sociological science

Measuring Paradigmaticness of Disciplines Using Text

Eliza D. Evans, Charles J. Gomez, Daniel A. McFarland

Stanford University

Abstract: In this paper, we describe new methods that use the text of publications to measure the paradigmaticness of disciplines. Drawing on the text of published articles in the Web of Science, we build samples of disciplinary discourse. Using these language samples, we measure the two core concepts of paradigmaticness—consensus and rapid discovery (Collins 1994)—and show the relative positioning of eight example disciplines on each of these measures. Our measures show consistent differences between the "hard" sciences and "soft" social sciences. Deviations in the expected ranking of disciplines within the sciences and social sciences suggest new interpretations of the hierarchy of disciplines, directions for future research, and further insight into the developments in disciplinary structure and discourse that shape paradigmaticness.

Keywords: sociology of science; computational linguistics; disciplinary paradigms; hierarchy of sciences

THOMAS Kuhn's (1962) concept of paradigmaticness had long been influential in studies of science and scientific inquiry. A paradigm is a guiding set of theories, methods, and questions, as defined by a previous scientific achievement or publication in the field. For example, Newton's *Principia* became a guiding set of principles for classical physics, while later discoveries by Einstein and Planck became the basis of quantum mechanics. Each paradigm provides a widespread, consensual approach to the questions and methods of a discipline and produces a "particular coherent tradition of scientific research" (Kuhn 1962:10). Paradigmatic disciplines are characterized by, first, consensus over the core knowledge, questions, and methods of the field, as defined by the paradigm. Second, consensus allows for rapid discoveries in paradigmatic fields, which are focused in their inquiry on the pursuit of new knowledge—not debating the validity of old claims (Kuhn 1962; Collins 1994).

The concept of paradigms and paradigmaticness has organized much social scientific and popular thinking about the structure, hierarchy, and importance of scientific fields. In particular, possessing a paradigm has come to be a marker of status and, importantly, value. Kuhn (1962:11) wrote that "acquisition of a paradigm. . . is a sign of maturity in the development of any given scientific field," and based on this argument, whether or not a field is paradigmatic has become a symbolic boundary (Lamont and Molnár 2002) that divides the "hard," high-status, paradigmatic natural sciences from the "soft," low-status, preparadigmatic social sciences (Lodahl and Gordon 1972; Smith et al. 2000; Peterson 2015).

This division between the social and natural sciences is enduring and may be contributing to real-world consequence for the social sciences. Across time and disciplines, academics consistently rank the natural sciences higher in status than

Citation: Evans, Eliza D., Charles J. Gomez and Daniel A. McFarland. 2016. "Measuring Paradigmaticness of Disciplines Using Text." Sociological Science 3: 757-778.

Received: March 4, 2016

Accepted: April 19, 2016

Published: August 31, 2016

Editor(s): Gabriel Rossman

DOI: 10.15195/v3.a32

the social sciences (Lodahl and Gordon 1972; Smith et al. 2000). Nonacademics seem to accept this ranking and value the social sciences less highly. For example, in 2015, the U.S. Congress cut National Science Foundation (NSF) funds for social science research by 45 percent, even though the overall allocation of funds to the NSF increased (Sides 2015). One congressman, describing a political science grant recipient's research question, said that "we can answer that question in about 5 minutes" (158 Congressional Record 2012), articulating the devaluation of knowledge contributions from the social sciences that is implicitly suggested by the funding cuts.

Although paradigms and paradigmaticness are important to the sociology of science and shape both academic and popular perceptions of the value of different disciplines, we have no consistent way of measuring this characteristic of academic disciplines. Prior attempts to measure and quantify paradigmaticness have gone poorly for the most part, as we review below. The lack of a measure of paradigmaticness stunts research into how paradigmaticness is linked to the nature of scientific discovery and the production and evaluation of scientific contributions across disciplines.

In this article, we propose three new metrics to measure paradigmaticness. Using text from published research, we measure paradigmaticness as described by Collins (1994) in three different characteristics of disciplines: disciplinary consensus, stability at the disciplinary core, and the rate of rapid discovery at the discipline's periphery. Using publication data from the Thomson Reuters Web of Science (WoS) database, we create disciplinary corpora containing the abstracts and titles of disciplinary publications. Others have mapped the social and structural barriers that divide disciplinary communities (e.g., Boyack, Klavans, and Börner 2005; Leydesdorff and Rafols 2009; Moya-Anegón et al. 2004); here, we focus on the cultural differences among disciplines as represented by the language they use (Biancani N.d.; Vilhena et al. 2014). At its root, a paradigm is cultural, in that it shapes acceptable actions, language, and ways of thinking. As Latour and Woolgar wrote (1979:54), what is "referred to as the 'culture' in anthropology. . . is commonly subsumed under the term *paradigm* when applied to people calling themselves scientists."

We gauge consensus in disciplinary culture by measuring the entropy of disciplinary language. We show that paradigmatic disciplines display greater concentration of disciplinary language in particular terms, while the language of preparadigmatic fields is more diffuse. We measure rapid discovery by comparing the similarity of disciplinary language over time. Through weighted and unweighted cosine similarity measures, we are able to show the consolidation of disciplinary knowledge at the core and the rapid turnover of ideas at the periphery in the paradigmatic sciences, as compared to the preparadigmatic social sciences.

By being text-based, our measures offer key advantages over previous measures. First, text is the primary source of scientific discourse. More than citations or other characteristics of publications, which inform or describe the ideas in a paper, text is the ideas in a paper. Text and vocabulary reveal paradigms (Kuhn 1962:148) and differentiate among academic communities (Toulmin 1972: Chapter 2–4; Vilhena et al. 2014). Second, a text-based measure can be applied to any body of academic

work, in any publication format, and at a variety of scales. While we use a curated collection of publications from academic journals to test our measures, these metrics could be easily applied to text scraped from the Internet or other, less-curated sites of academic reporting. Third, we take advantage of current computational social science techniques and bring measures of paradigmaticness up to date from the previous work that mostly occurred decades ago, prior to the development of these tools.

Literature Review

The notion of a hierarchy of sciences, ordered from "hard" to "soft," is old and surprisingly uncontested. The concept dates back to at least the early nineteenth century and the writing of Auguste Comte (1865). The most widely understood hierarchy of sciences places hard, natural sciences like physics at the top of the hierarchy and soft social sciences at the bottom.[1] When academics are asked to rank a list of disciplines from hardest to softest, the rankings are remarkably consistent across time and raters: both physicists and sociologists alike put physics and chemistry at the top of the hierarchy and sociology and political science at the bottom (Lodahl and Gordon 1972; Smith et al. 2000).

Hard disciplines are paradigmatic disciplines (Kuhn 1962). Paradigmatic disciplines have consensus: they agree upon a paradigm that defines legitimate theories, questions, and methods for practitioners (Kuhn 1962:148). In Kuhn's telling, having a paradigm leads to rapid discovery: the paradigm guides and structures academic inquiry, allowing for rapid and consistent progress as scientists build on the work of their predecessors. The less developed the paradigm, the more preparadigmatic and softer the discipline is perceived to be (Kuhn 1962). Preparadigmatic disciplines feature "a multiplicity of competing schools" (Kuhn 1962:163) that seek to gain primacy within the discipline. Debate among these groups slows progress, and the lack of a consensual goal makes it difficult to measure the progress that does occur (Kuhn 1962:162–3).

Collins (1994), too, describes the hardness and softness of disciplines in terms of consensus and the pace of discovery, but he sees the two characteristics as interdependent rather than causally linked. Like Kuhn, Collins argues that consensus allows for rapid discovery, but he also argues that rapid discovery encourages consensus. For Collins, both characteristics developed in response to the use of technology in scientific knowledge production: technology allowed scientists in hard fields to discover new, observable phenomena. New discoveries became highly valued in these disciplines, and contesting old discoveries came to merit few accolades and little attention. Spurred by technology and ambition, scientists in hard fields became disinclined to disagree, and consensus grew out of a desire to pursue new discoveries and ignore the past. Freed from the disagreement and debate that drove earlier and other types of knowledge production, scientists could solely pursue new knowledge, spurring the rate of discovery to an ever more rapid pace. Whether consensus leads to rapid discovery (Kuhn 1962) or the two are mutually generative (Collins 1994), both characteristics are indicators of the paradigmaticness of disciplines.

