network and from our dynamic approach. In years that the set of top authorities is unchanged, mean age increases by one. A smaller increase, or a decline, signifies that new papers became authoritative. If the mean age rises by more than one, older papers that were not authoritative are rediscovered as such.

Figure 3 displays this analysis for cumulative and dynamic windows in two cases—the carcinogenicity of tobacco and coffee. In the top panels, triangles represent the cumulative approach, and circles represent our dynamic

windows. The Y-axis reports mean age of the set of top authorities. The bottom panels simply show the slopes of the top panels, to highlight changes. Dark bars signal the cumulative approach, and clear bars represent dynamic windows. In both cases, the cumulative networks (represented by triangles) reveal early lock-in on a set of authorities.¹⁰

Relying on cumulative networks, one would conclude that no major shifts occurred in the research on coffee and cancer since 1984, and in research on smoking and cancer since 1953. The latter, of course, is false. Dynamic window networks, represented by circles and empty bars, tell a different story: These networks reveal critical points in which the set of top authorities *as discussed in a specific year* change. With respect to smoking, dynamic windows show an exuberant literature, evident by the changing slopes in the empty bars. Here we see that the set of top authorities was stable in only 3 of 53 years. This is not an artifact of the modeling structure. In sluggish literatures, like the carcinogenicity of coffee, the dynamic windows approach reports similar results to the cumulative approach. Our modeling also produces a relevant variable for future studies—the real-time meaningful observation period—and thus answers the critique that quantitative analyses of science are inherently ex-post. They need not be.

RESULTS

We test whether scientific consensus formation is observable as a reduction in the community salience of propositions' citation networks, measured by the network modularity score, scaled for size. When communities are no longer a defining characteristic of a citation network, the network works together as a black box (compared to a past state of contestation). To examine this idea, we present scaled modularity dynamics of five sets of scientific literatures, comparing modularity drops to expert reports¹¹ for

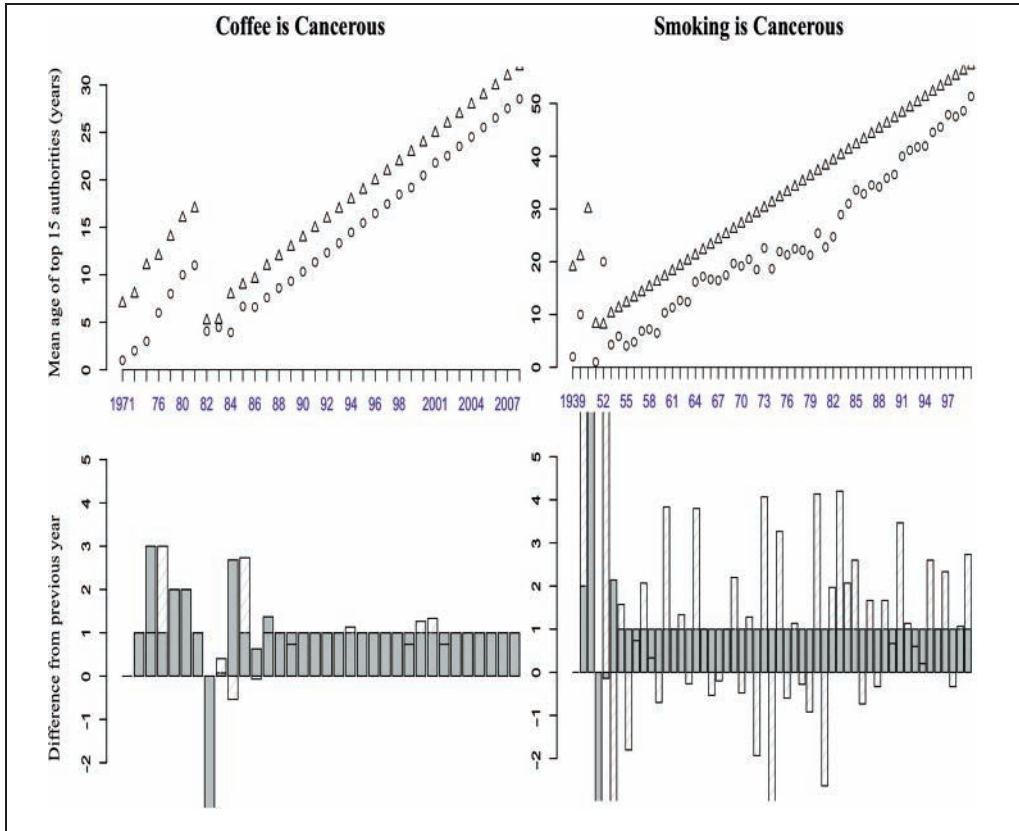


Figure 3. Top 15 Authorities across Time: Cumulative and Dynamic Time Frames

Note: The top panels plot the mean age of the top-15 ranking papers on Kleinberg (1999) authority scores. Triangles represent the cumulative approach, and circles represent dynamic windows. The bottom panels show the slopes of the top figures, with shaded bars for cumulative networks and empty bars for windowed networks. When the top authorities are fixed, each passing year increases their mean age by one year, and the slope is one. A naive cumulative conception of passing time produces the appearance of lock-in on leaders once a network is larger than the set size (15). Dynamic moving windows that capture papers relevant to a specific period (represented by circles and empty slope bars) reveal the difference between the cases. The literature on smoking and cancer shows many turning points, likely reversing significant contestation. The research on coffee and cancer is stable over time, maintaining its consensus.

calibration. We use monographs of the International Agency for Research on Cancer (IARC), an agency of the World Health Organization, as indicators of scientific consensus regarding suspected carcinogens.¹² Likewise, the Intergovernmental Panel on Climate Change (IPCC) supplies a consensus indicator regarding climate change. For controversies surrounding gravitational waves,

we rely on Collins's (2004) extensive analysis of the field.

