Ambiguity and Engagement[1]

Peter McMahan
Stanford University

James Evans
University of Chicago

Despite modernity's love affair with rationality and the precision that supports it, ambiguity persists not only in humor and politics but in all areas of contemporary life including scholarship and science. Here the authors explore how knowledge cultures differ in their precision of expression and the consequences of ambiguity for those cultures. They develop, estimate, and validate a model of ambiguous expression from large-scale publication data and then show that ambiguous scholarly language acts like a boundary object between researchers and their communities, drawing competing interpretations into conversation with one another as they build on it. Ambiguity, and the uncertainty that follows, stimulate social learning and so ironically play a crucial role in focusing modern knowledge and creating zones of social and intellectual engagement.

INTRODUCTION

With the rise of modern markets, organizations, states, technologies, and science came a necessary increase in the precision of expression that made these complex institutions possible. The emergence of coinage, commercial

[1] This work was supported in part by National Science Foundation grants SBE-0915730, SBE-1158803, NCSES-1422902, AFOSR grant FA9550-15-1-0162, and DARPA grant 14145043. We acknowledge Clarivate Analytics for allowing us to use their Science, Social Science, and Arts and Humanities Abstracts and Citation Indexes. We thank Jacob Foster, Walter Powell, members of the Center for the Study of Social Organization at Princeton University, participants in the Organizations and Markets workshop at the

0002-9602/2018/12403-0006$10.00

calculation, and rational valuation practices dramatically increased the volume and diversity of trade by relieving potential partners from the "coincidence of wants" that barter and reciprocity demand (Jevons 1876; Weber, Roth, and Wittich 1978). The invention of accurate clocks, watches, and interconnected manufacturing equipment gave rise to standardized factory shifts and larger, hierarchical organizations (Thompson 1967). Napoleon in France and the Jacksonian Democrats in the United States sponsored legal codification, seeing beneath ambiguity an arbitrariness that protected aristocratic privilege (Levine 1988). With the Reformation came logically articulated theologies, clear scriptural translations into the vernacular, and explicit "plain style" religious speech as in ascetic Puritan sermons (Levine 1988; Weber 1992). Despite the visibility of these modern precision movements, here we argue and empirically demonstrate that ambiguity continues to play a productive role in society by focusing attention and facilitating social and cognitive engagement.

In perhaps no other institution has precise articulation been praised and ambiguity derided than modern science and engineering. With the emergence of modern mathematical and experimental science in the 17th century, the infusion of precise methods, techniques, and data was coupled with new linguistic technologies that sought to transcend the vagaries of ambiguous common speech, integrate existing knowledge, and discover new ideas. In math and logic, Leibniz articulated the need for a universal language of science (Leibniz and Loemker 1976).[2] Frege later constructed this in the first work of modern logic to "break the domination of the word over the human spirit" by forging an unambiguous "concept script" (1879, p. x). In experimental and observational science, the *Philosophical Transactions of the Royal Society*, the first scientific serial, was characterized by a genre of prolix, precise description that enabled natural philosophers to "virtually witness" and replicate each other's experiments with something approximating "mathematical exactness" (Shapin, Schaffer, and Hobbes 1985; Hobbes 2009, p. 69). Lavoisier launched chemical description toward this ideal in the 18th century by introducing modern chemical nomenclature and reaction formulas, which not only facilitated precise articulation but also defined zones for discovery (Guyton de Morveau et al. 1787). In the 20th cen-

Booth School of Business, the Columbia University and Cornell University Sociology Colloquia, and the *AJS* reviewers for comments on earlier drafts. Direct correspondence to James Evans, Department of Sociology, 1126 East 59th Street, University of Chicago, Chicago, Illinois 60637. E-mail: jevans@uchicago.edu

[2] Leibniz hoped that this *Characteristica Universalis*, alongside his aspiring universal logical calculus, the *Calculus Raciocinator*, would accelerate the generation of new scientific knowledge. Leibniz's dissertation, "De arte combinatoria" (1923), forged a simple combinatorial calculus to operate on precisely defined scientific arguments and predicates.

tury, the Vienna Circle eschewed ambiguity by attempting to reduce all scientific description to systematic logical expression (Carnap 1928).

The social sciences matched the natural sciences in their assault on ambiguity. In his essay "On the Abuse of Words," John Locke argued that "all the artificial and figurative applications of words Eloquence hath invented are for nothing else but to insinuate the wrong Ideas, move the Passions and thereby mislead the Judgment; and so indeed are perfect cheats." Seekers of truth must deliver their messages "without Obscurity, Doubtfulness, or Equivocation, to which Men's Words are naturally liable" (Locke [1690] 1975, pp. 504–9). Hume and Bentham similarly attacked poetry, favoring description with unvarying symbolic associations to eliminate metaphor (Levine 1988).[3] Condorcet's social science curriculum abandoned the classics "to preserve the reason of citizens against the wiles of eloquence, hastening the transition towards a rational political science" (Baker 1975, p. 298). For Durkheim, social theory was fundamentally limited by ambiguities of common speech, which "risk[s] distinguishing what should be combined, or combining what should be distinguished, thus mistaking the real affinities of things and . . . misapprehending their nature" (1951, p. 41).

In summary, generations of natural and social scientists claimed that precise language would enable critical evaluation of truth claims, accelerate scientific discovery, and facilitate the accumulation of knowledge through reproducible description. Precision would coordinate the rationality of scientific investigation. Ambiguity, by contrast, was accused of impeding high-fidelity communication by failing to maintain a sharp boundary between truth and fiction, information and emotion, substance and style, and so increased the flow of false information. By mixing truth with error, ambiguity would lure audiences down conceptual dead ends and result in uncoordinated, fragmented understanding. As we will argue, many of these feared consequences hinge on an extrapolation of ambiguity's effect on individuals in social isolation.

Despite calls for precision, ambiguity remains a feature of discourse in all domains of social life, including contemporary science and scholarship. Some fields of science have been particularly resistant to efforts that would subdue ambiguity with imposed precision, especially those engaged in field or observational research where fluid comparisons of complex and dynamic systems fuel high rates of conceptual innovation (e.g., sociology, cosmology, meteorology, ecology). Differences in expressive precision result from not only the objects of study or mode of analysis, however, but also the historically contingent success of particular codes among scientists. As a result, some knowledge cultures sustain more ambiguity than others (Knorr-Cetina

[3] This was the same historical moment that Samuel Johnson began to standardize English words in sourcing the first dictionary.

1999). Moreover, recent research on scholarly rhetoric has enumerated constructive consequences of ambiguity for the enduring success of science and social science "classics" (Campbell 1975; Davis 1986), as well as synthetic projects that draw attention to new multidisciplinary problem areas (Ceccarelli 2001).

Substantial work has explored the role of ambiguous social action (Cohen, March, and Olsen 1972; DiMaggio and Powell 1983; Leifer 1983; Padgett and Ansell 1992), but ambiguity's role has been systematically studied neither through the medium of language nor in the domains of science, social science, and humanistic scholarship. In his treatise on ambiguity and modernity, Levine claims, "The disposition to flee from the ambiguities of human life and utterance has produced three characteristic failings in modern social science. These failings reflect (1) a trained incapacity to observe and represent ambiguity as an empirical phenomenon; (2) insufficient awareness of the multiple meanings of commonly used terms . . . and (3) where such an awareness exists, an inability to realize the constructive possibilities of ambiguity in theory and analysis" (1988, p. 8). Research on ambiguous expression has failed to directly model precision and its converse. This work has relied instead on shared interpretation (Levine 1965; Edelman 1992) or measures of varying word contexts, but not uncertainty about meaning (Hamilton, Leskovec, and Jurafsky 2016). The limited research that does explore ambiguity's role in social science and humanistic scholarship denies its application to the natural sciences (Davis 1986; Levine 1988). In the philosophy of science, Quine (1960) has argued that translation between languages is indeterminate. As every person represents a distinct combination of experiences and exposures to prior language, meanings conveyed through interdisciplinary or even interpersonal dialogue become ambiguous and uncertain.

In this article, we explore the relationship between ambiguous language and social engagement. Then we model and validate a notion of lexical ambiguity, demonstrating that it corresponds to perceptions of semantic uncertainty. Next, we estimate this model using millions of abstracts from science and scholarship, producing a dispersion of fields with the humanities using language most ambiguously, followed by the social sciences and then the natural sciences, with few exceptions. Finally, we investigate how individual articles differ in their precision of expression and show that more ambiguity systematically leads to greater engagement within and integration across knowledge communities. Ambiguous scientific language plays the role of boundary object between scientists, scholars, and fields, bringing ideas into conversation even as they are built on in new intellectual projects (Starr and Greisemer 1989). In this way, ambiguity ironically focuses modern knowledge and creates active zones of intellectual engagement. We explore the implications of these findings for science, scholarship, and other domains of social life.

AMBIGUOUS COMMUNICATION

From poetry to politics, the ambiguity of natural language and symbolic action is considered central to many aspects of social life. Ambiguity is most often associated with *polysemy*, or the way in which a symbol, word, phrase, sentence, or higher-level expressive unit such as a document or performance can simultaneously carry multiple meanings (Tuggy 1993; Ceccarelli 1998). Social actors have used ambiguous communication to efficiently say more with less (Piantadosi, Tily, and Gibson 2012). In poetry, humor, and persuasion, communicators often couple a focal subject with an evocative alternative meaning to arouse or distract the audience—from playful pun to bawdy double entendre to coupling of a banal subject with one of emotional charge (e.g., "death tax"). As underspecified meanings multiply, however, an ambiguous message becomes vague (Tuggy 1993), such that at the limit what is communicated remains so contradictory or irrelevant "the reader is forced to invent interpretations" (Empson 1966, p. vi).[4]

Multiple meanings conveyed through vague or ambiguous communication may lead to confusion, where audience members become individually or collectively uncertain about the meaning intended by the sender (Levine 1988). *Individual uncertainty* implies that individual recipients are themselves unclear about the intended meaning. *Collective uncertainty*, by contrast, suggests that different audiences, who may not themselves be individually uncertain about the meaning of a multivocal communication, disagree with one another (Solomon and McMullen 1991).

Information theory furnishes a formal model of communication in which uncertainty regarding intended meanings plays the central role. In Claude Shannon's *A Mathematical Theory of Communication* (1949), he modeled the accurate transmission of symbols over a noisy channel. Crucial to this theory is the concept of (Shannon) entropy, or the level of statistical uncertainty in a discrete probability distribution. Precisely, entropy is the expectation of the unique information content from a single draw of a random variable and so measures the unpredictability or uncertainty associated with that variable's outcomes. A message communicated with maximum Shannon entropy contains no redundancy and cannot be compressed. Fifteen years after its introduction, Shannon republished his theory, introduced with an essay by Warren Weaver, who placed it in context of interpersonal communication (Shannon and Weaver 1963). Weaver linked Shannon's "technical problem" of exact symbolic transmission to the "semantic problem" of uncertainty induced by ambiguity—how precisely symbols convey desired

[4] Audience interests play a crucial role in shaping the ambiguity experienced about an expression as when hearers seek diverse understandings (Gaonkar 1989) or ply subversive counterinterpretation (McKerrow 1989). Intended or unintended, a text or symbol's ambiguity may facilitate more less disagreement, and so constitute its polysemic potential (Ceccarelli 1998).

meaning (see fig. 1). To do this, Weaver invents the concept of "semantic noise," from which a "semantic receiver" attempts to decode the communicator's intended meaning, as shown in our figure 1. Semantic noise is the uncertainty unleashed by ambiguity as it interacts with the experience-borne "capacity of the audience" to understand (p. 116). Moreover, Weaver hinted that entropy could play a role in evaluating the distribution of meanings, such that it could answer the semantic question of "how precisely do the transmitted symbols convey the desired meaning?" (p. 114).

Information theory outlines a model of ambiguity, which we will draw upon in measuring ambiguous expression in the methods section below; but to our knowledge it has not been used before to explore the consequences of ambiguity. This is likely because Weaver framed all expressed ambiguity as "unintended" (p. 115) and all experienced "confusion" as a social problem (p. 116). Like the champions of precision quoted in the prior section, Weaver only imagined that ambiguity could yield individual uncertainty about communicated meaning by "overcrowd[ing] the capacity of the audience" (p. 117). As such, he missed the possibility of ambiguous performances *designed* to yield individual or collective uncertainty. Although this vision limited use of the model to analyze ambiguity, it does not reflect a limitation of the model itself, in which entropy can be used to trace semantic uncertainty intended or unintended, individual or collective.

AMBIGUOUS CONSEQUENCES

What is the consequence of ambiguity in scientific and scholarly discourse? Scholars have imagined two very different outcomes. Ambiguity could fragment a field if distinct interpretations are not reconciled, but rather avoid or ignore one another. An ambiguous characterization, like Marx's separation of society into "base and superstructure" or Simmel's paradoxical network characterization of "freedom," may yield several interpretations. Uncertain of the author's true intentions, analysts would interpretively fill in the gaps according to their own idiosyncratic intuitions, tastes, mental heuristics, and skills. The more ambiguous an interpreted work, the more likely individual interpretations would deviate from core text and from each other. At the limit, these interpreted readings would be incommensurable, irreconcilable, and mutually exclusive, categorically unable to engage with each other. And yet, for fragmentation to be the primary outcome of ambiguity in science and scholarship assumes that scientists and scholars independently interpret ambiguous works—that they are intellectually asocial. This cartoon of the interpretive process seems unrealistic and its consequences unlikely. Note, however, that it rests on the same assumptions as diatribes against ambiguity, technologies of communicative precision, and Weaver's interpretation of information theory.

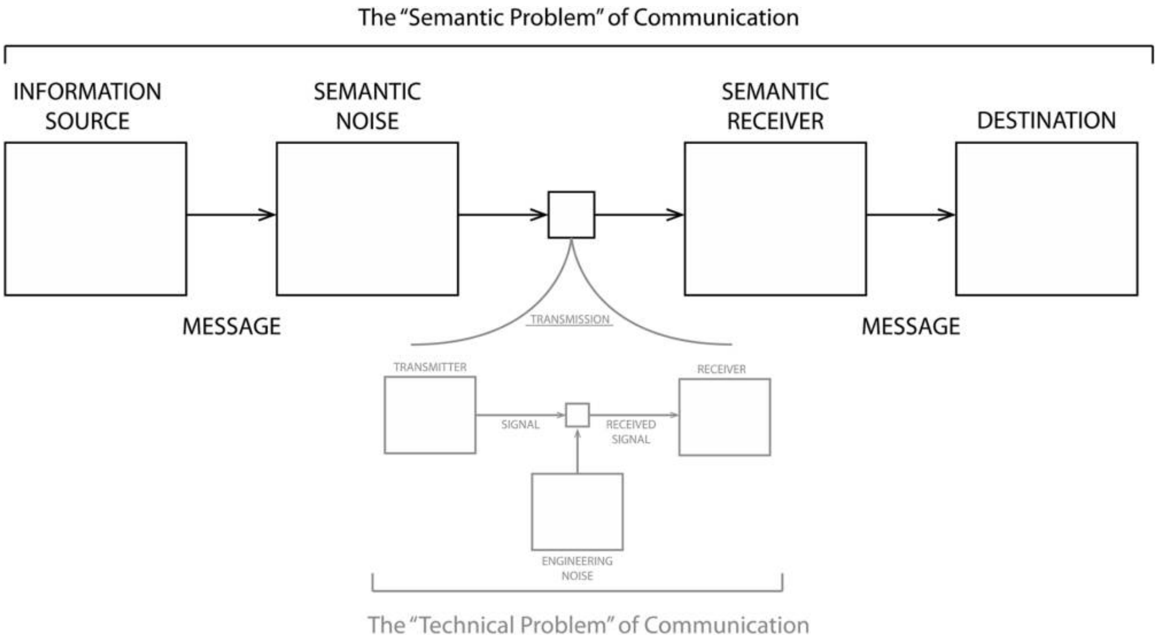

FIG. 1.—The Weaver model outlined both the semantic and technical problems of communication (Shannon and Weaver 1963, pp. 98, 115–16)

In contrast to this assumption of isolation, intellectuals are intensively engaged in dialogue at all stages of the interpretive process (Collins 1998). In Stark's *Sense of Dissonance* (2011), he illustrates how ambiguity can lead to competing interpretations. When this yields subsequent disagreement, however, it can ultimately benefit discovery. If ambiguity drives independent interpretations, then it could consolidate a field by generating and then drawing researchers and ideas into conversation who might not otherwise have found one another. DiMaggio, Powell, and the neo-institutionalists push social engagement back further into cognition itself, suggesting that "uncertainty is . . . a powerful force that encourages imitation" (DiMaggio and Powell 1983, p. 151). Even before scientists or scholars have interpreted a vague or ambiguously worded article, they seek out the interpretations of others. This may lead to agreement and imitation, or to rejection, reversal, and polarization. Either path elevates the influence of early interpretations on later ones.