Though these facets of paradigmaticness have been identified and described theoretically, they have proved difficult to measure empirically. Earliest attempts focused on the rapid discovery aspect of paradigmaticness. de Solla Price (1970) crafted an "Immediacy Index," which measured the rate of obsolescence of citations and appeared highly correlated with the paradigmaticness rankings of fields. However, later work showed that his results were a function of the growth rates of scientific literature rather than evidence of more rapid discovery in some fields than others (Cole, Cole, and Dietrich 1976). This suggests that citation age is not a good proxy for rapid discovery.

Subsequent efforts focused on measuring consensus. Cole (1983) found evidence of consensus in what he identifies as the core knowledge of disciplines—the knowledge that is codified in undergraduate textbooks. But his search for evidence of consensus at the knowledge frontier, where researchers are actively creating new knowledge, was unsuccessful. Examining peer evaluations of new research, citations, and other characteristics, Cole found few differences between disciplines that are consistently ranked at opposite ends of the hardness–softness scale. More recently, Shwed and Bearman (2010) were able to use citation networks to demonstrate consensus-building at the frontier of particular scientific debates, but they argue that their approach does not focus on or apply to whole disciplines, where "benign contestation" is a normal characteristic of scientific communities (Shwed and Bearman 2010:823–824).

Most recently, two measures of consensus have met with the greatest success. Both are based on the understanding that consensus allows a greater amount of information to be tacit and, as a result, ends the need for a "constant reiteration of fundamentals" (Kuhn 1962:18). This allows the texts of paradigmatic disciplines to be shorter than those of preparadigmatic disciplines, in which the fundamentals of the disciplines are contested and so must be reiterated in each publication. A first set of measures focused on this characteristic of consensus in the length of disciplinary texts (Ashar and Shapiro 1990; Pfeffer and Moore 1980; Salancik, Staw, and Pondy 1980) and used the length of disciplinary dissertations and their abstracts to measure the paradigmaticness of a discipline. In these studies, ordering the disciplines by the average length of dissertations and dissertation abstracts corresponded closely with the conventional hierarchy of sciences.

A second measure calculated fractional graph area (FGA), which describes the proportion of the page area of disciplinary journal publications that is occupied by graphs (Cleveland 1984; Smith et al. 2000). Smith et al. (2000) draw on Latour (1987) to argue that graphs encode consensual knowledge into compact forms. More graphs, then, suggest more consensual knowledge and higher paradigmaticness. Though more successful, these measures face obstacles and limitations. First, neither measure rapid discovery, a second key element of paradigmaticness. Second, as scientific publishing companies move increasingly towards open-access and online-only versions of their journals, figure-to-page-area ratios become increasingly difficult to measure, as articles feature embedded links to figures. Additionally, the far fewer space constraints of online media may shape the length of disciplinary publications. With these changes in publication technologies, these more recent measures may become obsolete.

We propose a text-based measure of paradigmaticness that is flexible in its application to different forms of publication media and addresses consensus, stability, and rapid discovery. We focus on published academic text, the site where consensus is achieved, discoveries are reported, and culture is circulated. Rhetoric is the key site of scientific communication and debate (Latour 1987), and our focus on the text of research publications (more so than metrics that examine citations, document lengths, or graphs) locates our measure in the center of scientific dialogue, language, and culture. With respect to culture, Vilhena et al. (2014:221) write that "underlying the apparent diversity of cultural objects is a common capacity to circulate. This suggests an analytical focus on human communication" because communication is the medium "through which values, tastes, styles, and logics are transmitted." Our focus on language locates our measure in the medium of scientific culture (Knorr-Cetina 1999; Toulmin 1972), and a paradigm, according to Latour and Woolgar (1979:54), is just a word for "culture" when talking about scientists.

Our measures offer methodological advantages over previous measures. We are able to differentiate between and measure both consensus and rapid discovery, allowing us to gauge the contribution of each to perceptions of paradigmaticness and explore the relationship between the two characteristics. In measuring consensus, we focus on the discipline's core. The core is defined as "having a relatively small number of theories and substantial consensus on the importance of these theories" (Cole 1983:114), which are viewed as essential to the discipline's canonical knowledge. We measure consensus by examining the concentration of disciplinary language in key words that represent the core.

Our other set of measures focuses on both the discipline's periphery and core. The periphery is the "research frontier" (Cole 1983) of disciplines, a cacophonous site of debate and discovery. In rapid discovery fields, there is very high knowledge generation and turnover through time at the periphery: only a fraction of the new discoveries in paradigmatic fields becomes significant and part of the core over time, while many are dropped from disciplinary discourse. We measure this characteristic of paradigmatic fields by calculating the proportion of distinct terms [2] that persist from one time period to the next in disciplinary text. When low, the measure suggests rapid discovery is occurring at the periphery, as many terms are dropping out of usage from one time period to the next. When high, the measure suggests a preparadigmatic state, in which debate around the same topics, using the same language, fills the disciplinary discourse over time.

The final aspect of a paradigmatic field concerns the continuity of its core knowledge. We measure stability in the core by calculating the concentration of disciplinary discourse in terms that persist over time. When disciplinary dialogue is concentrated in a particular set of persistent terms over time, it suggests that a stable—and thus concentrated—research focus is being sustained and that a research paradigm is present.

In what follows, we describe our data, which we draw from published articles found in WoS. We then operationalize our concepts—consensus, stability, and rapid discovery—as measures in the context of our focal samples of disciplinary discourse in eight scientific and social scientific disciplines.

Building Samples of Disciplinary Language

To obtain samples of disciplinary text on which we test our measures, we used publication titles and abstracts from Web of Science, a curated collection of publications in scientific and medical fields, primarily, but also containing publications from journals in the social sciences and humanities. Since 1991, coverage has been reliable. WoS tags every journal with two-letter disciplinary subject category tags (SCs), allowing us to associate each publication's text with particular disciplines. As others have done before us (Porter et al. 2007; Porter and Rafols 2009; Rafols and Meyer 2010), we extend the SC tags of a journal to the articles published within it. In the interests of space and clarity, we focus in this article on eight core scientific and social scientific disciplines: physics, chemistry, biology, math, psychology, economics, sociology, and political science. In our analyses, we show the differences in our measures among these disciplines and between the broader natural and social science classifications that contain them.

For each of our eight focal disciplines, we wanted to collect publications that were representative of the discipline. Because disciplines are sprawling areas of inquiry with poorly defined borders, the boundaries of disciplines are uncertain and shifting: even members of disciplines debate what lies inside or outside disciplinary boundaries. The most conservative sample of WoS publications representing the discourse of each discipline includes only the SC that is the name of the discipline (e.g., "Biology" [CU] or "Sociology" [XA]). For example, the most conservative sample of disciplinary text for psychology would contain articles tagged with the SC "Psychology" (VI) but not articles tagged with SCs such as "Psychology, Experimental" (VX) or "Psychology, Developmental" (MY). A broader definition of the discipline would allow for the inclusion of these additional subfield SCs.

For the main analysis in the article, we present results obtained using the most conservative sampling strategy. For six of our focal disciplines, the sample consists of articles tagged with the single SC tag that is the name of the discipline. However, neither chemistry nor physics has a single SC tag. For these two disciplines, we followed the next most conservative sampling strategy, in which we combined core subfield SCs following the pattern "Chemistry, <subfield>" and "Physics, <subfield>" into single SC tags representing each discipline.[3]

Each disciplinary corpus is a 5 percent random sample of all articles tagged with the discipline's SC(s) and published between 1991 and 2011, inclusive (N = 167,959). The online supplement contains a robustness check in which we compare the results obtained using this sampling strategy to results obtained with different and broader samples of disciplinary discourse. We find similar results across all samples. While our analyses here focus on the usage of unigrams, or single words, we also tested our main analyses with various combinations of n-grams, the combinations of n adjacent words. We reran our results using unigrams through 5-grams (i.e., every combination of five adjacent words in a corpus) and obtained similar results to those presented here. The results of all comparative analyses are presented in the online supplement.