We start with a simple case to illustrate the analysis framework: the proposition, rejected by experts, that coffee causes cancer. This issue was never really contested, and we expect modularity to be low throughout the proposition's history. The other cases examine modularity trends vis-à-vis experts' consensus

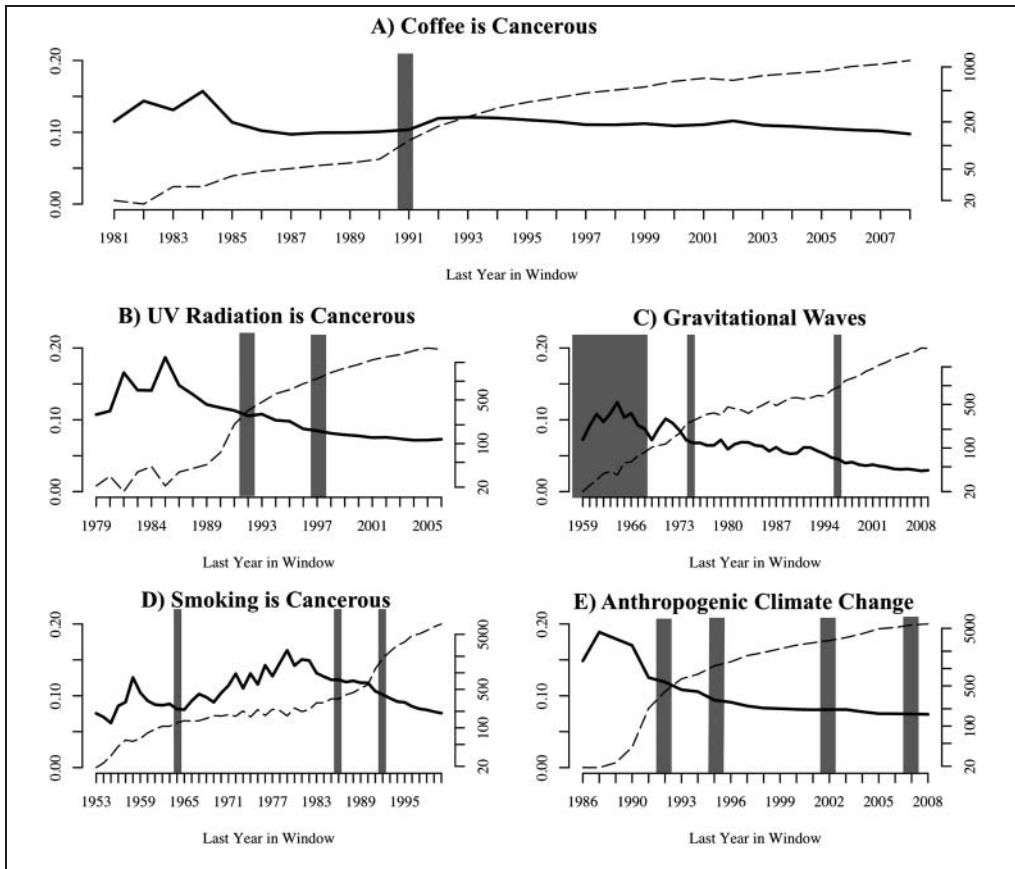


Figure 4. Epistemic Rivalry, Size, and Expert Reports in Five Validating Cases
Note: The dashed line refers to the number of papers in the dynamic window and to the logarithmic right-hand-side-axis. The solid line refers to the level of epistemic rivalry, estimated as the modularity score scaled for logged network size, on the left-hand-side-axis. The bars show years in which critical expert committees published a consensus report, or, in panel C, the years Collins identifies as marking the end of controversy and the emergence of consensus.

declarations: carcinogenicity of solar radiation (IARC 1992), gravitational waves (Collins 2004), tobacco’s carcinogenicity (IARC 1986; U.S. Surgeon General 1964), and anthropogenic climate change (IPCC 2007). Figure 4 presents modularity analyses, noting the timing of experts’ consensus declarations vis-à-vis our analysis.

Note that modularity trends are driven neither by time (or any time-dependent process, such as online archiving) nor by the number of papers in a window (N).¹³ This is evident in the simple case of coffee and cancer.

Coffee is not cancerous; this was never hotly debated. We thus expect the figure to show a flat pattern of consensus formation.

The solid line in Panel A presents modularity scores for the coffee and cancer literature with the number of papers represented by the dashed line. Except for a steep decline between 1984 and 1987, as the literature grew to more than 30 papers, the trend is stable and hovers around .1, even as the number of papers increases. In 1991, the IARC lumped coffee with several other drinks, announcing they are not carcinogenic (IARC

1991). The trend and level show no epistemic rivalry since 1985 and are driven by neither time nor *N*.

Panel B considers the proposition that solar radiation causes cancer. At the outset, this case seems to represent normal science—scholars find that the sun causes cancer and then disseminate that knowledge. We expect modularity trends to follow what we call a spiral—some initial epistemic rivalry is quickly resolved, and scholars then move to secondary questions. This leads to increasing numbers of papers linked to a common core, keeping modularity low. The history of skin cancer research confirms this interpretation. In the early 1980s, this literature was contentious and its network structure was well defined by communities, with scaled modularity fluctuating around .15 with a peak of .19 in 1985. Subsequently, modularity dropped, following a large-scale study relating melanoma to sun exposure (Elwood et al. 1985). By 1992, scaled modularity levels dropped to .1. We view such a significant decline over several years as consensus formation. That year, the IARC published its first report on solar radiation, stating sufficient evidence for carcinogenicity. In 1997, the IARC updated its report with evidence of carcinogenicity of tanning lamps. Note that marked drops in modularity preceded both expert reports.

The case of gravitational waves further validates modularity as a consensus index. Scholars of gravitational waves debate whether and which of their observation tools may detect the waves of gravitational energy emitted from distant astronomical events. This case, recorded by Collins (2004), allows one to track periods of relative consensus or contestation. Our analysis shows (see Panel C) that the history of gravitational waves had three periods of declining epistemic rivalry, in which consensus was formed: 1966 to 1969, 1970 to 1976, and 1992 to 1997, each marked by a steep and consistent decrease in modularity scores. The earlier periods are followed by a significant rise in scaled modularity,

signaling contestation, while the consensus obtained in the last period is maintained.

According to Collins (2004), the field underwent three major shifts: from attempts to measure gravitational waves with metal bars, to the use of cryogenic devices, and finally to expensive interferometers. Panel C maps well to these shifts. The first period¹⁴ of decreasing modularity (from .14 in 1966 to .12 in 1969) corresponds to the first experiments conducted by Joseph Weber, which consolidated the field of experimental gravitational waves. The reaction to Weber's papers instigated the controversy that occupied Collins's early publications on the field and is apparent in the increase of scaled modularity from 1969 to 1971 (to .16). Collins (2004) argued recently that this closure was clear by 1975, although he did not know it in real time. Our approach suggests that contestation decreased after 1971 and reached its local low (.1) by 1976. Collins (2004) calls the following period (late 1970s to late 1990s) "the bar wars"—a dispute over the use of cryogenic bars versus interferometers. Scaled modularity fluctuates between .1 and .12 in that period. Collins (2004) does not clearly state when this debate ended, although he points to the National Science Foundation review of 1996. We observe closure signs as early as 1992, when modularity declined and reached below .1 in 1995. Across the board, our analysis of changing contestation levels is consistent with Collins's narrative.

