American Sociological Review
1–40

DOI:10.1177/00031224231166955
journals.sagepub.com/home/asr

How New Ideas Diffuse in Science

Mengjie Cheng,[a] Daniel Scott Smith,[b] Xiang Ren,[c] Hancheng Cao,[b] Sanne Smith,[b] and Daniel A. McFarland[b]

Abstract

What conditions enable novel intellectual contributions to diffuse and become integrated into later scientific work? Prior work tends to focus on whole cultural products, such as patents and articles, and emphasizes external social factors as important. This article focuses on concepts as reflections of ideas, and we identify the combined influence that social factors and internal intellectual structures have on ideational diffusion. To develop this perspective, we use computational techniques to identify nearly 60,000 new ideas introduced over two decades (1993 to 2016) in the Web of Science and follow their diffusion across 38 million later publications. We find new ideas diffuse more widely when they socially and intellectually resonate. New ideas become core concepts of science when they reach expansive networks of unrelated authors, achieve consistent intellectual usage, are associated with other prominent ideas, and fit with extant research traditions. These ecological conditions play an increasingly decisive role later in an idea's career, after their relations with the environment are established. This work advances the systematic study of scientific ideas by moving beyond products to focus on the content of ideas themselves and applies a relational perspective that takes seriously the contingency of their success.

Keywords

sociology of science, sociology of ideas, science of science, sociology of knowledge, innovation, diffusion

Why do some new ideas have long and varied careers, diffusing extensively among diverse social actors and distinct cultural contexts, morphing all along the way, while others are short-lived, find limited use, and are quickly forgotten? This question underpins the nature, structure, and development of knowledge. And it is pressing to consider in our contemporary knowledge-based (Powell and Snellman 2004), internet-driven society (Castells 1998), where people of all backgrounds are inundated with a dizzying and ever-expanding array of new ideas competing for their finite attention online (Bail, Brown, and Wimmer 2019; Shifman 2013), in business (Beath et al. 2012; Fuller 2010), and in science (Adair and Vohra 2003; Huth 1989; Prasad et al. 2010). Indeed, new ideas and knowledge are foundational to sustained scientific and

[a]Harvard University
[b]Stanford University
[c]University of Southern California

Corresponding Author:
Daniel A. McFarland, Stanford University, 520 Galvez Mall, Stanford, CA 94305, USA
Email: mcfarland@stanford.edu

technological advance as well as to effective policymaking, organizational change, and economic progress. Yet, while many new publications, inventions, and products offer novel intellectual contributions, only some stick, spread, and get meaningfully integrated into our collective understanding and knowledge base (Heath and Heath 2007). Understanding how and why this happens is the focus of this article.

To be sure, this article is not the first to ask why some new ideas diffuse while others do not. Extensive literatures in the history of science, sociology of knowledge and science, and even in the management sciences are dedicated to the task of exploring the conditions that promote the genesis of new ideas and facilitate their widespread adoption. The bulk of this literature emphasizes factors external to the ideas themselves to explain their propensity to diffuse. Based on these "externalist" accounts, a new idea spreads for a host of reasons located in its social context: the individual actors who create, champion, frame, and manage it (Lakatos and Feyerabend 2010; Perry-Smith and Mannucci 2017; Rogers [1962] 2010); the interconnections these actors share with other actors in the wider community and network (Grannovetter 1985; Rodan and Galunic 2004; Uzzi and Gillespie 2002); and even mundane material matters such as funding, institutional resources, and differential status and prestige (Bloor 1976; Bourdieu 1988; Merton 1968). Rightfully, in explaining an idea's diffusion, these accounts look to the identities, relationships, resources, and aims of the people who carry the idea widely and who continue to work on it at length. These accounts, in other words, see diffusion in the first instance as a social phenomenon (Sorenson and Fleming 2004).

Yet, a complementary perspective looks to the cultural and ideational context of the idea itself (Fiegel 1970; Kaufman 2004; Quine 1951; Toulmin 1972). According to these "internalist" accounts, the diffusion of a new idea is predicated on its comprehensibility and coherence (McDonald and Mair 2010), its relationship to established paradigms and thought styles (Fleck [1935] 1979; Kuhn 1970), and its pragmatic usefulness to solve problems and resonate with the intellectual and cultural moment (Antons, Joshi, and Salge 2019; Hallett, Stapleton, and Sauder 2019). Obviously, it takes people to imagine, understand, interpret, and situate an idea in an established body of knowledge at career junctures critical to its later diffusion. But integral to these internalist accounts is the ideational, indeed, intellectual context that a new idea enters and variably relates to when considering the conditions facilitating its later adoption.

The reality is that both external and internal conditions matter jointly and dynamically, depending on the new idea's own idiosyncratic career. This is the guiding intuition of this study. Drawing on these extensive literatures, we build and test a theory of diffusion that combines both external and internal factors (i.e., social and ideational factors). In particular, we argue that a new idea—in this study's case, a new scientific idea introduced in published journal articles—diffuses when its authors have great social prominence, span diverse, distal research collectives, and have consistent champions. We additionally argue that an idea is likely to diffuse when it is linked to prominent scientific facts, when it is deeply integrated into extant research traditions, and when it achieves coherence through consistent conceptual linkages to the established body of knowledge.

We move beyond a simple additive theoretical model that combines social and ideational factors. We bring in the perspective that ideas have careers: that these ideational and social conditions change over time and that these changes, in turn, correspond with a new idea's propensity to diffuse at different stages of its lifetime. Our findings related to these social and ideational conditions suggest a developmental story where an idea's continued resonating appeal depends on a changing cast of characters and associated uses (and used associations). An idea's meaning and position within ideational and social contexts is a diachronic process, and historically contingent shifts in its meaning have consequences for its diffusion over time.

To test this theory of diffusion, we assemble a uniquely comprehensive dataset (~7 billion token words from 38,578,016 articles of the Web of Science [WoS], 1900 to 2016), using advanced natural language processing (NLP) techniques, socio-semantic network analysis, and over-dispersed Poisson regression adapted to a multilevel context. Specifically, we extract scientific ideas from articles, identify which are new, and then trace their diffusion across published articles over time. By identifying the authors and neighbor terms associated with these new ideas within and across articles, we capture each new idea's networks of collaborating authors and networks of interrelated concepts. We then aggregate these networks over distinct periods to reflect a new idea's changing social and ideational context. In so doing, we develop a longitudinal ecological depiction of how new scientific ideas vary in their diffusion and how such variation corresponds with their adapted use. Finally, we model which social and ideational conditions are associated with new ideas that diffuse more widely, and we use interaction models to identify when these conditions are associated with a new idea's propensity to diffuse.

This article's depiction of the diffusion of new ideas changes our understanding of scientific development in several ways. First, it offers a uniquely comprehensive view of how new scientific ideas diffuse by analyzing a large, longitudinal corpus of scholarship that spans heterogeneous fields of knowledge for over 20 years. Second, it measures ideas in a more refined way by observing them as discrete new terms and expressions (not as whole articles), and it traces diffusion as the incorporation of these ideas within later articles (not as mere citation). Third, it integrates externalist and internalist accounts, identifying conditions under which new ideas find both social and intellectual resonance (or lack thereof), and shows how this relates to their future adoption. In particular, new ideas find intellectual resonance when they are consistently and intelligibly used, and they address valued intellectual problems (e.g., core to science and embedded in research traditions); they find social resonance when they achieve social reach (e.g., with expansively networked scholars). Last, the article depicts adaption as an important feature of ideational diffusion. Scholarly ideas are slow to take hold, and it is only after they establish their social and intellectual placement that fit conditions, or conditions of intellectual and social resonance, have greatest influence.

This study offers an important new contribution to our understanding of why new ideas diffuse. It is not simply that ideas diffuse when they have champions or star entrepreneurs. And, more generally, it is not simply a question of the status and resources of an idea's authors. These matter, certainly. But what we find is that they matter in conjunction with an idea's cultural resonance (Hallett et al. 2019): its relations to the established body of knowledge, its coherence and consistency, and its association with prominent ideas. These internal and external ecologies matter for both minor and major ideas. And they set off different trajectories for their careers, helping us understand how some new ideas continue to diffuse in science, while others do not and peter out.

WHAT IS A NEW IDEA?

Various literatures concern the diffusion of new ideas, but they vary in what they designate as a new idea and the mechanisms driving adoption. For example, research on the diffusion of innovation focuses on new products as reflections of new ideas. From this perspective, products are cultural gestalts or artifacts, like new books (Dodson and Muller 1978) and published papers (Crane 1972), songs (Salganik, Dodds, and Watts 2006), drugs (Coleman, Katz, and Menzel 1957), and inventions (Rogers 1976). This line of work analyzes such products, finding they tend to be accepted and shared when they are determined by editors, reviewers, patent officers, and other cultural brokers to impart novelty. Product-based depictions of new ideas can be useful and applicable for following and understanding the diffusion of innovation. For example, many companies and cultural entrepreneurs are explicitly interested in whether and how a specific new drug, text, or invention gets adopted. In some instances in science, it also makes intuitive sense to synecdochally represent a new idea with a new product (Merton 1968; Small 1978), such as when Keuchenius, Törnberg, and Uitermark (2021) equate Granovetter's article "The Strength of Weak Ties" with the idea of "weak ties." In these instances, articles-qua-products are rendered into highly portable conceptual tokens that get freely and broadly passed around.

Yet, most scientific articles relate a variety of new and old ideas, and a focus on the article as a holistic product leads to generic and coarse representations of new ideas (Foster, Rzhetsky, and Evans 2015; Uzzi et al. 2013). This obstructs efforts to precisely locate and trace the diffusion of new ideas. For example, prior work consistently finds that references (i.e., diffused scientific ideas encased in published articles) are cited for a variety of reasons in ensuing scientific work, only some of which have to do with authors' attempts to establish dependence on them or attribute their provenance (see Jurgens et al. 2018; Teplitskiy et al. 2022). Moreover, many citations to published scientific articles do not target the true origin of an idea but instead capture modern practices of paraphrasing, whereby extant ideas across distinct articles are intermeshed ("palimpsestic syndrome"), and some articles are even treated interchangeably ("citation substitution") (see McCain 2014; McMahan and McFarland 2021; Merton 1965:xxiii; Zuckerman 1987).

Alternatively, the original article reporting on the initial discovery of the new idea may fade from citational practices altogether, as scholars and inventors take it for granted and thus do not cite it ("obliteration by incorporation"; McCain 2014; Merton 1988:622). Last, citations are frequently biased, with higher-status authors and papers cited for reasons beyond their topical relevance or even their claim to original authorship of the new idea (e.g., the "Matthew effect"; see Barabási and Albert 1999; MacRoberts and MacRoberts 1989; McCain 2014; Merton 1988). This emphasis on citations and the underlying privileging of the scientific article qua product as the analytic locus of novelty and diffusion misses authors' rhetorical moves as well as the conceptual development of science itself, thereby making it difficult to identify new ideas within, and track their spread across, scientific publications.

A more recent line of work offers a finer-grained identification of new ideas by shifting the focus to specific cultural elements or kernel ideas within product gestalts. So, instead of entire products as the units of observation and analysis, their constituent features are analyzed: specific characteristics of songs (Askin and Mauskapf 2017) or aspects of fashion (Godart and Galunic 2019); metadata associated with documents, such as categories, citations, and keywords (Denrell and Kovács 2020; Fleming, Mingo, and Chen 2007); or even specific words and phrases within documents (Carley 1997; Doerfel and Barnett 1999; Doerfel and Connaughton 2009; Hill and Carley 1999). From this perspective, new elements and even novel recombinations of old elements are seen as innovations, such as novel combinations of subjects, keywords, chemicals, or even words (Foster et al. 2015; Galunic and Rodan 1998; Hofstra et al. 2020; Leahey and Cain 2013; Leahey and Moody 2014; Uzzi et al. 2013). When new ideas are regarded within this element perspective, they can be observed and analyzed in interrelation and in competition for attention (Dawkins 1982; Heath, Bell, and Sternberg 2001; Weeks and Galunic 2003). In science, new ideas are often denoted by concepts or words, and these are typically interrelated in arguments of published reports (Carley 1997; Hill and Carley 1999). From these textual interrelations emerges a larger network of concepts, wherein some concepts and concept pairings emerge and gain frequent use, and others fail to take hold, fade, and disappear.

Various theoretical arguments about the body of scientific knowledge rely heavily on this notion of a conceptual or semantic network structure composed of ideas as elements. In general, the network or matrix of interrelated ideas is such that established hypotheses can have repercussions for new ideas or new ideational associations (Fiegel 1970; Quine 1951; Toulmin 1972). This means new ideas are contingent and must often build on and reaffirm prior relations or not be recognized at all. The preexisting complex or structure of scientific knowledge establishes their worth. For example, Fleck ([1935] 1979) and Kuhn (1970) describe how thought styles and paradigms often resist new ideas and empirical discoveries that do not fit extant interrelations and hypotheses on which established ideas depend (see also Godfrey-Smith 2003, 2010). This can diminish a new idea's propensity to take hold and diffuse. Conversely, should new ideas arise and gain influence, they may redirect attention and alter extant networks of association, potentially leading to a revolution or rewiring of conceptual relations (Collins 1998; Kuhn 1970). As such, science studies repeatedly find that new ideas, as elements within the network structure of science, either succeed or fail to garner sustained attention, depending on how well they mobilize and transform extant conceptual relations (Callon 1986; Callon, Law, and Rip 1986; Latour 1987).

An element perspective thus enables precision in identifying and tracing the diffusion of specific new ideas, but it also offers the building blocks to represent the structure of knowledge, which new ideas are introduced into, embedded in, and even sometimes reconstitute. It is this conceptualization of an idea—a constitutive *element* of a larger cognitive structure—that we move forward with in our arguments here.

WHY DO NEW IDEAS DIFFUSE?

Our central contention is that ideas have variable careers, where some take off and others fizzle out, and their changing adoption is heavily influenced by how they are adapted to and fit with social and ideational contexts (Denrell and Kovács 2020; Frickel and Gross 2005; Godart and Galunic 2019; Goldberg, Srivastava, et al. 2016). In particular, we argue that new ideas diffuse more when they are taken up by individuals who have greater social prominence (Gerow et al. 2018; Goel et al. 2016) and who span disparate collaborative research communities (Burt 2004; Moody 2004). They also diffuse when they have consistent intellectual leaders who persistently champion them over time (Merton 1968).

But we also maintain that ideas and their interrelations matter. We argue that the usage and placement of a new idea in the accruing network of scientific knowledge and concepts greatly affects its career (Gieryn 1978; Hallett et al. 2019; Kuukkanen 2008; McDonald and Mair 2010). New ideas take off when they resonate intellectually: they diffuse more when they are related to other prominent concepts, are deeply situated within focused research discourses and thought styles, and after they achieve coherent usage reflective in collective recognition. In what follows, we synthesize the literatures in the history of science and ideas, the sociology of science and knowledge, and management science to identify and characterize the influence of each of these factors on the propensity of new ideas to diffuse. We conclude this section by situating the ideational and the social conditions of an idea within its larger career, leading to a dynamic (diachronic) depiction of how ideas adapt as they get adopted, and how that, in turn, affects their diffusion.

Social Factors That Facilitate the Diffusion of New Ideas

Scholars have identified the individuality of social actors, such as managers, entrepreneurs, and luminary scientists, as a key mover in the diffusion of new ideas. For example, a multitude of biographies attest to the creative genius of various scholars and inventors and attribute their skills, training, attitudes, and aesthetic sensibilities as key to idea generation and diffusion (Lakatos and Feyerabend 2010; Rogers [1962] 2010). Here, the general argument is that certain characteristics make some individuals more influential, and influential individuals play outsized roles and get preferential treatment (or a "Matthew effect"; see Merton 1968). Prior work in the sociology of science focuses on this because a consistent finding is that research programs and communities tend to form around a few central scholars who can spread new ideas further afield (Barabási et al. 2002; Keuchenius et al. 2021; Newman 2009; Price 1976).

Luminaries have long played a decisive role in not only mobilizing a following behind new ideas, but also in redirecting the flow of new knowledge around themselves, including the diffusion of ideas (Azoulay, Fons-Rosen, and Graff Zivin 2019). Similar depictions can be found in the management literature (see Fleming et al. 2007), where a patent's association with highly central actors with expansive social network ties serves to garner more citations. This suggests an idea is more likely to diffuse when it is championed by authors integral to a domain. Likewise, the literature on threshold models (Coleman et al. 1957; Granovetter 1978) suggests that opinion leaders may pressure others to adopt a new idea, making it more palatable to yet more others, leading the idea to diffuse in an exponential fashion. Based on this insight, we expect that, as more prominent (more centrally collaborative) authors publish a new idea, it is more likely to be encountered, and therefore, more likely to diffuse to others. Stated formally:

Hypothesis 1: The more socially prominent its adopting authors, the more a new idea will diffuse.