This process of ambiguity-induced engagement occurs in many social domains. The ambiguous, multivocal action of Cosimo de Medici strengthened alliances to conflicting parties and facilitated the coordination critical to the rise of the Renaissance State in Florence (Padgett and Ansell 1992). When contemporary nonprofits in the arts experienced uncertainty about their future, they tended to mimic visibly successful exemplars (DiMaggio and Powell 1983). Harrison White (1981) modeled this process for art dealers, whose uncertainty about artistic value led them to attend more to the behavior of other dealers than of their customers. Uncertainty-induced social convergence also occurs in the popular consumption of cultural goods such as CDs and DVDs (Elberse 2008), downloads and replays of online music (Salganik and Watts 2008), and other markets in which customers look to each other to determine quality.

A social response to ambiguity is consistent with research on neural reactions to uncertainty. Human gambling experiments demonstrate that uncertainty leads to generalized cognitive arousal (Schultz et al. 2008). Similarly, in a study in which adult subjects were asked to make sense of nonliteral, metaphoric sentences, the mesial frontal regions of the social brain were stimulated in addition to the classical language areas (Uchiyama et al. 2012). In both studies, uncertainty inspired cognitive activation, which enabled subjects to scan the social environment in an attempt to pragmatically resolve it.

A similar process occurs in scholarship as researchers, uncertain about what to study, converge around what to read, build on, and cite (Evans 2008). But ambiguity is not merely an inevitable and accidental outcome of scholarship. Ambiguity in scholarly discourse can be designed to produce individual uncertainty that staves off firm commitments, preserves deniability, and ventures more than one can credibly claim (Leifer 1983; Levine

1988). Skilled scholarly actors can also use multivocal performances to engender collective uncertainty, activate diverse audiences, and catalyze larger conversations (Ceccarelli 2001).[5]

AMBIGUOUS WORDS AS BOUNDARY OBJECTS

Despite vocal scientists and scholars who aspire to precise communication and unambiguous understanding, many frequently and wittingly deploy polysemic prose. Gillian Beer (2009) writes of Darwin's *Origin of Species*: "Darwinian theory will not resolve to a single significance nor yield a single pattern. It is essentially multivalent. It renounces Descartian clarity, or univocality. Darwin's methods of argument and the generative metaphors of *The Origin* lead . . . into profusion and extension. The unused, or uncontrolled, elements in metaphors such as 'the struggle for existence' take on a life of their own. They surpass their status in the text and generate further ideas and ideologies. They include 'more than the maker of them at the time knew'" (p. 9). Even Darwin's central phrase "natural selection" simultaneously suggests passive and active selection, stimulating 19th-century theists to appropriate the scheme as a manifestation of the Divine (Bowler 1989) just as others read it as the keystone proof of atheism (Dawkins 2016).

Scientists and scholars deploy ambiguity to further their position in the competition for advance. "The protection of one's meanings through ambiguously opaque utterance . . . [is] useful as a protective ploy in scientific competition" (Levine 1988, p. 218). A study of industry funding in oncology research found that methods sections of pharmaceutical-sponsored papers contained less detail and were more ambiguous about exact procedures followed, possibly to limit competitors' ability to trail or bypass company science (Knox et al. 2000). Researchers have also used ambiguity to suggest more than they know. This is a common strategy for the "discussion" section of a research article, in which partial, inconclusive, or unsubstantial empirical traces are woven together to suggest vague and deniable possibilities.[6] In fields where interpretive analysis is a central activity, increased ambiguity and the unspecified use of "essentially contested concepts" (Gallie 1955) such as art, culture, or democracy may facilitate the perpetuation of argu-

[5] Scientists and humanists use ambiguity not only to generate uncertainty but also to say more with less and cultivate evocative associations that increase the appeal and persistence of their ideas. In social theory, Davis exhorts that "to become a classic . . . it is not enough for a social theory to be true; it must also be seductive" (1986, p. 298).

[6] Davis argued that ambiguity facilitates the reinterpretation and appropriation of classics to new generations of students through class discussion. "The classical social theorist's very incoherence, which makes him so difficult for students to understand, allows teachers to fill classroom time by synthesizing his scattered and unrelated ideas into a coherent whole. . . . Ambiguity can stimulate teachers to try to synthesize a social theory" (1986, p. 296).

ments and articulations. At the limit, scientists and scholars engage in a form of scholarly mysticism when they fixate on criticism or the unresolvable paradoxes of a theory or argument (Gaonkar 1989).[7]

In works of synthesis, scientists and scholars attempt to speak to multiple audiences, breaking down barriers between disciplines while attracting broader attention to their work than any one audience could supply (Ceccarelli 2001). In *Genetics and the Origin of Species* (1937), Theodosius Dobzhansky used metaphors to address taxonomists, who studied evolution through museum collections, but also geneticists, who studied inheritance through experiments with lab populations, ambiguously harmonizing the two styles of research as "evolutionary dynamics" and "statics." In *What Is Life? The Physical Aspect of the Living Cell* (1944), physicist Erin Schrödinger inspired a generation of physicists and biologists to forge molecular biology in the 1950s by simplifying and ambiguously reframing core elements of each other's fields, which he "immediately retracted . . . in the footnotes for the benefits of those who possessed more sophisticated knowledge" (Ceccarelli 2001, p. 90). This ability to speak to different audiences in different ways is underscored even when it fails. E. O. Wilson's *Consilience: The Unity of Knowledge* (1999) attempts to colonize the content of the social sciences and humanities with methods from natural sciences, which is read very differently by scientists and humanists—conqueror and conquered.

In writing about the rhetoric of social science classics, Murray Davis argued that "ambiguity in social science is . . . crucial to the social theorist's appeal. An ambiguous theory can appeal to different—even if hostile—divisions in its audience, allowing each subgroup to interpret the theory in congenial, if mutually incompatible, ways" (1986, p. 296). But ambiguous expressions do not only passively link communities. They actively stimulate engagement and coordination.

In Starr and Greisemer's (1989) classic study of cultural diversity and cooperation underlying the founding of Berkeley's Museum of Vertebrate Zoology, they coined the term *boundary object* to describe essentially ambiguous objects, which simultaneously inhabit intersecting social worlds. Such objects, including abstract concepts like biological species, are flexible enough to maintain distinct meanings in different domains, but with a sufficiently common structure to make them mutually recognizable to more than one social world. These objects facilitate the translation of meanings

[7] Scholars in the ethnomedological tradition of sociology often seek to destabilize social and cultural categories imposed on social life by researchers in an effort to recover the true ambiguity experienced and expressed by native actors' categories. Ambiguity has similarly been deployed by post-structuralist literary critics like Jacques Derrida for intuiting texts' "'irreducible' multiplicity of meanings" (Ceccarelli 1998), such that work in this tradition become "a galaxy of signifiers. . . . The codes it mobilizes extend as far as the eye can reach, they are indeterminable" (Barthes and Balzac 1974, p. 6).

and interests from one world to another. Star and Griesemer argue that because "each social world has partial jurisdiction over the resources represented by that object, . . . mismatches caused by the overlap become problems for negotiation" (1989, p. 412) such that the "management of boundary objects is a key process in developing and maintaining coherence across intersecting social worlds" (p. 393).

Here we argue that ambiguous words and scholarly concepts represent critical boundary objects that facilitate communication and coordination between fields. Ambiguous words and phrases allow ideas to travel further, as they are transformed by new, receiving audiences, and word meanings are adapted to fit local, representational needs and circumstances. When ambiguous phrases persist in multiple fields, they function as bridges that help to maintain continuing scholarly engagement and exchange. The ambiguous phrase "social capital," used widely within sociology and economics, has sometimes emphasized the social value relationships represent (Bourdieu 1986; Coleman 1988) and at others the possibility of rational relational investment (Knack and Keefer 1997; Schultz 1961). This boundary concept has launched a generation of awkward, conflicted, but nevertheless coordinating conversations between sociology and economics (Portes 1998).

When scientists and scholars interpret an intriguing but ambiguously worded article, they may seek to reduce uncertainty about the reception of their interpretation by engaging the interpretation of others. Here we propose that more ambiguity in science and scholarship will, on average, lead to greater integration and less fragmentation of subsequent work that references it. Insofar as social learning occurs over time, we should see duration since publication interact with ambiguity to yield greater intellectual engagement. Moreover, following Stark's claim that ambiguity yields disagreement and drives productive engagement in science and scholarship as elsewhere, we should see that the diversity of fields citing a focal work interacts with its ambiguity and yields still higher levels of engagement.

We note what is at stake in this investigation. Historical and contemporary projects to increase precise scientific expression and diatribes against ambiguity do not anticipate benefits from ambiguity in science. Moreover, even sociologists who do anticipate the impact of ambiguity on engagement consider it for social and humanistic scholarship, but not for the natural sciences. In his analysis of sociological classics, Davis argues that while ambiguity and incompleteness can motivate social science researchers, "natural scientists, engaged in what Kuhn calls 'normal science', only 'solve puzzles', applying their theories inside an already defined range of topics to fill in the blanks like support troops who 'mop up' behind battle lines. . . . Ambiguity in the social science is not the embarrassment Kuhn finds it in natural science" (1986, p. 295). Similarly, Levine (1988, p. 218) suggests that vagueness does not characterize common communication in the sciences and that formal models are "rigorous and consistent," unsoiled from the ambiguities associated with "verbal models" of the social sciences. We predict that while ambiguity may be more common in some fields of science and scholarship than others, its effect will be consistent in all areas.

In this article, we focus on the lowest level of ambiguous discourse: the word. Most early social and natural scientists who articulated the need for precision articulated this in the form of lexical precision to bolster the association between words and concepts. New theoretical concepts are often encased in older, ambiguous words (e.g., "social capital"), infusing those words with new meanings and spawning generative, sprawling theoretical possibilities. We will show that lexical ambiguity, when considered in aggregate, reveals significant, consistent patterns that trace the lower bound of ambiguity's influence on the discursive arenas of science and scholarship. In order to test the relationship between ambiguity and fragmentation, this article introduces two measures, both of which are based on the information-theoretic notion of entropy. The first measure describes the ambiguity of a term as it occurs in a larger corpus (Yao et al. 2011) and is an extension of distributional similarity metrics from computational linguistics (Justeson and Katz 1991). The second measure applies the concept of graph entropy (Corominas-Murtra et al. 2010) to the citation networks of articles in the corpus and so describes the overall pattern of intercitation among an article's citers.

By focusing on ambiguous words and concepts, our approach neglects ambiguity at higher levels of expression. Syntactic ambiguity in a theoretical claim often influences multiple concepts, as in Wittgenstein's characterization of his own field and the subject of this article: "Philosophy is a battle against the bewitchment of our intelligence by means of language" (1973, sec. 109). Is language philosophy's weapon in the battle over intelligence or is it the primary agent of bewitchment? What is Wittgenstein saying? Theoretical uncertainty also occurs at the level of the pragmatic ambiguity of an entire argument. Why is Wittgenstein saying that? We do not attempt to measure these higher levels of expressed ambiguity here. But we argue that lexical ambiguity generates uncertainty regarding the conceptual building blocks of scholarship and so contributes to the perception of a message's overall ambiguity.

PRECISELY MODELING IMPRECISION

To formalize our notion of ambiguity and interpretive uncertainty introduced above we propose a statistical model of language generation that focuses on the association between words and meanings. Polysemy is fundamentally concerned with the uncertainty of meanings communicated in language. In our case, we are interested in quantifying the uncertainty of meaning imparted by any given word as encountered in a text. Let us first assume that it is possible to think about "meanings" in language the same way we think about words: as a set of discrete, enumerable, and identifiable objects $\mathbf{M} \in \{m_0, m_1, ...\}$. Similarly, we represent the set of lexical tokens (words) with $\mathbf{T} \in \{t_0, t_1, ...\}$ and the set of linguistic contexts (specified formally below) with $\mathbf{C} \in \{c_0, c_1, ...\}$. The problem of measuring lexical ambiguity can be positioned in terms of the probability that a particular token in a particular linguistic context communicates each of a set of possible meanings. Formally, this is represented by the conditional probability distribution across all meanings: $\Pr(\mathbf{M}|t_i, c_j)$. This distribution represents the set of meanings held by a word as used in the corpus, along with the relative probability for each of those meanings conditional on each linguistic context.

We turn to Shannon's information theory to translate such distributions of meaning probabilities into a single measure of ambiguity. The level of statistical uncertainty in a discrete probability distribution can be concisely characterized using information-theoretic or Shannon entropy. For any categorical random variable X taking values $x_0, \dots, x_k$ with probabilities $p_0, \dots, p_k$, its entropy is defined as[8]

$$H(X) = -\sum_{i=0}^{k} p_k \log_2(p_k).$$

In this context, entropy is simply the expectation of the information content of a single draw from a random variable, $H(X) = E(I(X))$, which allows for its construal as a measure of the unpredictability or uncertainty associated with that variable. As a measure of ambiguity, information-theoretic entropy concisely summarizes both the probability and diversity of meaning distributions. The entropy of a word's possible meanings efficiently models a reader's uncertainty about its sense in a given context.

Consider the phrase "the new line cook got cut early and went home." Suppose that the term *cut* has exactly two possible meanings in this context: either cut with a blade or cut from the job. Even a careful reader would not have much hope of working out what exactly happened to the poor line cook. Formalizing, let m_0 represent the first meaning (cut with a blade) and m_1 represent the second (cut from the job). We might say that in this context c_0 the two meanings of the word "cut" have equal probability:

$$\Pr(\mathbf{M} = m_0|\text{cut}, c_0) = \Pr(\mathbf{M} = m_1|\text{cut}, c_0) = 0.5.$$

Calculating the entropy, we obtain

[8] Any outcomes with zero probability are left out of the summation, informally letting $0 \cdot \log_2(0) = 0 = \lim_{p \to 0}(p \cdot \log_2(p))$.

$$H(\mathbf{M}|\text{cut}, c_0) = -0.5 \cdot \log_2(0.5) - 0.5 \cdot \log_2(0.5) = 1.0,$$

measuring one *bit* of entropy. Altering the linguistic context, even slightly, can change the situation dramatically, however. Hearing that "the new relief pitcher got cut early and went home," there is little doubt discerning between the two meanings of *cut*. In this new linguistic context c_1 the conditional probability distribution is much less balanced. For example, $\Pr(\mathbf{M} = m_0|\text{cut}, c_1) = 0.05$ and $\Pr(\mathbf{M} = m_1|\text{cut}, c_1) = 0.95$. The new entropy calculation yields

$$H(\mathbf{M}|\text{cut}, c_1) = -0.05 \cdot \log_2(0.05) - 0.95 \cdot \log_2(0.95) \approx 0.286$$

bits of entropy. "Cut" in the context of baseball is far less ambiguous than it is in the context of restaurants. By measuring the degree to which a word's probability mass is evenly distributed among possible meanings, information-theoretic entropy identifies ambiguity between words and the linguistic contexts in which they appear.

Building on this formalization, we specify a statistical model of ambiguity. Using Bayes's rule, the probability of a specific meaning, given a term in its linguistic context, can be expressed as

$$\Pr(m_k|t_i, c_j) = \frac{\Pr(t_i|m_k, c_j)\Pr(m_k|c_j)}{\Pr(t_i|c_j)}. \tag{1}$$

Formulating the probability this way simplifies the model considerably. The denominator, $\Pr(t_i|c_j)$, is fixed and represents the distribution of words in a given context. The first numerator term, $\Pr(t_i|m_k, c_j)$, captures the probability of a particular term, given a context and specific meaning, and $\Pr(m_k|c_j)$ the a priori probability of that meaning appearing in that context.