Table 1 contains descriptive statistics about the disciplinary corpora, which includes the number of words, unique words, and documents for each of the eight

Table 1: Descriptive statistics of the eight disciplines used in the 5% sample.

Discipline Subfield SCs Included in Corpus[a]	Included Web of Science Subject Code(s)	Time Period	Number of Words	Number of Unique Words	Number of Documents
Biology	CU	1	79,399	10,291	1,977
		2	92,212	11,180	1,998
		3	105,481	11,702	2,021
		4	108,854	11,947	1,956
		5	120,651	12,439	2,204
		6	184,227	15,677	2,128
		7	372,843	22,570	2,538
Chemistry		1	345,362	23,319	4,676
Analytical	EA	2	442,570	26,485	5,399
Inorganic & Nuclear	EC	3	534,766	28,987	6,103
Organic	EE	4	569,138	30,847	6,324
		5	625,820	32,810	6,642
		6	735,885	35,133	7,534
		7	798,874	35,974	7,810
Economics	GY	1	35,956	4,805	1,322
		2	59,432	5,980	1,494
		3	78,953	6,332	1,575
		4	85,355	6,524	1,494
		5	91,867	6,580	1,590
		6	145,293	8,186	2,201
		7	185,383	9,318	2,652
Mathematics	PQ	1	39,733	3,997	1,594
		2	64,717	4,961	1,781
		3	82,169	5,705	1,945
		4	95,356	6,325	2,050
		5	109,754	6,657	2,042
		6	149,698	7,943	2,666
		7	185,262	9,035	3,117
Physics		1	452,734	16,461	5,992
Atomic, Molecular, & Chemical	UH	2	527,202	17,456	6,947
Condensed Matter	UK	3	675,908	21,092	7,717
Fluids & Plasmas	UF	4	739,883	22,247	8,252
Nuclear	UN	5	821,719	24,273	8,904
Particles & Field	UP	6	937,874	26,224	9,831
		7	983,417	27,119	9,928
Political Science	UU	1	17,324	4,587	1,388
		2	21,646	4,902	1,396
		3	28,710	4,979	1,437
		4	31,644	4,970	1,316
		5	30,935	4,712	1,025
		6	47,078	5,825	1,410
		7	59,781	6,435	1,531

Table 1 continued

Discipline Subfield SCs Included in Corpus[a]	Included Web of Science Subject Code(s)	Time Period	Number of Words	Number of Unique Words	Number of Documents
Psychology	VI	1	50,353	6,185	1,190
		2	71,041	7,041	1,230
		3	62,946	6,244	1,003
		4	60,030	5,821	940
		5	65,024	6,105	965
		6	87,719	6,974	1,340
		7	109,100	7,502	1,264
Sociology	XA	1	18,417	4,128	851
		2	30,016	5,140	841
		3	34,112	4,980	859
		4	35,586	4,795	796
		5	34,418	4,686	748
		6	45,778	5,520	962
		7	58,416	5,947	1,063

[a] Subfield SCs were used to create the corpora for chemistry and physics, each of which did not have a single, disciplinary SC in Web of Science. For all other disciplines, the SC used to build the disciplinary corpus is the SC bearing the discipline's name.

disciplines. We also include the WoS SCs used to build each disciplinary corpus. We used these disciplinary corpora as the data on which we test out measures of consensus, stability, and rapid discovery. In the following section, we formally define the three measures.

Measuring Paradigmaticness Using Text

Measuring Consensus

Consensus in a discipline legitimates particular theories, questions, and methods, narrowing the scope of what practitioners are legitimately able to ask and do. This leads to a narrowing of disciplinary language in paradigmatic disciplines, as disciplinary experts converge on the language of the discipline's core questions and methods. With greater consensus in a discipline comes greater concentration of disciplinary language in particular words.

To capture consensus in the language of disciplines, we measure the Shannon entropy of the language in each disciplinary corpus. Entropy tells us how the language of each discipline is distributed among the set of all words it contains. In Figure 1, we plot the frequency of words in two imaginary disciplinary corpora to show how entropy reveals consensus in the language of a discipline.

In the left pane, showing high entropy, the language of the discipline is spread evenly across most of the terms in the vocabulary of the discipline. This is an exaggeration of what occurs in a preparadigmatic discipline, where competing

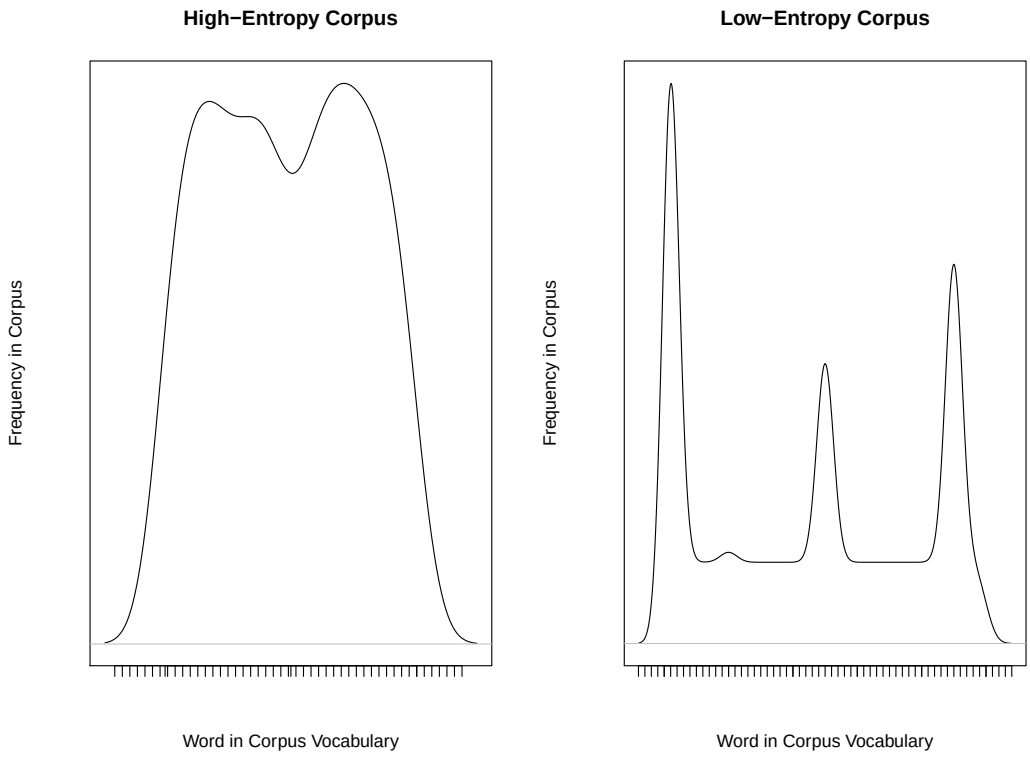

Figure 1: Hypothetical distributions of term frequency in high- and low-entropy corpora.

subfields each use different words to describe different methods, theories, and problems within the discipline. This spreads disciplinary language evenly across a broad set of terms. The more the language of the discipline is spread evenly over the set of unique words in the disciplinary corpus, the higher the entropy.

In the right pane, the language of the discipline is concentrated in particular sets of words. The greater the concentration of disciplinary discourse in a particular subset of words, the lower the entropy. This graph exaggerates what happens linguistically in a paradigmatic discipline, where disciplinary discourse is concentrated on the language of a particular set of theories, methods, and problems. Other approaches, described with other language, become peripheral to disciplinary discourse. With greater coherence and consensus in disciplinary language, entropy decreases.