Related to the above literature's emphasis on the role of individual thinkers and creators, another line of argument identifies the importance of having persistent champions for a new idea to diffuse (Rogers 1976). Here we turn to the scientific/intellectual movements (SIMs) literature, which emphasizes the substantial and concerted collective action needed for new ideas to become influential (Frickel and Gross 2005). To be sure, this coordinative work is not centralized but instead includes the sustained recruitment and mobilization of participants, representatives, and resources, as well as the relentless interfacing with and persuasion of reviewers, editors, publishers, and granters (Latour 1987). While hardly the work of any one individual, we nonetheless extend this line of reasoning in our observation that the success, coherence, and recognizability of the movement behind a new idea is partly a function of the consistency of its leadership and representatives, as well as the integrity of its long-run institutional memory.

Even within sociology, we see the persistence of social cores behind new schools of thought and ideas, where an esoteric intellectual elite (Fleck [1935] 1979; Lakatos and Feyerabend 2010)—real individual persons—come to represent the ideational core: for example, the third-wave historical/comparative sociologists, new institutionalists, the Carnegie School, the Chicago School, and so on (Abbott 1999). Other literature shows that the persistence of an expert core is associated with collaborative teams that are more successful, as it lends know-how and direction (Guimera et al. 2005). This literature, in short, suggests that a new idea might be more likely to diffuse if it gets consistent backing by authors, or when its social context is consistent. We expect that the more consistent the user base of a new idea is over time, the more likely it will get taken up. Stated formally:

Hypothesis 2: The more consistent its adopting authors, the more a new idea will diffuse.

Perhaps the most decisive external factor in a new idea's propensity to diffuse is the context of social relationships in which it is placed: its social embeddedness (Centola 2015; Grannovetter 1985; Rodan and Galunic 2004; Uzzi and Gillespie 2002). At least three arguments have been advanced to understand the diffusion of new ideas with respect to social embeddedness. Researchers have long argued that *communities* of scholars devote attention to an idea (Fleck [1935] 1979; Fujimura 1992; Kuhn 1970; Lakatos and Feyerabend 2010; Newman 2001, 2004). Less a function of a prophet-like scientist, these accounts emphasize the decentered, diffuse, and dense interrelations among many and varied collaborating scientists. Complementarily, in the management literature, dense and closed social networks of collaboration facilitate the efficient exchange of information and new insights (Fleming et al. 2007; Uzzi and Spiro 2005).

However, there is a diminishing return to such dense interconnection; with greater and overlapping interconnections among collaborating scientists comes redundancy of contacts and perhaps even competition (i.e., density dependence, see Hannan and Freeman 1989). Still, the insight here is that embeddedness in a densely interconnected, collaborative community predicts later uptake, because the collaborative social structure has more possibilities for new ideas to flow among authors. Based on this literature, we conceptualize a new idea's social embeddedness as a function of author collaboration. The more a new idea's authors collaborate, the more the idea is socially embedded within a scientific community and, therefore, the more likely it will diffuse. Stated formally:

> *Hypothesis 3a:* The more socially embedded among densely collaborating scientists, the more a new idea will diffuse.

Other literature emphasizes the inverse of the above argument: that expansive and open networks bring products and ideas to different audiences and into greater use (Burt 2004; Burt and Soda 2017; Fleming et al. 2007; Moran 2005; Rodan and Galunic 2004). Instead of densely and even redundantly interconnected communities of collaborating scientists promoting diffusion, social networks tend to bridge distinct and disparate subcommunities further afield to reach new audiences. This can be seen as an idea gaining a bigger network (Latour 1987) or acting as an object that spans the boundaries of distinct and diverse communities (Star and Griesemer 1989). It also relates to studies of scientific interdisciplinarity that find authors who bridge distinct disciplinary traditions can represent a high-risk, high-reward situation—with reward being greater diffusion and impact of the new idea, and risk being the authors' perceived incompetence or unrecognizability among participant communities (Shi, Leskovec, and McFarland 2010). Together, this literature suggests that new ideas are more likely to diffuse when they encounter different (versus the same and uniform) authors because their reach is more extensive (versus restricted). Stated formally:

> *Hypothesis 3b:* The more socially extensive its reach among disparate communities of collaborating scientists, the more a new idea will diffuse.

Ideational Factors That Facilitate the Diffusion of New Scientific Ideas

Scientists introduce new ideas within a broader intellectual environment of existing knowledge. How they situate their new idea, or how they frame and place it in the body of other interrelated ideas, has real ramifications for whether the new idea has intellectual resonance and cultural fit and whether and how it will diffuse (Hallett et al. 2019). For example, a new idea's interrelation with other prominent ideas can draw more attention to it and make it easier to find, situate, and understand (Denrell and Kovács 2020; Godart and Galunic 2019; Piazza and Castellucci 2014; Sorenson 2014). In the sociology of science, a key quality of influential ideas is that they get related to other foundational scientific concepts and problems and are thereby integrated into the core of paradigms and thought styles (Evans, Gomez, and McFarland 2016; Fleck [1935] 1979; Kuhn 1970). In other words, new ideas are more likely to diffuse when they are linked to well-established, central constructs within the intellectual structure of scientific knowledge. We expect that as a new idea gets linked to these more prominent and integrated ideas, it is more likely to be encountered, recognized, and understood within the established semantic space of science and, therefore, to diffuse. Stated formally:

> *Hypothesis 4:* The more linkage with prominent ideas, the more a new idea will diffuse.

A good deal of historical work on science argues that new ideas begin ontologically uncertain (Chen and Song 2017; Kuhn 1970; Toulmin 1972). A new idea's initial practice may be specific, intermittent, and inconsistent, as its linkages within and relevance to the extant structure of knowledge are not yet fully explicated and normalized. It may only have novel, idiosyncratic, and shuffling associations, which make comprehending the new idea difficult, reducing the immediacy of its applicability and therefore its propensity to diffuse. As the new idea's relations with other established ideas become clearer and more consistent, elaborated, and recognizable, the new idea achieves greater conceptual coherence, definition, and continuity (Ramiro et al. 2018; Toulmin and Goodfield [1961] 1999:164; Wang, Schlobach, and Klein 2011).

Prior work develops the idea of research focus and problem consistency in relation to scholars' careers (Braddon-Mitchell 2005; Gieryn 1978; Heiberger, Muñoz-Najar Galvez, and McFarland 2021), but we adapt it to new ideas and their consistent use over time. In fact, historians of ideas argue that most established concepts have stable definitions, and that relations change far more slowly at their core than at their margins (Kenter et al. 2015:1192; Kuukkanen 2008). Together, new ideas can get used in a variety of ways after they are introduced, when authors draw connections and interrelate them to diverse sets of already established ideas. We expect that the consistent use of new ideas, whereby authors relate them to a stable (versus ever-changing) constellation of established ideas, anchors their relative meaning, gives them more semantic integrity and conceptual coherence, and thereby facilitates their later adaption. Stated formally:

> *Hypothesis 5:* The more consistent its links to ideas, the more a new idea will diffuse.

Another potential dimension of a new idea's fit or resonance is the degree to which creators embed it in the established ideational context into which it is introduced. When an element fits the cultural environment, it is more readily understood and adopted (Goldberg, Srivastava, et al. 2016; McDonnell, Bail, and Tavory 2017; Wuthnow 1989). From this line of reasoning, it is not a question of simply linking a new idea to a famous and foundational one (Hypothesis 4); instead, fit entails elaborating expansive interconnections within an established theoretical core. This occurs when new ideas enter and associate with other terms and ideas that have been related before, or when new ideas complement and integrate into the extant space of interrelated ideas (Becker 1982; Hallett et al. 2019).

In science, this arises when new ideas fit into, consolidate, and "fill the gaps" of extant research topics or thought styles (Fleck [1935] 1979; Foster et al. 2015). Prior work on citations (Shi et al. 2010) and categorization (Goldberg, Hannan, and Kovács 2016; Kovács and Hannan 2015) argues that placement within a research community and a clear position in one category tends to increase an offer's appeal (Hannan et al. 2019:4). Some work contends cultural fit has diminishing returns, much like that for social connectedness (Askin and Mauskapf 2017). So, situating new ideas in dense intellectual spaces can result in them making sense, but it can also crowd them out and make it difficult for a new idea to get recognized as distinctive (Denrell and Kovács 2020). Hence, one might expect a curvilinear relation of cultural fit with adoption. In short, no new idea is completely new and unrelated to established knowledge. Instead, new ideas varyingly evoke and are embedded in networks of interrelating established scientific ideas. We expect new ideas that are more deeply interconnected (embedded) within the network of established ideas are also more comprehensible and recognizable and, therefore, more likely to diffuse. Stated formally:

Hypothesis 6a: The more embedded it is within densely interconnected extant ideas, the more a new idea will diffuse.

Yet a third complementing perspective on cultural fit foregrounds the importance of new ideas bridging hitherto disparate concepts and topics—not so much fitting into extant knowledge structures, but instead spanning cultural holes within them (Goldberg, Srivastava, et al. 2016; Pachucki and Breiger 2010; Vilhena et al. 2014). When ideas span cultural holes, they bring into contact ideas of different thought styles, or nonredundant ideas. Ideas garner attention and take off when they solve a problem or interrelate hitherto independent lines of thought (Hallett et al. 2019). However, some work finds that in the realm of ideational networks, spanning too distal a cultural hole makes it difficult to comprehend and translate the idea for use, and this inhibits its adoption (Hofstra et al. 2020). As such, one might expect, again, a curvilinear relation of spanning cultural holes and ideational diffusion. Together, this literature suggests that when new ideas bridge extensive semantic spaces to distinct thought styles and paradigms, they pragmatically fill cultural holes and therefore become more useful and likely diffuse. Stated formally:

Hypothesis 6b: The more it bridges unrelated ideas, the more a new idea will diffuse.

Ideational Adaptation with Diffusion

The element perspective gives us the analytic precision to trace the manifest adoption of new ideas across texts. Yet, the bulk of element-focused research from which we deduce our expectations tends to study the adoption of ideas without much theoretical concern for how they are *adapted* as they diffuse (Callon 1986; Keuchenius et al. 2021; Latour 1987). As such, new ideas, as elements, are often depicted as immutable cultural memes or tropes passed from place to place (Dawkins 1982; Gruhl et al. 2004; Heath et al. 2001; Leskovec, Backstrom, and Kleinberg 2009; Lieberson 2000; Weeks and Galunic 2003). Remiss is how a new idea's propensity to diffuse is conditioned by its changing position within its ideational and social contexts while it diffuses (Kuhn, Perc, and Helbing 2014; McLean 2016). Such a focus requires extending the element perspective to consider the development of ideas over time and how distinct structural conditions of the scientific community and scientific knowledge jointly come into play.

Prior work on the history of ideas (Kuukkanen 2008; Toulmin 1972), ideational careers (Bonifati 2010; Rogers [1962] 2010), and problem change in science (Foster et al. 2015; Gieryn 1978) suggest ways that new ideas adapt as they diffuse. The management literature on ideas and innovations suggests they have careers with distinct phases and changes (Perry-Smith and Mannucci 2017; Rogers [1962] 2010).[1] For example, in their review, Perry-Smith and Mannucci (2017) conceptualize ideational careers as having four phases: an idea's journey begins with its creation (generation); then it is defined and established via practice, bringing it into relation with other ideas (elaboration); then it is fit with a local environment (championing); and eventually it is fit with wider environments (implementation).

The first two phases of idea generation and elaboration resemble historical accounts of how new ideas begin as uncertain and find consistent use (Hypothesis 5, ideational consistency). The third and fourth career phases suggest means by which established ideas can be related to core concepts (Hypothesis 4, ideational prominence) and topics of increasingly larger scientific domains (Hypothesis 6, ideational embeddedness) as they are consistently championed by key players (Hypothesis 1, social prominence; Hypothesis 2, social consistency) in their specific collaborative contexts (Hypothesis 3, social embeddedness). In these later stages, ideas appear to retain a core pattern of use but are fit into and consolidated with increasingly broader epistemic cultures (Fujimura 1992; Knorr-Cetina 1999; Star and Griesemer 1989). In science studies, one can imagine these being increasingly broader domains like subfields, disciplines, thought styles, and paradigms. Taken together, this literature clearly points to incremental development of new ideas. Rather than static material, they take on careers as they accrete relevant ties to extant knowledge, get extended to new domains, and get taken up and advocated by diverse actors.

Yet, while these narratives helpfully underscore the dynamism of new ideas, it remains that few ideas diffuse widely and become well known precisely because the success of their careers is contingent on the varied ways they may or may not resonate both socially and ideationally. Prior accounts of ideational careers also understate the inherent path dependency of new ideas. Whether and how social or ideational conditions matter more at different stages of a new idea's career has not been systematically established. Much less a story of constant and linear influence over time, we speculate that some conditions may be more salient to diffusion at different stages of ideational careers. For example, the ontological uncertainty of new ideas makes it likely there is greater randomness early in an idea's adoption that makes its later diffusion and long-run success less predictable (Kenter et al. 2015; Kuukkanen 2008). Perhaps only after a concept has coherent use do the conditions of social and cultural embeddedness apply for later diffusion.

We therefore conjecture that time matters, and interpretation of new ideas varies as they age and get adopted. This, in turn, fundamentally alters their career trajectories. Such a view escapes teleological overdetermination and suggests the fundamental changes an idea undergoes during its initial diffusion may be decisive in its long-run success and later diffusion. With diffusion comes ideational adaptation, which may recursively influence its later diffusion. We proceed in our analyses, then, testing the above expectations (Hypotheses 1 to 6) and controlling for the effects of time independently. We then relax the assumption of time held constant and empirically explore the variable careers of ideas by investigating whether and how their adaptation, which we define as their interrelations and positioning within their respective social and ideational ecologies, shapes their diffusion earlier and later in their careers.

STUDY

Empirical Setting: Identifying New Ideas in Science

The element view we advance here allows us to see words or terms within texts as expressions of discrete ideas or concepts,[2] which may be particular or general—idiosyncratic or widely recognized. We conceptualize authors' uses of these terms as speech acts: namely, terms represent their authors' intentional choices to articulate specific meaning in a specific context ("rhetic act"; see Austin 1975:93–95). And we take seriously the effects of these acts—particularly how names and terms, along with their meanings, not only statically describe the world but also dynamically interact with people who use them (Hacking 2006). Our chief aim in conceptualizing and studying ideas through the language that expresses them—their names or terms—is to understand why the ideas that some terms express get broadly diffused, capaciously adapted, and diffused further, while other terms (and the ideas they express) do not. Our aim is not to evaluate whether a given term appropriately or completely expresses a specific idea, or whether a given term even expresses an idea widely held and shared. Instead, our approach analytically treats terms as floating signifiers, whose initial use and meaning (their ideational content), whatever it is, may diffuse and morph or not (Lévi-Strauss 1987:63–64).