Questions of smoking and climate change (see Panels D and E) are intrinsically more interesting because while experts today believe these propositions are true, they used to be very contested and were riddled with claims of inconclusive science (McCright and Dunlap 2000; Samet and Burke 2001). These cases are central to the concept of agnotology—that is, industries' deliberate hindrance of science (Proctor and Schiebinger 2008). Absent structural analysis, timing consensus formation on the hazards of smoking would likely point to 1964 and the first Surgeon General report, or 1986, when three

major reports were published by the IARC (1986), Surgeon General Koop (USDHHS 1986), and the National Academy of Sciences (NAS) (1986). Timing climate change consensus, one would likely point to the IPCC's (2007) fourth report or Oreskes's (2004b) paper that surveyed all relevant abstracts. Our approach identifies consensus earlier, refuting claims of inconclusive science and revealing that the cases are different.

Panel D presents analysis of the proposition that smoking causes cancer. This is the iconic cyclical case. Despite huge research efforts, consensus was hard to form. Early claims of carcinogenicity (e.g., Wynder, Graham, and Croninger 1953) spiked fierce debate; scaled modularity rose to .13 by 1958, only to drop to .08 by 1964. Looking at modularity trends, one could identify consensus as early as 1961 based on the continuous decline. Indeed, in 1962 the Royal Academy of Physicians declared that tobacco is carcinogenic, and in two years the U.S. Surgeon General (1964) joined this assessment. Following the Surgeon General's report, consensus was shattered. Modularity ascended from 1965 through the early 1980s and remained high even as it fluctuated. Here we can observe a combative literature, with research funded in part by the public and in part by tobacco companies. Historical accounts describe the period as a series of battles (Brandt 1998), which inspires our metaphor of a cyclical pattern. The question of tobacco's carcinogenicity was answered and reopened in different formulations, such as the possibility of safe cigarettes and the role of nicotine. Historians argue that the Koop (USDHHS 1986) and the NAS (1986) reports resolved the conflict, showing that smoking kills non-smokers. Starting in 1981 (when the first study to show the hazards of secondhand smoke was published), scaled modularity began to sharply drop, from .15 to .12 in 1985. Kabat (2008) describes how hazards of secondhand smoke remained controversial after the 1986 report, creating the need for the 1992 EPA report. By then, scaled

modularity was at .1. As with gravitational waves, modularity analysis conveys the general pattern of historians' account but identifies nascent consensus somewhat earlier.

The climate change case (see Panel E) reveals that scientific contestation was evident only until the early 1990s. While the public representation of this debate suggests it is similar to the tobacco case, Oreskes (2004b) shows that the scientific community reached consensus as early as 1993. We can observe earlier dynamics: Between 1986 and 1990, scaled modularity was relatively high, showing a significant, stable, but not ultimate decline toward 1992. IPCC's (1992) early report states consensus on climate change but not on its anthropogenic causes, which is not stated until the IPCC's (1995) second report, at which point scaled modularity drops below .1, echoed in 2001 and 2007. Our results reject the claim of inconclusive science on climate change and identify the emergence of consensus earlier than previously thought. Given the weight of this case in illustrations of political interventions in science, it is noteworthy that its scientific representation, derived solely from peer-reviewed articles, resembles the spiral pattern of cases like skin cancer far more than cyclical cases such as the hazards of smoking.

Two Currently Contested Cases

The patterns reported in Figure 4 support the idea that the community salience of scientific citation networks describes their epistemic rivalries. To validate the approach, we compared it with traditional ways of declaring consensus. Having validated it, we can now use our method to describe cases that still seem contested. Figure 5 presents analysis for the propositions that mobile phones' radiation is cancerous (Panel A) and that vaccinations cause autism (Panel B).

With respect to the proposition that cell phones cause cancer, Panel A shows relatively high scaled modularity of .15 in 1997. Since

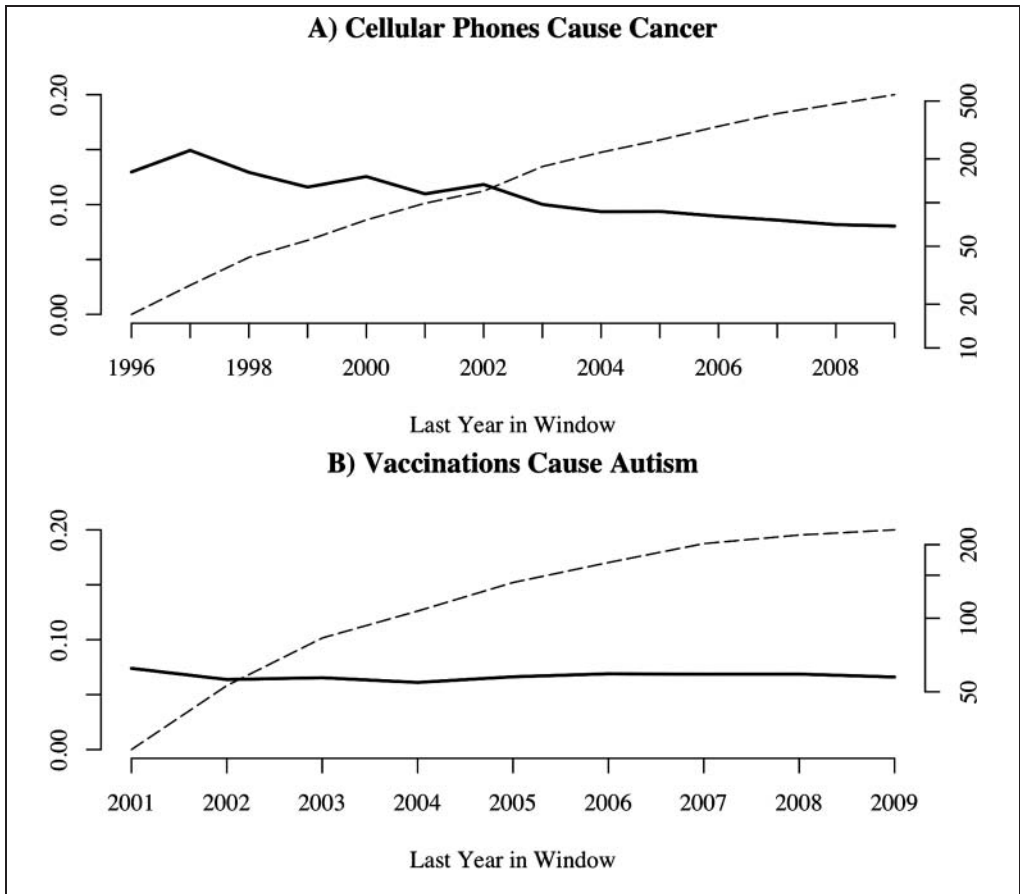


Figure 5. Epistemic Rivalries and Literatures’ Size in Two Publicly Contested Cases
Note: The dashed line refers to the number of papers in the dynamic window and to the logarithmic right-hand-side-axis. The solid line refers to the level of epistemic rivalry, estimated as the modularity score scaled for logged network size, on the left-hand-side-axis. Scaled modularity levels in the debate regarding the carcinogenicity of cellular phones decrease until 2004 from .15 to .09 and are more stable since. The scientific discussion of vaccinations as a cause of autism was never contested, as scaled modularity levels are very low, between .06 and .07 throughout the research period.