Semantics are fluid, however, and a model that casts meanings as distinct, equivalent categories seems to violate the nuance of language. Any attempt to enumerate "the set of all meanings" will necessarily be limited. Nevertheless, we argue that our model can reveal the contours of semantic uncertainty within a corpus by differentiating between some discretely represented meanings. It is important only that our set of meanings captures substantial semantic contrasts across the corpus. While there is no one correct way to discretize meanings, we propose a method that associates them with sets of words that encapsulate a particular semantic sense. Our identification of meanings is versatile enough to discern both subtle and stark distinctions in the connotations of encountered words.

Consider a semantic model that identifies meanings as tightly connected clusters of terms in a synonymy network built from curated, English-language thesauri. We construct a synonym graph G with one vertex for each word and an undirected edge (t_i, t_j) if t_j is listed as a synonym of t_i or

t_i is listed as a synonym of t_j in at least one thesaurus. We build on the intuition that terms in tightly interconnected clusters within this graph should overlap significantly in their semantic content and that an overlapping cluster of terms represents a particular meaning. The meaning associated with a set of terms can be thought of as the intersection of meanings for all terms in the set, suggesting that in at least one context, these words can be interchanged without a marked loss or alteration of meaning. In theory, such meaning sets could be identified as simply all of the k-cliques within the synonymy graph—any set of terms in which each is considered a synonym of all others. In practice, however, synonymy is imprecise. Thesauri vary substantially in their assertions (Blair et al. 2014). Appendix A describes a relaxation of the strict definition of a k-clique as meaning and outlines a methodology for estimating ambiguity based on such meanings. This model allows the straightforward analytic representation of the central probability distribution $\Pr(\mathbf{M}|t_i, c_j)$ mentioned above, but its explicit estimation presents significant computational challenges (again, see app. A for details). For this reason, we focus on a simplified, computationally efficient special case of this more general model, which we demonstrate is substantially informative regarding the presence and usage of ambiguity in writing.

In our simplified model, meanings are identified more tersely as dyads and neighborhoods within the directed synonymy graph.[9] A term with n synonyms is associated with $n + 1$ distinct meanings: one for each of its synonyms, which provide slightly different word senses, in addition to the term itself. Each of these words can substitute for t_i in at least one linguistic context. Specifically, for every term t_i in the aggregate thesaurus, let $S(t_i)$ represent the set of synonyms for t_i, including t_i itself. Then define one meaning cluster for each member of the set, $\{m_{ij} = [t_i, t_j] : t_j \in S(t_i)\}$, plus one containing the entire neighborhood, $m_{ii} = S(t_i)$. Furthermore, make a simplifying assumption about the probability of encountering term t_i in each of these contexts:

$$\Pr(\mathbf{T} = t_i|\mathbf{M} = m_{ij}, \mathbf{C} = c) = \Pr(\mathbf{T} = t_j|\mathbf{T} \in S(t_i), \mathbf{C} = c).$$

If we also let the a priori probability of a meaning occurring within a particular context be constant, $\Pr(\mathbf{M} = m|\mathbf{C} = c) = \Pr(\mathbf{M} = m'|\mathbf{C} = c)$, the calculation of the posterior probability from equation (1) is relatively simple. By assuming that a term and its synonyms each encode a slightly distinct word sense, this approach makes computation tractable. Ultimately, this estimates a term as more ambiguous or uncertain if its many senses are equally available in a given context such that an informed audience cannot predict which is intended.

[9] For the simplified model, we use a directed, rather than an undirected, synonymy graph. An edge (t_i, t_j) is included only if t_j is listed as a synonym of t_i in the thesauri.

Our model defines the few words preceding and following the occurrence of each word in our corpus as its *linguistic context*.[10] To calculate the posterior distribution of $\Pr(\mathbf{M}|t_i, c_j)$, we take a census of every time the same context appears in the corpus surrounding the term of interest or one of its synonyms. For terms that occur in many contexts, the corpus is searched for instances of that term or its synonyms occurring in any of the original term's contexts. Then we tally the frequency of each synonym in each context.

For the phrase "pikas don't hibernate through winter," if we were interested in the ambiguity or semantic uncertainty of *hibernate* (t_0) in this context, we would first identify *hibernate*'s synonyms in our thesaurus. These include *slumber* (t_1), *kip* (t_2), *rest* (t_3), *nap* (t_4), *sleep* (t_5), *bundle* (t_6), *estivate* (t_7), and many more, but we will imagine that it possesses just these to simplify our example. Scanning the corpus, we might find that the original phrase occurs five times, while "pikas don't sleep through winter" occurs twice, "pikas don't slumber through winter" occurs only once, and no other synonyms occur in this context anywhere. This leaves us with a frequency vector across eight terms (t_0 through t_7): (5, 1, 0, 0, 0, 2, 0, 0).

A simple count of occurrences like this can be modeled as a random draw from a multinomial distribution over the term's synonyms, or, according to our model, as a distribution over the set of possible meanings. Under this scheme, each time a term appears in its context, it is understood as a random sample from that unobserved categorical distribution. As each word occurs in a corpus it can be interpreted as the realization of a probabilistic process in which the word itself is selected over its approximate synonyms. In the example about pikas above, "hibernate" was pulled from a hat five times, "sleep" twice, and "slumber" once. Using these counts, it is straightforward to calculate a posterior distribution of the entropy associated with $\Pr(\mathbf{M}|t_i, c_j)$ in our corpus.[11] Note the asymmetry in this model. The term "sleep" will have a different, but likely overlapping, set of synonyms com-

[10] In practice, words in the corpora used are lemmatized (stemmed and normalized) and tagged with their parts of speech category (noun, verb, adjective, etc.). Thus the contexts consist of sets of lemmatized terms coupled with their part of speech. Furthermore, word order of occurrence in context is not considered, so a term's context should be understood as an unordered set of surrounding words. In this way, more of each context's content is captured, and more contexts in total register synonyms.

[11] Making the translation from existing textual data to entropy measures requires a correction to the observed frequencies. For example, an uncommon word may not appear frequently enough in any given context for us to observe it replaced by one of its synonyms in our corpus. Calculating substitution entropy from frequency alone, we would conclude that there is no probability that it could be substituted by the known synonym. This would artificially deflate its substitution entropy relative to more frequent words, failing to accurately measure the term's ambiguity. To correct for this, we calculate the posterior distribution of entropies for any given frequencies of synonym occurrence. See app. B for details.

pared with "hibernate" in a typical thesaurus. This derives less from sloppiness and more from cognitive asymmetries in analogy. For example, specific terms are more likely perceived to be the synonyms of more general, "prototype" terms than the converse (Tversky 1977). Thus an evaluation of the likelihood of synonym substitution within the context "pikas don't _____ through winter" will likely yield different results for "sleep," "slumber," and "hibernate."

To illustrate how this ambiguity measurement operates on an actual abstract, figure 2 highlights four terms from the abstract of a 1994 paper "Weld Lines and Mechanical Properties of Injection Molded Polyethylene/Polystyrene/Copolymer Blends" (Brahimi, Ait-Kadi, and Ajji 1994). These include two of the most precise and two of the most ambiguous terms from the abstract, listing the frequency of the terms' synonyms as they occur in the same subfield between 1989 and 1994. The terms "mechanical" and "properties" were among the most precise measured in the abstract, "me-

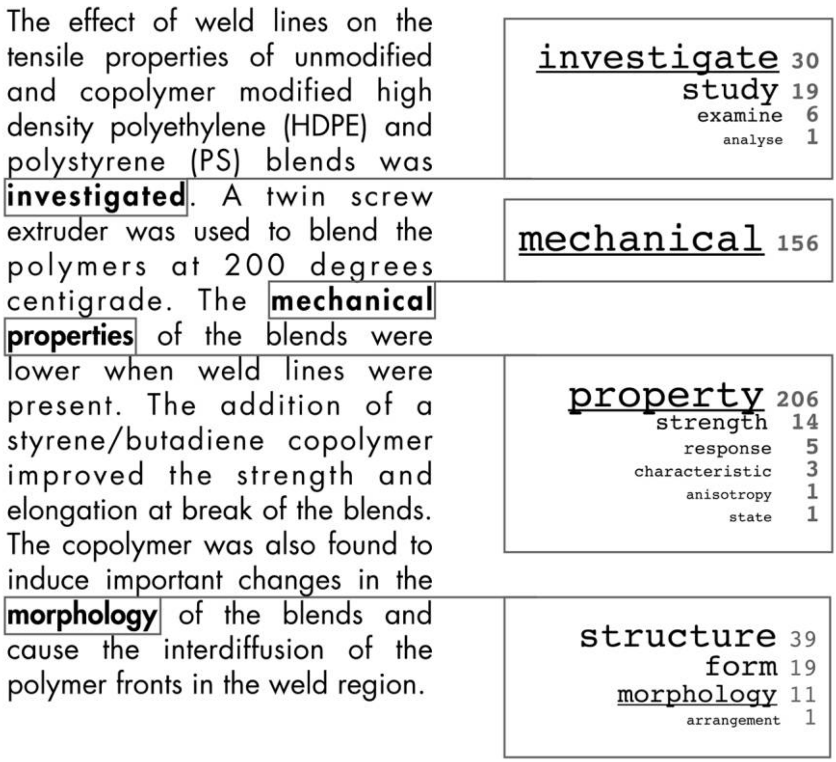

FIG. 2.—Sample abstract highlighting four words for which ambiguity is measured. "Mechanical" and "properties" are measured as relatively precise while "investigate" and "morphology" are measured as relatively ambiguous (Brahimi et al. 1994).

chanical" in fact being the unique member of its synonym set that was used within the relevant subsample. "Investigated" and "morphology," in contrast, are measured as much more ambiguous: none of the terms' synonyms are used much more frequently than any other, leading to an assessment of semantic uncertainty in context for the informed reader. The (stemmed) term "investigate" accounts for barely half of the substitutions for the first term, in contrast with "property," which is used almost 90% of the time in this context.

AMBIGUITY VALIDATION

We sought to validate our measure of linguistic ambiguity against human judgments of ambiguity by conducting a survey that asked respondents to rate the ambiguity of specific terms from a corpus of *New York Times* articles. Using the simplified model described above, we calculated ambiguity across the corpus for millions of terms-in-context. Details are recounted in appendix C (table C1). Summarizing this process, 102 respondents, recruited through Amazon's Mechanical Turk service, were asked to rate their individual uncertainty on the basis of whether or not they were confident that they understood what a given term meant in the context of a displayed sentence, with answers ranging from 1 to 7 on a Likert scale anchored by "extremely uncertain (ambiguous)" and "extremely certain (precise)." The survey used 1,020 sentences, each appearing in exactly three respondents' surveys, with items randomized to provide maximal overlap between respondents.

In order to handle the possibility of high variability between respondents, we employed a multilevel generalized linear model to estimate the relationship between our measure and respondent rankings of ambiguity. Estimates in table 2 below reveal a significant and substantial overall relationship between our measure of ambiguity and the rating reported by respondents. For every standard deviation increase in measured ambiguity, respondents' expected rating increases by an average of 32%, with a strong overall model fit measured by a pseudo-R^2 of .598.

This analysis adds support to the validity of our measure of ambiguity. The model detailed in appendix C does a good job of predicting human rankings of term ambiguity using only our entropy-based measure and random-effects terms as predictors. Respondents that identified ambiguity in survey items tended to strongly agree with our metric, which suggests that we capture most of the signal of individual uncertainty. This represents a lower bound of the total ambiguity we seek to measure. Variability in respondents' sensitivity to ambiguity suggests the possibility that we may also capture some proportion of collective uncertainty as well. Identifying ambiguity is difficult for some people and assumes enough relevant language exposure that many are oblivious to it. Many of our survey participants did not recognize significant ambiguity in any of the items provided them, consistently reporting individual certainty in their understanding of words surveyed, even though these individual certainties likely disagreed with one another. While our survey did not allow us to validate the potential for our measure to identify collective uncertainty, we nevertheless establish a strong signal of individual uncertainty from our survey respondents.

MEASURING INTEGRATION AND FRAGMENTATION

The second measure used in our analysis describes the structure of the graph of citations received by an article and its "descendants." It differentiates highly coherent citation networks that reflect mutual awareness and engagement from fragmented networks in which disparate branches of citations do not cite one another from obliviousness, irrelevance, or opposition. We use the notion of graph entropy described by Corominas-Murtra and Solé (2010) to characterize this particular notion of coherency. The intuition behind this measure rests on the idea of path reversibility in directed acyclic graphs (DAGs), sometimes called feed-forward networks. If a particular article's citation graph G has n maximal or "leaf" nodes and, by definition, one minimal or "root" node, then $\mathcal{H}(G)$, the graph's entropy, characterizes the level of uncertainty that would be encountered if trying to trace a path backward from any leaf to the root.[12] More formally, $\mathcal{H}(G)$ quantifies the amount of information that would be needed to unambiguously specify a path from each terminal citing article to the root cited article.

This measure succinctly describes the degree of fragmentation in a citation structure. Figure 3 shows three hypothetical citation graphs with their graph entropy scores. The leftmost graph is a simple tree in which each article cites exactly one other article in the graph and for which the entropy takes its theoretical minimum (zero). The center and right graphs illustrate two different ways of distributing 13 citations among the nine articles. The center graph displays two distinct clusters of articles, with high within-cluster but no between-cluster citations. The rightmost graph shows a graph with a much more even distribution of citations. Because entropy describes the amount of path certainty between the root and leaf nodes, it measures the extensiveness of intercitation in the citation graph. A graph with lower

[12] The entropy $\mathcal{H}(G)$ is defined as

$$\mathcal{H}(G) = \sum_{v_i \in V \setminus M} n^{-1} \sum_{v_k \in V \setminus M} \phi_{ik} \log(d_{\text{in}}(v_k)),$$

where V is the set of vertices in G, M is the set of maximal nodes, ϕ is the graph's transition matrix, and d_{in} is the in-degree of node v.

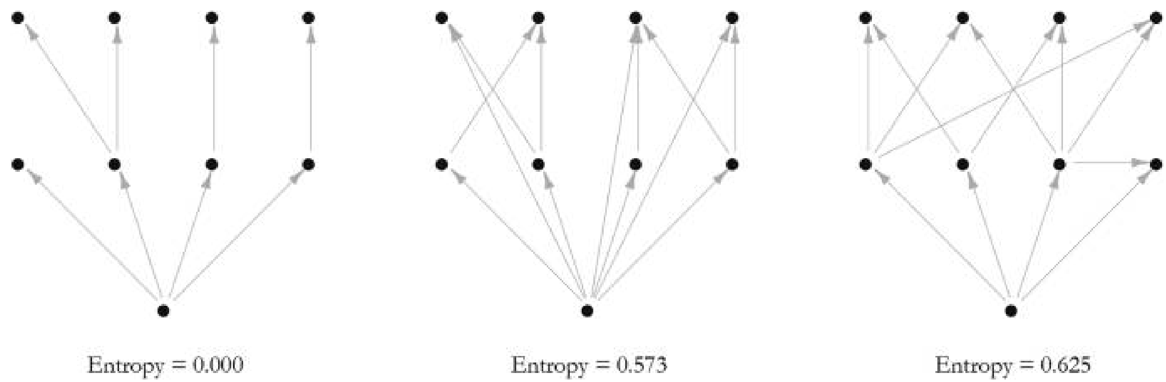

FIG. 3.—Three sample DAGs and their entropy measures (H). Trees such as the leftmost graph always have the theoretical minimum entropy of zero. The center and right graphs have the same numbers of edges, but the center graph has much more distinct clusters of interciting articles and thus a lower entropy.

entropy will be conditionally more predictable in its citation pattern. Thus, the middle graph, with clear "camps" of citing articles, has a lower entropy than the rightmost one in which all of the articles are engaged in the same conversation.