Thus, we conceptualize new ideas as new terms reflected in language, and their

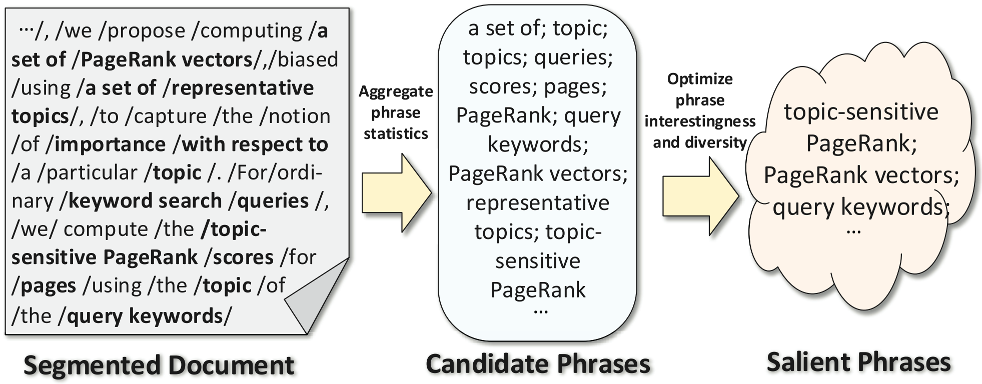

Figure 1. Concept Extraction Pipeline for Generating Concepts from Each Text Document

diffusion as captured by their manifest adoption across published texts. Such conceptualization of an idea fits science well, because scientific constructs and their hypothesized relations have long been related in language via words and phrases in published reports representing natural phenomena, methods, tools, tasks, and theories (Herfeld and Lisciandra 2019; Kuhn 1970; Toulmin 1972; Vilhena et al. 2014). These terms that scientists use to represent new ideas tend to be nouns and noun phrases, which refer to specific scientific content (Kuhn 1990). And these terms tend to be marshalled in summaries (e.g., titles and abstracts) of scientific studies and their respective novel contributions to the advancement of the field (Syed and Spruit 2017). Some terms are seldom or never used together, while others are heavily relied on, revealing where core concepts and conceptual relations exist. In this manner, an evolving structure of interrelated scientific ideas, new and old, and their roles can be represented as unfolding over time (Hill and Carley 1999). Based on this framing, we use "new idea" and "new term" interchangeably.[3]

The Web of Science (WoS) corpus from 1900 to 2016 is ideal to study the conditions that facilitate the diffusion of new ideas.[4] The corpus consists of more than 7 billion token words from 38,578,016 articles across journals in every scientific field, spanning STEM (science, technology, engineering, and mathematics), humanities and social science subjects (HSS), and their core specializations. As our primary data source, we analyze the titles and abstracts that summarize each article. We apply a data-driven phrase segmentation algorithm, AutoPhrase (Shang et al. 2018), to identify the most coherent phrases in these articles, thus operationalizing our construct of an idea (for details, see Figure 1 and Part S1 of the online supplement). We then post-process these ideas, removing cases that are clearly spurious and collapsing successive unigrams into identified ngrams ("social" "capital" = "social capital"). After post-processing, we identified 624,934 ideas.[5]

Our empirical focus remains on new ideas and their careers. However, our sample of 624,934 distinct ideas are pooled across time in the corpus, some of which emerged long ago and others more recently. Early papers (starting around 1900) identify many new terms, but this quickly decelerates over time and assumes a linear growth in vocabulary afterward (see Figure 2). The inflection point occurs around 1992, so we focus on the set of new terms arising from 1993 onward, which amounts to 56,540 new ideas (~9 percent of identified ideas).

It is possible a new idea existed before this date, and that we do not observe the true start of its career. We take several steps to make sure our results are robust to left-censoring issues. First, we use all the articles in the WoS, so that our idea histories globally

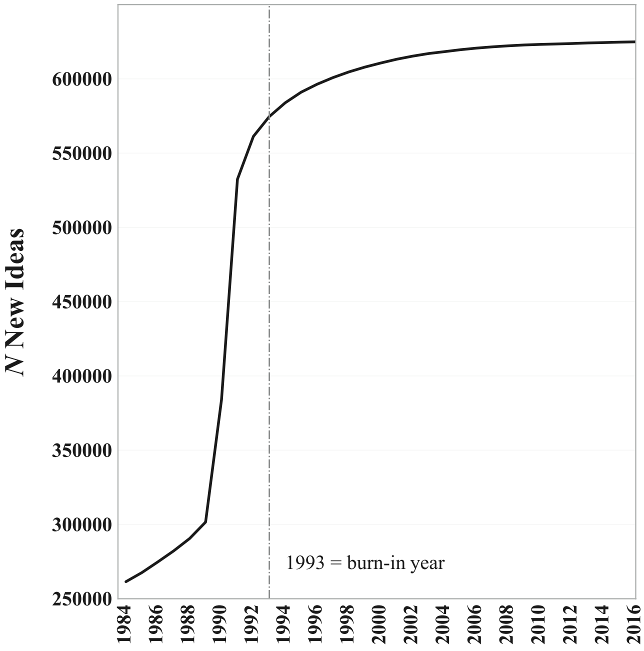

Figure 2. Number of New Ideas Introduced Each Year

reflect trends within the entirety of the WoS corpus. Second, we perform various robustness checks (further details are in Part S2 of the online supplement). In particular, we vary the burn-in year (e.g., to 1992, 1993, 1994, 1995, 1996) to see if our results are sensitive to the year at which we begin identifying new ideas. The results are stable and suggest we can reasonably assume our findings are robust to misclassifying some ideas as new when they were in fact old ideas. In addition, we randomly selected 100 popular ideas and found that their growth curves in the WoS looked similar to their growth curves in Google Ngram, a nonacademic corpus. This gives us confidence that our findings are not idiosyncratically representative of the WoS. Last, in our predictive models, we introduce additional control variables that account for potential sampling error (i.e., *journal abstract history*) and boundary issues (i.e., *WoS inward citation*) of the corpus (see the Control Variables section).

Another challenge is right censoring, which means we do not observe each idea's full life cycle. This is the case because WoS representativeness is skewed toward the recent period (post 1990), which means we have fewer observations of new ideas more recently introduced to the corpus. We address potential bias due to right censoring by controlling for age of the term and cohort effects (i.e., the year a concept was first observed). In addition, we used different burn-in years to examine different career windows, and the results were consistent, which strengthens our belief that our results are robust to right censoring. In general, the empirical focus of this article is mostly on the early development of ideas and their diffusion.

Table 1 presents examples of some of the most recognizable new ideas arising in different fields. Notably, most of the terms reflect constructs, concepts, procedures, and ontological entities specific to each field of research. In most cases these are unigrams, bigrams, and trigrams, but in some instances, they are phrases akin to propositions. These cases fit how philosophers and historians of science conceive of scientific terms and concepts (Kuhn 1990; Toulmin 1972). In addition, the set of new ideas we identify reflects

Table 1. Selected New Scientific Ideas (Year Introduced) by Field and Century

	Twentieth Century	Twenty-First Century
Agricultural sciences	***biofuel product** (1993), folin ciocalteu assay (1994), mega city (1995), citizen scienc (1997), dpph center dot (1999)	local ecolog knowledg (2000), disast resili (2001), googl earth (2005), water footprint (2006), fukushima nuclear accid (2011)
Biological and health sciences	prostat cancer pca (1992), fmri (1993), proteom (1995), ***gene ontolog** (2006), percutan coronari intervent pci (1998)	crispr (2002), next gener sequenc technolog (2007), whole exom sequenc (2009), afford care act (2010), vemurafenib (2011)
Education	student reflect (1993), english languag learner (1994), onlin discuss (1995), virtual learn (1997), concept inventori (1998)	onlin learn environ (2000), mobil learn (2001), ***blend learn** (2003), stem educ (2004), flip classroom (2011)
Engineering	nanofib (1993), biodiesel product (1994), lab on a chip (1996), bluetooth (1998), mobil ad hoc network (1999)	wireless sensor network (2000), ***smartphon** (2002), carbon footprint (2006), public cloud (2010), fukushima (2011)
Humanities	queer theori (1993), ***intim partner violenc** (1995), long eighteenth centuri (1997), harri potter (1998), sex traffick (1999)	gender mainstream (2000), anthropocen (2001), creativ economi (2002), youtub (2006), arab spring (2011)
Law	racial profil (1993), drug court (1994), eu law (1995), trip agreement (1996), intern crimin justic (1999)	ident theft (2000), dodd frank (2010)
Physical and mathematical sciences	nanowir (1993), singl wall carbon nanotub (1994), agent base model (1996), support vector machin (1997), ***dark energi** (1999)	metamateri (2000), dye sensit solar cell dssc (2001), click reaction (2004), topolog insul (2005), perovskit solar cell (2013)
Social and behavioral sciences	emot intellig (1993), ***transgend** (1994), autism spectrum disord asd (1995), social media (1997), eurozon (1999)	bridg social capit (2002), social network site (2004), hurrican katrina (2005), food sovereignti (2007), twitter data (2011)

Note: These new ideas were selected based on their success (*n* articles diffused into). Ideas in bold are plotted in Figure 4a. Education, law, humanities, and social and behavioral sciences are collapsed into one field in our predictive models due to their similar behaviors and smaller sample sizes in comparison with science and engineering fields.

relatively novel constructs that emerged in each field from 1993 to 2016, rather than constructs with long histories extending back 50 years or more. As evident in Table 1, our measurement procedure produces a list comprising mostly nouns and noun phrases that both narrowly and capaciously signify new ideas. For example, Hurricane Katrina was a discrete thing in one part of the world, yet it simultaneously represents many ideas

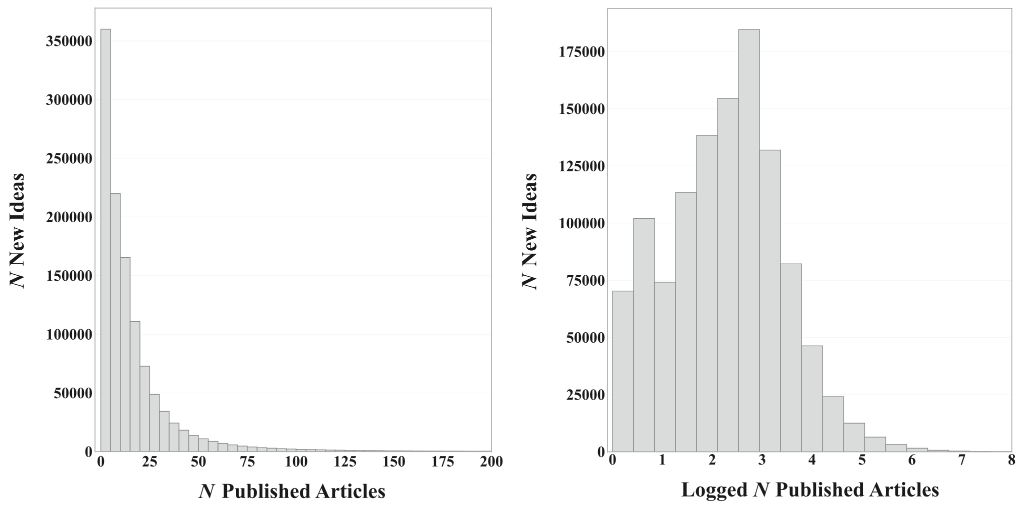

Figure 3. Distribution of New Ideas in Published Articles (Dependent Variable)

(natural disaster, effective governance, racial inequity) and thereby lives and grows in the scientific literature as an idea itself that takes on important meanings and associations contingent on its interrelations.

Outcome of Interest: The Diffusion of New Ideas

Our dependent variable is a measure of a new idea's diffusion. We measure a new idea's diffusion as the number of distinct articles in which the new idea appears each year after its first publication (*N published articles*). If the new idea occurs multiple times in a single article, we count it once. We construct our dependent variable as the number of articles a new idea diffuses into the year ahead (time $t + 1$). Given that new ideas arrive in different years and span different year ranges, these 56,540 new ideas sum to 995,945 observations at the idea-year level.

On average, once introduced, a new idea diffuses into 27 articles, with a standard deviation of 49 and a range from 0 to 4,703 articles. Diffusion is heavily right skewed, suggesting few new ideas go truly viral within science (see Figure 3). Figure 4a takes a random subsample of 100 new ideas introduced in 1993 and illustrates the changing number of published articles they reach over time. The bold line highlights the average trajectory of this subsample. For this subsample, a new scientific idea in 1993 diffuses to as many as 35 articles a year. Figure 4b shows the number of articles for an illustrative set of recognizable new ideas that come from multiple fields noted in Table 1. In both figures, the careers vary in their commencement, amplitude, and slope. Some new ideas, like "dark energy," gain popularity early, with a steep slope and higher uptake, whereas "smartphone" has its slope change at a later period. Over time, the concepts "smartphone" and "gene ontology" get taken up in 900 to 1,450 articles each year, whereas other concepts stay under 600 articles, like "blended learning." Even when we look at the new ideas with a maximum diffusion set of 600 articles, we still see some variation. For example, "dark energy" gains traction relatively early and grows at a stable speed, whereas "biofuel production" gets picked up later but grows at a faster speed and surpasses the growth of "dark energy" in the most recent periods. Such variation motivates our research question: Why do some new ideas have greater uptake than others, and why do they have such varied careers over time?

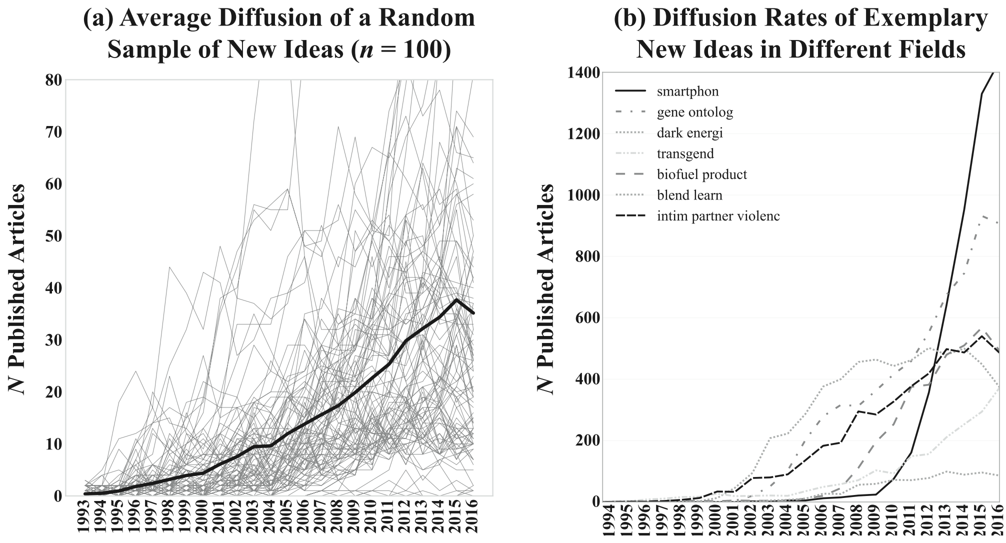

Figure 4. Ideational Careers

Focal Predictors

We use the articles' meta-data (author name, institution, subjects, year, citations) to construct indicators reflective of our hypotheses concerning diffusion mechanisms and conditions. These indicators are derived from a new idea's aggregated social networks of its authors who work on and elaborate it, and its ideational networks in which it is related to established scientific ideas. We measure each indicator in the current year (i.e., one year behind the dependent variable) and model them longitudinally (i.e., multilevel growth models) to test whether the social and ideational conditions of a new idea predict its later diffusion (see Tables 2 and 3).

Social prominence. We measure the social prominence of a new idea as the average weighted page rank of all its authors at time t, where the page rank, a measure of network centrality, is computed based on co-authorships in the previous 10 years and weighted by each author's number of publications based on the new idea at time t. We add the weight to better calibrate the measure to the authors who publish on the new idea most prolifically (Bonacich 1987; Freeman 1978).

Ideational prominence. We measure the ideational prominence of a new idea as the average weighted page rank of all the established ideas it is interlinked to at time t, where page rank is computed based on a new idea's interlinks among established ideas in articles published during the previous 10 years and weighted by each established idea's number of links to the new idea at time t. We add the weight to calibrate the measure to the extant ideas that are most frequently published in connection with the new idea.

Social consistency. We measure a new idea's social consistency as the cosine similarity between its authors' rate of publishing articles using the focal term at year $t - 1$ with their rate of publishing articles using the same focal term at year t. Only authors from $t - 1$ are the focal group, and new authors at time t are ignored. As such, it is a measure of past authors' consistent usage of the term. Should all the authors in $t - 1$ stop using the term in t, the cosine similarity is rendered as 0. Should there be no authors in $t - 1$ when there are some in t, then cosine similarity is again equal to 0. There is no author consistency score for the first year of a term's existence, as there is no prior history of author usage.

Table 2. Variable Descriptions

Variable Names	Variable Descriptions
Dependent Variable for Term Adoption	
N published articles	The number of unique published articles in Web of Science in which an idea is used in the future (time $t + 1$).
Key Independent Variables: Social and Ideational Conditions of Resonance	
Social prominence	For each focal term at time t, we measure the weighted average publication number of its related authors at time t, where weight is the number of times the author uses the term at time t. This captures the degree to which a focal term is used by highly productive authors (i.e., author page rank), and thus likely to be encountered in the social space.
Social consistency	For each focal term at time t, we focus on the authors in the prior year ($t - 1$) who used the term and then compare their rate of focal term usage (as number of term adoptions per author in $t - 1$) to rate of focal term usage in year t using cosine similarity. Should all the authors in $t - 1$ stop using the term in t, the cosine similarity is rendered as 0. Should there be no authors in $t - 1$ when there are some in t, then cosine similarity is again equal to 0.
Social embeddedness	For all authors associated with a focal term in year t, we estimate their density of collaboration with each other (number of observed ties divided by the total possible ties between them) in the prior 10 years of the WoS. We ignore papers with more than 15 authors. High values indicate a term is used by authors in an interconnected research community; low values indicate a term is used by unrelated and expansively located sets of authors.
Ideational prominence	For each focal term at time t, we measure the weighted average popularity of its neighbor terms at time t, where weight is the number of co-occurrences between them. This captures the degree to which a focal term is co-used with other highly used terms (i.e., term page rank), and thereby likely to be encountered in the semantic space.
Ideational consistency	For each focal term at time t, we focus on its neighbor terms co-used with the focal term in the prior year ($t - 1$), and then compare each neighbor terms' rate of co-usage with the focal term in year $t - 1$ to that observed in year t using cosine similarity. Should all the neighbor terms in $t - 1$ stop being co-used with the term in t, the cosine similarity is rendered as 0. Should there be no neighbor terms in $t - 1$ when there are some in t, then cosine similarity is again equal to 0.
Ideational embeddedness	For each focal term's neighbor terms at time t, we estimate their variation in semantic network positioning. We first take the cumulative 10-year semantic network of terms (valued ties by number of cooccurrence) and estimate network embeddings using word2vec (200 dimensions). We then take the neighbor terms associated with a focal term, and for all pairs of neighbors, we calculate their cosine similarity on these dimensional arrays. The average of this measures the degree to which a focal term is used with a set of neighbor terms with similar semantic placement (or conversely, used in a neighborhood composed of many distinctive neighbor terms, in a cultural hole).
Time Variables	
Age	How many years the term has been in usage since its first publication.
Age^2	The square of age (for polynomial growth in document term frequency).
Start year	The year a term appeared for the first time.