then, the literature is characterized by a fairly steady decrease in scaled modularity, reaching .09 in 2004 and decreasing since. Recall that our strategy measures consensus but does not point to its substance. For that, we examine abstracts of influential papers. By 2004, most authorities found no significant cancerous effects of mobile phones (e.g., Wakeford 2004). This particular literature is remarkably cordial; the few studies that find effects admit their methodological problems, while the

majority that do not find effects argue that more research is needed. Neither side conclusively argues that the issue is proven. Our results suggest that this case has been consensual since 2002. The prevailing representation in the field, exemplified by the comprehensive INTERPHONE study (Cardis et al. 2010), is that the science remains inconclusive. Yet, in opposition to some media representations of the subject that portray this inconclusiveness as an epistemic rivalry (e.g.,

Ketcham 2010), our analysis indicates that the scientific consensus is that no proof for cellular radiation hazards has been identified. The contestation tapped by journalists is entirely benign contestation, while the scientists who find no hazards of cellular radiation argue themselves that more research is needed.

Turning to Panel B where we consider the risk of autism posed by the triple vaccination for Measles, Mumps, and Rubella (MMR), it is evident that the scientific community has consensus refuting the relationship. Yet anecdotal information from parents of children with autism generates strong sentiment in many lay communities that vaccines are causally related to autism. In the case of cellular phones, doubt about the scientific consensus, or belief that it may soon change, may lead individuals to stop using them with few implications aside from one's social life. In the case of MMR vaccinations and autism, however, doubts about the scientific consensus lead individuals to withdraw from vaccinations, risking the loss of herd immunity for diseases once largely eradicated from the developing world (Glanz et al. 2009; Jansen et al. 2003; Salathe and Bonhoeffer 2008; Smith et al. 2008). As with other contested issues that are not really contested—for example, the effectiveness of abstinence pledges and DARE programs—identifying when science has got the story right may have important policy implications.

DISCUSSION

This article provides a new way of measuring scientific consensus. We suggested that consensus formation is a form of black boxing, traceable as a decline in the community salience of citation networks. Along the way, we developed a new approach to temporality in citation networks. Measuring community salience as modularity, we distinguish between the component of community salience created by normal fragmentation and specialization, which we name benign

contestation, and the epistemic rivalries that are the substance of severe contestation and around which consensus forms. The former is a product of the literatures' size, while we show that the latter identifies consensus in accordance with expert evaluations. We then analyzed two still-contested cases and revealed emerging (or consistent) consensus. Since 2004, the literature on cellular phone hazards has been consensual. Regarding the idea that MMR vaccinations cause autism, our analysis reveals that this issue has never carried any scientific contestation.

While our interpretation of the results provides excellent fit with existing evaluations, there is no single ahistorical decisive empirical threshold between consensus and contestation. Reifying any value to identify such a threshold would be ill advised, as our tool requires sensitivity to different citation styles, literature sizes, and periods. The results suggest that a ratio of .1 between raw modularity and logged network size may provide a useful rule of thumb. But just as blind adherence to the .05 threshold for statistical significance leads to substantive nonsense in interpreting relationships between variables in extremely large datasets, judgment is necessary here, as well, to make substantively meaningful statements. In these analyses, some consensus formation processes did not always remain below .1. For example, scientific discussion on the carcinogenicity of coffee hovers on either side of .1 after 1985. Like other analyses (e.g., Bhutani, Johnson, and Sivieri 1999), marking a threshold of .1 does not mean that coherent literatures with scaled modularity of .09 are always consensual, and those with .11 are always contested. One should focus on the trend and the context more than the number.

A quantitative measure of scientific consensus reinstates a sociological niche in the field defined by science policy analysts on the one hand and STS scholars on the other. We utilize different approaches. Following Oreskes, we seek consensus in scientific inscriptions; following Latour and Collins, we model consensus as black boxing and show

that internal divisions are observable in citation networks as competing communities. Noting the unified structural implication of Merton's and Latour's theories of citations, we offer a structural measure of real-time scientific consensus, with a minimal latency period induced mostly by journal response time. We depart from ANT and SSK's qualitative empirical orientation and subject their insights to quantitative modeling, enabling us to evaluate consensus within arcane scientific fields.

The utility of our approach is evident merely by considering the scope of the scientific issues considered, covering a century of research in several different disciplines. Modularity trends not only identify transition periods but also show that scientific literatures adapt to new findings quickly. In cases previously considered by science scholars, the quantitative trends fit the narrative reports.

Scientific findings such as modularity never completely speak for themselves. A new method cannot be calibrated without external judgments. Our validation of the analysis presented here has two potential sources of bias. The first is minor: the choice of the IARC and the IPCC as calibration measures. Other benchmarks may exist. The second source of bias is more challenging: populating the dataset by analyst's keywords selection. As discussed earlier, this has the advantage of defining a cognitive domain through its substance. It is also robust across different formulations (see the online supplement), but like any method it is at risk to malfeasance. The deepest challenge arises from the fact that a change in science induces change in nomenclature. For scientific contestation dynamics operating over the long term, sensitivity to shifting keywords is critical.

Public Understanding and the Sociology of Science

One of the many virtues of contemporary STS and Public Understanding of Science

(PUS) studies is their attention to different mechanisms that may limit the scope of our strategy. Consensus may emerge if one side of a controversy strategically changes its language (Simon 2002), or consensus may veil contestation by actors with no access to peer-reviewed journals (Wynne 1996). For PUS scholars, scrutinizing science is only a part of scrutinizing the public engagement with science (Nelkin 1995). Our strategy is limited to peer-reviewed journals. It is by no means a panacea to scrutinizers of science. It could, however, help PUS scholars evaluate the academic side of their story. For example, for issues such as climate change and smoking, where scholars argue that a minority of hired experts created a distorted view of the scientific literature, our strategy may offer a precautionary comment.¹⁵ Future studies might implement our strategy for the blogosphere, extending it beyond peer-reviewed papers.

