Research Policy 43 (2014) 927–940

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Do innovation measures actually measure innovation? Obliteration, symbolic adoption, and other finicky challenges in tracking innovation diffusion

Andrew Nelson[a,*], Andrew Earle[b], Jennifer Howard-Grenville[a], Julie Haack[c], Doug Young[d]

[a] Lundquist College of Business, University of Oregon, Eugene, OR 97403, United States
[b] Peter T. Paul College of Business and Economics, University of New Hampshire, Durham, NH 03824, United States
[c] Department of Chemistry, University of Oregon, Eugene, OR 97403, United States
[d] Department of Chemistry, Lane Community College, Eugene, OR 97403, United States

ARTICLE INFO

Article history:
Received 14 May 2013
Received in revised form 5 November 2013
Accepted 18 January 2014
Available online 8 February 2014

Keywords:
Innovation
Diffusion
Measurement
Labeling
Green chemistry

ABSTRACT

Although innovation diffusion is a central topic in policy and strategy, its measurement remains difficult – particularly in cases where the innovation is a complex and possibly ambiguous practice. In this paper, we develop four theoretical mechanisms that may bias diffusion markers by leading to the understatement and/or overstatement of diffusion at different points in time. Employing the case of "green chemistry," we then compare three different diffusion markers – keywords, database index terms, and domain expert assessments – and we demonstrate how they lead to differing conclusions about the magnitude and timing of diffusion, organizational demography, publication outlets, and collaboration. We also provide suggestive evidence of extensive "greenwashing" by particular organization types and in particular countries. Building on these findings, we point to potential challenges with existing diffusion studies, and we make a case for the incorporation of practitioners in construct measurement and for the integration of comparative metrics in diffusion studies.

1. Introduction

Innovation diffusion is a central topic in public policy and strategic management (e.g., Abrahamson and Rosenkopf, 1993; David, 1986; Geroski, 2000; Rosenberg, 1986; Stoneman and Diederen, 1994). At their core, diffusion studies seek to explain "how things. . .get from here to there" (Katz, 1999: 145). Essential to such explanations are accurate data about whether and when a "thing" actually has moved from "here" to "there." Unfortunately, common measurement methods can both over- and under-state diffusion, and these errors can vary in prominence and magnitude over time.

Researchers occasionally benefit from good data on innovation diffusion, as when sales records, census data, or other markers exist to trace a given product or practice. In many cases, however, well-formed, reliable and complete diffusion records are not available. In particular, novel strategies, practices and other innovations that are not tied directly to an artifact can be especially difficult to measure. Moreover, such innovations may be more likely to be ambiguous in their labeling and are especially prone to adoption fads (e.g., David and Strang, 2006; Fiss and Zajac, 2004; Hendricks and Singhal, 1997; Wang and Bansal, 2012).

Measurement is further challenged by the fact that participants in diffusion processes can manipulate, intentionally or not, diffusion indicators. For example, organizations sometimes claim to adopt practices that they have not in fact adopted, believing that adoption confers legitimacy or other benefits (Carpenter and Feroz, 2001; Elsbach and Sutton, 1992; Fiss, 2008; Fiss and Zajac, 2006; Oliver, 1991; Westphal and Zajac, 1994). Alternatively, adopters sometimes fail to report their adoption because they fear that it might signal *illegitimacy* (Colyvas and Jonsson, 2011; Granqvist et al., 2013; Terlaak and Gong, 2008) or they presume that a practice is so widespread that it no longer warrants specific mention (Lederberg, 1977; Merton, 1968). For all of these reasons, scholars struggle to obtain reliable diffusion data. In turn, data limitations can hamper efforts to build theory around innovation diffusion itself.

In this paper, we contribute to a growing literature that addresses the correspondence between measures and the activities or phenomena that these measures are attempting to capture

* Corresponding author. Tel.: +1 541 346 1569.
E-mail address: ajnelson@uoregon.edu (A. Nelson).

http://dx.doi.org/10.1016/j.respol.2014.01.010
0048-7333/

(e.g., Aguinis and Edwards, 2014; Boyd et al., 2013; Chatterji et al., 2009; Colyvas and Powell, 2009; Freeman and Soete, 2009). Specifically, we develop four mechanisms that can account for divergence between "actual" diffusion patterns and patterns derived from two commonly employed measures: keywords and index terms. We then quantitatively examine the diffusion of "green chemistry" by comparing results obtained from keywords and index terms against those obtained from careful assessments by domain experts.