We measure consensus using a version of Shannon entropy. The measure is defined by the following equation,

$$consensus \ =: 1 - \frac{\sum_{w \in d} \left[p_w \times \log(p_w) \right]}{\sum_{w \in d} \left[\frac{1}{V} \times \log(\frac{1}{V}) \right]} \tag{1}$$

where w is each word in a disciplinary corpus d. p_w is the probability of observing word w in corpus d, and all p_w sum to equal 1. We turned this value into a proportion by dividing the obtained entropy value by the maximum possible entropy value for the disciplinary corpus d.

To calculate the maximum possible entropy value for d, we calculated the Shannon entropy in which $p_w = \frac{1}{V}$. V is the vocabulary size, which is the total number of different words that appear in the corpus d; when calculating V, each word is counted once, whether it appears one or many times in the corpus. By making $p_w = \frac{1}{V}$, we establish conditions in which each word w in corpus d has an equal probability of being observed. This is the condition of maximum entropy because the language of the corpus is spread exactly evenly across all terms. We divided our observed entropy by the maximum possible entropy value to obtain a measure of proportional entropy. We then subtracted the proportional entropy value from 1 to get our measure of consensus.

Measuring Rapid Discovery at the Periphery and Stability at the Core of a Discipline

Rapid discovery is reflected in how disciplinary language changes over time at its periphery. In general, paradigmatic disciplines, driven by rapid discovery, demonstrate great linguistic change over time: new ideas and data enter disciplinary discourse at an accelerated pace as practitioners pursue and report ever more new discoveries, but many of these ideas are discarded and never discussed again, leading to high linguistic turnover at the periphery. On the other hand, preparadigmatic disciplines produce knowledge and evolve through longitudinal debate and discussion, and so the same topics appear in disciplinary discourse repeatedly over time as new thinkers repeatedly rehash older ideas (Cole 1983). Though topics may wax and wane over time, the periphery of preparadigmatic fields does not change at the rapid pace of paradigmatic fields.

At the core, stability and continuity are a necessary condition for rapid discovery at the periphery (Kuhn 1962; Collins 1994). The entropy measure of consensus shows concentration in any words over time, but the measure we describe below shows continuity and concentration of disciplinary language in the same words over time. The rapid linguistic changes at the periphery of paradigmatic disciplines are facilitated and enabled by the continuity and stability of the same ideas and same language at the disciplinary core. Continuity in the core of a paradigmatic discipline suppresses longitudinal debate and leads to high consensus when evaluating new knowledge, facilitating the rapid dismissal of many ideas and the incorporation of a few new ideas into the core. In preparadigmatic disciplines, on the other hand, the churn of topics over time leads to an ever-shifting disciplinary core: the most central issues of the discipline change over time as scholars emphasize different topics over time.

In paradigmatic disciplines, we expect greater stability at the linguistic core over time but greater turnover at the periphery. In paradigmatic disciplines, there will be low similarity over time in the whole of what is being discussed, as the rapid turnover at the periphery adds and subtracts lots of new words over time. For

preparadigmatic fields, we expect the opposite: as topics are hashed and rehashed in the disciplinary discourse, the whole of the disciplinary discourse will be similar over time, but the most central terms will change. To capture the peripheral aspect of this mechanism in disciplinary language, we measured the raw proportion of terms that appear in disciplinary discourse across time. We expect a greater proportion of terms to persist over time in preparadigmatic fields than in paradigmatic fields.

However, for the language that does persist over time in paradigmatic disciplines, we expect it to occupy a lot of the disciplinary conversation: the words that survive over time are likely to be part of the disciplinary core, which is much-discussed in paradigmatic disciplines. In preparadigmatic disciplines (because of their lack of focus and consensus), a greater number of distinct words may be carried forward across time, but none are likely to be as central to the disciplinary conversation as the core terms and concentrated dialogues of a paradigmatic discipline. To capture this mechanism in disciplinary language at the core, we examined the frequency of the terms that persist over time within the disciplinary discourse and expect that in paradigmatic disciplines, conversation is concentrated in these terms.

To capture the mechanisms of rapid discovery at the periphery and stability at the core, we used cosine similarity to compare disciplinary discourse over time. We implemented two cosine similarity measures to capture the different dynamics of rapid discovery and stability in the language of disciplines. For both cosine similarity measures, we divided each disciplinary corpus into seven separate bodies of text, each of which contained the abstracts and titles from disciplinary publications in a particular three-year window (Figure 2). For each three-year window of text, we calculated the cosine similarity between that text and the text of the preceding three-year window (shown by the arrows in Figure 2).

Cosine similarity is a common natural language processing method that uses the occurrence of overlapping words in two bodies of text to calculate a similarity score between them, independent of the length of the two corpora (Jurafsky and Martin 2010). The measure can be weighted and scaled in a variety of ways.

Our first measure implements cosine similarity in its most basic form, in which the result is a proportion of words that appear across time periods, normalized by the total number of words in the two time periods. This measure describes the linguistic characteristics of rapid discovery at the periphery of disciplines, where paradigmatic disciplines carry forward a smaller proportion of terms over time than preparadigmatic disciplines. The calculation is represented by the equation

$$rapid\ discovery\ \left(\overrightarrow{d_j}, \overrightarrow{d_{j+1}}\right) =: 1 - \frac{\sum_{w \in d_j, d_{j+1}} \left[\ tp_{w,d_j} \ \times \ tp_{w,d_{j+1}}\right]}{\sqrt{\sum_{w \in d_j} \left(tp_{w,d_j}\right)^2} \times \sqrt{\sum_{w \in d_{j+1}} \left(tp_{w,d_{j+1}}\right)^2}} \quad (2)$$

where d_j is the disciplinary text from time window j, d_{j+1} is the disciplinary text from time window $j+1$, and w is each unique word in the union of the text from the two time periods. $tp_{w,j}$ is the presence of word w in d_j, and $tp_{w,j+1}$ is the presence of word w in d_{j+1}. Each value is "1" if the word is present in d_j or d_{j+1}, respectively, and "0" if the word is not present. It does not matter how many times a word is used, only that it is present in both consecutive time periods. The denominator

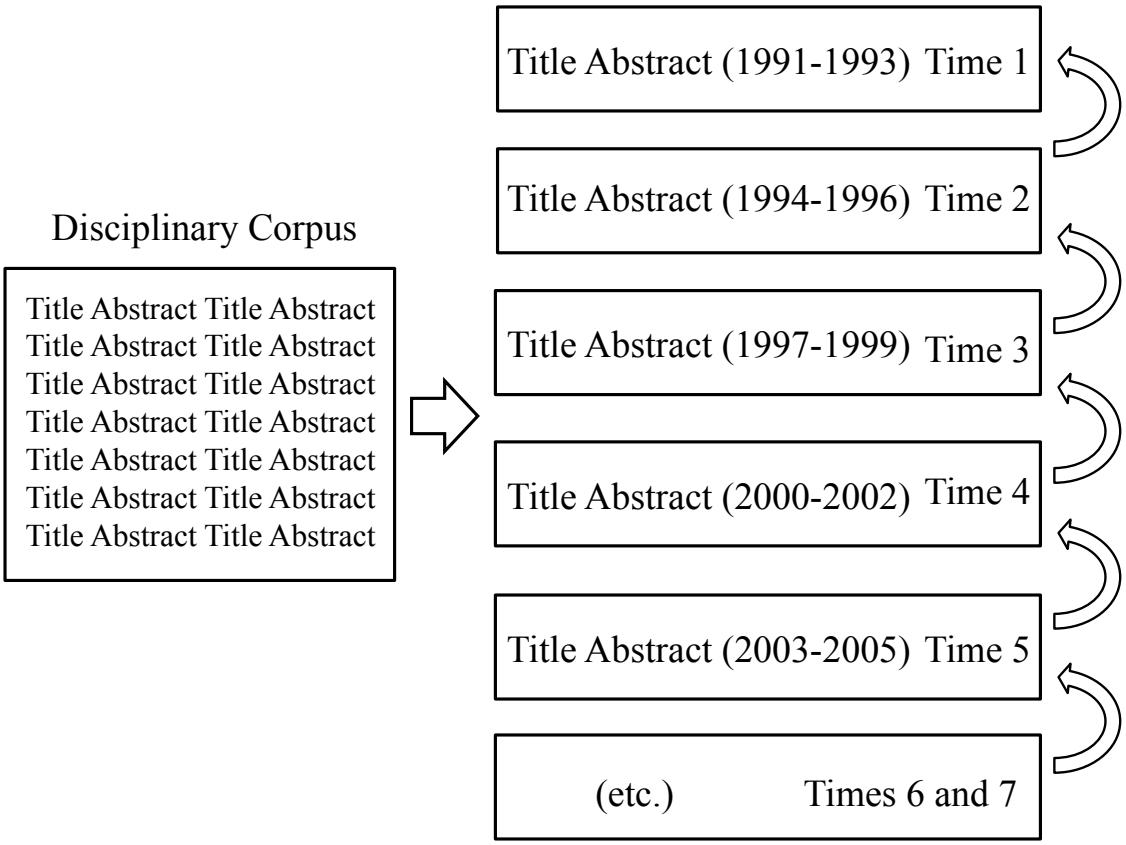

Figure 2: Schematic for over-time comparisons of text in metrics of rapid discovery and stability at the core.