Determining scientific consensus without relying on structural tools typically required expert knowledge. We do not aim at rendering experts obsolete; rather, we offer a complementary strategy designed to help experts and their audiences. In the future, our approach could be refined, implemented in online search engines, and used by everyone. By allowing anyone to define a literature and assess its dynamics quantitatively, sociology can partake in the effort to make science public and more democratic.

Patterns of Consensus Formation

Assessing consensus, of course, has nothing to do with "the truth." It is thus encouraging to find that when consensus is achieved, networks grow exponentially. More studies are published in peer-reviewed journals that use the keywords attached to the recent consensus. Evans (2007a) shows that discursive consensus increases scientific production. This anti-intuitive claim comes into clear focus here: If consensus was obtained with

fragile evidence, it will likely dissolve with growing interest, which is what happened at the onset of gravitational waves research. If consensus holds, it opens secondary questions for scrutiny. This observation gives rise to the three different trajectories of scientific propositions—what we call flat, spiral, and cyclical.

It seems trivial that some people do not drink coffee for fear of cancer, even though the scientific community considers coffee to be a non-carcinogen (see Figure 4, Panel A); the belief that MMR vaccinations cause autism (see Figure 5, Panel B), however, which leads some people to reject vaccination for their children, is not trivial. Despite this difference, both cases show no epistemic rivalry. The world of flat science can mean two main things. In the coffee case there is no coherent research agenda. The coincidence of coffee and cancer in papers is largely the accidental byproduct of large research efforts in cancer and coffee respectively. Contention around the carcinogenicity of coffee does not arise (IARC 1991), and articles cite other articles seemingly at random. By contrast, the scientific flatness around vaccinations and autism is different—here science speaks with a single voice in opposition to a lay critique. The drive to new studies arises exogenously, but there is no real debate. Articles that refute the connection cite other similar articles. Here too, as a consequence, communities of contention within science fail to arise.

The propositions that the sun causes cancer, people cause climate change, and mobile phones do not cause cancer unfold in a spiral trajectory. In spiral trajectories, initially intense contestation generates rapid settlement and induces a spiral of new questions to which scientists become oriented. Here the settlements of earlier contestation provide scaffolding for new communities of research. Consequently, the modularity of the foundational question—do people cause climate change, for example—remains low because the communities of contestation organized around secondary issues are bridged through

citation to a historically evolving core of accepted knowledge. This spiral conforms to our cultural ideal of science in which scientists are left to their own devices. The drive to new research arises endogenously, as if a Kuhnian machine were operating just as it should.

The case of smoking and cancer looks like Abbott's (2001) description of the social sciences—a constant return to initial states. Each reduction in contestation levels was followed by reoccurring contestation—on the same plane, reformulating the same issue of public interest. In cyclic trajectories, reframing requires new consensus formation, alignments, and goal settings. This trajectory—in the case of smoking and cancer it was sustained for years around the controversial quest for safer cigarettes—eventually transitioned to a spiral pattern around secondhand smoke in the 1980s.

One further sociological insight that our analysis emphasizes is unanticipated differences between cases—for example, between tobacco's carcinogenicity and climate change. Qualitative analyses of each in tandem are rare because of the expertise required to analyze them. The few attempts at comparison frame both cases similarly—as cases in which powerful groups created public doubts in otherwise consensual science (Michaels 2006; see also Proctor and Schiebinger 2008). Our analysis, by contrast, clearly shows that the formation of consensus took different paths: Climate change followed a spiral trajectory, while tobacco research was (for most of its history) trapped in a cycle of persistent repetition. Scientific consensus in this literature was solidified only in the late 1980s, more than half a century after initial evidence was published. Early consensus on tobacco's carcinogenicity formed in 1959 to 1964 and led to the controversial search for a safer cigarette. Scientific consensus on climate change, on the other hand, formed in the early 1990s as evidence was still being gathered. Our findings suggest that commentators on these cases should be aware of the different pathways

agnetology takes. Tobacco firms directly invested in scientific research (Bero 2003), while climate change skeptics used the media and political office holders to cultivate doubt (Jacques, Dunlap, and Freeman 2008). Both strategies are effective in creating contestation—in the short run.

Our research strategy first asked the question—are the dynamics of consensus formation the same across these hotly contested propositions—and then answered it. Hopefully, development of approaches like ours within the sociology of science will lead us out of the cycle of persistent repetition by identifying new problems and new answers.

Acknowledgments

Comments from the editors of *ASR*, anonymous reviewers, and Harry Collins, Jonathan Cole, Gil Eyal, Bruno Latour, Douglas White, Balazs Vedres, Shamus Khan, Dan Lainer-Voss, Martha Poon, Rozlyn Redd, Dan Navon, William McAllister, and participants of the Lazarsfeld-Mellon interdisciplinary workshop and SUNBELT have greatly improved previous drafts. We thank Elizabeth Leicht and Mark Newman for their invaluable help in implementing the algorithm.

Funding

Support for this project came from the Paul Lazarsfeld Center for the Social Sciences, the Andrew Mellon Foundation, and the NIH Director's Pioneer Award program, part of the NIH Roadmap for Medical Research, through grant number 1 DP1 OD003635-01.