(continued)

Table 2. *(continued)*

Variable Names	Variable Descriptions
Control Variables	
Root term	The total number of other terms and phrases in the corpus that entail the focal term as part of their phrasing.
Interdisciplinary	This is computed as a term's average entropy across NRC discipline subject codes (e.g., sociology, math, economics). In a given year, "SVM" used in subject1 k1 times, subject2 k2 times, and subject3 k3 times. Therefore, the probability of "SVM" used in subject1 is p(s1) = k1 / (k1 + k2 + k3). Then subject diversity is calculated as -p(s1)log(p(s1))-p(s2)log(p(s2))-p(s3)log(p(s3)).
Physical sciences and math	Percent physical sciences and mathematics papers associated with a term.
Biological and health sciences	Percent biological and health sciences papers associated with a term.
Engineering	Percent engineering papers associated with a term.
Agricultural sciences	Percent agricultural sciences papers associated with a term.
Humanities and social sciences	Percent humanities and social sciences papers associated with a term.
Woman author	The percentage of authors associated with a term that have probable women's first names.
Journal impact factor	The average impact factor of journals using the focal term, across all instances of the term's use in year *t*.
Abstract length	The average word count or length of abstracts in which the term is embedded.
Abstract readability	The degree to which a focal term is situated in abstracts with accessible language. We measure this via a modified Dale-Chall Formula that uses Google Book Ngrams' most common 10,000 unigram list (http://norvig.com/ngrams/count_1w.txt). This is measured for each abstract in which a term is embedded and taken as an average across abstracts.
Abstract positivity	The degree to which a focal term is situated in abstracts with greater proportion of positive or negative terms. We use VADER (Valence Aware Dictionary and sEntiment Reasoner), a widely used lexicon and rule-based sentiment analysis tool that labels lexical features (e.g., words) according to their semantic orientation to get the compound score as a measure for emotional valence. Compound scores larger than .05 denote positive emotion.
Abstract negativity	If compound score is less than –.05, it denotes negative emotion.
Journal abstract history	We first compute the first year the journal starts to have abstracts in Web of Science and estimate the abstract history as the difference between year *t* and first year. Then for each focal term, we estimate the average abstract history across all the journals that the term appears in year *t*.
WoS inward citation	For all the papers using the term, we estimate the percentage of all their citations that are within the Web of Science.

Table 3. Summary Statistics

Variable Names	Mean	SD	Min.	Max.
Dependent Variable				
N published articles	21.063	49.603	0	4703
Key Independent Variables				
Social prominence (log)	.581	.344	0	1.642
Social embeddedness	.197	.261	0	1
Social consistency (log)	.181	.180	0	.693
Ideational prominence (log)	.924	.262	0	3.434
Ideational embeddedness	.320	.039	.133	.937
Ideational consistency (log)	.444	.171	0	.693
Time Variables				
Age	14.034	5.920	1	23
Age^2	231.989	159.767	1	529
Start year	1996.014	3.373	1993	2015
Control Variables				
Root term	.110	.341	0	4.788
Interdisciplinary	1.379	.807	0	4.799
Physical sciences and math	.191	.299	0	1
Biological and health sciences (baseline comparison for fields)	.566	.416	0	1
Engineering	.128	.252	0	1
Agricultural sciences	.055	.156	0	1
Humanities and social sciences	.060	.187	0	1
Woman author	.325	.198	0	1
Journal impact factor (log)	1.357	.493	0	4.016
Abstract length (log)	5.193	.418	.693	7.278
Abstract readability (log)	2.048	.190	.118	3.938
Abstract positivity	.645	.228	0	1
Abstract negativity	.451	.300	0	1
Journal abstract history	5.436	5.349	–23	91
WoS inward citation	.659	.226	0	1

Note: There are 995,945 observations across 56,540 new concepts. Some variables are log transformed due to their skewness.

Ideational consistency. To measure the ideational consistency of a new idea, we compute the cosine similarity between the number of articles the new idea shares with established ideas at $t - 1$ and then the frequency for that same set at time t. Should all the associated ideas in $t - 1$ stop their association in time t, the cosine similarity is rendered as 0. Should an idea not exist at time $t - 1$ when it does at time t, then the cosine similarity is 0. As with social consistency, there is no score for a term's first year of existence, as there is no history of associated terms prior.

Social embeddedness. We measure a new idea's social embeddedness as the number of observed collaborations occurring in the past 10 years (i.e., $t - 9$ to t) among all the authors publishing on the new idea at time t, divided by the total number of possible collaborations among these authors. High values indicate the new idea is more socially embedded within a collaborating scientific community; low values indicate the new idea is taken up, but by authors who are brokering or spanning distinct research communities.[6]

Ideational embeddedness. To measure an idea's ideational embeddedness, we construct a semantic network for the prior decade (i.e., $t - 9$ to t) for all ideas (their links valued as the number of published articles they share), which represents a broader context of scientific knowledge. Next, we encode the semantic network into a 200 latent dimension embedding space using word2vec (Church 2017), which gives us a vector representation summarily describing each established idea's position and interrelations within the knowledge context. Next, for our focal term, we identify its co-used terms or neighbors occurring in articles at time *t*, and we compute cosine similarities of these embedding dimensions on each neighbor-pair. Finally, we average these pairwise similarities, which results in a measure gauging the degree to which the focal term is situated among terms that have similar semantic placement within the network of established ideas (or conversely, used in a neighborhood of distinctive terms, or cultural hole).

An Idea's Birth and Age

As shown in Figures 3 and 4, new ideas arrive at different years and have variable careers. To account for this, we construct two measures: *start year* and *age*. *Start year* denotes the cohort effect, or the year in which the term appears in the WoS for the first time. *Age* denotes the age of a term. We compute this as the particular year in which a term is used minus the first year when the term arrived in the WoS. Because the diffusion of most new ideas follows a quadratic curve, we square age to best fit our model to the data (with varying intercept [or average] and slope [growth rate]). Both *start year* and *age* help address issues of right censoring, or the fact that some terms may not be observed over their full life. It also considers the fact that later terms enter a larger WoS corpus that has more papers than in the prior year. As such, we expect later cohorts and older terms to have greater document frequencies, and including these variables helps us compare terms of similar ages and cohorts.

Control Variables

We use a variety of control variables that account for alternative explanations of diffusion (see Tables 2 and 3). For example, the propensity of a new idea to diffuse may be related to its idiosyncratic qualities, its institutional contexts, the traits of its authors, or qualities of articles in which the new idea is embedded and read.

Prior work finds that terms with larger morphological family size (number of words with the same root) are easier to process and recall (Bertram, Baayen, and Schreuder 2000; Kuperman, Bertram, and Baayen 2008). These terms are often constitutive of other more specialized terms and phrases and therefore are more common (Searle 1998:122–4). This may be similar for nouns and noun phrases in our corpus of scholarly ideas. To capture this, we ascertain the extent to which a new idea is a *root term* by counting the number of compound nouns or noun phrases that use the same noun head or head noun compound. For example, "capital" is a root term that can be found in more specialized concepts of "social capital," "human capital," and so on. And "social capital" is also a root term, as it can be found in "bonding social capital" or "bridging social capital." We measure the rootedness of a new idea as the number of unique other ideas in the corpus it can be found within, and we include it to control for the possibility that new ideas that are root terms might find greater uptake than new ideas using specialized terms by virtue of their inherent linguistic portability (versus social or ideational ecologies).

Another possible explanation for new idea uptake has to do with the disciplines in which they are used. Some disciplines and fields are larger, have greater status, and have more developed conceptual languages than others (Evans et al. 2016), and this may affect a new idea's diffusion. Some new ideas may land between disciplines, leading them to find even wider diffusion. We control for this by locating new ideas in journals, and then identifying the broad field codes of those journals, reflecting *biological and health sciences* (baseline comparison in all models), *physical sciences and math*, *engineering*, *agricultural sciences*, and *humanities and social sciences*. To control for possible disciplinary and epistemic differences in diffusion patterns, we measure a new idea's adoption in these fields as the percentage of articles it appears in for a given year.[7]

When a new idea appears in multiple fields, it spans institutional holes and extends its reach to a larger pool of potential adopters, perhaps promoting its diffusion the next year. Various authors have argued that diversity metrics can capture interdisciplinarity (Leahey, Beckman, and Stanko 2017; Porter and Rafols 2009; Rafols and Meyer 2009; Stirling 2007). Similar to Leahey and colleagues, we measure the spread of an idea across discipline subject categories as *interdisciplinarity*, and it is computed as a new idea's average entropy across the National Research Council's (NRC) discipline subject codes (e.g., sociology, mathematics).[8]

In some prior work, author traits are regarded as important factors in the diffusion of new ideas. One salient trait would be authors' gender, which we measure via *women authors*, or the proportion of women out of all the authors using the new idea in each year. The idea here is that women scientists' work might be implicitly discounted and less likely to diffuse, what scholars call the "Matilda effect" (Rossiter 1993). Therefore, due to lower-status positions and empowerment, the higher the association with women authors (Bourdieu 1988; England et al. 1988; McDonald and Mair 2010), the lower a new idea's propensity to diffuse. Our technique for identifying women relies on names and machine learning methods. It was initially developed by Hofstra and colleagues (2017) and generalized to international gender based on *genderize api*.[9]

Another confounder in our understanding of diffusion may be the prestige of the journal in which an idea is published (McMahan and McFarland 2021; Teplitskiy et al. 2018). When a new idea is placed in a high-status venue, it will likely garner a larger readership, be attributed greater value, and find greater recognition, thereby facilitating its uptake more generally (Latour 1987). To capture this, we use *journal impact factor*, which measures the average impact factor of journals that publish the new idea across all instances of the term's use in year *t*.

Some articles may be long and dense and bury the new idea term. Because longer abstracts use more words, the new idea may be hard to identify, and long texts may obscure the effects of a new idea's association with emotion words, accessible language, and our metrics for ideational conditions on its propensity to diffuse. We estimate this via *abstract length*, which measures the average length of abstracts in which a term is situated at time *t*.

Abstracts vary in how accessible they are in terms of language difficulty or technicality. Prior work discusses how news articles vary in the accessibility of their language (Berger and Milkman 2012). The more the text has a vocabulary and grammar that is accessible to most people, the more viral it can become (Berger 2013). This same notion can be extended to ideas and the texts in which they are embedded. We modify a classic and robust frequency-based method, the Dale-Chall Formula (Chall and Dale 1995), to compute a score for *abstract readability*.[10] We replace the formula's original use of 3,000 words designed for 4th-grade U.S. students with the most common 10,000 unigram list from Google Book Ngram.[11] This widens the word set so it appeals to slightly more educated audiences and more topical areas reflective in the WoS.

Prior work has argued that products, authors, and their ideas disseminate broadly when they contain emotionality. When ideas (or any cultural object) are associated with emotionally charged language and interaction, they spark interest (Collins 1998). In addition, emotional language on social media is strongly associated with viral views of posts (Bail 2014, 2021; Bail, Brown, and Mann 2017). In science, the more an article uses emotion words, the more it may be recognized and shared, particularly because it reflects well on the identity of the potential sharer (Berger and Milkman 2012; Milkman and Berger 2014). We expect positive emotion may facilitate adoption (praise), and negative emotion may dampen it (critique) in scientific communities. We measure the emotionality of texts surrounding a new idea by studying the abstracts in which it is situated and identifying the extent they comprise more positive or negative expressions. We adopt the VADER neutral score and establish thresholds on the score to capture *abstract positive emotion* and *abstract negative emotion* (see Table 2). We use this measure due to its accuracy and accessibility (Hutto and Gilbert 2014).[12]

Finally, we include additional variables to control boundary effects in the Web of Science corpus: *journal abstract history* and *WoS inward citation*. It is possible that our identification of new ideas and their diffusion across articles is an artifact of some journals only recently using abstracts. As such, the new idea may be an established one that the WoS corpus misses when there is no abstract in which to observe it. This can be seen in history journals, which lacked abstracts for years and only recently instituted them. To control for this, we develop a measure of *journal abstract history*, which captures how much a new idea lands in journals that had abstracts in the past as the number of prior years each journal has had abstracts. Journals with, say, five years of prior abstracts are given a value of 5, and journals with five years of publications and no abstracts are given a value of –5. For each new idea, we then average these values across their journal instances.

Another potential error in observing the diffusion of new ideas may arise from the fact that they can be published in articles whose main referenced sources lie outside the WoS. When a new idea is published in an article heavily citing references outside the WoS, it may be more likely to diffuse in these external venues, which we do not measure. Conversely, when a new idea is published in an article that cites other articles published in journals contained in the WoS, it may be more likely to diffuse within the WoS, which we measure. We control for this partial observation of diffusion via a measure of *WoS inward citations*. For all the papers using the new term in year *t*, we estimate the percentage of all their citations that are located within the WoS.

Tables 2 and 3 show variable definitions and summary statistics, respectively. For some variables, we log transform them in our predictive models because they are skewed. Table 3 presents unstandardized sample statistics and log transformations where applicable. In the ensuing regressions, we report standardized coefficients to help readers interpret the magnitudes of reported results across measures with diverse scales and to better capture substantive significance, as the research design is sufficiently powered to detect even the smallest of relationships.

Statistical Approach

Our analyses have two inferential aims. First, we want to understand why some ideas diffuse more than others. For this, we model the diffusion of new ideas into published articles in the year ahead, all else equal, as a function of our main question predictors measured in the present year: social prominence, social embeddedness, social consistency, ideational prominence, ideational embeddedness, and ideational consistency. Second, we want to understand how the impact of these ideational and social conditions on a new idea's diffusion depends on the idea's career stage. For this aim, we estimate interactions between the new idea's age and our main predictors, asking, *How does the lifespan of a new idea depend on its evolving social and ideational contexts?*

As suggested by the curve in Figure 3, the average number of articles a new idea diffuses into (mean = 27) is smaller than its standard deviation (SD = 49), which means our outcome of interest is an over-dispersed count variable. Therefore, we use an over-dispersed Poisson regression model to model the data (Gardner, Mulvey, and Shaw 1995).[13] We observe an idea multiple times across the years of its life span. Therefore, we adopt a multilevel (panel or longitudinal) version of the over-dispersed Poisson model with random intercepts (Hox, Moerbeek, and van de Schoot 2017), where years are nested within ideas. A multilevel model leads to more accurate estimation of the standard errors when independent variables are analyzed at both levels, as it takes the clustered nature of the data into account (i.e., time points within terms are more alike than time points between terms). We fit the model using the *lme4* package in R (Bates et al. 2015) and report standardized coefficients to facilitate comparisons among the magnitudes of effect sizes.

Because negative binomial regression is also a common way to model over-dispersed count data (Ver Hoef and Boveng 2007), we explore other model specifications (e.g., log-linear and negative binomial models) as robustness checks. Those results are nearly identical to the Poisson model results and reinforce our findings further (see Part S2 of the online supplement). Finally, we check the multicollinearity of our full set of variables using the variation inflation index and find no collinearity aside from the obvious cases of age and age^2 (see Part S2 of the online supplement).

RESULTS

The results in Table 4 and Figure 5 answer the question as to why some ideas diffuse widely while others do not. To gauge how well this outcome is explained by our above expectations, we present a taxonomy of fitted models in Table 4, where Model 1 (Time) is nested in Model 2 (Time + Predictors), Model 3 (Time + Controls), and Model 4 (Full). Model 1, which includes a new idea's start year, age, and squared age explains around 51 percent of the variation in the number of articles that new ideas diffuse into. Model 2 builds off Model 1 but includes our six key predictors, improving model fitness by 24 percentage points. Model 3 also builds off Model 1 and includes our 14 control variables, improving the variance explained by 17 percentage points, significantly less than our key predictors. Model 4, the full model, fits the data best, explaining around 80 percent of the variance in the diffusion of new ideas into scientific articles. Shifts in AIC and deviance tell a similar story and suggest our ecological variables reflecting social and intellectual conditions are especially relevant.