The calculation of topological entropy is computationally expensive, and so we explored its relationship with an alternative, more efficient measure of citation clustering, *maximal network modularity* (Newman 2006). Modularity is the fraction of edges that fall within the groups produced by a network partitioning, minus the expected fraction if edges were distributed at random. The network partitioning is chosen to maximize this quantity, and so high maximal modularity suggests a network in which ties within groups far exceed what would be expected at random. Shwed and Bearman (2010) used the modularity of citation networks to identify consensus periods in the scientific literature surrounding debate. They found that peaks in modularity corresponded to cliques whose members cite each other but who do not cite their epistemic rivals. Dips in modularity, by contrast, identified broad periods of agreement typically before any expert study proclaimed consensus. We calculated the correlation of topological entropy and modularity across three-level citation graphs for a random sample of 1,235 scientific and scholarly abstracts, across all years, and found a strong, significant negative correlation of $-.7559$ ($P < .001$):[13] high topological entropy equates with low maximal modularity. As a result, we calculated maximal modularity as a fragmentation measure for the remainder of our citation networks.

TEXT AND CITATION DATA

We used the methodology just described to analyze a corpus of 1,942,424 academic article abstracts from all fields of science and scholarship included in Clarivate's Science, Social Science, and Humanities Citation Indexes between 1974 and 1995 (inclusive)—hereafter referred to as the Web of Science. To calculate discipline-level ambiguity, abstracts were assigned to at least one of 14 subject areas: agriculture, biology, business and management, chemistry, computers and information technology, engineering, environmental and earth sciences, humanities, law, mathematics, medicine, multidisciplinary sciences, physical sciences, and social sciences.

Abstracts were tokenized and lemmatized, and each term was tagged with its part of speech drawing on the Penn Treebank tag set (Marcus, Marcinkiewicz, and Santorini 1993). Any term that occurred 100 or fewer times in the corpus was replaced with a simple "rare term" marker, with part of

[13] We use the agglomerative method of Clauset, Newman, and Moore (2004) to maximize modularity.

speech preserved (Yao et al. 2011). After the text was transformed in this way, the data set was constructed by indexing each term that occurs both in the corpus and as a headword in the thesaurus, together with its four-word context. The current analysis uses the union of two thesauri: the unmodified Moby Thesaurus II, an extensive public-domain thesaurus, and WordNet, a large networked dictionary that organizes words into semantic "synsets."[14] Supporting analysis revealed remarkably little sensitivity to the thesaurus used. Many terms used in the abstracts do not occur in the combined thesauri, leading to potentially unreliable results for abstracts in which ambiguity can be measured for only a handful of terms. To mitigate this potential bias, the regression analyses below were performed only on the subset of abstracts with at least five usable terms.

Article-level ambiguity was calculated for our sample of abstracts by comparing each abstract to the complete corpus of articles from a five-year window within the same detailed subject area. The Web of Science assigns up to seven nonexclusive subject classifications or "journal categories" from a collection of 325 to each journal in the sample. Our ambiguity measurements are therefore determined in the context of articles published in similar journals and similar time frames. The citation graph of these articles was gathered by recursively following citations for each article to a maximum of three iterations and up to 1,000 citations. We also required the graphs to have at least 15 "nonleaf" nodes, so that a singular or sparse chain of citations would not lead to a highly cited downstream article whose citations mischaracterize the influence of the focal, upstream article. This led to a removal of approximately 40% of the sample. The final sample used for the analyses contained 1,101,766 articles.

DESCRIPTIVE STATISTICS

Our assessments of ambiguity are imperfect in that they do not capture the entire lexicon of scholarship and science, but rather the precise or ambiguous use of words common to the broader English language. Nevertheless, visual inspection of many randomly selected articles demonstrates that our measures capture field- and author-inspired differences in the ambiguous use of language. Consider three more precise (P1–P3) and three more ambiguous (A1–A3) abstracts:

Abstract P1.—"Herbicides were evaluated for control of field violet in strawberries in North Carolina and in Nova Scotia, Canada. DCPA at 11.0 kg ai ha^{-1} and terbacil at 0.5 kg ai ha^{-1} PRE controlled field violet 80 to

[14] Moby Thesaurus II is available at http://www.gutenberg.org/ebooks/3202. WordNet is available at http://wordnet.princeton.edu/.

95%. Control with simazine PRE at 1.0 kg al ha^{-1} was 70 to 75%. Oxyfluorfen at 0.6 to 1.1 kg ai ha^{-1} controlled 95 to 100% of the established field violet in dormant strawberries following a mid-winter application. Control with acifluorfen and lactofen applied during mid-winter was 75 to 90%. Oxyfluorfen applied to dormant plantings at rates in excess of 0.60 kg al ha^{-1} caused foliar damage; however, yield was not reduced at rates up to 4.5 kg ha^{-1}. Oxyfluorfen at 0.25 kg ha^{-1} caused severe injury to newly-transplanted strawberries" (Doohan, Monaco, and Sheets 1993, p. 185).

Abstract P2.—"A behavioral study was performed in an attempt to understand the neurological mechanism involved in yawning in rats. Injections i.p. of low doses (0.25 mg/kg) of apomorphine, which preferentially activate presynaptic dopamine autoreceptors, elicited yawning. Whereas apomorphine, at a high dose of 2 mg/kg, produces stereotypy which was thought to be mediated by stimulation of postsynaptic dopamine receptors. The yawning and stereotypy did not occur simultaneously in the rat. The apomorphine-induced yawning was completely inhibited by pretreatment with fluphenazine (9 mg/kg, intermuscular) or scopolamine (0.5 mg/kg i.p.), but markedly increased by reserpine (5 mg/kg, s.c.), however it was not affected by methylscopolamine (0.5 mg/kg, i.p.). Both physostigmine (0.2 mg/kg, i.p.), an indirect acetylcholine agonist, and pilocarpine (4 mg/kg, i.p.), a direct acetylcholine agonist, also induced yawning. This was abolished by scopolamine (0.5 mg/kg, i.p.) and increased by reserpine (5 mg/kg, s.c.). Fluphenazine (9 mg/kg, i.p.) did not affect the pilocarpine-induced yawning but increased the physostigmine-induced yawning. Apomorphine elicits yawning by stimulating presynaptic dopamine receptors and dopaminergic inhibition and cholinergic activation are concomitantly involved in the yawning" (Yamada and Furukawa 1980, p. 39).

Abstract P3.—"A national sample of 1944 white menopausal women greater-than-or-equal-to 55 years old from the epidemiologic follow-up of participants in the National Health and Nutrition Examination Survey was reviewed to investigate the role of hormone therapy in altering the risk of death from cardiovascular disease. Women in the study were observed for up to 16 years after the baseline survey in 1971 to 1975. By 1987 631 women had died; 347 of these deaths were due to cardiovascular disease. History of diabetes (relative risk, 2.38; 95% confidence interval 1.73 to 3.26), previous myocardial infarction (relative risk, 2.12; 95% confidence interval 1.56 to 2.86), smoking (relative risk, 2.18; 95% confidence interval, 1.69 to 2.81), and elevated blood pressure (relative risk, 1.49; 95% confidence interval, 1.14 to 1.94) were strong predictors of cardiovascular disease-related death in this cohort. After adjusting for known cardiovascular disease risk factors (smoking, cholesterol, body mass index, blood pressure, previous myocardial infarction, history of diabetes, age) and education, the use of postmenopausal hormones was associated with a reduced risk of death from cardiovascular disease (relative risk, 0.66; 95% confidence interval, 0.48 to 0.90). The same protective effect provided by postmenopausal hormone therapy was seen in women who experienced natural menopause (relative risk, 0.69; 95% confidence interval, 0.45 to 1.06)" (Wolf et al. 1991, p. 489).

Abstract A1.—"Welfare is a measure of the well-being of society. It is sustainable if it can be maintained steadily for many years. According to neoclassical economic theory, value—that is, the flow of welfare attributed to the totality of available outputs—is generated, distributed, and consumed solely within the economy. When environmental externalities are absent, or after they have been internalized, environment-derived welfare is constant, and no environmental costs need be incurred to sustain economy-derived welfare. Hence, resource accumulation and technological progress enable society to experience sustainable development futures where its welfare always increases, possibly exponentially. Thermodynamics generates an alternative theory of value. Drawing on this theory, I show that the sum of the economy-derived and environment-derived welfare is not affected by the location of the boundaries between the economy and the environment. Hence, to recognize their contribution to welfare, environmental resources need not be internalized into the economy. According to thermodynamics-based theory, economy-derived welfare may be sustained only because the economy is able to transport net value from its environment to restore the value that is necessarily consumed within it. Hence, the environmental costs of sustaining welfare are always positive. They increase with the welfare to be sustained and decrease as a consequence of technological advance. Since part of these costs will never disappear, the room for better technologies cannot be exhausted. Once solar power is fully utilized, future increase of welfare will be guided by the rate of adoption of more efficient technologies that slow down the economic process. Although welfare is bounded asymptotically and its steady growth is precluded, economy-derived welfare and the corresponding environmental costs are determined by social decisions rather than by immutable laws of nature" (Amir 1995, p. 27).

Abstract A2.—"The present column lists commercially available zinc and zinc alloy reference materials. Pure zinc, various zinc base alloys, zinc ores and concentrates, and setting-up samples have been considered. Included are three tables that provide an easy-to-use survey. The following information is covered: the name of the material, the sample code, the producer, the reference to certification, the names and addresses of the suppliers from whom the reference material may be obtained, and specific remarks" (Roelandts 1993, p. 461).

Abstract A3.—"Using Bi-Digital O-Ring test resonance phenomenon, accuracy of the widely used organ representation areas, currently used in different schools of foot and hand reflexology was evaluated. In general, results show that when a specific organ is abnormal and its function is abnormally reduced, the size of that organ representation area on the foot and hand diminishes. When codeine is taken, organ representation area on the foot and hand diminishes, particularly the areas where codeine is deposited. Even if no abnormalities are found by visual examination or palpation, the Bi-Digital O-Ring Test shows abnormality in the corresponding organ representation area" (Omura 1994, p. 153).

The first three abstracts are riddled with particularist jargon that seemingly can be understood in no way other than intended: "Intrinsic neurons in the interhemispheric cortex (IHC) were studied by the rapid Golgi method in the young mouse." The second set of three abstracts suggest a number of places that feel less precise: "It is sustainable if it can be maintained steadily for many years." "Organ representation area on the foot and hand diminishes." Sustainable? Organ representation? Note that these entropies were established within the context of other papers from the same field and era. As a result, high-entropy/ambiguity abstracts should feel semantically looser not only to the lay reader but also to those with experience in those fields.

Figure 4 provides another illustration of the validity of our ambiguity measure. This figure plots six terms whose ambiguities vary by field. Circles represent the point estimates of ambiguity and dots represent actual instances of the terms on which the estimates are based (jittered randomly in the vertical dimension). Fields with too few occurrences of a term for reliable estimation are omitted. This figure shows that fields in which a term is "native" tend to have lower ambiguity or entropy for that word than other fields. For example, "beam" (panel A) is much more precisely used in the physical sciences and engineering than it is in agriculture, medicine, or the social sciences. A "beam" in the physical sciences refers most often to a ray of particles or energy rather than a physical structure. "Prove" (panel B) has a much more precise and consistent meaning in mathematics and computer science than in the biological sciences. Finally, "subject" has a particular meaning in the social sciences and humanities and a different precise meaning in medicine and biology rather than its broad semantic usage in other fields.

Figures 5 and 6 illustrate the usage of our citation fragmentation measure in the context of the science, social science, and humanities data in the Web of Science. Figure 5 graphs 12 randomly drawn examples of low-fragmentation citation graphs, with network modularity scores less than 0.35. These citation graphs, typically three citation "generations" deep, all reveal many cross-cutting citations and the lack of discrete submodules. By contrast, figure 6 graphs 12 high-fragmentation citation graphs, with network modularity scores greater than 0.65. Many of these graphs are perfect (or near-perfect) trees. They are highly modular, exhibiting strong, exclusive clusters of co-citing articles.

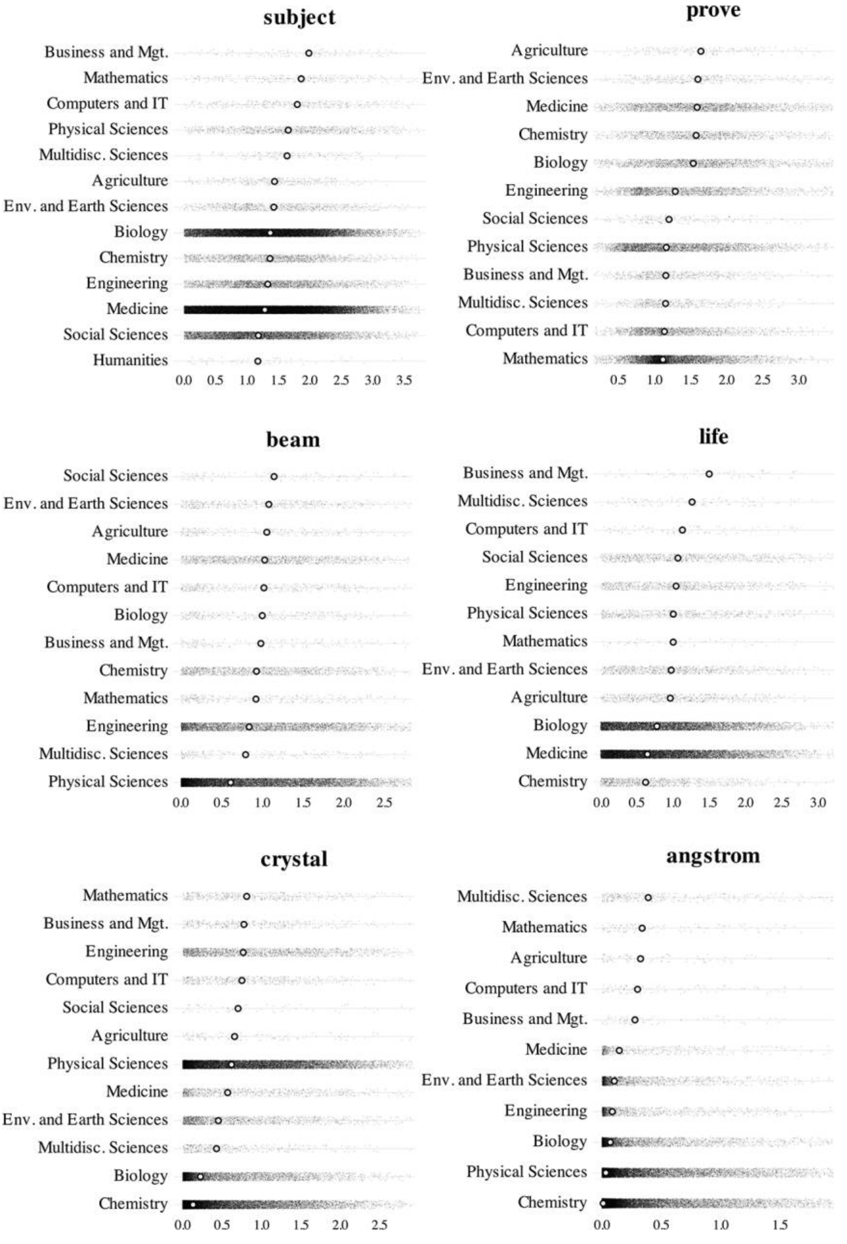

FIG. 4.—A sample of terms and their distribution of ambiguity across various disciplines. The circles represent the mean ambiguity of a term in a field, and the scattered dots represent the actual individual ambiguity of each term (randomly jittered in the vertical dimension for greater visual discrimination).