Notes

1. The distinction between benign contestation and epistemic rivalries harkens back to Kuhn's (1970) distinction between normal science and periods of crisis/revolution. Whereas Kuhn's focus is on different periods, we distinguish between forms of contestation that co-exist temporally.
2. At this time, consensus holds that smoking and solar radiation cause cancer and that humans are causing climate change. In *Sleeper* (1973), Woody Allen wakes up in a future where the consensus is that smoking improves health. This article is concerned with understanding when scientific consensus is established, not if it is true in some absolute way.
3. Merton (1957) suggests that citations are acts of debt payment, following a market metaphor. Latour (1987) argues that citations are rhetorical acts of mobilization, following a military metaphor. For current purposes, both lead to the same conclusion: citations are more likely to signal agreement (shown empirically in Hanney et al. 2005).
4. Of course, papers have a life of their own. For example, while Merton never explicitly argued that sociology of science should avoid analyzing knowledge, his 1973 book was read in this manner, and the string "Merton (1973)" was used to assert this (Hargens 2004). Reducing knowledge to its practitioners is not solely the vice of Mertonians. SSK added knowledge to the study of scientists, but they accept knowledge only as a property of a human actor. This view prevails in contemporary studies of expertise as Collins and Evans (2007) and Weinel (2008) indicate.
5. Community salience is also low when papers do not promote the same views but discuss different things and have little to do with each other. The nature of the data suggests that this too signals consensus.
6. While we reached this idea via Latour, it can be framed in Kuhnian (1970) terms, as the emergence of a paradigm, or in Lakatosian (1970) terms, that is, in new research programs that have yet to develop a core, auxiliary hypotheses form separate communities. When science matures, practitioners agree on a core and community demarcations dissolve.
7. Ending a word with a * means that all characters following the previous string are accepted. Keywords are connected by OR within parentheses and AND between parentheses, so that a paper needs at least one word from *each* parentheses to be included in the dataset.
8. At the 100th percentile, the network is the cumulative network. At the 1st percentile, the network is similar to a sliding window with a width of 1. The median is akin to citation half-life, which is not only an intuitive cutoff point but also provides intuitive widths, usually of 3 to 6 years with outliers of 2 and up to 11.
9. Authority is a centrality measure for directed networks, assigning nodes with an authority score based on in-degree and a hub score based on out-degree. Each score's calculation weighs neighbors' scores on the other measure. This calculates papers' importance not only by their citation counts but by considering who is citing them.
10. Initial periods of fluctuating authorities occur when networks are very small. The slopes obtained by the cumulative approach converge to 1 when networks have more than 17 papers.
11. Experts' reports are not the only mechanism for consensus declaration, but they are the most

- authoritative and frequently used mechanism, and the one most often cited as proof.
12. Because the IARC was founded in 1965, we also note the first U.S. Surgeon General (1964) report on smoking. We also note the 1992 EPA report because of its importance (see Kabat 2008).
 13. See also the online supplement.
 14. Collins's detailed review of the field starts at 1966. We grayed out the earlier years that our data covers.
 15. As one reviewer noted, if everyone used our strategy it may be subjected to Goodhart's law: interested parties who in the past needed only to hire a well-respected scholar may attempt to publish papers and manipulate citations to create a false sense of consensus. We should be so fortunate. The fact that a useful measure may be manipulated in the future provides little reason to abandon it.

References

- Abbott, Andrew. 2001. *The Chaos of Disciplines*. Chicago: University of Chicago Press.
- Allen, Woody. (Director). 1973. *Sleeper*. Rollins-Joffe Productions.
- Ashar, Hanna and Jonathan Z. Shapiro. 1990. "Are Retrenchment Decisions Rational? The Role of Information in Times of Budgetary Stress." *Journal of Higher Education* 61:121–41.
- Bero, Lisa. 2003. "Implications of the Tobacco Industry Documents for Public Health and Policy." *Annual Review of Public Health* 24:267–88.
- Bhutani, Vinod, Lois Johnson, and Emidio Sivieri. 1999. "Predictive Ability of a Predischarge Hour-Specific Serum Bilirubin for Subsequent Significant Hyperbilirubinemia in Healthy Term and Near-Term Newborns." *Pediatrics* 103:6–14.
- Biglan, Anthony. 1973. "Characteristics of Subject Matter in Different Academic Areas." *Journal of Applied Psychology* 57:195–203.
- Bourdieu, Pierre. 1975. "The Specificity of the Scientific Field and the Social Conditions of the Progress of Reason." *Social Science Information* 14:19–47.
- Boykoff, Maxwell T. and Jules R. Boykoff. 2004. "Balance as Bias: Global Warming and the U.S. Prestige Press." *Global Environmental Change: Human and Policy Dimensions* 14:125–36.
- Brandt, Allan M. 1998. "Blow Some My Way: Passive Smoking, Risk and American Culture." Pp. 164–88 in *Ashes to Ashes: The History of Smoking and Health*, edited by S. L. Lock, A. Reynolds, and E. M. Tansey. Atlanta: Rodopi.
- Cardis, E. et al. 2010. "Brain Tumour Risk in Relation to Mobile Telephone Use: Results of the INTERPHONE International Case-Control Study." *International Journal of Epidemiology* 39:675–94.
- Cole, Jonathan R. and Harriet Zuckerman. 1975. "The Emergence of a Scientific Specialty: The Self Exemplifying Case of the Sociology of Science." Pp. 139–74 in *The Idea of Social Structure: Papers in Honor of Robert K. Merton*, edited by L. A. Coser. New York: Harcourt Brace Jovanovich.
- Cole, Stephen. 1983. "The Hierarchy of the Sciences." *American Journal of Sociology* 89:111–39.
- Collins, Harry M. 1974. "The TEA Set: Tacit Knowledge and Scientific Networks." *Science Studies* 4:165–86.
- . 1975. "The Seven Sexes: A Study in the Sociology of a Phenomenon, or the Replication of Experiments in Physics." *Sociology* 9:205–224.
- . 1992. *Changing Order*. Chicago: University of Chicago Press.
- . 2004. *Gravity's Shadow: The Search for Gravitational Waves*. Chicago: University of Chicago Press.
- Collins, Harry M. and Robert Evans. 2002. "The Third Wave of Science Studies: Studies of Expertise and Experience." *Social Studies of Science* 32:235–96.
- . 2007. *Rethinking Expertise*. Chicago: Chicago University Press.
- Csardi, Gabor and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal Complex Systems*:1695.
- Elwood, J. Mark, Richard P. Gallagher, G. B. Hill, and J. C. G. Pearson. 1985. "Cutaneous Melanoma in Relation to Intermittent and Constant Sun Exposure: The Western Canada Melanoma Study." *International Journal of Cancer* 35:427–33.
- Epstein, Steven. 1998. *Impure Science: AIDS, Activism and the Politics of Knowledge*. Berkeley: University of California Press.
- Evans, James A. 2007a. "Consensus and Knowledge Production in an Academic Field." *Poetics* 35:1–21.
- . 2007b. "Start-Ups in Science: Entrepreneurs, Diverse Backing, and Novelty Outside Business." Pp. 261–307 in *The Sociology of Entrepreneurship*, edited by M. Ruef and M. Lounsbury. Oxford: JAI Press.
- . 2008. "Electronic Publication and the Narrowing of Science and Scholarship." *Science* 321:395–99.
- Frickel, Scott and Neil Gross. 2005. "A General Theory of Scientific/Intellectual Movements." *American Sociological Review* 70:204–232.
- Frickel, Scott and Kelly Moore, eds. 2006. *The New Political Sociology of Science*. Madison: University of Wisconsin Press.
- Fujimura, Joan. 1996. *Crafting Science: A Sociohistory of the Quest for the Genetics of Cancer*. Cambridge: Harvard University Press.
- Garfield, Eugene. 1972. "Citation Analysis as a Tool in Journal Evaluation: Journals Can Be Ranked by