allows us to normalize the values by the length of the two bodies of text being compared in order to obtain a proportion of overlapping words rather than a raw count. We subtract this similarity value from 1 to illustrate rapid discovery; thus, fields that have low cosine similarity values over time are high in rapid discovery. We expect that paradigmatic disciplines with rapid discovery will have low cosine similarity over time using this measure, as their language changes quickly over time at the periphery.

In some regards, this measure can capture a process of "obliteration by incorporation" (Garfield 1977; Merton 1968:27–29, 35–38). For preparadigmatic fields, the general language and vocabulary persists through dialectical debate over time. By contrast, paradigmatic fields focus on core concepts that persist in language, while secondary terms turn over or disappear. These peripheral terms may disappear because they get related to and encapsulated in the language of the paradigm's core theories and perspective. This is similar to the process of obliteration by in-

corporation for citations[4] because they are discarded in the face of rapid, ongoing discovery.

That said, in other analyses we find evidence that terms are used persistently over time, even as they are cited differently. Especially in the case of references to methodological concepts, core ideas receive less explicit recognition over time but are often still represented in the language of a discipline. As one example, Francis Galton (1894) is often credited as the creator of linear regression. Though this citation now rarely appears in the literature, the term "regression" remains common. Our measure may be susceptible to obliteration by incorporation of terms, but changes in language over time are more likely to be driven by the focus (or lack of focus) on particular topics, especially in the relatively short 20-year time window we analyze. It is possible that over longer periods of time, we will observe terms becoming implicit and then eventually being dropped, but that does not appear to be the case in our 20 year time window.

Our second measure addresses stability at the core. Again, we calculate the cosine similarity between the corpora of consecutive time periods, but instead of using a binary representation of the presence and absence of words, we count the frequency of the words in each time period to gauge the extent to which overlapping terms occupy a more central role in the discourse of disciplines. Much of the discourse of paradigmatic disciplines takes place in the periphery, but even if there is rapid turnover at the periphery (as measured by the basic, binary calculation of cosine similarity), we expect that the terms that do survive will be more central to disciplinary discourse, which this measure captures. We capture this mechanism with a term frequency (tf) weighting scheme.[5] Tf weighting gives greater weight to overlapping words that are frequent in the two corpora being compared. The tf-weighted cosine similarity measure that we use to compare disciplinary text from different time windows is described by the following equation,

$$stability\left(\overrightarrow{d_j}, \overrightarrow{d_{j+1}}\right) =: \frac{\sum_{w \in d_j, d_{j+}} \left[tf_{w,d_j} \times tf_{w,d_{j+1}} \right]}{\sqrt{\sum_{w \in d_j} \left(tf_{w,d_j}\right)^2} \times \sqrt{\sum_{w \in d_{j+1}} \left(tf_{w,d_{j+1}}\right)^2}} \quad (3)$$

where d_j is the disciplinary text from time j, d_{j+1} is the disciplinary text from time $j+1$, and w is each unique word in the union of the text from the two time periods. tf_{w,d_1} is the frequency (count) of word w in the text of d_j. $tf_{w,d_{j+1}}$ is the frequency of word w in the text of d_{j+1}. Words that are oft-used in both time periods receive the highest weights in this measure. As a result, in paradigmatic fields where core language is often used and stable over time, we expect to see higher cosine similarity values than for preparadigmatic fields, where even the most central terms occupy less of the disciplinary discourse.

This measure of stability at the core of disciplines is closely related to the entropy measure described above: both show the extent of consensus and linguistic concentration at the core of a discipline. However, this measure of stability is different in that it shows continuity and concentration of disciplinary language in the same words over time, while the entropy measure of consensus shows concentration in any words over time.[6]

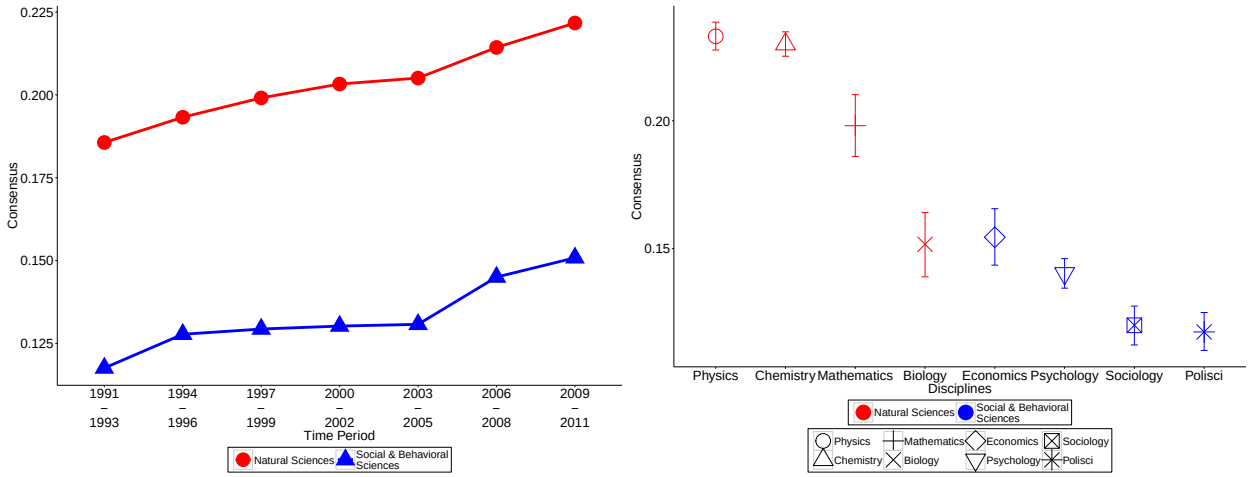

Figure 3: Consensus values over time by science/social science group and discipline.

Results

Figure 3 contains two panes showing the values of the consensus measure when performed on the text corpora of each of our eight focal disciplines. The left pane splits the eight disciplines into two groups—the natural sciences (i.e., mathematics, physics, chemistry, biology) and the social and behavioral sciences (i.e., economics, psychology, political science, and sociology)—and plots their average consensus values over seven three-year time windows: t_1 as 1991–1993, t_2 as 1994–1996, t_3 as 1997–1999, t_4 as 2000–2002, t_5 as 2003–2005, t_6 as 2006–2008, and t_7 as 2009–2011. The right pane plots each discipline as a dot and confidence interval, which represents the mean consensus value for each discipline averaged across all seven time periods.

The relative positioning of the disciplines parallels the ordering of hard to soft (or paradigmatic to preparadigmatic) that academics typically ascribe to the disciplines. Broadly, the scientific and mathematic disciplines do exhibit greater consensus than the social scientific disciplines. The ordering within the natural sciences does match the typical order assigned by academics. Within the social sciences, the ordering matches expectations: the discourses of economics and psychology—the "harder" of the social sciences—exhibit more consensus than sociology or political science. It is also important to remember that paradigmaticness is a multifaceted concept, represented by consensus, stability, and rapid discovery.