Our results point to considerable differences between measurement approaches, especially in the early years of the field. Specifically, we detail how different measures reflect different magnitudes of activity, publication outlets, organizational demography, and collaboration patterns. We also show how certain organization types, countries, and author-team compositions may be particularly subject to measurement errors. To conclude, we describe circumstances when the use of domain expert assessments might be particularly valuable to researchers; we discuss why diffusion studies must be cautious to interpret their results in light of the particular measure used; we make the case for exploiting differences between measures; and we offer several policy implications and recommendations.

2. Measurement of innovation diffusion and associated challenges

Researchers have employed a number of approaches to measure innovation diffusion (see Rogers, 2003, for a useful summary). One of the most common methods is to examine patent and/or publication citations (e.g., Cockburn and Henderson, 1998; Jaffe, 1989; Nelson, 2009). Citation-based analyses face a fundamental problem, however, in that many innovations do not have a clearly defined originating patent or publication from which citations can be traced. Moreover, even if an initial record exists, subsequent adopters may not leave a paper trail to indicate the innovation's diffusion. For example, Total Quality Management (TQM) is a highly influential and widespread management practice. Constructing a TQM paper trail is difficult, however, since many adopters may not cite the original developers of the practice (Westphal et al., 1997), others may differ in what they label TQM (Giroux, 2006; Zbaracki, 1998), and still others may not produce any public records of their adoption (David and Strang, 2006).

In light of these challenges, many researchers have employed other methods of measurement. Two of the most common alternatives are to search for traces of an innovation by its "label," using either keywords or index terms. In a keyword search, a researcher sorts through a set of documents to find instances of a label(s) associated with an innovation. Previous studies have conducted keyword searches of patent databases (e.g., Beaudry and Allaoui, 2012; Hu and Jaffe, 2003; Jamasb and Pollitt, 2011; Lee et al., 2011; Xie and Miyazaki, 2013), other databases (e.g., Gay and Dousset, 2005; Hendricks and Singhal, 1997; Mogoutov and Kahane, 2007), annual reports (e.g., Fiss and Zajac, 2004), websites (e.g., Wang and Bansal, 2012), and company directories (e.g., Zaheer and Mosakowski, 1997).

In an index-term search, a researcher relies on another party's identification of specific codes, as with JEL codes assigned to articles that appear in *Research Policy*. As with keyword searches, the researcher still consults a corpus of documents such as a publication database (e.g., Abrahamson and Fairchild, 1999) or an industry directory (e.g., David and Strang, 2006; Greve, 1995, 1996; Jonsson, 2009). In an index-term search, however, the search terms themselves are defined in advance rather than open-ended. Nevertheless, the search strategies underlying keywords and index-terms are very similar, with diffusion researchers using search "hits" on the label associated with an innovation in order to construct a list of adopters and to analyze adoption patterns.

Of course, an outstanding question with any approach to diffusion measurement concerns the correspondence between the data obtained from a particular measurement approach and the "actual" diffusion pattern that a researcher hopes to investigate and explain. In turn, there are a number of mechanisms that may be at work, individually or collectively, to distort the correspondence between actual patterns and those patterns derived from keywords or index terms.

First, in some cases, a label for an innovation might not yet exist. For example, researchers studying nanotechnology have traced the term "nanotechnology" to a 1986 publication by Eric Drexler, but they are quick to note that much research that qualifies as nanotechnology took place prior to 1986 (Granqvist et al., 2013). Scholars in the social construction of technology emphasize that actors contest boundaries and definitions around innovations, especially in the early years of emergence (Kaplan and Radin, 2011). In turn, reliance upon existing labels to assess diffusion, especially in an early period, may result in "false negatives" or missed data. We refer to this early lack of a coherent label as "pre-labeled emergence."