- Frequency and Impact of Citations for Science Policy Studies." *Science* 178:471.
- Gieryn, Thomas. 1996. "Policing STS: A Boundary-Work Souvenir from the Smithsonian Exhibition on 'Science in Everyday Life.'" *Science, Technology and Human Values* 21:100–115.
- . 1999. *Cultural Boundaries of Science*. Chicago: Chicago University Press.
- Glanz, Jason M., David L. McClure, David J. Magid, Matthew F. Daley, Eric K. France, Daniel A. Salmon, and Simon J. Hambidge. 2009. "Parental Refusal of Pertussis Vaccination Is Associated With an Increased Risk of Pertussis Infection in Children." *Pediatrics* 123:1446–51.
- Hanney, Steve, Lain Frame, Jonathan Grant, Martin Buxton, Tracy Young, and Grant Lewison. 2005. "Using Categorizations of Citations when Assessing the Outcomes from Health Research." *Scientometrics* 65:357–79.
- Hargens, Lowell L. 1988. "Scholarly Consensus and Journal Rejection Rates." *American Sociological Review* 53:139–51.
- . 2004. "What is Mertonian Sociology of Science?" *Scientometrics* 60:63–70.
- Hargens, Lowell L. and Lisa Kellywilson. 1994. "Determinants of Disciplinary Discontent." *Social Forces* 72:1177–95.
- Hess, David J. 1995. *Science and Technology in a Multicultural World*. New York: Columbia University Press.
- . 2006. "Antiangiogenesis Research and the Dynamics of Scientific Fields: Historical and Institutional Perspectives in the Sociology of Science." Pp. 122–47 in *The New Political Sociology of Science*, edited by S. Frickel and K. Moore. Madison: University of Wisconsin Press.
- International Agency for Research on Cancer (IARC). 1986. "IARC Monographs on the Evaluation of Carcinogenic Risks to Humans." *Tobacco Smoking*, Vol. 38. Lyon, France: World Health Organization.
- . 1991. "IARC Monographs on the Evaluation of Carcinogenic Risks to Humans." *Coffee, Tea, Mate, Methylxanthines and Methylglyoxa*, Vol. 51. Lyon, France: World Health Organization.
- . 1992. "IARC Monographs on the Evaluation of Carcinogenic Risks to Humans." *Solar and Ultraviolet Radiation*, Vol. 55. Lyon, France: World Health Organization.
- Intergovernmental Panel on Climate Change (IPCC). 1992. *Climate Change: The 1990 and 1992 IPCC Assessment*. Canada: IPCC.
- . 1995. *Summaries for Policymakers of the Three Working Groups Report*. Geneva, Switzerland: IPCC.
- . 2007. *Climate Change 2007: The Physical Science Basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.
- Jacques, Peter J., Riley E. Dunlap, and Mark Freeman. 2008. "The Organization of Denial: Conservative Think Tanks and Environmental Skepticism." *Environmental Politics* 17:349–85.
- Jansen, Vincent, N. Stollenwerk, H. J. Jensen, M. E. Ramsey, W. J. Edmunds, and C. J. Rhodes. 2003. "Measles Outbreaks in a Population with Declining Vaccine Uptake." *Science* 301:804.
- Jasanoff, Sheila, ed. 2004. *States of Knowledge*. New York: Routledge.
- Kabat, Geoffrey. 2008. *Hyping Health Risks: Environmental Hazards in Daily Life and the Science of Epidemiology*. New York: Columbia University Press.
- Ketcham, Christopher. 2010. "Warning: Your Cell Phone May Be Hazardous to Your Health." *GQ*, February, pp. 60–63.
- Kleinberg, Jon M. 1999. "Authoritative Sources in a Hyperlinked Environment." *Journal of the ACM* 46:604–632.
- Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lakatos, Imre. 1970. "Falsification and the Methodology of Scientific Research Programmes." Pp. 91–197 in *Criticism and the Growth of Knowledge*, edited by I. Lakatos and A. Musgrave. Cambridge: Cambridge University Press.
- Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge: Harvard University Press.
- . 1999. *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge: Harvard University Press.
- . 2004. *Politics of Nature: How to Bring the Sciences into Democracy*. Cambridge: Harvard University Press.
- . 2009. "Is Walter Lippmann's Phantom Public More Visible on the Web?" Presented at The Changing Dynamics of Public Controversies, Columbia University, New York, February 7.
- Leahey, Erin. 2005. "Alphas and Asterisks: The Development of Statistical Significance Testing Standards in Sociology." *Social Forces* 84:1–24.
- Leicht, Elizabeth A., Gavin Clarkson, Kerby Shedden, and Mark E. J. Newman. 2007. "Large-Scale Structure of Time Evolving Citation Networks." *European Physical Journal* 59:75–83.
- Leicht, Elizabeth A. and Mark E. J. Newman. 2008. "Community Structure in Directed Networks." *Physical Review Letters* 100:118703.
- Martin, Brian and Evelleen Richards. 1995. "Scientific Knowledge, Controversy, and Public Decision Making." Pp. 506–526 in *Handbook of Science and Technology Studies*, edited by S. Jasanoff, G. E. Markle, J. C. Peterson, and T. Pinch. Thousand Oaks, CA: Sage.