Figure 4 displays the results of the measure of rapid discovery at the periphery of disciplines—which measures the raw proportion of overlapping words across time periods—using a binary representation of each word's presence or absence. Like Figure 3, Figure 4 plots the natural and social science group means (left pane) and the disciplinary means (right pane) for rapid discovery at the periphery. Higher values indicate that disciplinary language exhibits more change over time, indicating more rapid discovery and greater inclusion and deletion of words over time. Again, we see clear differentiation between the scientific and social scientific

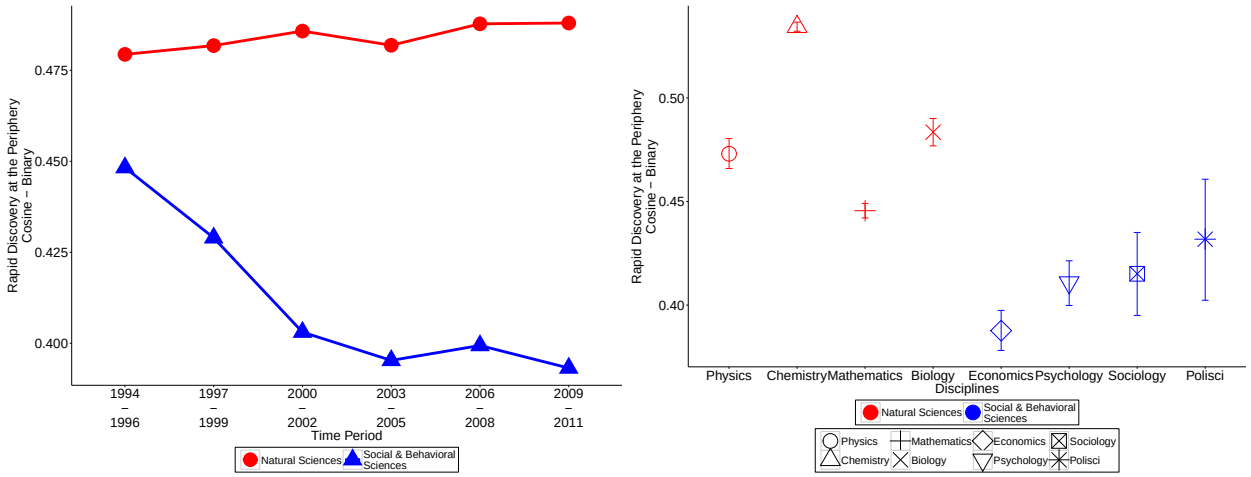

Figure 4: Rapid discovery values over time by science/social science group and discipline.

disciplines. The discourse of scientific disciplines is more dissimilar over time (i.e., a higher rapid discovery score), suggesting the turnover of ideas and language that occurs in rapid discovery at the periphery. The discourse of social scientific disciplines is more similar over time (i.e., a lower rapid discovery score), suggesting lower rates of rapid discovery and the churn and dialectical debate that drives progress in these disciplines.

The ordering of disciplines by our measure of rapid discovery at the periphery most differs from the conventional ordering of disciplines from hard to soft, or paradigmatic to preparadigmatic. Within the sciences, chemistry and biology demonstrate greater evidence of rapid discovery through language change over time than physics. Within the social sciences, the ordering is also different from what we would expect: economics and psychology exhibit less linguistic change over time as compared to sociology and political science.

Figure 5 displays the values for stability at the core of disciplines. In this measure, we use term frequency weighting in the cosine similarity metric to gauge whether the language that is retained in a discipline over time is central to disciplinary discourse. Higher values indicate stability at the core, as ideas that are retained over time are oft-used in disciplinary discourse. Like Figures 3 and 4, Figure 5 plots the natural and social science group means (left pane) and disciplinary means (right pane) for stability at the core. In this measure of consolidation at the core, we see that in physics, words that are retained over time occupy the disciplinary core most prominently. Physics is followed by, in order, chemistry, math, economics, biology, psychology, sociology, and political science. The disciplines closely follow the expected theoretical ordering based on this measure, except for economics. The relatively high positioning of economics on this value may be related to the increasing emphasis on econometrics in the field, as we discuss in greater detail in the discussion section below.

Figure 6 summarizes our key findings by plotting stability at the core and rapid discovery at the periphery, respectively, by the degree of consensus on the

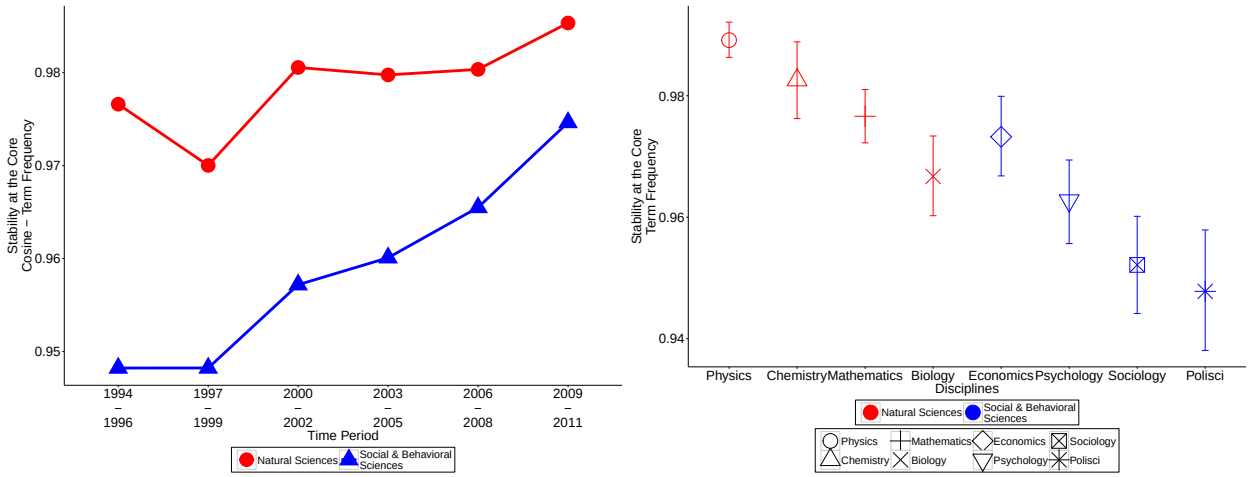

Figure 5: Stability values over time by science/social science group and discipline.

x-axis. The lines and shading in in Figure 6 show the correlations (with 95 percent confidence intervals) between the pairs of measures. The eight dots represent each of our eight disciplines' average value for our measures of rapid discovery, stability, and consensus across the seven time periods (they are the same values as those in the right panes in Figures 3, 4, and 5). Figure 6 tells a clear story. Stability (left) and rapid discovery (right) are significantly, positively related to consensus, just as described by Kuhn's concept of paradigmaticness. Cronbach's alpha ($\alpha = 0.73$) and a principal components analysis (details in the online supplement) suggest that the three measures gauge a single underlying latent concept: paradigmaticness.

Discussion and Conclusions

In this article, we proposed three new metrics that address multiple facets of paradigmaticness: consensus, rapid discovery, and stability. Using publication data from WoS, we created samples of disciplinary discourse from the abstracts and titles of disciplinary publications. Using these corpora, we measured consensus by measuring the entropy of disciplinary language, in which lower entropy shows greater concentration of disciplinary language in particular words. This concentration of disciplinary language suggests greater consensus, as more disciplinary discourse revolves around the particular questions, theory, and methods represented by these words. We measured rapid discovery by assessing the continuity of disciplinary language over time at both the periphery and the core of disciplines. Through weighted and unweighted cosine similarity measures, we assessed the concentration of disciplinary discourse in the same terms over time and the rapid turnover of terms at the periphery in paradigmatic disciplines as compared to preparadigmatic disciplines.