Second, actors may intentionally avoid using a label, even if it accurately describes their adoption – a practice that we call "strategic avoidance." One motivation for strategic avoidance lies in concerns over illegitimacy or negative perceptions by others. As Colyvas and Jonsson (2011) point out, diffusion and legitimacy are distinct concepts; thus, some things can diffuse widely without providing legitimacy. "Adult shops," for example, may be common in many cities, yet many of the people associated with such shops might obscure or disclaim their involvement with them. Granqvist et al. (2013) cite examples of executives at nanotechnology firms who did not use the term "nano" because they feared that it would signal a lack of commercially viable products (see also Hudson and Okhuysen, 2009; Zuckerman, 1999). For innovations that will eventually become legitimate, Terlaak and Gong (2008) propose that strategic avoidance (or "concealed adoption," as they label it) will be most prevalent in earlier diffusion stages when an innovation's value is unclear, as are external reactions to it.

Competitive concerns also can lead to strategic avoidance. In their study of patenting in the electricity sector, Jamasb and Pollitt (2011: 315) note, "Patentees might hence behave strategically and use [or avoid] particular words in the titles/abstracts in order to reduce the probability that others find their patents and reproduce their invention and market [it] in another country where the invention is not protected." Whatever the motivation, strategic avoidance produces a measurement problem of false negatives or missed data since searches for an innovation by its label will not identify all adopters.

On the other hand, "symbolic adoption," a third mechanism, may result in false positives, or the *overstating* of adoption. Symbolic adoption expands the classic notion of decoupling, developed most thoroughly by institutional theorists who note separation between formal structures and actual practice (e.g., Scott, 2000; Powell and DiMaggio, 1991; Bromley and Powell, 2012). Often, the motivation for symbolic adoption lies in an individual's or organization's desire to obtain legitimacy or some other benefit derived from conformity to a practice, without incurring the cost that may be incurred by actually engaging in the practice. Thus, as Elsbach and Sutton (1992), Oliver (1991), Fiss (2008) and others note, individuals and organizations can take "calculating, manipulative, or even deceptive actions" (Fiss, 2008: 397) to show apparent conformity even as the reality is non-conformity. For example, Westphal and Zajac (1994) found that many firms adopted long-term incentive plans for management even though they never actually implemented these plans. Similarly, Carpenter and Feroz (2001) found that state

Table 1
Summary of mechanisms.

	Pre-labeled emergence	Strategic avoidance	Symbolic adoption	Obliteration by incorporation
Description	A coherent label does not yet exist	Practitioners avoid association with a label	Individuals or organizations claim the label, but do not engage in the activity	Individuals or organizations are not explicit in their labeling since they believe it is taken for granted
Measurement problem for researchers	False negatives (especially early on)	False negatives	False positives	False negatives (especially later)

governments claimed to adopt certain accounting reforms in order to increase their legitimacy, but they limited implementation of these reforms. Mogoutov and Kahane (2007) found that as nanotechnology gained credibility and resources, authors could "re-label existing research to get the benefit from the 'serendipity' effect." In fact, Fiss and Zajac (2006) showed that organizations often accompanied their implementation failures with rhetoric that assured observers of adoption. Symbolic adoption, therefore, suggests that diffusion measurements based upon keywords and index terms may result in false positives as individuals or organizations claim an innovation label even absent adoption.

Finally, Merton (1968) proposed a fourth mechanism, "obliteration by incorporation," that can confound the correspondence between label-based measures and reality. Merton explained that some knowledge becomes taken-for-granted over time as part of the conduct of normal science and, therefore, no longer warrants explicit citation. As Merton described the process:

> The number of explicit references to the original work declines in the papers and books making use of it. Users and consequently transmitters of that knowledge are so thoroughly familiar with its origins that they assume this to be true of their readers as well. Preferring not to insult their readers' knowledgeability, they no longer refer to [it]. . . . As intellectual influence becomes deeper, it becomes less readily visible (Merton, 1979).