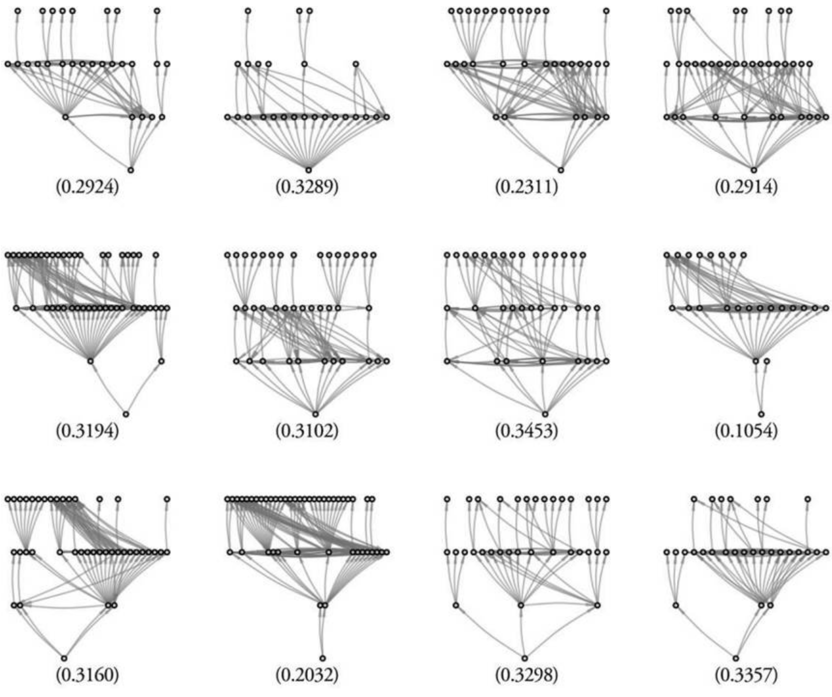

FIG. 5.—Random sample of low-fragmentation/high-entropy graphs

Cross-cutting citations are rare, and when they do occur, they minimally violate the closed conversations among citing articles.

MODELING STRATEGY AND AUXILIARY VARIABLES

Our modeling strategy uses linear regression to predict citation modularity ("fragmentation") with synonym entropy ("ambiguity"). We control for a number of auxiliary quantities that might influence citation fragmentation for other reasons. Our hypothesis is that scientific and scholarly communication induces ambiguous articles to have more integrated and less fragmented citation graphs. This effect may be moderated by the presence of distinctive fields. More fields may lead to integration within but not across fields, such that fragmentation rises with the number of articles from different fields participating in the citation graph. We captured this dynamic by calculating the number of different fields represented among the journals, as classified by the Web of Science journal categories described above, and also the entropy of articles across these subject areas. In this *subject entropy* measure,

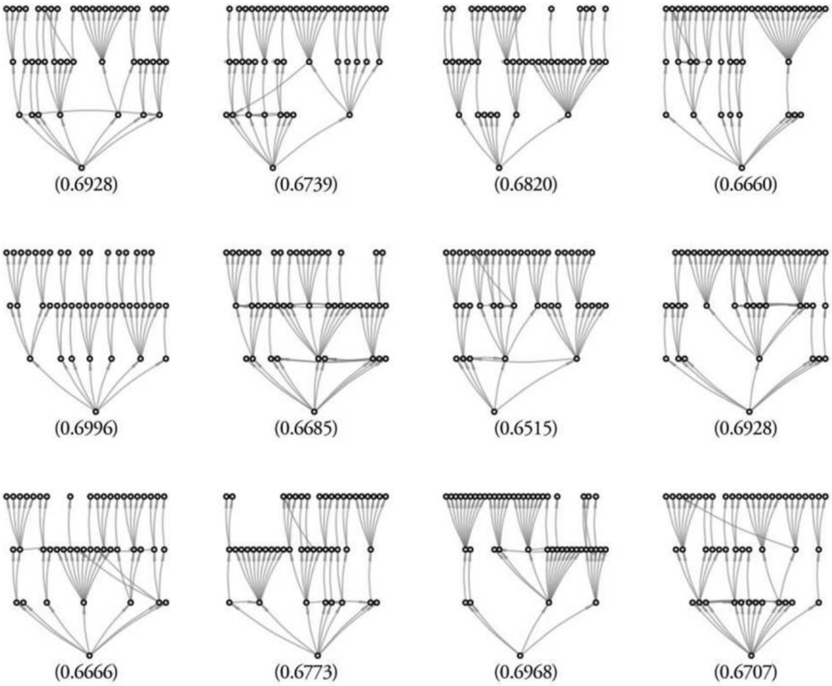

FIG. 6.—Random sample of high-fragmentation/low-entropy graphs

citation graphs dominated by journals from a single field like "ornithology," for example, would have a low subject entropy, and those evenly referenced by articles from journals in several fields would have high entropy.

We also captured the influence of subjects on the structure of citations with a more direct measure that assesses the modularity of the citation graph if partitioned into subject-specific articles. Precisely, we calculated the fraction of citations that fell within subject-specific articles (e.g., one "condensed matter, physics" article citing another), minus the expected fraction if citations were distributed at random. We call this *subject citation modularity*, and we expect that more modularity by subject will naturally lead to more modularity or fragmentation overall.

The temporal distribution of articles in a citation graph will also likely affect its consolidation or fragmentation. If articles are published many years apart from one another, later authors might be less aware of earlier work and cite it more selectively, leading to greater fragmentation overall. To capture this dynamic, we introduce a variable that tallies the number of distinct publication years present among articles in the citation graph and a second that calculates the variance of years in that distribution in order to capture the degree to which publication is spread over longer periods. We expect that *years* and *year variance* will increase citation fragmentation.

We also incorporated an interaction between *ambiguity* and *subject citation modularity*, *subject entropy*, and *time variance* to test how ambiguity influences researchers crossing field barriers and time periods. Our expectation is that ambiguity leads to greater consolidation within fields and time periods. Following Stark (2011), we anticipate that more citing fields and greater ambiguity will be associated with increased citation integration as interpretations across field boundaries are experienced as distinctively informative to one another. One could also imagine the opposite effect in which ambiguity leads to greater integration within but not across fields because communication and mutual awareness are impeded.

Broad interdisciplinary differences present a potential confounder to our model. It is conceivable that just as lexical ambiguity varies between disciplines (as shown below in fig. 7), different disciplines tend to have more or less fragmented citation graphs on average. If more ambiguous disciplines have less fragmented citation graphs but the ambiguity-fragmentation link is not present within fields, then our anticipated findings would be valid only across disciplines and not within them. We therefore include dummy variables in our analysis for each of the 14 broad disciplines mentioned above.[15]

Our model could also be confounded by the number of coauthors listed on a particular article. If many authors contribute to the writing of a paper, each with their own predilections for vocabulary and style, it is easy to imagine such a paper being dominated by polysemous words and phrases. A research project with many authors could also be naturally more likely to engage diverse academic communities. We therefore control for number of authors in all of our regression models.

Another important possible confounder is a potential "career effect." One could imagine that as authors advance in their career they may experience coinciding changes in both their writing style and the pattern of citation their work provokes. Senior scholars may find it easier to receive citations while simultaneously becoming less meticulous and more ambiguous in their argumentation. Unfortunately the Web of Science data contain inadequate information on article authors to account for this potential confounding in the model. Each author in the database is represented only by a last name and first initial, which makes precise disambiguation between all authors in the sample impossible. To check the confounding effect of maturing authors, we culled the data set using conservative rules to identify names less likely to refer

[15] Recall that discipline assignments are at the level of the journal and are nonexclusive. Although about three-quarters of the articles are assigned a single disciplinary label, two- and three-discipline journals are not uncommon, and about 1% of the articles in the sample are labeled with four or more disciplines.

to more than one author (e.g., names with fewer than 30 publications over less than 25 years and associated with fewer than 10 organizations). Such culling dramatically reduces the number of observations in the sample and selects on a number of potentially relevant factors. The resulting subsample is inappropriate for final analysis, but it does allow author tenure to be included in the model. Including a linear and quadratic term on the number of years since first publication did not substantively change our main findings presented below: that more ambiguous abstracts are associated with more integrated citation graphs. We are therefore confident that author seniority is not a significant confounder and is safe to omit from the full model.

Finally, we controlled for two quantities we observed to vary with our core dependent and independent variables: the number of terms in the focal abstract on which ambiguity was calculated and the number of nonleaf articles in the citation graph on which fragmentation was calculated.[16] Summary statistics of these variables are included in table 1.

FINDINGS

Figure 7 reveals striking differences between the precision of language in different fields. The average ambiguity of terms in each article is calculated, and the distribution of these average ambiguities is summarized for different fields of research. Diamonds represent the mean ambiguity for an article in the field, while the horizontal lines represent interquartile ranges. Note that differences between most adjacent field-level distributions are significant at the 0.05% level, using pairwise one-sided Kolmogorov-Smirnov tests, and every alternate field distribution in figure 7 is significantly different at the 0.01% level. (See fig. C1 in the appendix for a similar visualization disaggregating subject differences at the term rather than article level and in which every successful field is significantly different from every other.)

Humanities, law, and environmental and earth sciences exhibit the highest ambiguity and biology, medicine, and chemistry the lowest. This means that natural, common language—specifically, the terms present as headwords in our thesauri—is used most consistently in the biomedical and chemical sciences and least consistently in the humanities, law, and business. The social sciences and multidisciplinary sciences are in the middle. Recall that technical terms and formalisms within articles are unaccounted for with our measure. As a result, mathematics and engineering are measured as less precise in their use of natural, common language than the biomedical and chemical sciences, which use natural language to consistently convey much of their subject-

[16] As stated above, these two measures (number of terms with calculated ambiguity and number of nonleaf articles in the citation graph) were used to create the subsample for analysis.

TABLE 1
DESCRIPTIVE STATISTICS OF UNTRANSFORMED VARIABLES

Variable Name (Definition)	Min	Mean	Max	SD	1Q	Median	3Q
Dependent:							
Fragmentation (citation modularity) . . .	.10	.74	.92	.09	.69	.75	.80
Independent:							
Ambiguity (synonym substitution entropy)	.00	4.55	10.49	1.07	3.86	4.57	5.27
Subject modularity of citations	−.30	.05	.40	.05	.02	.05	.08
Entropy of subject counts.	.00	2.64	4.27	.51	2.37	2.71	3.00
Control:							
Year variance	.47	9.41	104.88	6.93	5.17	7.34	11.11
Observed terms in abstract.	5.00	9.20	173.00	5.05	6.00	8.00	11.00
Nonleaf articles in citation graph.	15.00	59.90	382.00	34.47	33.00	53.00	79.00
Number of authors	1.00	6.04	719.00	7.02	3.00	5.00	8.00

NOTE.—1Q and 3Q refer to the first and third quantities, respectively.

specific content (e.g., “protein x binds with protein y”). Also note that books, the central scholarly outlet for humanistic research, were excluded from this analysis but are widely considered a place in which more speculative, broad-brush arguments can be made than in articles. As a result, the field-level distribution of ambiguity shown in figure 7 is more concentrated and our assessment of disciplinary difference is more conservative than it would likely be if more completely captured. If equations and books were included in the analysis, mathematics and the physical sciences would likely prove more precise and the humanities more ambiguous.

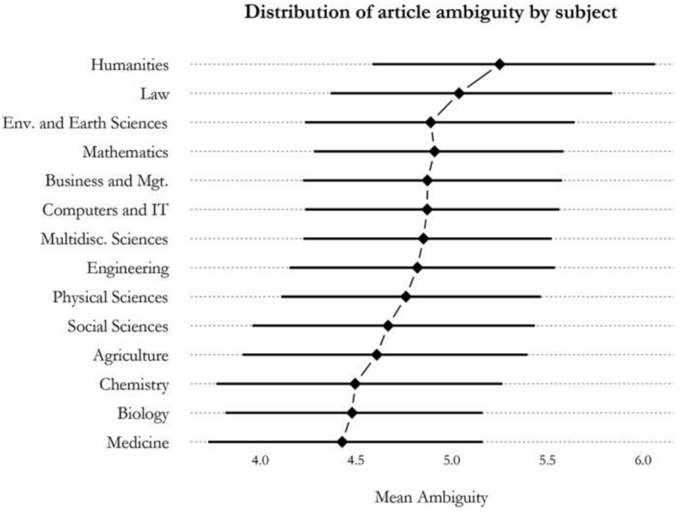

FIG. 7.—Median and interquartile range of abstract ambiguity across subjects

Table 2 summarizes results from seven models that assess the relationship between ambiguity and fragmentation.[17] All seven models exhibit that greater ambiguity leads to less citation fragmentation among articles in the citation graph. In model 1, with only the controls as covariates, this effect is −0.01373, meaning that a 1-SD increase in an article's average ambiguity leads to an expected decrease of about 1.4% SD in (the transformed) fragmentation. In models with more covariates, the coefficient on ambiguity ranges from a minimum of −0.01841 to a maximum of −0.01354. These effects are highly statistically significant ($P < 10^{-8}$) but comparatively modest in overall size: several of the (standardized) covariates have coefficients five to 10 times greater in magnitude than the coefficient on ambiguity. Nonetheless, an article's ambiguity maintains an important negative effect on the fragmentation of its resulting citation graph.

When number of journal subjects, subject entropy, and subject citation modularity are added to the model (models 2–5), they all post a positive influence on citation fragmentation. When more fields, a wider distribution of fields, or citation-insulated fields cite a core article, they are likely to increase the modularity or discipline-wide clustering of cites in that citation graph. The coefficients on these variables range between 0.03 and 0.43.

As anticipated, articles published in a larger number of distinct years in the citation graph, like the number of subjects, increases fragmentation of the graph. Somewhat surprisingly, an increase in the variance of those years decreases citation fragmentation such that a wider distribution of years contributes more to a coherent citation structure than a closely packed year distribution.

Subject citation modularity, when interacted with ambiguity in models 2 and 3, decreases the fragmentation of the citation. This means that when articles from new fields cite the focal article or one of its citation "branches," they are also more likely to cite articles from other branches of the citation tree and so integrate the citation graph if the focal article is more ambiguous. The same pattern holds, though to a somewhat weaker extent, with subject entropy (models 4 and 5). Increased interdisciplinary interest in an article intensifies the cohering effect of ambiguity and the fragmenting effect of precision. This supports the Stark (2011) proposition that diversity begets engagement but nevertheless remains surprising and underscores the power of uncertainty and the ethic of caution in scientific and scholarly life. If a scientist or scholar is citing an ambiguous article from a different field, then she is more likely to cite different branches of that article's citation structure, which likely correspond to distinct takes on the focal article.

[17] All variables in the linear regressions are standardized to have a mean of zero and a standard deviation of one. The outcome variable, fragmentation, is first transformed using a logistic transformation to be appropriate for a linear model.

TABLE 2
OLS Coefficient Estimates for Models Predicting Citation Fragmentation (Logistic Transformation)

Variable (Standardized)	Model 1 (×100)	Model 2 (×100)	Model 3 (×100)	Model 4 (×100)	Model 5 (×100)	Model 6 (×100)	Model 7 (×100)
α	−11.02*	−1.77*	−3.07*	−.23	−3.14*	−11.09*	−3.14*
	(.26)	(.23)	(.22)	(.24)	(.22)	(.25)	(.22)
N terms in article	2.50*	2.56*	4.04*	2.34*	4.07*	4.13*	4.01*
	(.10)	(.09)	(.09)	(.09)	(.09)	(.10)	(.09)
N nonleaf articles	5.15*	−21.13*	−22.23*	−5.85*	−22.18*	1.21*	−22.21*
	(.10)	(.09)	(.10)	(.10)	(.10)	(.10)	(.10)
N authors	−.06*	.08*	.27*	−.06*	.27*	.19*	.28*
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
N subjects in citation graph		44.38*	38.75*	31.40*	38.69*		38.74*
		(.10)	(.16)	(.16)	(.16)		(.16)
Subject citation modularity		36.41*	37.86*		37.80*		37.80*
		(.09)	(.09)		(.09)		(.09)
Subject entropy			1.04*	15.43*	1.12*		.98*
			(.14)	(.14)	(.14)		(.14)
Distinct years in citation graph			26.10*		26.15*	41.71*	26.18*
			(.21)		(.21)	(.22)	(.21)
σ(years)			−17.64*		−17.68*	−32.59*	−17.79*
			(.20)		(.20)	(.22)	(.20)
Ambiguity	−1.45*	−1.52*	−1.76*	−1.15*	−1.79*	−1.74*	−1.85*
	(.10)	(.08)	(.08)	(.09)	(.08)	(.10)	(.08)
Ambiguity × subject citation modularity		−1.74*	−1.64*				
		(.08)	(.08)				
Ambiguity × subject entropy				−.96*	−1.11*		
				(.09)	(.08)		
Ambiguity × σ(Years)						.93*	.58*
						(.09)	(.07)
R^2	.03	.26	.28	.16	.28	.06	.28

Note.—All models are estimated on 1,101,766 article abstracts. Controls for subject areas are presented in table 1. SEs are in parentheses.
* $P < .01$.