Turning to the specific results, we begin with the intercept. Because we standardize our variables beforehand, the intercept describes, on average, the number of articles a new idea has diffused into once it has reached the average age of 14 years old, which is roughly 9 to 12 articles (i.e., $e^{2.24}$ to $e^{2.47}$ articles), depending on the model specification.

The time variables tell an expected story: for Models 1 and 4, Figure 5 shows the plotted incidence rate of how often an average new term is used in articles each year of their age. The plot shows that when new ideas are introduced into the WoS corpus in later years, they have a slightly higher rate of diffusion, most likely due to the expansion of the corpus (more articles are published) over time. Generally, new ideas gain adoption slowly at first and then find accelerated use (broader diffusion) over time (akin to Jin et al. 2019).[14] As we control for confounders and the main predictors, as expected, the bivariate relationship between diffusion and time is weakened (i.e., the difference between the two lines).

The Effects of Ideational and Social Ecology

Next, we focus on our question predictors, which reflect conditions of social and intellectual resonance. These results can be found in Models 2 and 4 of Table 4. As shown in Table 4, a one standard deviation change in social embeddedness is associated with a 15 percent ($e^{-.16}-1$) decrease in the predicted number of articles it diffuses into. The inverse of this variable suggests the reach and bridging of authors into unrelated social communities is associated with higher term uptake. Smaller but significant effects are found when new terms are associated with prominent or core authors (social prominence; b =.05). We see mostly negligible effects for social consistency (b = .02). Together, we interpret this as evidence consistent with two of our hypotheses: new ideas tend to diffuse more widely when they have socially prominent authors (Hypothesis 1) and span diverse research communities (Hypothesis 3b).

Notably, our internalist measures of a new idea's ideational ecology are more predictive of its diffusion than are social, externalist ones. A one standard deviation increase in ideational consistency of a new idea is associated with a 53 percent ($e^{.43}-1$) increase in the estimated number of articles it diffuses

Table 4. Multilevel Over-Dispersed Poisson Regression Explaining the Diffusion of New Ideas in *N* Published Articles (Standardized Coefficients Reported)

	(1)	(2)	(3)	(4)
Time Variables				
Age	1.234***	.663***	.550***	.579***
	(.002)	(.002)	(.002)	(.002)
Age^2	–.505***	–.156***	–.231***	–.155***
	(.002)	(.002)	(.002)	(.002)
Start year	.092***	.071***	.120***	.087***
	(.002)	(.002)	(.003)	(.002)
Key Independent Variables				
Social prominence (log)		.069***		.053***
		(.001)		(.000)
Social embeddedness		–.262***		–.158***
		(.001)		(.001)
Social consistency (log)		.019***		.021***
		(.001)		(.000)
Ideational prominence (log)		.135***		.136***
		(.001)		(.000)
Ideational embeddedness		.183***		.222***
		(.001)		(.001)
Ideational consistency (log)		.540***		.427***
		(.001)		(.001)
Control Variables				
Root term			.187***	.171***
			(.003)	(.002)
Interdisciplinary			.548***	.277***
			(.001)	(.001)
Physical sciences and math			–.018***	–.016***
			(.001)	(.001)
Engineering			–.026***	–.019***
			(.001)	(.001)
Agricultural sciences			–.023***	–.005***
			(.001)	(.001)
Humanities and social sciences			–.020***	–.013***
			(.001)	(.001)
Woman author			.017***	.007***
			(.001)	(.001)
Journal impact factor (log)			.081***	.065***
			(.001)	(.001)
Abstract length (log)			–.077***	.034***
			(.002)	(.002)
Abstract readability (log)			.108***	.082***
			(.002)	(.002)
Abstract positivity			.058***	.006***
			(.001)	(.001)
Abstract negativity			.047***	.003***
			(.000)	(.001)
Journal abstract history			–.009***	–.018***
			(.003)	(.002)
WoS inward citation			.142***	.077***
			(.001)	(.001)
Intercept	2.472***	2.270***	2.377***	2.244***
	(.003)	(.002)	(.003)	(.002)
Model Fitness				
AIC	10148244	8147865	8825262	7928868
Deviance	10148234	8147843	8825224	7928818
Pseudo R^2	.508	.752	.682	.797

Note: *N* observations = 995,945, *N* terms = 56,540. Standard errors are in parentheses. This model was run on the "loosely" identified version of terms. We had near identical results using a more strictly identified version of terms (see the online supplement).
$^*p < .05$; $^{**}p < .01$; $^{***}p < .001$ (two-tailed tests).

Figure 5. Plots of Incidence Rates by Age for Models 1 and 4 in Table 4

into. This is the average effect of increased ideational consistency, holding age and all other variables constant.[15] Less influential but important are the effects of ideational embeddedness and ideational prominence. In terms of the number of published articles on a new idea, a one standard deviation increase in its ideational embeddedness[16] is associated with a 25 percent ($e^{.22}-1$) increase, and a one standard deviation increase in its ideational prominence is associated with a 15 percent ($e^{.14}-1$) increase. Therefore, consistent with Hypotheses 4, 5, and 6a, a new idea is significantly likely to be adopted in more articles when it achieves intellectual resonance: when authors link it to a consistent constellation of already established ideas; connect it to specific, prominent ideas; and place it within recognized core research topics and paradigms. Together, this suggests the success of a new idea depends on how it relates to the intellectual character and content of established knowledge. On average, then, new terms find greater diffusion and adoption when they are consistently and cogently used in the semantic space, associated with other core ideas, and receive expansive use across communities of authors (Goldberg, Srivastava, et al. 2016).

Our control variables (Models 3 and 4) mostly behave as expected. Notably, new ideas expressed as root terms are more likely to find greater uptake (b = .17). New ideas published in interdisciplinary journals are also more likely to diffuse (b = .28). The effect of interdisciplinarity seems partly mediated by social embeddedness and suggests both variables are identifying the positive additive effects that spanning structural and institutional holes has on ideational diffusion. Fields have modest negative effects on term adoption when compared to the biomedical sciences (i.e., the largest field in the WoS corpus). New ideas advanced by women authors, net of other factors, show a very small positive return (b = .01). By contrast, we find more sizeable effects of a new idea's journal impact factor (b = .06). As expected, new ideas tend to diffuse more broadly after they land in more influential journals. Features of abstracts are modestly predictive of diffusion. Abstract readability, in particular, facilitates the diffusion of a new idea (b = .08). However, the length and emotionality of abstracts have weak effects. Abstract emotionality is likely weakly related because most scientific abstracts do not contain metaphorical and emotive language. Abstract length is initially negative (b = –.08) in Model 3, but then becomes modestly positive in Model 4, likely because the abstract length determines the number of neighbor terms that are used in ideational embeddedness and ideational consistency.

Finally, our controls for journal abstract history and WoS inward citation have the expected effects. When new terms land in journals with a history of abstracts, they have lower rates of adoption (b = –.02) than do terms in journals with missing abstracts years back. We see this as a sign our control captures some modest error in left censoring. We find a more sizable effect for WoS inward citation (b = .08). When terms land in articles that are inwardly facing, their rate of adoption is higher, most likely because their thought community and style are more represented in the corpus (Fleck [1935] 1979). Again, we interpret this as a sign our control variable captures some error due to boundary issues in the corpus (something all corpora likely have).

How the Adaptation of Ideas Affects Their Later Diffusion

The results in Table 5 and Figure 6 reflect interactions between our focal variables and the age of a new idea. These interactions answer the question as to whether the effects of social and ideational conditions on the diffusion of a new idea depend on its age. Results in Table 5 build off Model 4 in Table 4 and perform interactions one at a time. Interactions can be hard to interpret, so we plot them in Figure 6 in ways consistent with prior work on multilevel over-dispersed Poisson models (Keppel 1982; McFarland 2001).[17] In Figure 6, the x-axis represents age from 1 to 23 years. The y-axis represents the incidence rate ratio (i.e., the expected number of new articles an idea is taken up in each year).

The lines in Figure 6 illustrate how prominence, embeddedness, and consistency in social and ideational conditions influence term adoption with age, when each variable is at a high (mean + one standard deviation), average (mean), or low (mean – one standard deviation) level.[18] The average curve is mostly consistent across models, as that is the case where the focal condition (X) is the mean, or effectively zero when standardized. The coefficient of age and the model intercept slightly change due to different model specifications and which focal variable and age interactions are included, but it is generally stable.

Panels a, b, and c in Figure 6 (Models 1, 2, and 3 in Table 5) show how the social conditions of associated authors can alter a new idea's diffusion across its career. Of particular interest is social embeddedness. A new idea is increasingly adopted as it authors bridge and connect the term to distal communities (lowest social embeddedness) later in its career. A one standard deviation decrease in social embeddedness is likely to increase the diffusion of a new idea each year by one additional published article early in the term's life (years 1 to 7), but this grows to over 15 additional articles later (years 21 to 23) (see Figure 6b). In contrast, the effects for social prominence and social consistency are notably smaller and seem to matter most in terms of diffusion later in their careers.

Turning to panels d, e, and f in Figure 6 (Models 4, 5, and 6 in Table 5), we see how ideational conditions can alter a term's rate of adoption in documents. In general, the magnitude of effects for ideational conditions are consistently stronger than those for social conditions. This can be seen in the coefficients of Table 5 and in the greater difference between low and high lines for each ideational condition—they are generally wider at most term ages. Should a term be co-used with the same set of neighbor terms, a one standard deviation change in ideational consistency is likely to increase term adoption each year by one to three additional documents early in a term's life (1 to 7 years) and 15 additional documents later (years 21 to 23) (see Figure 6f).

New ideas also appear to benefit from being situated in a research topic or thought

Table 5. Interaction of Social and Ideational Conditions with Term Age (Standardized Coefficients Reported)

	Social Conditions			Ideational Conditions		
	(1)	(2)	(3)	(4)	(5)	(6)
Prominence	.044***			.137***		
	(.001)			(.000)		
Prominence × Age	–.055***			–.061***		
	(.002)			(.002)		
Prominence × Age²	.075***			.125***		
	(.002)			(.002)		
Embeddedness		–.383***			.232***	
		(.001)			(.001)	
Embeddedness × Age		.245***			–.090***	
		(.003)			(.002)	
Embeddedness × Age²		–.521***			.166***	
		(.003)			(.002)	
Consistency			.016***			.487***
			(.000)			(.001)
Consistency × Age			–.005**			–.029***
			(.002)			(.002)
Consistency × Age²			.057***			.195***
			(.002)			(.002)
Age	.613***	.867***	.597***	.567***	.337***	.806***
	(.002)	(.003)	(.002)	(.002)	(.003)	(.003)
Age²	–.200***	–.580***	–.202***	–.183***	.095***	–.501***
	(.002)	(.003)	(.002)	(.002)	(.003)	(.002)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
(Intercept)	2.241***	2.140***	2.237***	2.255***	2.304***	2.179***
	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)

Note: N observations = 995,945, N terms = 56,540. Standard errors are in parentheses. This model was run on the "loosely" identified version of terms.
*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests).

style (e.g., langue and parole, see Saussure 1916). A one standard deviation change in ideational embeddedness is likely to increase term adoption each year by one additional document early in an idea's life (1 to 7 years) and 10 additional documents later (years 21 to 23) (see Figure 6e). In our analyses, this appears when an idea is meaningfully embedded in a research tradition, topic, or discourse (Heiberger et al. 2021). Of lesser relevance is ideational prominence, but it still appears to matter across a term's first 23 years. Bringing an idea into relation with core concepts has later stage effects. As a new idea ages, a one standard deviation change in ideational prominence grows in effect, increasing term adoption an additional five documents per year in years 21 to 23 (see Figure 6d).

These effects are net of and additive with other factors. They reflect yearly returns and likely compound over the course of an idea's career. As such, what may seem like small returns snowball over time and are additive across the variables. For example, considering just the top four hypothesized determinants of diffusion—ideational prominence, ideational embeddedness, ideational consistency, and social embeddedness—a new idea is likely to diffuse into 45 more articles per year than typical. That is a twofold gain over

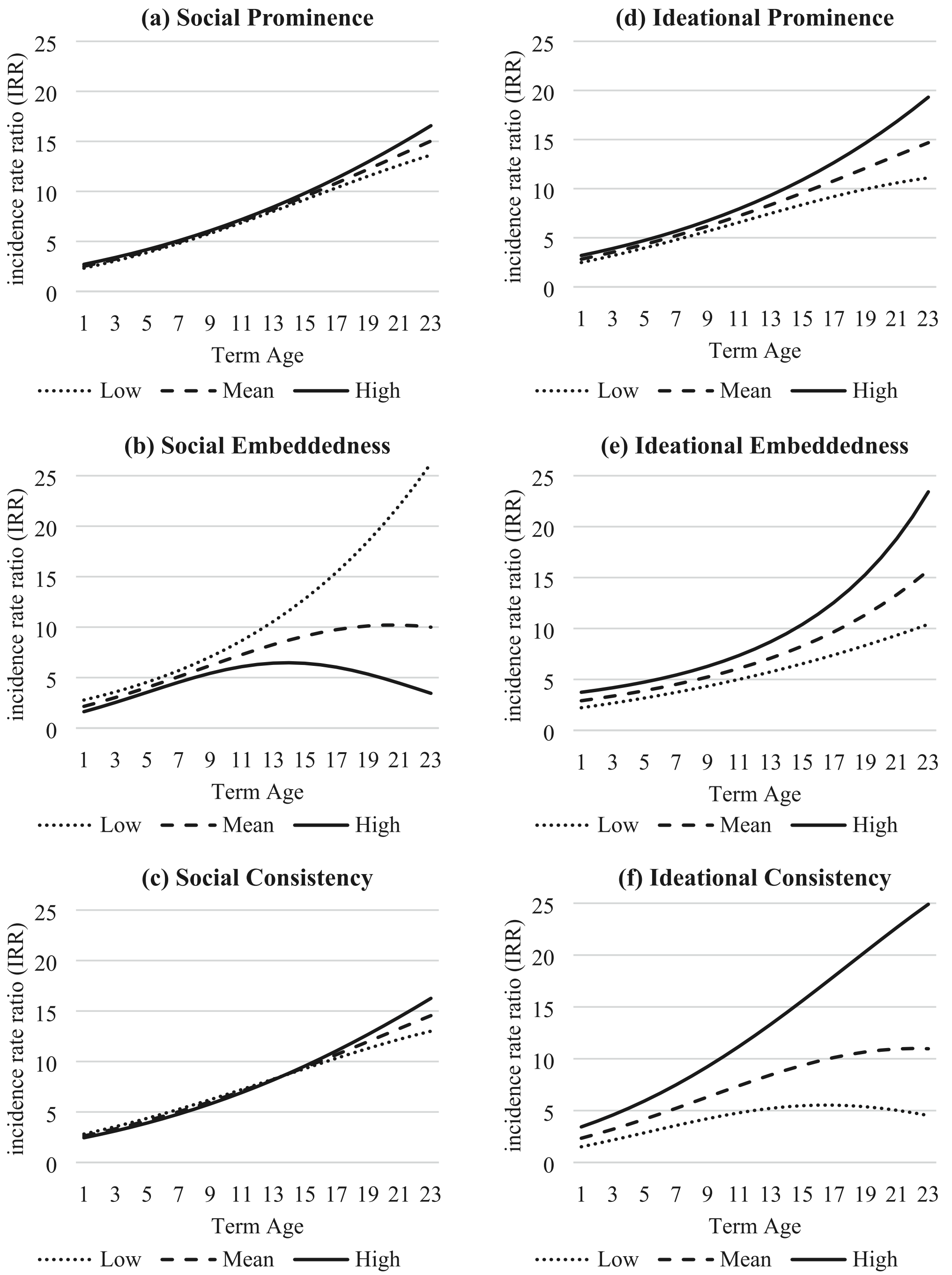

Figure 6. Interactions: Predicted Rate of Term Use by Age and Change in Focal Variable

the average rate of uptake (mean of number of articles is ≈ 21), and almost a one standard deviation gain overall (SD of *n* articles ≈ 50). If these effects occur yearly, we see means by which new vague terms can grow into core concepts and influential scholarly ideas.

Robustness Checks

The online supplement contains analyses designed to assess the robustness of our results. In Part S1, we document the procedure we used to validate the phrase quality extracted from our hybrid phrase segmentation algorithm AutoPhrase. In Part S2, we validate our results on another model specification (log-linear model and negative binomial model), different sets of terms (loose and strict versions of terms), and different burn-in years (1992 to 1996) to make sure our results are robust. We also explore other metrics for embeddedness and test whether our results hold when variables are lagged two and three years out. In all the analyses, our results remain very consistent, except for some slight changes in the significance and magnitude of some control variables' coefficients.