As Lederberg summarized the phenomenon: "The work that *everybody* knows. . .is hardly cited at all!" (Lederberg, 1977). Label-based measures such as keywords and index terms require that someone call attention to the innovation to be labeled in the first place. In turn, obliteration by incorporation suggests that such explicit labeling may dwindle even as – in fact, because – diffusion proceeds. Thus, obliteration by incorporation suggests that label-based measures may undercount diffusion – especially in later periods. Table 1 summarizes these different mechanisms and associated measurement challenges.

Of course, it is possible for all four of these mechanisms to operate simultaneously, resulting in both false negatives and false positives at any given point in time and raising the possibility that results obtained from keywords and index terms may differ widely from the actual diffusion experiences of individuals and organizations. Multiple studies, in fact, recognize this possibility. For example, describing the use of search terms to trace the emerging field of nanotechnology, Mogoutov and Kahane (2007: 894) write:

> While tracking emergent sciences and technologies may be of great importance for researchers, social scientists and decision makers, it often relies on poorly defined data which may be both too large on the one hand (including false positives, publications that are included but should not be) [and] incomplete on the other (including false negatives, publications that are missing but should have been there).

In their study of the electricity sector in the UK, Jamasb and Pollitt (2011: 315) take aim at the use of keywords, specifically. They argue that the use of keywords to search the patent database, "might hence leave out relevant patents or pick up some irrelevant ones." Similarly, Wu and Mathews (2012) write, "using keyword search alone on patent datasets is likely to be of limited value, because many patent documents do not contain expected keywords, while those patents containing them are not always relevant."

Nonetheless, keyword and index-term searches are widely used empirical approaches to assess innovation diffusion. Table 2 presents a selection of influential research papers published in top journals that employ keyword- and/or index-based approaches to diffusion measurement. To be clear, there are good reasons to employ these approaches – primarily tied to data availability – and these study authors themselves are careful to note the limitations of their data. For example, David and Strang (2006: 221–222), who use index terms in a directory to trace TQM consulting services, note, "[The directory] first included the category 'TQM' in 1992, however, which means that the directory could shed light on consulting dynamics during and after the TQM fashion boom, but not on the field's dynamics before the boom started." Similarly, Wang and Bansal (2012:148) write that a limitation of their use of keywords on websites to assess Corporate Social Responsibility lies in the fact that, "our data were essentially 'self-reported.'" In turn, they "encourage future researchers to seek third-party sources and longitudinal data to build further reliability in the data and validity in the findings."

Despite these acknowledged limitations of keyword and index-term based searches, we unfortunately have very little evidence of the extent and direction of biases arising from their use; in other words, we have little work that examines whether and how these measures may be accurate, noisy, or systematically distorting. In this paper, we offer a starting point for discerning these effects by investigating whether and how patterns derived from keywords and index terms differ from those based upon assessments by "domain experts" or individuals with deep knowledge of the innovation of interest. We focus specifically on the magnitude and timing of diffusion, along with three primary variables in the diffusion literature: organizational demography (e.g., Damanpour, 1991), publication outlets (e.g., Strang and Soule, 1998), and collaboration patterns (e.g., Powell et al., 1996).

3. Setting, data and methods

3.1. Setting

To conduct this investigation, we focus on the diffusion of an innovative practice called "green chemistry." As Arora (1997: 401) notes, "the chemical industry is the oldest of the 'high tech' industries, and remains one of the largest manufacturing industries." In turn, a number of scholars have documented the extensive innovations in this industry (e.g., Achilladelis et al., 1990; Arora et al., 1998; Walsh, 1984). Green chemistry is an approach to chemistry that emphasizes human health and safety (Anastas and Warner, 1998; Nameroff et al., 2004). The focus, as Woodhouse and Breyman (2005: 200) note, "is on the prevention of problems before they occur by (re)designing chemicals and chemical production processes at a molecular level." Thus, green chemistry approaches health, safety and pollution prevention by focusing on the entire production process, starting with molecular design and

Table 2
Examples of diffusion research using keywords and indexing.