- McCright, Aaron M. and Riley E. Dunlap. 2000. "Challenging Global Warming as a Social Problem: An Analysis of the Conservative Movement's Counter-Claims." *Social Problems* 47:499–522.
- Merton, Robert K. 1957. "Priorities in Scientific Discovery." *American Sociological Review* 22:635–59.
- . 1973. *The Sociology of Science*. Chicago: University of Chicago Press.
- Michaels, David. 2006. "Protecting Public Health in the Age of Contested and Product Defense." *Annals of the New York Academy of Science* 1076:149–62.
- Moody, James. 2004. "The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999." *American Sociological Review* 69:213–38.
- Moore, Kelly. 1996. "Organizing Integrity: American Science and the Creation of Public Interest Organizations, 1955–1975." *American Journal of Sociology* 101:1592–1627.
- Mulkay, M. J., G. N. Gilbert, and S. Woolgar. 1975. "Problem Areas and Research Networks in Science." *Sociology* 9:187–203.
- National Academy of Sciences (NAS), National Research Council Committee on Passive Smoking. 1986. *Environmental Tobacco Smoke: Measuring Exposure and Assessing Health Effects*. Washington, DC: National Academy of Sciences.
- Nelkin, Dorothy. 1995. "Science Controversies: The Dynamics of Public Disputes in the United States." Pp. 444–56 in *Handbook of Science and Technology Studies*, edited by S. Jasanoff, G. E. Markle, J. C. Peterson, and T. Pinch. Thousand Oaks, CA: Sage.
- Newman, Mark E. J. 2006. "Modularity and Community Structure in Networks." *Proceedings of the National Academy of Sciences of the United States of America* 103:8577–82.
- Oreskes, Naomi. 2004a. "Science and Public Policy: What's Proof Got To Do With It?" *Environmental Science and Policy* 7:369–83.
- . 2004b. "Beyond the Ivory Tower: The Scientific Consensus on Climate Change." *Science* 306:1686.
- Owen-Smith, Jason and Walter W. Powell. 2008. "Networks and Institutions." Pp. 276–98 in *The SAGE Handbook of Organizational Institutionalism*, edited by R. Greenwood, C. Oliver, R. Suddaby, and K. Sahlin-Andersson. London: Sage.
- Pickering, Andrew. 1993. "The Mangle of Practice: Agency and Emergence in the Sociology of Science." *American Journal of Sociology* 99:559–89.
- Pinch, Trevor J. and Wiebe E. Bijker. 1984. "The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology might Benefit Each Other." *Social Studies of Science* 14:388–441.
- Powell, Walter W. and Jeannette A. Colyvas. 2008. "Microfoundations of Institutional Theory." Pp. 276–98 in *The SAGE Handbook of Organizational Institutionalism*, edited by R. Greenwood, C. Oliver, R. Suddaby, and K. Sahlin-Andersson. London: Sage.
- Powell, Walter W., Douglas R. White, Kenneth W. Koput, and Jason Owen-Smith. 2005. "Network Dynamics and Field Evolution: The Growth of Inter-organizational Collaboration in the Life Sciences." *American Journal of Sociology* 110:1132–205.
- Price, Derek J. de S. 1965. "Networks of Scientific Papers." *Science* 149:510–515.
- Proctor, Robert and Londa Schiebinger, eds. 2008. *Agnotology: The Making and Unmaking of Ignorance*. Stanford: Stanford University Press.
- R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reichardt, Jorg. 2009. *Structure in Complex Networks*. Berlin: Springer.
- Rowe, Gene, and Lynn J. Frewer. 2005. "A Typology of Public Engagement Mechanisms." *Science, Technology and Human Values* 30:251–90.
- Salathe, Marcel, and Sebastian Bonhoeffer. 2008. "The Effect of Opinion Clustering on Disease Outbreaks." *Journal of The Royal Society Interface* 5:1505–1508.
- Samet, Jonathan M. and Thomas A. Burke. 2001. "Turning Science into Junk: The Tobacco Industry and Passive Smoking." *American Journal of Public Health* 91:1742–44.
- Shapin, Steven. 1994. *A Social History of Truth: Civility and Science in Seventeenth-Century England*. Chicago: University of Chicago Press.
- . 1996. *The Scientific Revolution*. Chicago: University of Chicago Press.
- Simon, Bart. 2002. *Undead Science*. New Brunswick: Rutgers University Press.
- Small, Henry. 2006. "Tracking and Predicting Growth Areas in Science." *Scientometrics* 68:595–610.
- Smith, Michael J., Susan S. Ellenberg, Louis M. Bell, and David M. Rubin. 2008. "Media Coverage of the Measles-Mumps-Rubella Vaccine and Autism Controversy and Its Relationship to MMR Immunization Rates in the United States." *Pediatrics* 121:e836–43.
- Smith-Doerr, Laurel, Ivan M. Manev, and Polly Rizova. 2004. "The Meaning of Success: Network Position and the Social Construction of Project Outcomes in an R&D Lab." *Journal of Engineering and Technology Management* 21:51–81.
- Star, Leigh S. 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* 19:387–420.
- U.S. Department of Health and Human Services (USDHHS). 1986. *The Health Consequences of Involuntary Smoking: A Report of the Surgeon General*. Rockville, MD: U.S. DHHS, Public Health

- Service, Centers for Disease Control, Center for Health Promotion and Education, Office on Smoking and Health. DHHS Publication No. (CDC) 87-8398.
- U.S. Environmental Protection Agency (EPA). 1992. "Respiratory Health Effects of Passive Smoking (Also Known as Exposure to Secondhand Smoke or Environmental Tobacco Smoke ETS)." Washington, DC: U.S. EPA, Office of Research and Development, Office of Health and Environmental Assessment, EPA/600/6-90/006F.
- U.S. Surgeon General. 1964. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington, DC: Department of Health, Education, and Welfare.
- Wakeford, Richard. 2004. "The Cancer Epidemiology of Radiation." *Oncogene* 23:6404–6428.
- Wallace, A. 1994. "Soil Science, Pesticides, and Risk Analysis." *Communications in Soil Science and Plant Analysis* 25:143–48.
- Weinel, Martin. 2008. "Counterfeit Scientific Controversies in Science Policy Contexts." *SOCSI Working Paper* 120.
- Whittington, Kjersten Bunker, Jason Owen-Smith, and Walter W. Powell. 2009. "Networks, Proximity, and Innovation in Knowledge-Intensive Industries." *Administrative Science Quarterly* 54:90–122.
- Wray, Brad K. 2005. "Rethinking Scientific Specialization." *Social Studies of Science* 35:151–64.
- Wynder, Ernst L., Everts A. Graham, and Adele B. Croninger. 1953. "Experimental Production of Carcinoma with Cigarette Tar." *Cancer Research* 13:855–64.
- Wynne, Brian. 1996. "May the Sheep Safely Graze? A Reflexive View of the Expert-Lay Divide." Pp. 104–137 in *Risk, Environment and Modernity*, edited by S. Lash, B. Szerszynski, and B. Wynne. London: Sage.
- Zuckerman, Harriet. 1988. "The Sociology of Science." Pp. 511–74 in *The Handbook of sociology*, edited by N. J. Smelser. Newbury Park, CA: Sage Publications.

Uri Shwed is an Assistant Professor of sociology and anthropology at Ben Gurion University of the Negev in Israel. He received his PhD from Columbia University in 2010. His current work focuses on networks in science and social networks of Israeli schoolchildren.

Peter S. Bearman is the Cole Professor of the Social Sciences at Columbia University. His current work focuses on understanding the increased prevalence of autism, analytic sociology, and modeling temporal dynamics in complex historical contexts.