When time rather than disciplinary boundaries diversifies a citation graph, as represented in models 6 and 7, then the association between ambiguity and cohesion is weakened. The coefficients on the interaction between ambiguity and year-variance are small but push in the opposite direction of those for ambiguity on its own. Unsurprisingly, scientists and scholars behave differently when citing work from a different field rather than a different time. Citing an ambiguous article from the distant past dampens authors' tendency to cite diverse interpretations for clarification or legitimation, as suggested by Gerow et al. (2018).

Coefficient estimates for subject area dummy variables (presented in table C1) show significant variability between disciplines. Controlling for ambiguity, an article from humanities or mathematics is expected to have the most integrated citation structure, while one from chemistry or business and management is expected to be much more fragmented. This pattern of between-subject variation in citation structure is not, however, immediately connected to the interdiscipline hierarchy we saw in ambiguity (fig. 7). This nonrelationship suggests that the potential confounding effect of disciplinary differences is not present. We ran the same analysis on a model without including these discipline-level dummies and yielded nearly identical results. We also reran the models individually for each broad field (e.g., chemistry, medicine, social science), and the relationship between ambiguity and integration held in every estimable case.[18] The relationship between lexical ambiguity and citation fragmentation exists within disciplines as much as between them.

DISCUSSION

Ambiguity is ubiquitous in natural language and unavoidable in scientific and scholarly discourse, despite the many "precision" and "transparency" projects that have sought to evacuate it. Our goal has been to better understand how it manifests in language and to identify its consequences for scholarly life. Building on existing literature about the types and properties of ambiguity, we focus on the uncertainty inherent in ambiguous language. Whether through vagueness or multiplicity of meaning, our work suggests that the lack of clear resolution in ambiguous language has real consequences for the communities that produce and consume it. Indeed we find strong evidence that ambiguity differs substantially from discipline to discipline in anticipated ways, ranging from the precise use of natural language in chemistry and biomedicine to its looser, more metaphorical use in humanities, law, and business. These differences substantially shape the contours of discourse and resolution of knowledge claims in these disciplines.

As with most methods addressing the subtleties of human language, our measure of ambiguity has some notable limitations. First, because we rely on general-purpose thesauri for our synonyms, our measure misses many technical or field-specific terms. Although sensitivity testing has shown that our measure is remarkably stable across different thesauri, more exhaustive coverage of

[18] Because of the overlapping structure of subject categories, we were unable to estimate a simple multilevel model that would have simultaneously produced both slope and intercept effects for ambiguity on fragmentation. To check for slope differences, we instead ran the field-level models independently. For a few of the fields (e.g., law) we were unable to obtain estimates because of the limited number of qualifying articles with estimates; but for all field models we could estimate, the effect held strong and with roughly the same slope.

the terms in the corpus would greatly increase the measure's precision. Furthermore, the definition of semantic context used in the measure is necessarily simplistic. Even though our initial data sample is quite large, the diversity of language used means that a more nuanced notion of context would yield shared contexts too sparse for analysis.

Nevertheless, our central findings that illuminate the consequences of ambiguity for knowledge communities are robust and compelling. We show how articles that use more ambiguous language tend to result in more integrated streams of citations tracing intellectual engagement. This pattern underscores the interpretation of ambiguity not only as a limitation but also as a potentially fruitful characteristic of language. Ambiguity leads to individual and collective uncertainty about communicated meanings in academic discourse. Uncertainty drives social interaction and friction, which yields coordination. If the knowledge community engaging with an article is more diverse, then the cohering effect of ambiguity amplifies. This finding suggests that the internal partitions between scientific and scholarly fields may be more permeable than many commentators have imagined, actively negotiated through the boundary objects of ambiguous, shared concepts.

The purpose of our investigation was to explore the distribution and consequence of ambiguity in science and scholarship. Ambiguity should be seen as beneficial and even necessary for the development of scientific and scholarly communities, but we do not argue that ambiguity incurs no costs. According to the information-theoretic cartoon of communication from Weaver (Shannon and Weaver 1963) pictured in figure 1, ambiguity may slow down communication by confusing the audience. Intellectual engagement and integration, ambiguity's chief consequences, incur costs beyond inefficiency: speaking and understanding across disciplines require a large and growing vocabulary (Vilhena et al. 2014). Specialization can be a powerful dimension of "fragmentation" and facilitate the rapid and focused accumulation of knowledge. Nevertheless, as fields fragment, specialists within them learn more and more about less and less. Ambiguity unleashes the uncertainty and insecurity required to help reweave specialties into an intellectual whole.

These broad findings are consistent with theories of information uncertainty found in neuroscience and across social and organizational sciences. As discussed above, when social actors are confronted with unresolved situations, theory and experiment suggest they will engage with others to reach resolution. Our findings are highly consistent with this standpoint. Articles citing work that is open to diverse interpretation are more likely to engage one another in sustained discourse.

Thomas Kuhn and others have noted his own polysemic use of the core term "paradigm" in *Structure of Scientific Revolutions*, the most-cited 20th-century work of social science or humanities. Similarly, Darwin's *Origin of Species* deployed a variety of ambiguous phrases such as "natural selection," "struggle for existence," and "survival of the fittest" that have been the subject of persistent debate, engagement, and extension. What Davis posited for social theory appears to apply more broadly: "Had each classical social theorist exhausted the implication of his fundamental factor, as some classical philosophers exhausted theirs, one could only admire the theory, not add to it" (1986, p. 297). In sociology, Simmel, Durkheim, and Weber each presented a sprawling collection of social insight, stitched together with deeply debated ambiguities.

Synthetic works that draw distinctive research areas together have a success likelihood partially in proportion to the polysemic potential of their prose. By unleashing collective uncertainty, they draw disparate researchers into a common conversation. Fields are not formed out of canonical findings, but canonical ambiguities and living debates.[19] We find specific support for the contention that "ambiguity is the critical resource out of which new ideas emerge. . . . The cell phone emerged in the space created by the ambiguity about whether the product was a radio or a telephone; by playing with that ambiguity, the device became something that was different from either" (Lester and Piore 2006, p. 54). This rings true not only for technological recombination but for the friction between diverse interpretations that stem from an ambiguous claim.

Our findings also hold implications for other domains of social life. For example, politicians are at times derided for being untrustworthy, two-faced opportunists but at others lauded as community builders who bridge divides and bring people together. Our findings from science and scholarship suggest that in some cases, the only difference between these two competing interpretations of the same discursive ambiguity is a post hoc assessment of success. If brokering multiple communities with evocative vagueness had failed, then politicians fall to their baser interpretation (Fine 2014). This suggests similar risks and benefits associated with ambiguity for leaders in organizations of all types.

Finally, our findings are suggestive for the realm of rumor, gossip, and common knowledge. In Tamatsu Shibutani's *Improvised News* (1966), he illustrates how ambiguous rumor leads to uncertainty and engagement, as community members develop elaborate theories and fill in unspoken details. Our findings suggest the possibility that these uncertainties, in a small town or a global network of scientists and scholars, become the nucleus around which community revolves.

[19] Relatedly, following the advice of philosopher Allesandro Gambera, "write a book with a black hole in the middle, into which the reader can lose themselves" (David Nirenberg, personal communication).

APPENDIX A

Here we detail a more sophisticated modeling approach for the direct estimation of lexical ambiguity in context.

Meaning

This model builds from our basic formulation of meaning probabilities $\Pr(\mathbf{M}|t_i, c_j)$, where $\mathbf{M} = \{m_0, m_1, ...\}$ is the set of all meanings or word senses, $\mathbf{T} = \{t_0, t_1, ...\}$ is the set of tokens, and $\mathbf{C} = \{c_0, c_1, ...\}$ is the set of linguistic contexts found in the corpus. We leave the specifics of what constitutes a linguistic context vague here but assume a many-to-many relationship between tokens and contexts. We also assume that contexts provide some information about how a particular term was used in relation to the rest of the text. Any model that uses this formulation will need to enumerate all possible "meanings" and their relationship to words in the corpus. Here, we build meanings from synonymy data pulled from a set of English-language thesauri.

Construct a synonym graph of all of the terms that appear in the thesauri: G with one vertex for each (stemmed) word and an undirected edge (t_i, t_j) if t_j is listed as a synonym of t_i or t_i is listed as a synonym of t_j in at least one thesaurus. We relax the strict definition of a k-clique to find sets of terms that are connected with uniform density but that allow for some uncertainty in the edge structure of the network. To do so, we define a restricted set of k-clique percolation communities for which k, the connectivity of the cliques, is no less than some proportion p of the total size of the community. Using $p = .75$, for example, yields a set of overlapping term sets in which each term is a synonym of at least 75% of other terms in its set. These communities can be seen as a relaxation of network cliques that allows for "missing" edges in the graph. As a final step, we remove any such communities that overlap at least 90% with another community. Each community so identified is treated as a discrete meaning.

Figure A1 is a subgraph of our synonymy network, induced from six overlapping meanings extracted from the full graph. While there are clearly two coarse meanings displayed here, the large one relating broadly to "precision" and the small one to "extortion," the methodology distinguishes several more fine-grained meanings:

1. (*exact*, *explicit*, *correct*, *formal*, *precise*)
2. (*exact*, *set*, *true*, *firm*, *correct*, *right*)
3. (*nice*, *correct*, *rigid*, *narrow*, *prim*, *precise*)
4. (*unbending*, *fixed*, *inflexible*, *rigid*, *formal*, *precise*)
5. (*exact*, *gouge*, *extort*, *screw*, *blackmail*, *indent*)
6. (*exact*, *gouge*, *screw*, *mulct*, *milk*)

The method identifies nuanced semantic distinctions even within larger clusters, such as the difference between meaning 2 (relating to correctness or concreteness) and meaning 3 (relating to parsimony or precision). Furthermore, the process is tunable using two parameters: the threshold of connectedness, here $p = .075$, and the threshold of overlap, $q = 0.9$.

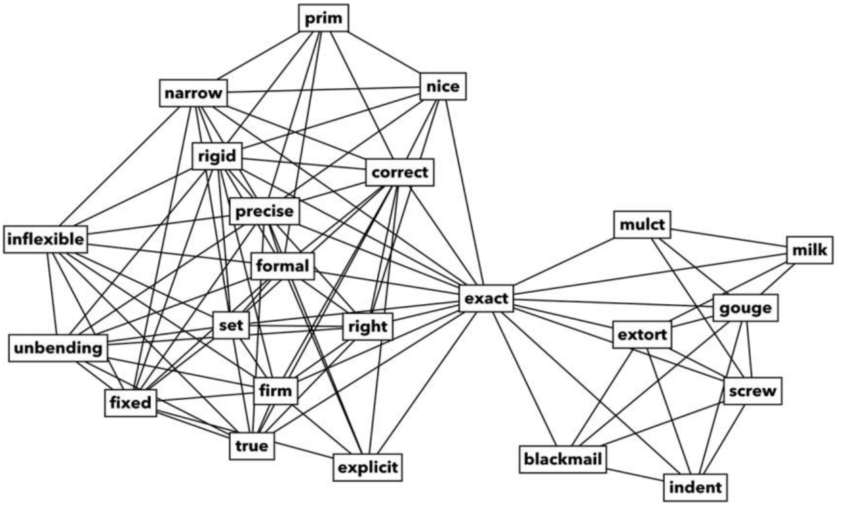

FIG. A1.—A subgraph of G representing six distinct meanings

Model

Given a set M of explicit meanings, defined as overlapping sets of terms, we can specify an explicit, generative model of text based on terms, contexts, and meanings. For clarity, we will define the *linguistic context* of a term as an unordered set of up to four terms, along with their parts of speech, preceding and following the focal word in text. For the phrase *these are the terms exacted by the contract*, the context of the term *exact* is {*by(prep)*, *term(n)*, *the(det)*, *the(det)*}. Recall that $C = \{c_i\}$ is the set of contexts found in the corpus.

We model each context c_i as being associated with a distribution $\theta_i = (p_{i,1}, \dots, p_{i,|M|})$ across meanings, where $|M|$ is the size of the total set of meanings. Similarly, each meaning m_j has a distribution of term probabilities $\phi_j = (q_{j,1}, \dots, q_{j,|W|})$, with $|W|$ denoting the size of the set W for all terms. Note that because each meaning is associated with only a very small subset of all terms in W, we can set most $q_{j,k} = 0$ a priori. Finally, for a given context c_i, let $T_i = (\tau_{i,1}, \dots, \tau_{i,|W|})$ be the vector of term counts that occur in that context across the corpus, so $\tau_{i,k}$ is the number of times term t_k occurs in context c_i. We model

τ_i as a draw from a multinomial distribution across terms, with probabilities $\Pi_i = (\pi_{i,1}, \dots, \pi_{i,|W|})$ from mixture θ_i. The probability of a particular use of term t_k in context c_i is

$$\pi_{i,k} = \sum_{j=1}^{|M|} p_{i,j}\, q_{j,k}.$$

We define Dirichlet prior distributions α and β_j by $\theta_i \sim \mathrm{Dir}(\alpha)$ for all i and $\phi_j \sim \mathrm{Dir}(\beta_j)$. Note that α is a fixed prior for meaning distributions across all contexts, while each meaning has its own β_j of probabilities across terms. We let α define a symmetric prior across all meanings, but β_j is symmetric across only the terms that define the meaning: the Dirichlet parameter associated with each other term is set to zero. The plate notation below (fig. A2) shows that this model is similar to a latent Dirichlet allocation (LDA) model with meanings taking the role of topics and contexts acting as documents. A consequential difference between a standard LDA and this model is that each meaning (topic) is assigned its own unique, highly informative prior. Furthermore, the number of meanings and contexts is considerably larger than in most LDA applications. While specification of such a model is relatively straightforward in an LDA context, estimation is computationally challenging. Furthermore, most techniques for estimating LDA models with large numbers of topics (or large numbers of meanings) rely on marginalizing the likelihood across the θ_i and ϕ_j terms, but our measure of ambiguity requires explicit estimates for both of those terms (see below). The authors are currently in the process of developing a computational methodology for estimation of this model.

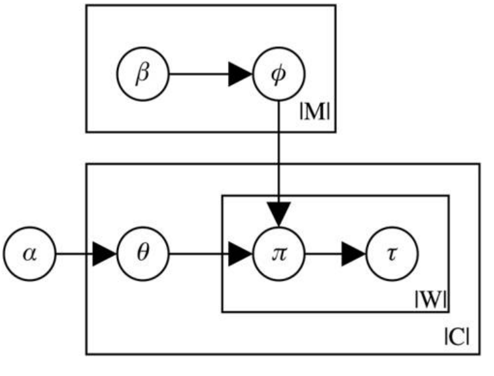

FIG. A2.—Plate notation for the model

Ambiguity

With posterior estimates of Π, θ, and ϕ, it is simple to calculate an exact posterior probability of a meaning conditional on term and context:

$$\Pr(m_j|t_k, c_i) = \frac{\Pr(t_k|m_j, c_i)\Pr(m_j|c_i)}{\Pr(t_k|c_i)} = \frac{\pi_{i,k}p_{i,j}}{\sum_{l=1}^{|M|}p_{i,l}\,q_{l,k}}.$$

This equation allows explicit calculation of the conditional entropy of meanings for a given term in its linguistic context. By calculating these entropies across a large posterior sample of the model parameters, a researcher can obtain a posterior estimate of the ambiguity of each term-in-context.