Study	Focal concept	Measurement challenge	Measurement approach	Potential mechanisms
Abrahamson and Fairchild (1999), *Administrative Science Quarterly*	Use of quality circles	Measuring the adoption of a managerial practice, specifically the "fashionability" of Quality Circles as management technique	*Indexing:* Used ABI Inform database subject headings to measure number of articles pertaining to Quality Circles. Authors recognize the left censoring of the data due to the creation of the Quality Circle subject-heading category in 1977	*Pre-labeled emergence* due to Quality Circles being adopted before the creation of a subject heading • *Symbolic adoption* as Quality Circles became fashionable • *Strategic avoidance* as the Quality Circles fashion waned
Beaudry and Allaoui (2012), *Research Policy*	Spread and impact of funding and collaborations on nanotechnology researchers	Determining which researchers are "nanotechnology scientists"	*Keywords:* Searched for keywords in USPTO; SCOPUS; and Quebec Ministry of Education, Leisure and Sport databases. Individuals who appear in all three databases are nanotechnology scientists. Acknowledge that not all identified individuals may be "nanotechnology scientists"	• *Pre-labeled emergence* due to nanotechnology activity prior to creation of label • *Symbolic adoption* as nanotechnology became fashionable and attracted increased funding
David and Strang (2006), *Academy of Management Journal*	TQM Consulting services	Determining if a firm offers a certain service, specifically if a given consulting firm offered TQM consulting services and whether firm was a "specialist" or "generalist"	*Indexing:* Relied on Kennedy Information's "Directory of Management Consultants" and its indexing of services offered by listed firms. Authors recognize the left censoring of the data due to the creation of a TQM category in 1992	• *Pre-labeled emergence* due to TQM services being offered before creation of TQM category • *Symbolic adoption* as TQM "fashion boomed" • *Strategic avoidance* as TQM boom turned to bust
Fiss and Zajac (2004), *Administrative Science Quarterly*	Shareholder value orientation	Capturing a given company's "orientation" toward distributing value – specifically, German firms' shareholder value orientation	*Keywords:* Searched annual reports of 112 largest firms in Germany for the phrase "shareholder value" or its German equivalent, "Unternehmenswertsteigerung"	• *Strategic avoidance* prior to a Shareholder Value Orientation (SVO) becoming legitimate • *Symbolic adoption* as a SVO becomes legitimate • *Obliteration by incorporation* after a SVO is taken for granted
Greve (1995, 1996), *Administrative Science Quarterly*	Adoption (1995) and abandonment (1996) of radio station formats	Capturing changes in market position, specifically adoptions (abandonment) of formats by radio stations	*Indexing:* Used notification of format changes published in a weekly industry newsletter, the M Street Journal. The newsletter collects reports of format changes from a network of industry insiders who alert the editor when they judge a station has changed its format	*Pre-labeled emergence* due to lag in reporting and creation of new format categories
Hendricks and Singhal (1997), *Management Science*	TQM awards	Determining if a firm had successfully implemented a practice – specifically, determining which public companies had won one of the various quality awards related to TQM	*Keywords:* Performed keyword search of online databases using terms "quality" and "award." Supplemented these results with lists of award winners from various award-sponsoring organizations. Authors recognize that initial adoption of TQM is unobservable	• *Pre-labeled emergence* due to TQM being adopted before the creation quality awards • *Strategic avoidance* in order to not be associated with TQM as it lost popularity
Jonsson (2009), *Organization Science*	Diffusion of socially responsible investing (SRI) funds	Determining which investing funds are socially responsible	*Indexing:* Relied on description of products from mutual funds and associated "socially responsible" labels	• *Pre-labeled emergence* due to socially responsible investment funds prior to SRI labeling • *Symbolic adoption* if firm wanted to feature SRI fund emphasis due to perceived social benefit • *Strategic avoidance* if firm wanted to hide SRI fund emphasis due to perceived lower performance
Kalev et al. (2008), *Administrative Science Quarterly*	Joint productivity councils	Determining the diffusion of a new managerial model, the Joint Productivity Council (JPC)	*Keywords:* Relied on mentions and descriptions of JPCs in archival reports of the National Productivity Council. Recognize that there was decoupling between adoption of JPCs and actual use	• *Pre-labeled emergence* when firms employed similar structures prior to the naming of JPCs • *Symbolic adoption* if firm wanted to feature JPCs without changing actual managerial model • *Strategic avoidance* if firm wanted to hide JPCs
Wang and Bansal (2012), *Strategic Management Journal*	Corporate Social Responsibility (CSR) Activities	Capturing firms' CSR activities – specifically, CSR activities of new ventures in areas of community, employee relations, environment, products and production and other stakeholders	*Keywords:* Counted the number of discrete CSR-oriented activities listed on firms' websites. Authors recognize that CSR activities on websites are self-reported, and encourage future research to use third-party sources	• *Strategic avoidance* if CSR thought to show distraction and resource wasting by new venture • *Symbolic adoption* if CSR activities thought to show legitimacy and attract resources for new venture
Zaheer and Mosakowski (1997), *Strategic Management Journal*	Trading rooms	Identifying whether a firm engages in a certain activity, specifically if financial firms operate market-making trading rooms	*Keywords:* Relied on self-reporting of trading room activity to Hambros Bank Directory. Authors recognize firms can accidently fail to self report and so ignored one year gaps in listings (two year gaps if trading rooms are subsequently relisted in the same location)	• *Symbolic adoption* if firm wanted to feature market-making activity • *Strategic avoidance* if firm wanted to hide market-making activity • *Obliteration by incorporation* if trading rooms became taken for granted