The rank-ordering of disciplines by each of the measures of consensus and stability corresponds well with prior theoretical claims about the paradigmaticness of disciplines and their places in a hierarchy of disciplines. The scientific disciplines

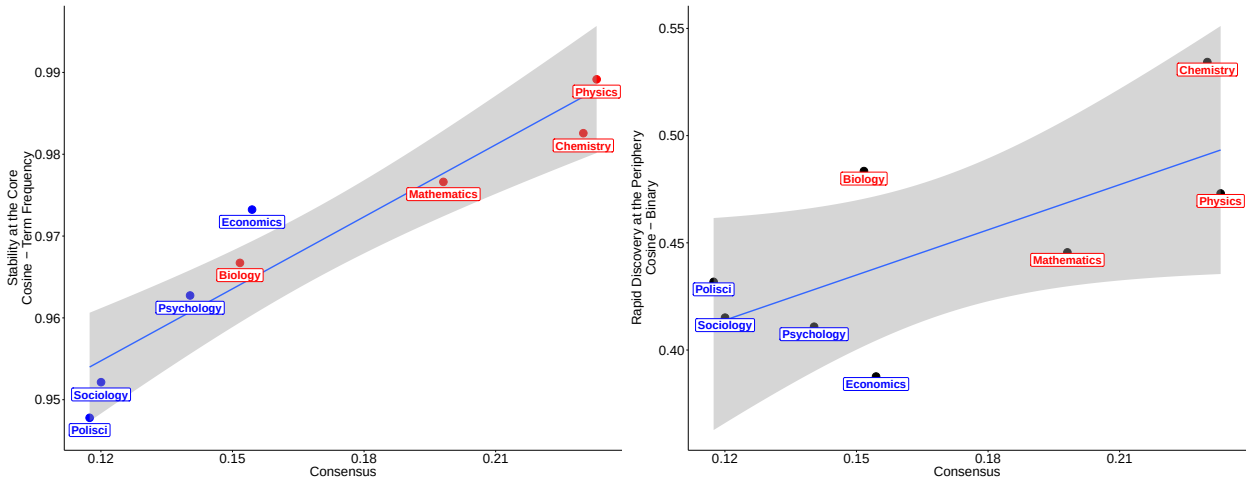

Figure 6: Graphed correlation lines between consensus and each of stability at the core and rapid discovery at the periphery (stability at the core with an R^2 of 0.95 and rapid discovery at the periphery with an R^2 of 0.69).

displayed higher consensus, more rapid discovery, and greater stability, while the social sciences exhibited lower consensus, less rapid discovery, and less stability. Within these broad academic cultures (Kagan 2009), the ordering of disciplines mostly followed the patterns previously suggested in the literature. However, our measure of rapid discovery showed an ordering different from this typical hierarchy, with chemistry demonstrating the fastest rate of rapid discovery as opposed to physics or mathematics. For the measure of rapid discovery, the order within the social sciences is also different from expected. The deviations in this measure suggest that rapid discovery may be contributing less to perceptions of an academic hierarchy than consensus and stability in disciplines. Future research using the rapid discovery measure may map variation in this measure to scientific advancements and investigate whether rapid discovery can be used to model and perhaps predict new breakthroughs and even paradigm shifts (Kuhn 1962).

In the social sciences, economics demonstrated interesting deviations from our expectations. The discipline was notably high-ranking in our measure of stability at the core, as it was more stable than biology and nearly as stable as math (Figure 5). This may be related to the increasing influence of mathematical models in the discipline (Leijonhufvud 1973) and the shoring up of social boundaries around the discipline through citations and cultural norms of behavior (Fourcade, Ollion, and Algan 2015). The quantification of economics had steered the discipline towards mathematical models and methods-based patterns of discourse, which may have led to greater consolidation of information at the core (Collins 1994). Fourcade et al. (2015) describe the socioacademic norms that economists have developed to separate themselves from the other social sciences and define themselves as more scientific and mathematical than sociology or political science. Our measures suggest that differences in the language of economics as compared to other social sciences may also reflect these disciplinary differences.

Our measures have vulnerabilities that offer future opportunities for research into the relationship between language use and the construction and maintenance of academic cultures. For example, token words are sometimes polysemous and can carry different, changing, or conflicting meanings across time and for different disciplines. For this reason, our measures may fail to capture the different meanings that a term like “culture” has for anthropologists and biologists. That said, our measures gauge commonality in word usage for large arrays of terms, from many persons and documents, within disciplines and across time. At an aggregate level, most of these differences in meaning will be signaled by different ancillary vocabulary and the changed meaning brought by their inclusion. For instance, the usage of "culture" might be stable over time in both sociology and biology, but for sociology, the use of “culture” in close proximity to "art" might be used less frequently over time, while usage of the term with "meaning" or "toolkit" may increase over time. In the biology corpus, use of “petri” alongside “culture” may decrease, while the pairing of “culture” and “assay” increases. Considering the occurrence (or not) of both “culture” and its ancillary terms helps to differentiate between biology’s and sociology’s usage of the term while also showing shifts in meaning and application over time. In this way, deceptive stability and similarity in a particular unigram is accompanied by more conspicuous shifts and differences elsewhere.[7]

Multiword concepts and propositions are less prone to polysemy because they qualify meanings. As a robustness check, we compared a bigram model (in addition to the cumulative model presented in the online supplement) to our unigram model. For each measure of paradigmaticness—consensus, rapid discovery at the periphery, and stability at the core—the correlations between the unigram and bigram-only models were 0.94, 0.94, and 0.99, respectively. We also ran cumulative n-gram models through $n = 5$ (outlined in the online supplement) and compared those results. In all cases, the same general pattern of results holds, suggesting our initial unigram measures are robust in spite of the potential for polysemy. In many regards, this is consistent with intuitions in natural language processing in which the simplest model (unigrams)—in the aggregate—often captures a great deal of the complex variation in language usage while reducing computational costs.

In spite of potential limitations, our measures offer key advantages over previous work. Our text-based measures can be applied to any body of academic work and at a variety of scales. While we use a curated collection of publications from academic journals to test our measures, these metrics could be easily applied to text scraped from the Internet or other less-curated sites of academic reporting, such as textbooks or conference proceedings. These methods take advantage of current computational social science techniques and bring measures of paradigmaticness up to date from the prior work that occurred before the development of these tools.

Notes

1 Comte’s hierarchy does not match what is typically described now. Comte described sociology (social physics) as the top of the hierarchy, with the other scientific disciplines (astronomy, physics, chemistry, and physiology) falling below (1865:46).

2 Recently, Vilhena et al (2014) argue that scientific concepts are really propositions. Thus, a method that allows for groups of words ranging from unigrams (i.e., single words) to n-grams (i.e., multiple words grouped together) may offer more at propositions and concepts (e.g., "social capital") than unigrams (e.g., "evolution"). The online supplement describes the extension of these measures to n-grams, across which we found similar results.

3 For physics and chemistry, we excluded subfield SCs that were "Applied" or "Multidisciplinary" because these categories suggest work that spans across disciplines. Included subfield SCs had pairwise cosine similarities above 0.9 in each time period. The online supplement contains analysis using all subfields, for which we obtained similar results (see online supplement for details).

4 The process of tacit knowledge creation is important and sufficiently complex in its own right to warrant separate study, and it is the focus of another developing article.

5 Often, this is combined with inverse document frequency weighting (idf), in which rare words in the whole set of corpora are given greater value. We found that idf weighting changed the absolute values but not the relative order of disciplines according to our measure, and so we do not include it in our model.

6 While our focus here is to empirically capture a hierarchy of sciences using text, if a discipline were to show low entropy and high consensus in each of two time periods but a low value of stability at the core, this could indicate a paradigm shift (Kuhn 1962), as the language of the discipline remains concentrated in particular terms but these terms have shifted over time. Future work can explore the applicability of these measures to capture these shifts.

7 We thank Gabriel Rossman for this insight.

References

158 Congressional Record H2543 (May 9, 2012) (statement of Sen. Flake), Available from: GPO Access, https://www.gpo.gov/fdsys/pkg/CREC-2012-05-09/pdf/CREC-2012-05-09-pt1-PgH2515-3.pdf#page=29; Accessed: 2/9/16.