APPENDIX B

Posterior Distribution of Ambiguity

Here we detail our approach to estimating the posterior distribution of lexical ambiguity. This is critical for realistically modeling ambiguity because some synonyms may never appear in a given context within the corpus, but this does not mean that they have zero probability of appearing in the future. By treating observed counts as draws from an underlying multinomial distribution, we confer on unseen synonym-context pairs a small but nonzero probability and debias our measure. For a given set of word frequencies or counts $f_0, \dots, f_k$ across $k + 1$ possible synonyms, we estimate the categorical probabilities $p_0, \dots, p_k$ from which observed frequencies are drawn. In this case the posterior distribution of probabilities is Dirichlet-distributed with parameters $f_i + \alpha$ for $i = 0, \dots, k$:

$$(p_0, \dots, p_k) \sim \text{Dirichlet}(f_0 + \alpha, \dots, f_k + \alpha).$$

Here α represents the concentration parameter of the symmetric-Dirichlet prior distribution, which we set to one.[20] For every sample of categorical probabilities it is possible to define a distribution of entropies associated with those probabilities. Formally, let h_{tc} be a random variable representing the entropy of some term t in context c. Then $h_{tc} = H(P_{tc})$, where $P_{tc} \sim \text{Dirichlet}(f_{tc} + \alpha)$, the parameter vector $f_{tc} + \alpha = (f_0 + \alpha, \dots, f_k + \alpha)$, and f_i is the observed frequency of the ith synonym of term t. While it is intractable to calculate the resulting distribution of entropy analytically, it is triv-

[20] We use $\alpha = 1$ because it represents the least informative prior in our context; a priori, any set of probabilities is equally likely as any other. We opted against the Jeffrey's prior of $\alpha = 1/(k + 1)$ because we found a strong association between the number of synonyms a term has (k) and the disciplines that tend to use those terms, which would result in a de facto informative prior on interdisciplinary differences.

ial to draw samples from that distribution for any set of frequency observations.

We note that the Dirichlet distribution is a "distribution over distributions." Because information-theoretic entropy is itself a function of distributions, it follows that every term-context pair in a corpus yields a distribution of entropies. Understanding the synonym-substitution entropy of each term occurrence in terms of a posterior probability distribution may seem overly complex, but it is necessary to avoid biased results.

A specific example helps to clarify the translation from words-in-context to distributions of entropy. In the hibernating pikas example used in the text ("pikas don't hibernate through winter"), we found that the synonym substitution frequencies for the term "hibernate" in that context across eight synonymous terms were (5, 1, 0, 0, 0, 2, 0, 0). These frequencies define the posterior distribution of multinomial probabilities that produced them:

$$p \sim \text{Dirichlet}(5+1, 1+1, 0+1, 0+1, 0+1, 2+1, 0+1, 0+1).$$

The curve in figure B1 shows the entropy of the posterior distribution of multinomial probabilities, with a modal value of approximately 2.39 (we calculate the log with a base of 2). We also see the bias that would be introduced through maximum-likelihood estimates of entropy: in this case the ML estimate is only about 1.30. An analogous process is used to yield a meaningful measure of ambiguity for any term occurrence in the corpus.

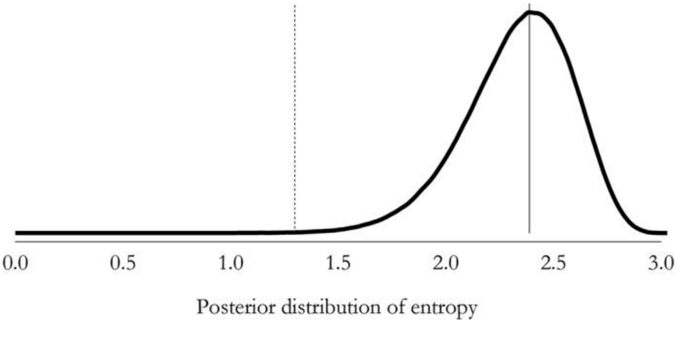

FIG. B1.—Example posterior distribution of entropy. The dotted (left) line represents the ML estimate of entropy based solely on the observed frequencies.

APPENDIX C

Survey Validation of Ambiguity

We conducted a survey asking respondents to rate the ambiguity of specific terms from a corpus of *New York Times* articles in order to assess individual uncertainty regarding word meanings.[21] We calculated ambiguity across the corpus using the simplified model of ambiguity, yielding approximately 4.2 million measured terms-in-context, a small sample of which are listed in table C1. There were 1,024 individual terms and their containing sentences drawn at random from this population using a weighted sampling method to oversample terms with most- and least-measured ambiguity. The items represented a broad array of terms.

TABLE C1
Sample Items from the *New York Times* Corpus Used in the Survey

Sample Items	Measured Ambiguity
About 250,000 people under the *age* of 19 went to emergency rooms with concussions in 2009 compared with 150,000 in 2001	.02
I'm not a huge fan, Ibisevic, the Stuttgart and Bosnia-Herzegovina striker, said in a *telephone* interview from Germany.	.55
Over the *course* of their professional careers, Jordan's team sent Ewing's team home from the playoffs five times	1.60
Many of the *details* of the negotiations remain cloaked	3.15
So if you get in there it's going to sit on the *bottom*	4.51
That was an improvement from 2001, when the state was ranked fourth, according to the *group*	5.12

Responses were collected from a survey of 102 participants recruited through Amazon's Mechanical Turk service. Participants were shown sentences from the corpus of *New York Times* articles with the focal word highlighted and provided the following prompt: "Written text is not always totally clear. In many cases, the words that we read could mean more than one thing. In the sentences below, please rank how certain you are that you know exactly what the highlighted term means as it is used in the sentence." Each sentence was accompanied by a 7-point Likert scale, ranging from "extremely uncertain (ambiguous)" to "extremely certain (precise)."

[21] Articles were drawn from a corpus of 77,406 *New York Times* articles written and published online between 2012 and 2015. The corpus comes from a database of articles that received at least one postprint correction to their text. We used the latest or corrected version of each article.

We randomized the surveys using a $\binom{102}{3}$ design, such that each respondent was provided a unique set of 30 questions with maximal overlap between respondents, and the 1,020 sentences used in the survey each appeared in exactly three surveys.[22]

The left panel of figure C1 shows a scatter plot with ambiguity as measured by our algorithm on the horizontal axis and respondent ratings of ambiguity on the vertical axis (transformed to a 0–6 scale and jittered for visual clarity). The right panel shows the marginal distribution of user-rated ambiguity across the entire sample. Respondents had an overall tendency to rate terms as precise, ranking 55% of all items as "extremely precise" (0 on the transformed scale). Fewer than one-quarter of respondents (24) rated any item as most ambiguous, and a similar number (22) failed to rate any of the examples presented more than a 2 on the 0–6 scale. This suggests variability in our respondents with respect to their sensitivity to ambiguity.

[22] Survey items were distributed among respondents using the following process: (1) Items (sentences) were divided into 30 groups of 34 items each. (2) Within each group, items were distributed randomly across the 102 respondents, each item appearing three times. (3) Single items assigned to a particular respondent from each of the 30 groups were collected into a 30-item survey for that respondent. This process ensures that the marginal probability that a particular respondent is shown a particular item is constant for all pairs of respondents and items. That is, if we let $q_{ij} = 1$ if respondent i was shown question j and $q_{ij} = 0$ otherwise, then $\Pr(q_{ij} = 1) = 1/34$. (Note, however, that the conditional probability $\Pr(q_{ij} = 1 | q_{ik} = 1)$ is not constant across items k.)

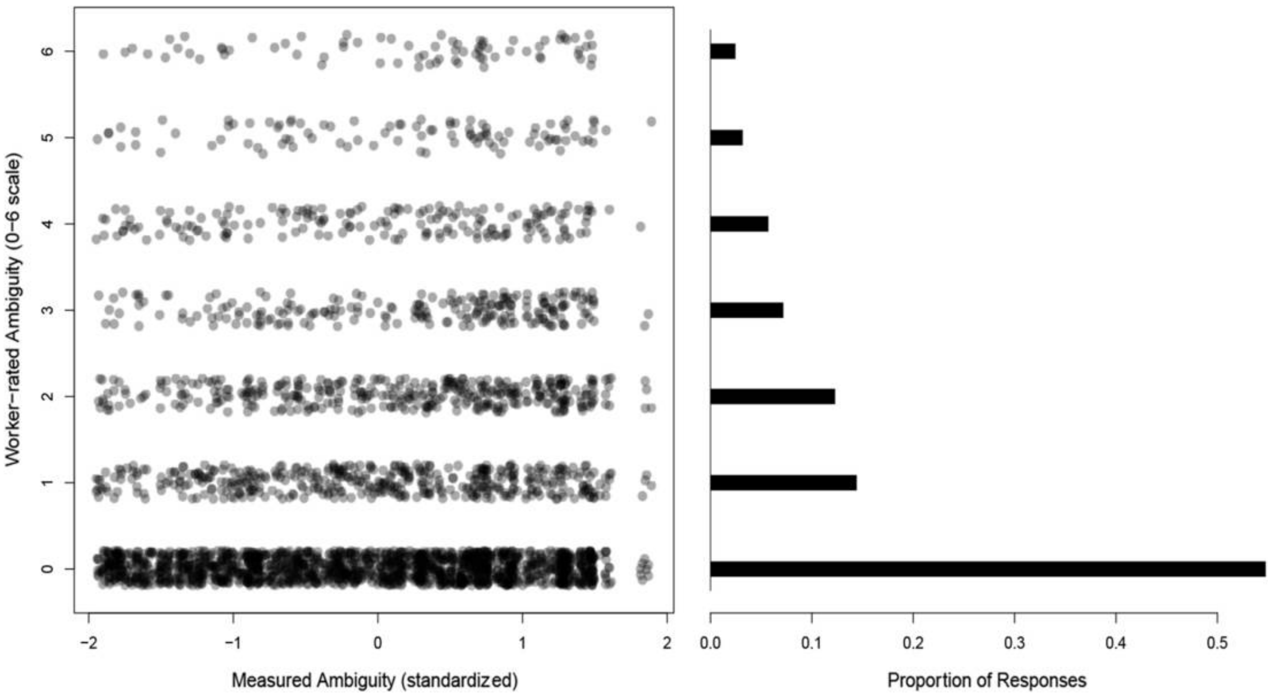

FIG. C1.—Left panel: Scatter plot with ambiguity as measured by the model (standardized to have a mean of 0.0 and a standard deviation of 1.0) on the horizontal axis and user ratings of ambiguity on a 0–6 scale on the vertical axis. Right panel: Marginal distribution of user-rated ambiguity.

We employed a multilevel generalized linear model to estimate the relationship between our measure and respondent rankings of ambiguity to handle the possibility of high variability between respondents. By design, each survey response is cross-classified by both the Mechanical Turk worker answering the question and the specific corpus item to which they are responding, as each corpus item appeared on three different surveys. We therefore model the responses using a cross-classified multilevel Poisson model:

$$\begin{aligned}\log(\lambda_{ijk}) &= \beta_{0jk} + \beta_{1jk}X_{ik},\\ \beta_{0jk} &= \alpha_0 + \eta_{0j} + \nu_k,\\ \beta_{1jk} &= \alpha_1 + \eta_{1j},\\ Y_{ijk} &\sim \mathrm{Pois}(\lambda_{ijk}),\\ \nu_k &\sim \mathrm{N}(0, \tau_0),\\ \eta_j &\sim \mathrm{N}(\mathbf{0}, \tau_1).\end{aligned}$$

Here, i indexes survey responses, j indexes survey participants, k indexes survey items, Y_{ijk} is the ith response (which was made by participant j in response to item k), and X_{ik} is the ambiguity of the ith survey question as measured by our model.[23] Terms η_{0j} and η_{1j} represent respondent-level random intercept and slope components, jointly normally distributed, and ν_k is a normal random intercept component at the level of the corpus item. In addition to modeling the overall relationship between our ambiguity measure and respondent ratings of ambiguity, this model accounts for two other sources of variation in the way participants encounter terms-in-context from the corpus. First, because each term is given its own random effect (ν_k), the model allows certain terms to be consistently understood as more or less ambiguous than our measure would suggest. Second, and more importantly for our analysis, the multilevel model estimates the degree to which respondents differ individually from the average predicted relationship.

[23] For clarity, we abuse notation slightly and treat measured ambiguity as a level-1 variable. In fact, it varies only between corpus items, and not between individual survey questions.

TABLE C2
Point Estimates and Confidence Intervals for Model Coefficients and Random-Effect Standard Deviations

Parameter	Estimate	95% Confidence Interval
Intercept (α_0)	−.64	(−.94, −.36)
Measured ambiguity (α_1)	.27	(.18, .38)
SD(ν_k)	.44	(.38, .49)
SD(η_{0j})	1.38	(1.18, 1.65)
SD(η_{1j})	.28	(.20, .37)

Note.—Each estimated parameter is significant at $P = .05$, and $\text{Cor}(\eta_{0j}, \eta_{1j}) = -.65$.

Estimates in table C2 show that there is a strong overall relationship between our measure of ambiguity and the rating reported by respondents. Respondents' expected rating increases by an average of 32% ($\exp(0.27462) = 1.31603$). This relationship is variable between respondents as indicated by variation in the random slope component η_{1j}. Figure C2 summarizes this variability, displaying the predicted relationship between measured ambiguity and rated ambiguity for each individual respondent, as well as the average relationship between variables. The figure underscores heterogeneity across survey participants, but because this is explicitly accounted for in the multilevel framework, the overall model fit is quite good. The deviance pseudo-R^2 is $R^2_{\text{DEV}} = .5980$.[24] This indicates that the clustering of survey responses around zero ("extremely precise") is the result of a subset of respondents displaying very little variability in their ratings—who lacked sensitivity to linguistic uncertainty that others were able to detect.

[24] Deviance pseudo-R^2 generalizes R^2 to (multilevel) generalized linear models and can be interpreted as the proportion of the statistical deviance explained by a model in comparison with a null model (in this case a simple Poisson model with no covariates). See Cameron and Trivedi (2013) for a discussion.

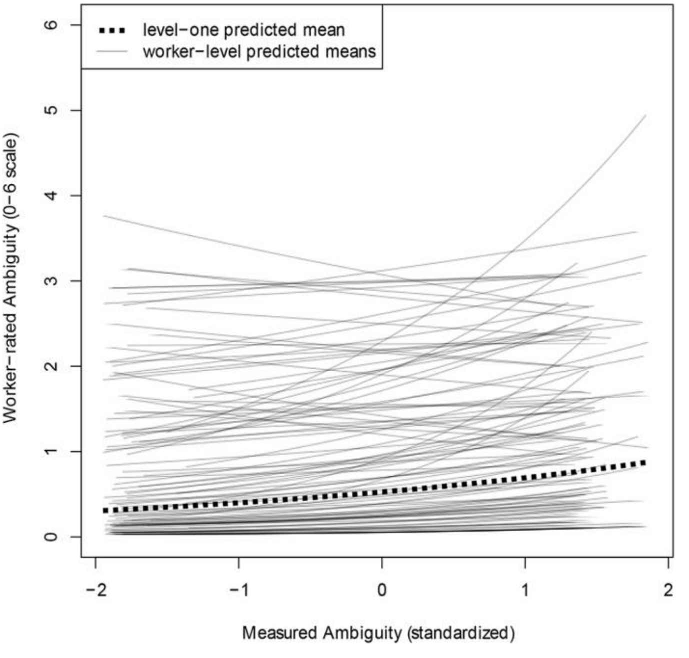

FIG. C2.—Respondent-level predicted means for all 102 participants (solid lines) and level-1 average predicted mean (dashed line).

In addition to providing support for the validity of our measure of ambiguity, the multilevel Poisson model does a good job predicting human rankings of term ambiguity using only our entropy-based measure and random-effects terms as predictors. Respondents that identified ambiguity in the survey items agreed with our measure. Our results also point to another, somewhat surprising aspect of ambiguity, as elicited by our survey. Nearly half of our respondents (50 out of 102) posted a mean response of less than 1.0 and the same number posted a response standard deviation of less than 1.0 on the 7-point scale. In other words, nearly half of our respondents did not recognize significant ambiguity in the items provided. Unfortunately, our survey collected no information on respondents (outside of their ambiguity rankings) and is therefore ill equipped to address questions of whether some respondents were "better" at identifying ambiguity than others because of topical expertise, English literacy, education, or demographics. Moreover, our survey neglected to identify particular word meanings that respondents felt certain about and so could not allow us to measure the degree to which our measure also captured collective uncertainty. Nevertheless, we find a strong support for our association between measured ambiguity and individually perceived semantic uncertainty.

APPENDIX D

Field-Level Differences

Figure D1 demonstrates robust differences between fields, here disaggregated by individual terms used in field-specific articles (fig. 7 in the main text reveals similar differences between fields by the average ambiguity within articles).