Limitations

Our research design choices have several limitations we wish to note. First, all the ideas are extracted from article titles and abstracts, not from the main text itself. This means we observe only those ideas that make it into the most visible and calculable parts of the scientific report. Like prior work (Hofstra et al. 2020), we argue that abstracts are a valid summary representation and best approach to getting document coverage (Schuemie et al. 2004; see Part S1 of the online supplement), but certainly future work could draw on representative samples of full texts. Second, we do not address all mechanisms of diffusion, including attitudes of entrepreneurs and further features of the new ideas themselves (Rogers [1962] 2010). Instead, we focus on six mechanisms stemming from both social and ideational ecologies of new ideas. While our focus is both novel and important because it draws on uniquely comprehensive contextual information and the ways it varies, later work could nonetheless test additional mechanisms. Third, we explore diffusion within only a single domain and corpus: we do not explore how an idea translates across domains or corpora. This was necessary, as different domains and genres of texts demand differentiated mechanisms respective to each of those different contexts. Different mechanisms likely amplify uptake across other domains and into the public sphere than what we describe here in academic journals. Scientific writing is probably less sensitive to the effect of emotion and accessible language than is writing driving the adoption of cultural memes, news, and commercial products.

DISCUSSION: THE CASE OF GENE ONTOLOGY'S SUCCESS

Noticeably, all the effects on the diffusion of new ideas are more pronounced as they age. What might explain this delayed effect? Lamont (1987) argues that the legitimation of theories results more from a complex environmental interplay than from the intrinsic qualities of the theories themselves. Based on this insight, one explanation may be that scientific ideas take time to find ideational and social resonance and diffuse in scholarly circles. Schudson (1989) defines "resonance" as the extent to which an object fits "with the life of the audience." McDonnell and colleagues (2017:3) conceptualize resonance as an experience that develops as people act to "puzzle out, or 'solve,' practical situations." All these notions of resonance suggest an idea needs elaboration and time for its fit to be recognized.

In this study's case, every new idea we investigate is already published. As such, these ideas are successful innovations. Yet this does not necessarily guarantee their broad diffusion, as our empirical analyses show. All start out with minimal uptake, yet only some achieve remarkable rates of diffusion.

The common trend for all new scientific ideas to start out tentatively can be understood by way of contrast with pop culture. Random new terms (neologisms) used in public blogs, emails, or texts (Cole, Ghafurian, and Reitter 2019) are not screened or vetted against collective standards before they are introduced, so they can achieve viral status quite quickly. In contrast, at the start, new scientific ideas are heavily selected for fit and suitability and are therefore less susceptible to faddish initial growth and decline. Once published, these same processes of initial evaluation apply to downstream uptakes of new scientific ideas. In addition to evaluation, academic research also moves slower in terms of production compared to pop culture, news cycles, and perhaps even product lines (Perry-Smith and Mannucci 2017). Indeed, this slow process of academic progress turns out to be decisive.

The broad-based diffusion of a new idea—what we loosely term as its success—depends on the scholarly process that ensues as later adopters work with the idea. To find an audience and outlet, ensuing researchers must read, interpret, and situate their own use-cases of new ideas against the body of extant knowledge. It is in the course of this intellectual labor that they set out to find and construct ideational resonance. To get past subsequent review, the new idea must be recognizable within the matrix of current and growing ideas and body of knowledge (langue) and may even need to be instantiated within a particular topic, research strand, or discourse (parole). Early on, with limited instances of recognized (published) applications, the new idea has a less established or even vague ideational precedent, which makes it difficult to resonate as useful or even recognizable. However, insofar as other scholars meaningfully and consistently adopt and adapt it, and the new idea steadily accrues stable relationships with other ideas and gets taken up by diverse researchers, it has greater facility in diffusing yet further. Indeed, as the idea is related to and integrated into constellations of core ideas, its usefulness, transferability, and influence increase as it enters further and more diverse fields. In this way, new ideas become successful as they grow into recognized concepts and ideas with wide appeal. It is this development that we interpret in the marked growth of the curves seen in Figure 6. New ideas are wildly successful and diffuse ever more broadly later in their careers only when, through the slow burn of the scholarly process, they have already achieved ideational and social resonance in the scientific community.

We anchor this discussion in empirical cases of new ideas. For example, "gene ontology" is a star concept in biological and health sciences that exhibits many of the qualities our statistical models identify. "Gene ontology" is a WoS term first used in 1999, and it diffuses to 907 published articles per year by 2016. "Gene ontology" refers to a vocabulary of hierarchically related biological terms used to characterize a gene product or the biochemical material that results from expressing a gene. "Gene ontology" describes gene products in terms of their molecular functions, cellular components, and biological processes, and with a unified cross-species vocabulary that makes discerning functions and making inferences about relations across gene products more feasible. Before the term "gene ontology" was invented, biologists had separate ontologies to characterize species and their gene products, and this made communication, synthesis, and cross-species inference about gene function difficult to determine.

The effort to establish a gene ontology began in 1998, when a group of researchers formed the Gene Ontology Consortium to develop gene ontologies for particular species of yeast, fruit flies, and mice (Ashburner et al. 2000). Publications in the WoS soon followed (Ashburner et al. 2000; Gene Ontology Consortium 2001; Jenssen et al. 2001; Lewis 2005; Shaw et al. 1999). Shortly thereafter, researchers explored gene ontology in many other organisms and the search was on to elaborate extant ontologies, to interrelate them and the genes/gene products in databases, and to facilitate the use and development of a more consistent gene ontology. Since then, gene ontology research has continued to expand and has been used to assign gene functions, identify similar gene functions, identify genes that co-express in development, and computationally model and verify genetic and metabolic networks. In short, it began with focused application in several species to grow into a larger framework and tool for identifying gene functions more generally.

We see evidence of gene ontology's development in our data. In the first years, the term is somewhat vaguely used—more of a promise—and over time it coheres into a widely resonant concept. At the beginning year of the concept (i.e., 1999), it has only seven neighbor concepts, such as genome database ("genom databas"), biological process ("biolog process"), and gene product ("gene product"). As gene ontology diffused into nearly 1,000 published articles in 2016, the number of other ideas it was interrelated with grew from seven to 19,935 unique neighbors. To illustrate how its ideational and social ecologies changed as it diffused (and how this had larger effects on its later diffusion), we plot the top 20 neighbors that co-occur with gene ontology most frequently in 2007 and in 2016. Panels a, b, and c in Figure 7 show that the connections among the focal concept and their neighboring community grow in number and become denser, which suggests gene ontology is also used in a focused research topic and thought style across all the years. This is reflected in the value of ideational embeddedness, which ranges from .28 to .35 and centers around the mean value of ideational embeddedness (i.e., .32 of the entire sample). We also see this in our metric of ideational consistency, which grows from .51 to .98. In later years, gene ontology is increasingly used in consistent ways with other neighbor concepts, such as genome database ("genom databas"), protein ("protein"), and transcriptome ("transcriptom"). At the same time, gene ontology is increasingly associated with other core concepts, such as protein, gene, and database over time, so ideational prominence grows from 1.0 to 3.4.

Gene ontology's research communities also grew since its introduction. In 1999, 11 authors collaborated on the first article studying "gene ontology"; by 2016, gene ontology's research community had grown to 6,154. We plot the top 50 authors adopting "gene ontology" most frequently in 2007 and in 2016. Panels d, e, and f in Figure 7 show that the relationship between the focal concept and the academic community became more spread out over time. We also see this in our metric of *social embeddedness*, which drops from 1.0 in 1999 to .0009 in 2016. This suggests that spanning more research communities over time helps an idea diffuse more broadly.

We also see some key authors who continue to work on gene ontology as a topic. For example, Michael Ashburner, a professor of genetics at the University of Cambridge, consistently published multiple works on this topic from 1999 to 2009. This is also reflected in our value of *social consistency*, which ranges from .00 to .74. Gene ontology is also associated with more productive authors over time, as *social prominence* grows from 1.0 to 1.47 (i.e., the median value of the entire sample is 1.0). Together, both the ideational and social conditions helped the growth of gene ontology, and this concept became increasingly key to many of the core concerns of biology. As these ideational and social conditions developed over the idea's career, its subsequent diffusion became more dramatic.

CONCLUSIONS

Every year, tens of thousands of scholarly articles are published reporting scientific innovations. These new ideas, carefully articulated and logically situated among established scientific facts and paradigms, span a dramatic number and array of diverse scholarly domains and modes of investigation. Yet, getting published is just the first stage of an idea's public career. Indeed, our starting contention motivating this study was that, despite their diversity and sheer numbers, most new ideas get little play after they make their debut in the scientific community. The vast majority are published only a few times and rarely incorporated as part of the progressive

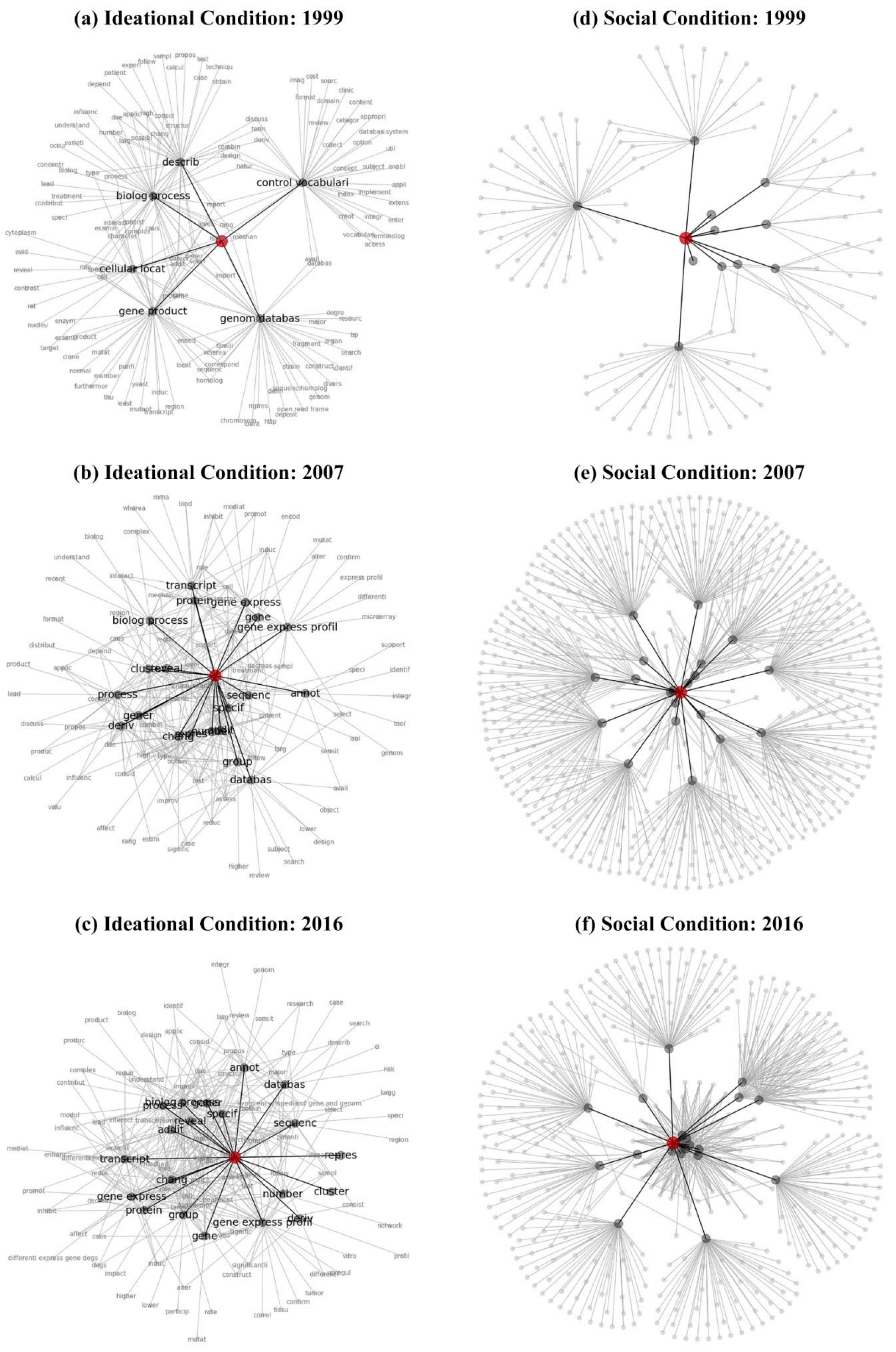

Figure 7. Social and Ideational Condition Change of Example Concept
Note: Red node = gene ontology.

advance of scientific knowledge. Some are incorporated, however, as we have shown with the case of gene ontology: they not only persist in our collective knowledge, but they fundamentally transform ways of conducting and thinking about new science.

In this study, we identified and followed the careers of nearly 60,000 new ideas over two decades to discover why some diffuse widely while most do not. We offered and tested a set of novel explanations—eclectically interweaving both internalist and externalist, ideational and social factors—that drew on established theories of diffusion in the sociology of science, management science and engineering, and the history of ideas. We found that the ways ideas become relationally situated in their ideational ecology is especially decisive for their long-run career. Ideas tend to diffuse more broadly when their authors link them to prominent, foundational ideas within science, consistently interrelate and position them within an established network of ideas, and thickly interrelate them—filling out research topics and networks with greater links. We also found that the successful diffusion of ideas depends on their social ecology, that is, the communities of scholars who take them up. Specifically, new ideas tend to diffuse when focally central researchers publish on them and when those authors span diverse research communities. Table 4 highlights this story.

Importantly, however, we also discovered that many of these ecological conditions tend to be even more salient at later stages of an idea's career (Table 5). For example, an idea's continued diffusion, even a decade after its initial introduction, is especially contingent on its persistently thick ties to research traditions (Figure 6e), its expansive spanning of diverse research communities (Figure 6b), and its consistent, coherent use (Figure 6f). In other words, the success of a new idea is not simply a story of monotonically linear diffusion as it ages; indeed, ideas that remain siloed in insular research communities (Figure 6b) and are incoherent and inconsistently used (Figure 6f) tend not to diffuse and peter out. Instead, the continued, long-run diffusion of new ideas—their resonant staying power in our collective understandings—depends on how they relate to their social and ideational worlds.

These findings express an emergent story of the development of new scientific and scholarly ideas. They begin in initial publications as terms with tentative meanings. Their ensuing widespread diffusion into new science depends on whether their meaning in relation to the body of knowledge gets worked out, systematized, elaborated, and broadly shared among diverse research communities. Insofar as these terms are elaborated, taken up, and championed by later researchers, they become more widely recognized as portable ideas that have resounding utility, much like a fact or machine that Latour (1987) describes. The idea begins to bridge institutional and structural holes to come into contact with different authors who translate and fit it to their needs (Perry-Smith and Mannucci 2017), perhaps becoming akin to a boundary object that is related to the same topic but selectively interpreted and distinctively repurposed for and by each community (Fujimura 1992; Star and Griesemer 1989). In this way, an initially narrow idea can grow into a concept with multiple recognized discrete senses and applications, as well as communities of practice, without undermining its meaningfulness.

An idea can accrete further interrelations within the network of scientific knowledge and thereby become a core concept as it becomes thickly related and relevant to other preexistent core concepts. In this manner, new ideas enter as focused objects with vague meaning in our intellectual and social communities, offered up as having specific, tenuous meanings and relations that need elaboration and recognition to take hold and diffuse broadly. Some ideas garner the intellectual and social ecological conditions that broaden and cohere their meaning, anchoring them in research traditions, but which get transported across communities and further elaborated, becoming increasingly complex, abstract ideas of core relevance, while others flitter and fail to resonate in scholarly texts and lives.

We were able to discover this emergent story because of our emphasis on both ideational and social ecologies. This is a real improvement on the extant literature, which tends to analytically preclude one to systematically investigate the other. By doing both simultaneously, we take seriously the social factors that condition the construction of new scientific facts while also foregrounding the conditions internal to the progressive change and development of science itself. Indeed, just as our theoretical eclecticism helped us discover a more complete picture, our analytic focus on ideas as elements of larger products—terms in discourses—powered a high-fidelity systematic empirical analysis that could observe the introduction of nearly 60,000 new ideas and follow their careers over two decades.