Ashar, Hanna, and Jonathan Z. Shapiro. 1990. "Are Retrenchment Decisions Rational?: The Role of Information in Times of Budgetary Stress." *The Journal of Higher Education* 61:121–41. http://dx.doi.org/10.2307/1981958

Biancani, Susan. N.d. "A Hard Science is Good to Find: Textual Similarity as a Measure of Scientific Paradigm Development, A Preliminary Investigation." Workshop paper.

Boyack, Kevin W., Richard Klavans, and Katy Börner. n.d. "Mapping the Backbone of Science." *Scientometrics* 64:351–74.

Cleveland, William S. 1984. "Graphs in Scientific Publications." *The American Statistician* 38:261–69. http://dx.doi.org/10.1080/00031305.1984.10483223

Cole, Stephen. 1983. "The Hierarchy of the Sciences?" *American Journal of Sociology* 89:111–39. http://www.jstor.org/stable/2779049

Cole, Stephen, Jonathan Cole, and Lorraine Dietrich. 1976. "Measuring the Cognitive State of Scientific Specialties." Pp. 209–58 in *Toward a Metric of Science*. New York: John Wiley & Sons, Wiley-Interscience.

Collins, Randall. 1994. "Why the Social Sciences Won't Become High-Consensus, Rapid-Discovery Science." *Sociological Forum* 9:155–77. http://dx.doi.org/10.1007/BF01476360

Comte, Auguste. 1865. *The Positive Philosophy of Auguste Comte*. New York: C. Blanchard.

de Solla Price, Derek John. 1970. "Citation Measures of Hard Science, Soft Science, Technology, and Nonscience." Pp. 3–22 in *Communication Among Scientists and Engineers*. Lexington Books, D.C. Heath and Company.

Fourcade, Marion, Etienne Ollion, and Yann Algan. 2015. "The Superiority of Economists." *Journal of Economic Perspectives* 29:89-114. http://dx.doi.org/10.1257/jep.29.1.89

Galton, Francis. 1894. *Natural Inheritance*. Macmillan and Company.

Garfield, Eugene. 1977. "The 'Obliteration Phenomenon' in Science—and the Advantage of Being Obliterated!" Pp. 396-398 in *Essays of an Information Scientist, vol. 2*. Philadelphia: ISI Press .

Jurafsky, Daniel, and James H. Martin. 2010. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* 2nd ed. Upper Saddle River, N.J.: Pearson Prentice Hall.

Kagan, Jerome. 2009. *The Three Cultures: Natural Sciences, Social Sciences, and the Humanities in the 21st Century*. Cambridge University Press.

Knorr-Cetina, Karin. 1999. *Epistemic Cultures - How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions.* 3^{rd} edition. Chicago: The University of Chicago Press.

Lamont, Michèle, and Virág Molnár. 2002. "The Study of Boundaries in the Social Sciences." *Annual Review of Sociology* 28:167–95.

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

Latour, Bruno, and Steve Woolgar. 1979. *Laboratory Life: The Construction of Scientific Facts.* Princeton, New Jersey: Princeton University Press.

Leijonhufvud, Axel. 1973. "Life Among the Econ." *Economic Inquiry* 11: 327–37. http://dx.doi.org/10.1111/j.1465-7295.1973.tb01065.x

Leydesdorff, Loet, and Ismael Rafols. 2009. "A Global Map of Science Based on the ISI Subject Categories." *Journal of the American Society for Information Science and Technology* 60:348–62.

Lodahl, Janice Beyer, and Gerald Gordon. 1972. "The Structure of Scientific Fields and the Functioning of University Graduate Departments." *American Sociological Review* 37:57–72. http://www.jstor.org/stable/2093493

Merton, Robert K. 1968. *Social Theory and Social Structure*. New York: The Free Press.

Moya-Anegón, Félix, Benjamín Vargas-Quesada, Victor Herrero-Solana, Zaida Chinchilla-Rodríguez, Elena Corera-Álvarez, and Francisco J. Munoz-Fernández. 2004. "A New Technique for Building Maps of Large Scientific Domains Based on the Cocitation of Classes and Categories." *Scientometrics* 61:129-45. http://dx.doi.org/10.1023/B:SCIE.0000037368.31217.34

Peterson, David. 2015. "All That Is Solid Bench-Building at the Frontiers of Two Experimental Sciences." *American Sociological Review* 80:1201–25. http://dx.doi.org/10.1177/0003122415607230

Pfeffer, Jeffrey, and William L. Moore. 1980. "Average Tenure of Academic Department Heads: The Effects of Paradigm, Size, and Departmental Demography." *Administrative Science Quarterly* 25:387–406. http://dx.doi.org/10.2307/2392259

Porter, Alan L., Alex S. Cohen, J. David Roessner, and Marty Perreault. 2007. "Measuring Researcher Interdisciplinarity." *Scientometrics* 72:117–47. http://dx.doi.org/10.1007/s11192-007-1700-5

Porter, Alan, and Ismael Rafols. 2009. "Is Science Becoming More Interdisciplinary? Measuring and Mapping Six Research Fields over Time." *Scientometrics* 81:719–45. http://dx.doi.org/10.1007/s11192-008-2197-2

Rafols, Ismael, and Martin Meyer. 2010. "Diversity and Network Coherence as Indicators of Interdisciplinarity: Case Studies in Bionanoscience." *Scientometrics* 82:263–87. http://dx.doi.org/10.1007/s11192-009-0041-y

Salancik, Gerald R., Barry M. Staw, and Louis R. Pondy. 1980. "Administrative Turnover as a Response to Unmanaged Organizational Interdependence." *Academy of Management Journal* 23:422–37. http://dx.doi.org/10.2307/255509

Shwed, Uri, and Peter S. Bearman. 2010. "The Temporal Structure of Scientific Consensus Formation." *American Sociological Review* 75:817-40. http://dx.doi.org/10.1177/0003122410388488

Sides, John. 2015. "Why Congress Should Not Cut Funding to the Social Sciences." *The Washington Post* 10 June, 2015. Retrieved from https://www.washingtonpost.com/blogs/monkey-cage/wp/2015/06/10/why-congress-should-not-cut-funding-to-the-social-sciences/

Smith, Laurence D., Lisa A. Best, D. Alan Stubbs, John Johnston, and Andrea Bastiani Archibald. 2000. "Scientific Graphs and the Hierarchy of the Sciences: A Latourian Survey of Inscription Practices." *Social Studies of Science* 30:73–94. http://dx.doi.org/10.1177/030631200030001003

Toulmin, Stephen. 1972. *Human Understanding, Volume I: The Collective Use and Evolution of Concepts* (1st). Princeton: Princeton University Press.

Vilhena, Daril A., Jacob G. Foster, Martin Rosvall, Jevin D. West, James Evans, and Carl T. Bergstrom. 2014. "Finding Cultural Holes: How Structure and Culture Diverge in Networks of Scholarly Communication." *Sociological Science* 1(June): 221–38. http://dx.doi.org/10.15195/v1.a15

Acknowledgements: This project has been generously funded by the Brown Magic Grant, Dean of Research at Stanford University, and NSF Award #0835614. This material is based upon work supported by the NSF Graduate Research Fellowship Program (No. DGE-114747). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. This work was also supported by the Stanford Graduate Fellowship Program and by a generous grant from the Global Development and Poverty (GDP) exploratory project, sponsored by the Stanford Institute for Innovation in Developing Economies (SEED) and the Freeman Spogli Institute for International Studies. These data were collected by the Mimir Project conducted at Stanford University by Daniel McFarland, Dan Jurafsky, and Jure Leskovec. Access to these data was approved by the Mimir Project, and usage followed IRB guidelines.

Eliza D. Evans*: Graduate School of Education, Stanford University.
E-mail: elizae@stanford.edu.

Charles J. Gomez*: Graduate School of Education, Stanford University.
E-mail: cjgomez@stanford.edu.

Daniel A. McFarland: Graduate School of Education, Stanford University.
E-mail: dmcfarla@stanford.edu.

* Co-first authors and corresponding authors.