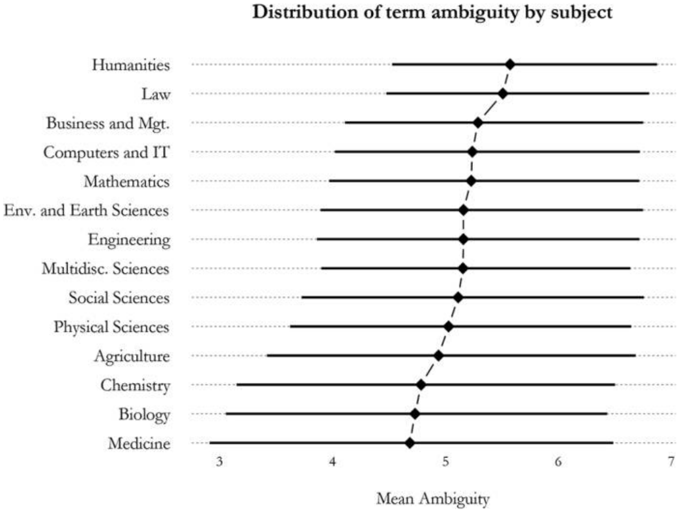

FIG. D1.—Median and interquartile range of term ambiguity across subjects. The order of each consecutive pair of subjects is significant at the 1% level (using a one-tailed Kolmogorov-Smirnov test).

Table D1 reports ordinary least squares (OLS) coefficient estimates for discipline indicator dummy variables associated with the models reported in table 2.

TABLE D1
OLS COEFFICIENT ESTIMATES AND STANDARD ERRORS FOR DISCIPLINE DUMMIES

Variable (Standardized)	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Agriculture	−.12*	−.07*	−.09*	.08*	−.10*	−.10*	−.10*
	(.01)	(.00)	(.00)	(.00)	(.00)	(.01)	(.00)
Engineering	−.10*	−.11*	−.09*	−.11*	−.09*	−.07*	−.09*
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
Environmental and earth sciences	−.06*	−.01	.01*	.01	.01	−.03*	.01
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
Multidisciplinary sciences	−.04*	−.03*	.00	−.12*	.00	−.01	.00
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Biology	.19*	−.04*	−.04*	−.06*	−.04*	.15*	−.04*
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
Chemistry	.09*	.10*	.12*	.17*	.12*	.12*	.12*
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
Physical sciences	−.13*	.22*	.18*	.16*	.18*	−.13*	.18*
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
Medicine	.15*	−.02*	−.01*	−.02*	−.01*	.14*	−.01*
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
Social sciences	.02*	−.06*	−.03*	−.03*	−.03*	.04*	−.03*
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Humanities	.02	−.18	−.10	−.02	−.10	.16	−.10
	(.08)	(.07)	(.07)	(.08)	(.07)	(.08)	(.07)
Computers and information technology	.15*	.27*	.27*	.12*	.28*	.14*	.28*
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Business and management	.06*	.13*	.08*	.12*	.08*	.02	.08*
	(.02)	(.01)	(.01)	(.01)	(.01)	(.02)	(.01)
Law	−.24	−.06	.04	(.01)	.03	−.04	.04
	(.13)	(.11)	(.11)	(.12)	(.11)	(.12)	(.11)
Mathematics	−.38*	−.40*	−.37*	−.34*	−.37*	−.32*	−.37*
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
R^2	.03	.26	.28	.16	.28	.06	.28

NOTE.—All models are estimated on 1,101,766 article abstracts. SEs are in parentheses.
* Significant at $P < .01$.

REFERENCES

Amir, Shmuel. 1995. "The Environmental Cost of Sustainable Welfare." *Ecological Economics* 13 (1): 27–41.

Baker, Keith Michael. 1975. *Condorcet: From Natural Philosophy to Social Mathematics*. Chicago: University of Chicago Press.

Barthes, Roland, and Honoré de Balzac Sarrasine. 1974. *S/Z*. New York: Hill & Wang.

Beer, Gillian. 2009. *Darwin's Plots: Evolutionary Narrative in Darwin, George Eliot and Nineteenth-Century Fiction*, 3d ed. Cambridge: Cambridge University Press.

Blair, David R., Kanix Wang, Svetlozar Nestorov, James A. Evans, and Andrey Rzhetsky. 2014. "Quantifying the Impact and Extent of Undocumented Biomedical Synonymy." *PLoS Computational Biology* 10 (9): e1003799.

Bourdieu, Pierre. 1986. "The Forms of Capital." Pp. 46–58 in *Handbook of Theory and Research for the Sociology of Education*, edited by J. Richardson. New York: Greenwood.

Bowler, Peter J. 1989. *Evolution: The History of an Idea*, rev. ed. Berkeley and Los Angeles: University of California Press.

Brahimi, B., A. Ait-Kadi, and A. Ajji. 1994. "Weld Lines and Mechanical Properties of Injection Molded Polyethylene/Polystyrene/Copolymer Blends." *Polymer Engineering and Science* 34 (15): 1202–10.

Cameron, A. Colin, and Pravin K. Trivedi. 2013. *Regression Analysis of Count Data*, vol. 53. New York: Cambridge University Press.

Campbell, John Angus. 1975. "The Polemical Mr. Darwin." *Quarterly Journal of Speech* 61 (4): 375–90.

Carnap, Rudolf. 1928. *Der logische Aufbau der Welt*. Hamburg: Felix Meiner.

Ceccarelli, Leah. 1998. "Polysemy: Multiple Meanings in Rhetorical Criticism." *Quarterly Journal of Speech* 84 (4): 395–415.

———. 2001. *Shaping Science with Rhetoric: The Cases of Dobzhansky, Schrödinger, and Wilson*. Chicago: University of Chicago Press.

Clauset, Aaron, Mark E. J. Newman, and Cristopher Moore. 2004. "Finding Community Structure in Very Large Networks." *Physical Review E* 70 (6): 066111.

Cohen, Michael D., James G. March, and Johan P. Olsen. 1972. "A Garbage Can Model of Organizational Choice." *Administrative Science Quarterly* 17 (1): 1–25.

Coleman, James S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 94:S95–S120.

Collins, Randall. 1998. *The Sociology of Philosophies: A Global Theory of Intellectual Change*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Corominas-Murtra, Bernat, Carlos Rodríguez-Caso, Joaquín Goñi, and Ricard Solé. 2010. "Topological Reversibility and Causality in Feed-Forward Networks." http://arxiv.org/abs/1007.1829.

Corominas-Murtra, Bernat, and Ricard V. Solé. 2010. "Universality of Zipf's Law." *Physical Review E* 82 (1): 011102.

Davis, Murray S. 1986. "'That's Classic!' The Phenomenology and Rhetoric of Successful Social Theories." *Philosophy of the Social Sciences* 16 (3): 285–301.

Dawkins, Richard. 2016. *The God Delusion*. New York: Random House.

DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Journal of Sociology* 48:147–60.

Doohan, Douglas J., Thomas J. Monaco, and Thomas J. Sheets. 1993. "Management of Field Violet (*Viola arvensis*) in Strawberries (*Fragaria* × *anassa*)." *Weed Technology* 7 (1): 185–89.

Durkheim, Émile. 1951. *Suicide: A Study in Sociology*. Glencoe, Ill.: Free Press.

Edelman, Lauren B. 1992. "Legal Ambiguity and Symbolic Structures: Organizational Mediation of Civil Rights Law." *American Journal of Sociology* 97 (6): 1531–76.

Elberse, Anita. 2008. "Should You Invest in the Long Tail?" *Harvard Business Review* 86 (July): 88–96.

Empson, William. 1966. *Seven Types of Ambiguity*. New York: New Directions.

Evans, J. A. 2008. "Electronic Publication and the Narrowing of Science and Scholarship." *Science* 321 (5887): 395–99.

Fine, Gary Alan. 2014. *Difficult Reputations: Collective Memories of the Evil, Inept, and Controversial*. Chicago: University of Chicago Press.

Frege, Gottlob. 1879. *Begriffsschrift Und Andere*, 2d ed. Hildesheim: Lubrecht & Cramer.
Gallie, W. B. 1955. "Essentially Contested Concepts." *Proceedings of the Aristotelian Society*, n.s., 56 (January): 167–98.
Gaonkar, Dilip Parameshwar. 1989. "The Oratorical Text: The Enigma of Ambiguity." Pp. 255–76 in *Texts in Context: Critical Dialogues on Significant Episodes in American Political Rhetoric*. Davis, Calif.: Hermagoras Press.
Gerow, Aaron, Yuening Hu, Jordan Boyd-Graber, David M. Blei, and James A. Evans. 2018. "Measuring Discursive Influence across Scholarship." *Proceedings of the National Academy of Sciences* 115 (3): 3308–13.
Guyton de Morveau, Louis-Bernard, Antoine Laurent Lavoisier, Claude-Louis Berthollet, Comte Antoine-François de Fourcroy, Jean-Henri Hassenfratz, and Pierre-Auguste Adet. 1787. *Méthode de nomenclature chimique*. Paris: Chez Cuchet.
Hamilton, William L., Jure Leskovec, and Dan Jurafsky. 2016. "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change." ArXiv Preprint ArXiv:1605.09096.
Hobbes, Thomas. 2009. *Leviathan*. Lexington, Ky.: CreateSpace Independent Publishing Platform.
Jevons, William Stanley. 1876. *Money and the Mechanism of Exchange*, 2d ed. London: H. S. King.
Justeson, John S., and Slava M. Katz. 1991. "Co-occurrences of Antonymous Adjectives and Their Contexts." *Computational Linguistics* 17 (1): 1–19.
Knack, Stephen, and Philip Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *Quarterly Journal of Economics* 112 (4): 1251–88.
Knorr-Cetina, K. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, Mass.: Harvard University Press.
Knox, K. S., J. R. Adams, B. Djulbegovic, T. J. Stinson, C. Tomor, and C. L. Bennet. 2000. "Reporting and Dissemination of Industry versus Non-profit Sponsored Economic Analyses of Six Novel Drugs Used in Oncology." *Annals of Oncology* 11 (12): 1591–95.
Leibniz, Gottfried Wilhelm. 1923. "Dissertatio de arte combinatoria." Pp. 163–241 in *Sämtliche Schriften und Briefe*, A VI 1:163. Berlin: Akademie Verlag.
Leibniz, Gottfried Wilhelm, Freiherr von, and Leroy E. Loemker. 1976. *Philosophical Papers and Letters*, 2d ed. Synthese Historical Library, vol. 2. Dordrecht: Reidel.
Leifer, Eric Matheson. 1983. *Robust Action: Generating Joint Outcomes in Social Relationships*. Cambridge, Mass.: Harvard University.
Lester, Richard K., and Michael J. Piore. 2006. *Innovation—the Missing Dimension*. Cambridge, Mass.: Harvard University Press.
Levine, Donald Nathan. 1965. *Wax and Gold: Tradition and Innovation in Ethiopian Culture*, vol. 11. Chicago: University of Chicago Press.
———. 1988. *The Flight from Ambiguity: Essays in Social and Cultural Theory*. Chicago: University of Chicago Press.
Locke, John. (1690) 1975. *An Essay Concerning Human Understanding*, edited by Peter Nidditch. Oxford: Clarendon Press.
Marcus, Mitchell P., Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. "Building a Large Annotated Corpus of English: The Penn Treebank." *Computational Linguistics* 19 (2): 313–30.
McKerrow, Raymie E. 1989. "Critical Rhetoric: Theory and Praxis." *Communication Monographs* 56:91–111.
Newman, M. E. J. 2006. "Modularity and Community Structure in Networks." *Proceedings of the National Academy of Sciences* 103 (23): 8577–82.
Omura, Y. 1994. "Accurate Localization of Organ Representation Areas on the Feet & Hands Using the Bi-Digital O-Ring Test Resonance Phenomenon: Its Clinical Implication in Diagnosis & Treatment—Part I." *Acupuncture and Electro-Therapeutics Research* 19 (2–3): 153–90.

Padgett, John, and Chris Ansell. 1992. "Robust Action and the Rise of the Medici." *American Journal of Sociology* 98:1259–1330.

Piantadosi, Steven T., Harry Tily, and Edward Gibson. 2012. "The Communicative Function of Ambiguity in Language." *Cognition* 122 (3): 280–91.

Portes, Alejandro. 1998. "Social Capital: Its Origins and Applications in Modern Sociology." *Annual Review of Sociology* 24:1–24.

Quine, William Van Orman. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.

Roelandts, Iwan. 1993. "Zinc and Zinc Alloy Reference Materials." *Spectrochimica Acta Part B: Atomic Spectroscopy* 48 (3): 461–64.

Salganik, Matthew J., and Duncan J. Watts. 2008. "Leading the Herd Astray: An Experimental Study of Self-Fulfilling Prophecies in an Artificial Cultural Market." *Social Psychology Quarterly* 71 (4): 338–55.

Schultz, Theodore W. "Investment in Human Capital." *American Economic Review* 51, no. 1 (1961): 1–17.

Schultz, Wolfram, Kerstin Preuschoff, Colin Camerer, Ming Hsu, et al. "Explicit Neural Signals Reflecting Uncertainty." *Philosophical Transactions of the Royal Society of London*, ser. B: *Biological Sciences* 363, no. 1511 (2008): 3801–11.

Shannon, Claude E. 1948. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Shannon, Claude E., and Warren Weaver. 1963. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Shapin, Steven, Simon Schaffer, and Thomas Hobbes. 1985. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life: Including a Translation of Thomas Hobbes, Dialogus Physicus de Natura Aeris by Simon Schaffer*. Princeton, N.J.: Princeton University Press.

Shibutani, Tomotsu. 1966. *Improvised News: A Sociological Study of Rumor*, vol. 10. Oxford: Bobbs-Merrill.

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

Solomon, Martha, and Wayne J. McMullen. 1991. "Places in the Heart: The Rhetorical Force of an Open Text." *Western Journal of Speech Communication* 55 (4): 339–53.

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.

Stark, David. 2011. *The Sense of Dissonance: Accounts of Worth in Economic Life*. Princeton, N.J.: Princeton University Press.

Thompson, E. P. 1967. "Time, Work-Discipline, and Industrial Capitalism." *Past and Present*, no. 38 (December), 56–97.

Tuggy, David. 1993. "Ambiguity, Polysemy, and Vagueness." *Cognitive Linguistics (Includes Cognitive Linguistic Bibliography)* 4 (3): 273–90.

Tversky, Amos. 1977. "Features of Similarity." *Psychological Review* 84 (4): 327–52.

Uchiyama. Hitoshi T., Daisuke N. Saito, Hiroki C. Tanabe, Tokiko Harada, et al. "Distinction between the Literal and Intended Meanings of Sentences: A Functional Magnetic Resonance Imaging Study of Metaphor and Sarcasm." *Cortex* 48, no. 5 (2012): 563–83.

Vilhena, Daril, Jacob Foster, Martin Rosvall, Jevin West, James Evans, and Carl Bergstrom. 2014. "Finding Cultural Holes: How Structure and Culture Diverge in Networks of Scholarly Communication." *Sociological Science* 1:221–38.

Weber, Max. 1992. *The Protestant Ethic and the Spirit of Capitalism*. New York: Routledge.

Weber, Max, Guenther Roth, and Claus Wittich. 1978. *Economy and Society: An Outline of Interpretive Sociology*. Berkeley: University of California Press.

White, Harrison C. 1981. "Where Do Markets Come From?" *American Journal of Sociology* 87 (3): 517–47.

Wilson, Edward O. 1999. *Consilience: The Unity of Knowledge*. New York: Vintage.
Wittgenstein, Ludwig. 1973. *Philosophical Investigations: The English Text of the Third Edition*. New York: Macmillan.
Wolf, Pamela H., Jennifer H. Madans, Fanchon F. Finucane, Millicent Higgins, and Joel C. Kleinman. 1991. "Reduction of Cardiovascular Disease—Related Mortality among Postmenopausal Women Who Use Hormones: Evidence from a National Cohort." *American Journal of Obstetrics and Gynecology* 164 (2): 489–94.
Yamada, Katsushi, and Tatsuo Furukawa. 1980. "Direct Evidence for Involvement of Dopaminergic Inhibition and Cholinergic Activation in Yawning." *Psychopharmacology* 67 (1): 39–43.
Yao, L., A. Divoli, I. Mayzus, J. A. Evans, and A. Rzhetsky. 2011. "Benchmarking Ontologies: Bigger or Better?" *PLoS Computational Biology* 7 (1): e1001055.