We can thus contribute an account of ideas as they are articulated in their initial introductions to the scientific community at large: as unique and discrete terms expressing newly advanced scientific meaning. Our contribution, in this regard, is to offer a uniquely exhaustive view of the structure of science, not simply a detail of a few touchstones, star ideas, or people (Stinchcombe 1982). Our story therefore offers more general insight about science than do previous findings focused on focal ideas, citations, or product gestalts. Finally, our methods, steeped in relational perspectives, enabled us to see historical contingency and path dependency in the evolving structure of science. Ideas are relationally situated to other ideas as well as to people. Ideas are not static—and neither are how people use them. Changes in these relations over time have big effects on later success. Our unique integration of social and ideational networks enabled us to discover the relational conditions that facilitate the diffusion of new ideas.

Acknowledgments

We are grateful for the feedback we received presenting this at ASA, University of Leiden, University of Minnesota, and Stanford's Computational Sociology Workshop. We are grateful to Jure Leskovec and the Stanford Infolab for allowing us to run some of our more taxing computations on their servers.

Author Contributions

D.A.M. designed research; M.C., D.A.M. performed research; M.C., D.A.M., S.S., X.R., H.C. contributed to measures and methods; M.C., D.A.M., S.S. analyzed data; and D.A.M., M.C., and D.S.S. wrote the paper.

Funding

This work has been supported by funds from NSF grants SMA-1829240 and 2022435.

ORCID iDs

Mengjie Cheng https://orcid.org/0000-0002-8890-7648
Daniel Scott Smith https://orcid.org/0000-0002-0979-4009
Hancheng Cao https://orcid.org/0000-0001-7231-1076
Sanne Smith https://orcid.org/0000-0002-1251-4442
Daniel A. McFarland https://orcid.org/0000-0002-6805-0798

Notes

1. Rogers ([1962] 2010) argues that the adoption of innovations progresses via a sequence of five subprocesses: (1) agenda-setting (problem definition and recognition); (2) matching (linking to other ideas and concerns); (3) redefining (to context and local community); (4) structuring (establishing the idea within the community); and (5) interconnection (relate to other communities).
2. "More widely: any word or phrase expressing a particular idea or concept, or denoting a particular object" (*Oxford English Dictionary* 2022: see "term, n.11b.").
3. Thus, our analyses center on words as expressions of ideas. These words are related to other words and authors over time reflecting their adaptation, which influences their later diffusion. In particular, our analyses focus on newly introduced words and follow their diffusion, but we consider them in context with extant ideas.
4. Although the entire WoS main corpus ranges from 1900 to 2017, 99 percent of papers with abstracts are from after 1985. We also add a variable *journal abstract history* to help control for how much a term lands in journals that have abstracts. This helps remove the boundary effects in the WoS corpus.
5. We develop two versions of our list of new ideas that move in two different directions of quality—

accuracy and recall—to ascertain the robustness of our results (i.e., higher accuracy contents with lower recall, we refer to as a "strict" version, and lower accuracy contents with higher recall, we refer to as a "loose" version). Part S1 in the online supplement explains how we determine these two sets of ideas. The main analyses presented here rely on the loose version, and the strict version is used for robustness checks in the online supplement.

6. It would have been ideal to use node2vec (Grover and Leskovec 2016) here for social embeddedness and to make the measure parallel to that of ideational embeddedness, but the sheer number of WoS authors was too large for the procedure to run even on advanced supercomputers. As such, we use density as a proxy.
7. We use the National Research Council's (NRC) broad field codes and map Web of Science journals' subject codes to NRC discipline codes. We use the NRC codes because they are determined by a large panel of leading disciplinary experts. We collapse law, education, and business into humanities and social sciences as the *humanities and social sciences* field.
8. To obtain entropy across disciplines, for each term's associated papers in year *t*, we calculate the frequency of different associated NRC codes. If a paper is associated with several subject codes, we assign weights to the subject codes accordingly.
9. https://rdrr.io/cran/genderizeR/man/genderizeAPI.html
10. The Dale-Chall Formula is widely used today (Yen and Wiseman 2019), but we explored additional readability measures (Gunning Fog Index, Automated Readability Index), and all relate similarly to the outcome variable.
11. http://norvig.com/ngrams/count_1w.txt
12. We also experimented with LIWC and Stanford CoreNLP to conduct sentiment analysis and compare our results. All three methods were consistent in results.
13. In an over-dispersed Poisson regression, the response variable is assumed to have a quasi-Poisson distribution with variance equaling the dispersion parameter multiplied by its mean. In our case, we assume the dispersion parameter to be 1.81, considering that is the ratio of the variance and mean in our dependent variable.
14. The functional form of the logged incident rates reported in Tables 4 and 5 suggests an inverted parabola, initially suggesting a later decline. Yet, when we exponentiate and restrict on the observed range of terms' ages, the slope is consistently positive and, in fact, increases (see Figure 5).
15. As the age function illustrates, the average term diffuses into a small number of articles early on but gains over time following a polynomial curve (see Figure 4a). Early in a term's career, then, a 53 percent gain corresponding to a one standard deviation increase in ideational consistency is small in terms of total additional articles using it (one to two additional documents). But later in the term's career, a 53 percent increase due to a one standard deviation increase in ideational consistency is more sizable (10 to 15 additional documents). The over-dispersed Poisson model in Table 4 shows the average rate of return across all ages. Later models will show the variable rate of return (Table 5).
16. Supplemental models find the polynomial form (squared) has the expected relation for ideational embeddedness. There is a diminishing return to diffusion for increases in ideational embeddedness. For social embeddedness, we likewise find a curvilinear relation, but the inverse. Lessened social density (or increased structural hole) has positive returns to diffusion, but this diminishes and reverses at more extreme values. We refrained from presenting polynomial results here because they greatly complicate the results for interactions with the polynomial for age. Interacting two polynomials with standardized values is difficult to follow and detracts from the main storyline, for which the linear relation is sufficient. Polynomial results are available upon request.
17. The difference between the predicted mean line (in all panels in Figure 6) and the observed value in Figure 4a is due to the cohort effect (i.e., the year in which the term appears in the WoS for the first time) we control for in the model. The panels in Figure 6 thus plot the within-cohort concept growth curve.
18. Each line for the conditions of resonance (X) are calculated as follows (b* = standardized coefficient). We standardized each variable beforehand, so the mean value of each variable X is zero:
(a) Mean X – 1 SD = exp [intercept + b1* (X = –1 SD) + b2* (Age varied) + b3* (Age^2 varied) + b4* (Age varied)(X = –1 SD) + b5* (Age^2 varied) (X = –1 SD)]
(b) Mean X = exp [intercept + b1* (X = 0) + b2* (Age varied) + b3* (Age^2 varied) + b4* (Age varied)(X = 0) + b5* (Age^2 varied)(X = 0)]
(c) Mean X + 1 SD = exp [intercept + b1* (X = +1 SD) + b2* (Age varied) + b3* (Age^2 varied) + b4* (Age varied)(X = +1 SD) + b5* (Age^2 varied) (X = + 1SD)].

References

Abbott, Andrew. 1999. *Department and Discipline*. Chicago: University of Chicago Press.

Adair, John G., and Neharika Vohra. 2003. "The Explosion of Knowledge, References, and Citations: Psychology's Unique Response to a Crisis." *American Psychologist* 58(1):15–23.

Antons, David, Amol M. Joshi, and Torsten Oliver Salge. 2019. "Content, Contribution, and Knowledge Consumption: Uncovering Hidden Topic Structure and Rhetorical Signals in Scientific Texts." *Journal of Management* 45(7):3035–76.

Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, et al. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25(1):25–29.

Askin, Noah, and Michael Mauskapf. 2017. "What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music." *American Sociological Review* 82(5):910–44.

Austin, John Langshaw. 1975. *How To Do Things with Words*, 2nd ed. Cambridge, MA: Harvard University Press.

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S. Graff Zivin. 2019. "Does Science Advance One Funeral at a Time?" *American Economic Review* 109(8):2889–920.

Bail, Christopher A. 2014. *Terrified: How Anti-Muslim Fringe Organizations became Mainstream*. Princeton, NJ: Princeton University Press.

Bail, Christopher A. 2021. *Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing*. Princeton, NJ: Princeton University Press.

Bail, Christopher A., Taylor W. Brown, and Marcus Mann. 2017. "Channeling Hearts and Minds: Advocacy Organizations, Cognitive-Emotional Currents, and Public Conversation." *American Sociological Review* 82(6):1188–213.

Bail, Christopher A., Taylor W. Brown, and Andreas Wimmer. 2019. "Prestige, Proximity, and Prejudice: How Google Search Terms Diffuse across the World." *American Journal of Sociology* 124(5):1496–548.

Barabási, Albert-László, and Réka Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286(October 15):509–12.

Barabási, Albert-László, Hawoong Jeong, Zoltan Néda, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek. 2002. "Evolution of the Social Network of Scientific Collaborations." *Physica A: Statistical Mechanics and its Applications* 311(3–4):590–614.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67(1):1–48.

Beath, Cynthia, Irma Becerra-Fernandez, Jeanne Ross, and James Short. 2012. "Finding Value in the Information Explosion." *MIT Sloan Management Review* 53(4):18–20.

Becker, Howard S. 1982. *Art Worlds*. Berkeley: University of California Press.

Berger, Jonah. 2013. *Contagious: Why Things Catch On*. New York: Simon & Schuster.

Berger, Jonah, and Katherine L. Milkman. 2012. "What Makes Online Content Viral?" *Journal of Marketing Research* 49(2):192–205.

Bertram, Raymond, R. Harald Baayen, and Robert Schreuder. 2000. "Effects of Family Size for Complex Words." *Journal of Memory and Language* 42(3):390–405.

Bloor, David. 1976. "The Strong Programme in the Sociology of Knowledge." Pp. 3–23 in *Knowledge and Social Imagery*, by D. Bloor. Chicago: University of Chicago Press.

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

Bonifati, Giovanni. 2010. "'More is Different,' Exaptation and Uncertainty: Three Foundational Concepts for a Complexity Theory of Innovation." *Economics of Innovation and New Technology* 19(8):743–60.

Bourdieu, Pierre. 1988. *Homo Academicus*. Stanford, CA: Stanford University Press.

Braddon-Mitchell, David. 2005. "Conceptual Stability and the Meaning of Natural Kind Terms." *Biology and Philosophy* 20(4):859–68.

Burt, Ronald S. 2004. "Structural Holes and Good Ideas." *American Journal of Sociology* 110(2):349–99.

Burt, Ronald S., and Giuseppe Soda. 2017. "Social Origins of Great Strategies." *Strategy Science* 2(4):226–33.

Callon, Michel. 1986. "The Scallops of St. Brieuc Bay: Some Elements of a Sociology of Translation." Pp. 196–223 in *Power, Action and Belief: A New Sociology of Knowledge?* edited by J. Law. London, UK: Routledge & Kegan Paul.

Callon, Michel, John Law, and Arie Rip. 1986. *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*. London, UK: Macmillan Press.

Carley, Kathleen M. 1997. "Extracting Team Mental Models through Textual Analysis." *Journal of Organizational Behavior* 18(S1):533–58.

Castells, Manuel. 1998. *The Information Age: Economy, Society and Culture*, Vol. I. Oxford, UK: Blackwell Publishers.

Centola, Damon. 2015. "The Social Origins of Networks and Diffusion." *American Journal of Sociology* 120(5):1295–338.

Chall, Jeanne Sternlicht, and Edgar Dale. 1995. *Readability Revisited: The New Dale-Chall Readability Formula*. Havertown, PA: Brookline Books.

Chen, Chaomei, and Min Song. 2017. *Representing Scientific Knowledge: The Role of Uncertainty*. New York: Springer.

Church, Kenneth Ward. 2017. "Word2Vec." *Natural Language Engineering* 23(1):155–62.

Cole, Jeremy, Moojan Ghafurian, and David Reitter. 2019. "Word Adoption in Online Communities." *IEEE Transactions on Computational Social Systems* 6(1):178–88.

Coleman, James, Elihu Katz, and Herbert Menzel. 1957. "The Diffusion of an Innovation among Physicians." *Sociometry* 20(4):253–70.

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

Crane, Diana. 1972. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. Chicago: University of Chicago Press.

Dawkins, Richard. 1982. "Replicators and Vehicles." Pp. 45–64 in *Current Problems in Sociobiology*, edited by King's College Sociobiology Group. Cambridge, UK: Cambridge University Press.

Denrell, Jerker, and Balázs Kovács. 2020. "The Ecology of Management Concepts." *Strategy Science* 5(4):293–310.

Dodson, Joe A., and Eitan Muller. 1978. "Models of New Product Diffusion through Advertising and Word-of-Mouth." *Management Science* 24(15):1568–78.

Doerfel, Marya L., and George A. Barnett. 1999. "A Semantic Network Analysis of the International Communication Association." *Human Communication Research* 25(4):589–603.

Doerfel, Marya L., and Stacey L. Connaughton. 2009. "Semantic Networks and Competition: Election Year Winners and Losers in US Televised Presidential Debates, 1960–2004." *Journal of the American Society for Information Science and Technology* 60(1):201–18.

England, Paula, George Farkas, Barbara Stanek Kilbourne, and Thomas Dou. 1988. "Explaining Occupational Sex Segregation and Wages: Findings from a Model with Fixed Effects." *American Sociological Review* 53(4):544–88.

Evans, Eliza D., Charles J. Gomez, and Daniel A. McFarland. 2016. "Measuring Paradigmaticness of Disciplines Using Text." *Sociological Science* 3(32):757–78.

Fiegel, Herbert. 1970. "The Orthodox View of Theories: Remarks in Defense as Well as in Criticism." Pp. 3–16 in *Analyses of Theories and Methods of Physics and Psychology*, edited by M. Radner and S. Winokur. Minneapolis: University of Minnesota Press.

Fleck, Ludwik. [1935] 1979. *Ludwik Fleck: Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press.

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

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

Freeman, Linton C. 1978. "Centrality in Social Networks Conceptual Clarification." *Social Networks* 1(3):215–39.

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

Fujimura, Joan H. 1992. "Crafting Science: Standardized Packages, Boundary Objects, and 'Translation.'" Pp. 168–211 in *Science as Practice and Culture*, edited by A. Pickering. Chicago: University of Chicago Press.

Fuller, Jack. 2010. *What Is Happening to News: The Information Explosion and the Crisis in Journalism*. Chicago: University of Chicago Press.

Galunic, D. Charles, and Simon Rodan. 1998. "Resource Recombinations in the Firm: Knowledge Structures and the Potential for Schumpeterian Innovation." *Strategic Management Journal* 19(12):1193–201.

Gardner, William, Edward P. Mulvey, and Esther C. Shaw. 1995. "Regression Analyses of Counts and Rates: Poisson, Overdispersed Poisson, and Negative Binomial Models." *Psychological Bulletin* 118(3):392–404.

Gene Ontology Consortium. 2001. "Creating the Gene Ontology Resource: Design and Implementation." *Genome Research* 11(8):1425–33.

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(13):3308–13.

Gieryn, Thomas F. 1978. "Problem Retention and Problem Change in Science." *Sociological Inquiry* 48(3–4):96–115.

Godart, Frédéric C., and Charles Galunic. 2019. "Explaining the Popularity of Cultural Elements: Networks, Culture, and the Structural Embeddedness of High Fashion Trends." *Organization Science* 30(1):151–68.

Godfrey-Smith, Peter. 2003. *Theory and Reality: An Introduction to the Philosophy of Science*. Chicago: University of Chicago Press.

Godfrey-Smith, Peter. 2010. "David Hull." *Biology & Philosophy* 25(5):749–53.

Goel, Sharad, Ashton Anderson, Jake Hofman, and Duncan J. Watts. 2016. "The Structural Virality of Online Diffusion." *Management Science* 62(1):180–96.

Goldberg, Amir, Michael T. Hannan, and Balázs Kovács. 2016. "What Does it Mean to Span Cultural Boundaries? Variety and Atypicality in Cultural Consumption." *American Sociological Review* 81(2):215–41.

Goldberg, Amir, Sameer B. Srivastava, V. Govind Manian, William Monroe, and Christopher Potts. 2016. "Fitting In or Standing Out? The Tradeoffs of Structural and Cultural Embeddedness." *American Sociological Review* 81(6):1190–222.

Granovetter, Mark. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology* 83(6):1420–43.

Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91(3):481–510.

Grover, Aditya, and Jure Leskovec. 2016. "node2vec: Scalable Feature Learning for Networks." Pp. 855–64 in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, August 13–17.

Gruhl, Daniel, Ramanathan Guha, David Liben-Nowell, and Andrew Tomkins. 2004. "Information Diffusion through Blogspace." Pp. 491–501 in *Proceedings of the 13th International Conference on World Wide Web*. New York: Association for Computing Machinery.

Guimera, Roger, Brian Uzzi, Jarrett Spiro, and Luis A. Nunes Amaral. 2005. "Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance." *Science* 308(5722):697–702.

Hacking, Ian. 2006. "Making Up People." *London Review of Books*, August 17 (https://www.lrb.co.uk/the-paper/v28/n16/ian-hacking/making-up-people).

Hallett, Tim, Orla Stapleton, and Michael Sauder. 2019. "Public Ideas: Their Varieties and Careers." *American Sociological Review* 84(3):545–76.

Hannan, Michael T., and John Freeman. 1989. *Organizational Ecology*. Cambridge, MA: Harvard University Press.

Hannan, Michael T., Gaël Le Mens, Greta Hsu, Balázs Kovács, Giacomo Negro, László Pólos, Elizabeth Pontikes, and Amanda J. Sharkey. 2019. *Concepts and Categories: Foundations for Sociological and Cultural Analysis*. New York: Columbia University Press.

Heath, Chip, Chris Bell, and Emily Sternberg. 2001. "Emotional Selection in Memes: The Case of Urban Legends." *Journal of Personality and Social Psychology* 81(6):1028–41.

Heath, Chip, and Dan Heath. 2007. *Made to Stick: Why Some Ideas Survive and Others Die*. New York: Random House.

Heiberger, Raphael H., Sebastian Muñoz-Najar Galvez, and Daniel A. McFarland. 2021. "Facets of Specialization and Its Relation to Career Success: An Analysis of U.S. Sociology, 1980 to 2015." *American Sociological Review* 86(6):1164–92.

Herfeld, Catherine, and Chiara Lisciandra. 2019. "Knowledge Transfer and Its Contexts." *Studies in History and Philosophy of Science* 77:1–10.

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

Hofstra, Bas, Rense Corten, Frank Van Tubergen, and Nicole B. Ellison. 2017. "Sources of Segregation in Social Networks: A Novel Approach Using Facebook." *American Sociological Review* 82(3):625–56.

Hofstra, Bas, Vivek V. Kulkarni, Sebastian Muñoz-Najar Galvez, Bryan He, Dan Jurafsky, and Daniel A. McFarland. 2020. "The Diversity-Innovation Paradox in Science." *Proceedings of the National Academy of Sciences* 117(17):9284–91.

Hox, Joop J., Mirjam Moerbeek, and Rens van de Schoot. 2017. *Multilevel Analysis: Techniques and Applications*. New York: Routledge.

Huth, Edward J. 1989. "The Information Explosion." *Bulletin of the New York Academy of Medicine* 65(6):670–72.

Hutto, Clayton, and Eric Gilbert. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the International AAAI Conference on Web and Social Media* 8(1):216–25.

Jenssen, Tor-Kristian, Astrid Lægreid, Jan Komorowski, and Eivind Hovig. 2001. "A Literature Network of Human Genes for High-Throughput Analysis of Gene Expression." *Nature Genetics* 28(1):21–28.

Jin, Ching, Chaoming Song, Johannes Bjelland, Geoffrey Canright, and Dashun Wang. 2019. "Emergence of Scaling in Complex Substitutive Systems." *Nature Human Behaviour* 3(8):837–46.

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

Kaufman, Jason. 2004. "Endogenous Explanation in the Sociology of Culture." *Annual Review of Sociology* 30:335–57.

Kenter, Tom, Melvin Wevers, Pim Huijnen, and Maarten De Rijke. 2015. "Ad Hoc Monitoring of Vocabulary Shifts over Time." Pp. 1191–200 in *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*.

Keppel, Geoffrey. 1982. *Design and Analysis: A Researcher's Handbook*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Keuchenius, Anna, Petter Törnberg, and Justus Uitermark. 2021. "Adoption and Adaptation: A Computational Case Study of the Spread of Granovetter's Weak Ties Hypothesis." *Social Networks* 66:10–25.

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

Kovács, Balazs, and Michael T. Hannan. 2015. "Conceptual Spaces and the Consequences of Category Spanning." *Sociological Science* 2(13):252–86.

Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*, 3rd ed. Chicago: University of Chicago Press.

Kuhn, Thomas S. 1990. "The Road since Structure." Pp. 3–13 in *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*.

Kuhn, Tobias, Matjaž Perc, and Dirk Helbing. 2014. "Inheritance Patterns in Citation Networks Reveal Scientific Memes." *Physical Review X* 4(4):041036 (https://doi.org/10.1103/PhysRevX.4.041036).

Kuperman, Victor, Raymond Bertram, and R. Harald Baayen. 2008. "Morphological Dynamics in Compound Processing." *Language and Cognitive Processes* 23(7–8):1089–132.

Kuukkanen, Jouni-Matti. 2008. "Making Sense of Conceptual Change." *History and Theory* 47(3):351–72.

Lakatos, Imre, and Paul Feyerabend. 2010. *For and Against Method*. Chicago: University of Chicago Press.

Lamont, Michele. 1987. "How to Become a Dominant French Philosopher: The Case of Jacques Derrida." *American Journal of Sociology* 93(3):584–622.

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

Leahey, Erin, Christine M. Beckman, and Taryn L. Stanko. 2017. "Prominent but Less Productive: The Impact of Interdisciplinarity on Scientists' Research." *Administrative Science Quarterly* 62(1):105–39.

Leahey, Erin, and Cindy L. Cain. 2013. "Straight from the Source: Accounting for Scientific Success." *Social Studies of Science* 43(6):927–51.

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

Leskovec, Jure, Lars Backstrom, and Jon Kleinberg. 2009. "Meme-Tracking and the Dynamics of the News Cycle." Pp. 497–506 in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Lévi-Strauss, Claude. 1987. *Introduction to the Work of Marcel Mauss*. Translated by Felicity Baker. London, UK: Routledge & Kegan Paul.

Lewis, Susanna E. 2005. "Gene Ontology: Looking Backwards and Forwards." *Genome Biology* 6(1):1–4.

Lieberson, Stanley. 2000. *A Matter of Taste: How Names, Fashions, and Culture Change*. New Haven, CT: Yale University Press.

MacRoberts, Michael H., and Barbara R. MacRoberts. 1989. "Problems of Citation Analysis: A Critical Review." *Journal of the American Society for Information Science* 40(5):342–49.

McCain, Katherine W. 2014. "Assessing Obliteration by Incorporation in a Full-Text Database: JSTOR, Economics, and the Concept of 'Bounded Rationality.'" *Scientometrics* 101(2):1445–59.

McDonald, Steve, and Christine A. Mair. 2010. "Social Capital across the Life Course: Age and Gendered Patterns of Network Resources." *Sociological Forum* 25(2):335–59.

McDonnell, Terence E., Christopher Bail, and Iddo Tavory. 2017. "A Theory of Resonance." *Sociological Theory* 35(1):1–14.

McFarland, Daniel A. 2001. "Student Resistance: How the Formal and Informal Organization of Classrooms Facilitate Everyday Forms of Student Defiance." *American Journal of Sociology* 107(3):612–78.

McLean, Paul. 2016. *Culture in Networks*. Cambridge, UK: Polity Press.

McMahan, Peter, and Daniel A. McFarland. 2021. "Creative Destruction: The Structural Consequences of Scientific Curation." *American Sociological Review* 86(2):341–76.

Merton, Robert K. 1965. *On the Shoulders of Giants: A Shandean Perspective*. Chicago: University of Chicago Press.

Merton, Robert K. 1968. *Social Theory and Social Structure*. New York: Simon & Schuster.

Merton, Robert K. 1988. "The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property." *Isis* 79(299):606–23.

Milkman, Katherine L., and Jonah Berger. 2014. "The Science of Sharing and the Sharing of Science." *Proceedings of the National Academy of Sciences* 111(Supplement 4):13642–49.

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

Moran, Peter. 2005. "Structural vs. Relational Embeddedness: Social Capital and Managerial Performance." *Strategic Management Journal* 26(12):1129–51.

Newman, Mark E. J. 2001. "The Structure of Scientific Collaboration Networks." *Proceedings of the National Academy of Sciences* 98(2):404–09.

Newman, Mark E. J. 2004. "Coauthorship Networks and Patterns of Scientific Collaboration." *Proceedings of the National Academy of Sciences* 101(Supplement 1):5200–05.

Newman, Mark E. J. 2009. "The First-Mover Advantage in Scientific Publication." *EPL (Europhysics Letters)* 86(6):68001 (https://doi.org/10.1209/0295-5075/86/68001).

Oxford English Dictionary. 2022. Oxford University Press (https://www.oed.com/).

Pachucki, Mark A., and Ronald L. Breiger. 2010. "Cultural Holes: Beyond Relationality in Social Networks and Culture." *Annual Review of Sociology* 36:205–24.

Perry-Smith, Jill E., and Pier Vittorio Mannucci. 2017. "From Creativity to Innovation: The Social Network Drivers of the Four Phases of the Idea Journey." *Academy of Management Review* 42(1):53–79.

Piazza, Alessandro, and Fabrizio Castellucci. 2014. "Status in Organization and Management Theory." *Journal of Management* 40(1):287–315.

Porter, Alan L., and Ismael Rafols. 2009. "Is Science Becoming More Interdisciplinary? Measuring and Mapping Six Research Fields over Time." *Scientometrics* 81(3):719–45.

Powell, Walter W., and Kaisa Snellman. 2004. "The Knowledge Economy." *Annual Review of Sociology* 30:199–220.

Prasad, M. N. V., Helena Freitas, Stefan Fraenzle, Simone Wuenschmann, and Bernd Markert. 2010. "Knowledge Explosion in Phytotechnologies for Environmental Solutions." *Environmental Pollution* 158(1):18–23.

Price, Derek de Solla. 1976. "A General Theory of Bibliometric and Other Cumulative Advantage Processes." *Journal of the American Society for Information Science* 27(5):292–306.

Quine, Willard V. 1951. "Two Dogmas of Empiricism." *Philosophical Review* 60(1):20–43.

Rafols, Ismael, and Martin Meyer. 2009. "Diversity and Network Coherence as Indicators of Interdisciplinarity: Case Studies in Bionanoscience." *Scientometrics* 82(2):263–87.

Ramiro, Christian, Mahesh Srinivasan, Barbara C. Malt, and Yang Xu. 2018. "Algorithms of Historical Word Sense Emergence." *Proceedings of the National Academy of Sciences* 115(10):2323–28.

Rodan, Simon, and Charles Galunic. 2004. "More Than Network Structure: How Knowledge Heterogeneity Influences Managerial Performance and Innovativeness." *Strategic Management Journal* 25(6):541–62.

Rogers, Everett M. [1962] 2010. *Diffusion of Innovations*. New York: Simon & Schuster.

Rogers, Everett M. 1976. "New Product Adoption and Diffusion." *Journal of Consumer Research* 2(4):290–301.

Rossiter, Margaret W. 1993. "The Matthew Matilda Effect in Science." *Social Studies of Science* 23(2):325–41.

Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science* 311(5762):854–56.

Schudson, Michael. 1989. "How Culture Works: Perspectives from Media Studies on the Efficacy of Symbols." *Theory and Society* 18(2):153–80.

Schuemie, Martijn, Marc Weeber, Bob J. A. Schijvenaars, Erik M. van Mulligen, C. Christiaan van der Eijk, Rob Jelier, Barend Mons, and Jan A. Kors. 2004. "Distribution of Information in Biomedical Abstracts and Full-Text Publications." *Bioinformatics* 20(16):2597–604.

Searle, John R. 1998. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, UK: Cambridge University Press.

Shang, Jingbo, Jialu Liu, Meng Jiang, Xiang Ren, Clare R. Voss, and Jiawei Han. 2018. "Automated Phrase Mining from Massive Text Corpora." *IEEE Transactions on Knowledge and Data Engineering* 30(10):1825–37.

Shaw, D. R., M. Ashbumer, J. A. Blake, R. M. Baldarelli, D. Botstein, A. P. Davis, J. M. Cherry, et al. 1999. "Gene Ontology: A Controlled Vocabulary to Describe the Function, Biological Process and Cellular Location of Gene Products in Genome Databases." *American Journal of Human Genetics* 65(4):A419.

Shi, Xiaolin, Jure Leskovec, and Daniel A. McFarland. 2010. "Citing for High Impact." Pp. 49–58 in *Proceedings of the 10th Annual Joint Conference on Digital Libraries*.

Shifman, Limor. 2013. *Memes in Digital Culture*. Cambridge, MA: MIT Press.

Small, Henry. 1978. "Cited Documents as Concept Symbols." *Social Studies of Science* 8(3):327–40.

Sorenson, Olav. 2014. "Status and Reputation: Synonyms or Separate Concepts?" *Strategic Organization* 12(1):62–69.

Sorenson, Olav, and Lee Fleming. 2004. "Science and the Diffusion of Knowledge." *Research Policy* 33(10):1615–34.

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

Stinchcombe, Arthur L. 1982. "Should Sociologists Forget Their Mothers and Fathers?" *American Sociologist* 17:2–11.

Stirling, Andy. 2007. "A General Framework for Analysing Diversity in Science, Technology and Society." *Journal of The Royal Society Interface* 4(15):707–19.

Syed, Shaheen, and Marco Spruit. 2017. "Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation." Pp. 165–74 in *Proceedings of the 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. Tokyo, Japan: IEEE.

Teplitskiy, Misha, Eamon Duede, Michael Menietti, and Karim Lakhani. 2018. "Why (Almost) Everything We Know about Citations is Wrong: Evidence from Authors." Pp. 1488–92 in *Proceedings of the 23rd International Conference on Science and Technology Indicators (STI 2018)*, September 12–14, Leiden, The Netherlands.

Teplitskiy, Misha, Eamon Duede, Michael Menietti, and Karim Lakhani. 2022. "How Status of Research Papers Affects the Way They are Read and Cited." *Research Policy* 51(4):104484 (https://doi.org/10.1016/j.respol.2022.104484).

Toulmin, Stephen E. 1972. *Human Understanding*, Vol. 1, *The Collective Use and Evolution of Concepts*. Princeton, NJ: Princeton University Press.

Toulmin, Stephen, and June Goodfield. [1961] 1999. *The Fabric of the Heavens: The Development of Astronomy and Dynamics*. Chicago: University of Chicago Press.

Uzzi, Brian, and James J. Gillespie. 2002. "Knowledge Spillover in Corporate Financing Networks: Embeddedness and the Firm's Debt Performance." *Strategic Management Journal* 23(7):595–618.

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

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

Ver Hoef, Jay M., and Peter L. Boveng. 2007. "Quasi-Poisson vs. Negative Binomial Regression: How Should We Model Overdispersed Count Data?" *Ecology* 88(11):2766–72.

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(15):221–38.

Wang, Shenghui, Stefan Schlobach, and Michel Klein. 2011. "Concept Drift and How to Identify It." *Web Semantics: Science, Services and Agents on the World Wide Web* 9(3):247–65.

Weeks, John, and Charles Galunic. 2003. "A Theory of the Cultural Evolution of the Firm: The Intra-organizational Ecology of Memes." *Organization Studies* 24(8):1309–52.

Wuthnow, Robert. 1989. *Meaning and Moral Order: Explorations in Cultural Analysis*. Berkeley: University of California Press.

Yen, Paul P., and Sam M. Wiseman. 2019. "Poor Readability of Online Patient Resources Regarding Sentinel Lymph Node Biopsy for Melanoma." *Cureus* 11(1):e3877 (https://doi.org/10.7759/cureus.3877).

Zuckerman, Harriet. 1987. "Citation Analysis and the Complex Problem of Intellectual Influence." *Scientometrics* 12(5–6):329–38.

Mengjie Cheng is currently a PhD student in Business Administration at Harvard Business School. Her research focuses on innovation management, creator economy, and computational social science. She received her BS in Finance from Chu Kochen Honors College at Zhejiang University and MS in Management Science and Engineering from Stanford University.

Daniel Scott Smith is a postdoctoral scholar at Stanford University studying the historical and cultural dimensions of science and nation-state formation using computational techniques. He holds a PhD from Stanford University.

Xiang Ren is an Assistant Professor of computer science at the University of Southern California. He completed his PhD at the University of Illinois Urbana-Champaign. His current work focuses on topics in human-center natural language processing and commonsense reasoning.

Hancheng Cao is currently a PhD candidate in computer science at Stanford University with a minor in management science & engineering; he is also a Stanford Interdisciplinary Graduate Fellow. He works on computational social science, data science, and human computer interaction. He received his bachelor's degree in electronic engineering from Tsinghua University.

Sanne Smith is a Lecturer and Master's Program Director at Stanford University. She teaches methods, statistics, and data science courses at the Graduate School of Education. Her PhD research about adolescent friendship networks was completed at Utrecht University.

Daniel A. McFarland is a Professor of Education, and by courtesy, Sociology and Business, at Stanford University. He completed his PhD at the University of Chicago. His current work focuses on topics in the sociology and philosophy of science, as well as relational sociology.