transformation, rather than focusing on the "end of the pipe" by addressing environmental challenges only after substances have been created.

The drafting of 12 "principles" in 1998 codified the practice of green chemistry (see Appendix 1). These principles offer guidance on how to improve processes and products for different stages in a chemical's lifecycle. For example, one principle advises against using solvents (or encourages selecting innocuous ones) for chemical synthesis. Another principle asserts that chemicals should degrade into benign by-products at the end of their useful lives, rather than persist in the environment. These 12 principles are straightforward to practice and are easily identifiable by any knowledgeable chemist (Anastas and Warner, 1998; Linthorst, 2010). Green chemistry is not, however, based upon a single (or few) breakthrough patent(s) or publication(s). As such, it is comparable to the diffusion of other innovative practices, such as TQM, that researchers cannot trace through simple citation-based approaches.

To practicing chemists, green chemistry's role and legitimacy has varied over time. Green chemistry arose, most immediately, from a shift by the US Environmental Protection Agency (EPA) in the mid-1980s to "pollution prevention" as opposed to solely "end of pipe" emission controls (Anastas and Kirchhoff, 2002; Linthorst, 2010; Stephan and Atcheson, 1989). The 1990 Pollution Prevention Act in the US and a number of other initiatives in the US and Europe around the same time were indicative of an emerging sense that chemists could approach molecular design and synthesis in new, environmentally sensitive ways (Linthorst, 2010). Nevertheless, most chemists in the 1990s viewed green chemistry as a marginalized field. As one chemist shared with us, "You didn't get tenure [in the mid-1990s] by making your name in green chemistry" (we interviewed several chemists in order to learn more about the field).

By the end of the 1990s, however, green chemistry began to exhibit the trappings of a distinct scientific specialty with growing legitimacy: The Clinton administration launched the US Presidential Green Chemistry Challenge Awards in 1996 to recognize both academic and commercial contributions to green chemistry; the Green Chemistry Institute (GCI) was founded in 1997; and the peer-reviewed journal *Green Chemistry* was first published in 1999. Chemistry's major professional organizations recognized green chemistry by merging the Green Chemistry Network (GCN) with Britain's Royal Chemical Society (RCS) in 1998 and absorbing the Green Chemistry Institute (GCI) within the American Chemical Society (ACS) in 2001. In 2005, the Nobel Prize in chemistry was awarded on the basis of the science underlying this approach – a strong signal of legitimation. In turn, however, other chemists worry that the label is not always applied appropriately. For example, one practitioner shared with us, "[today,] there is a lot of stuff getting published under the name green chemistry that really shouldn't have been published under that name or, perhaps, at all. And that's hurting us." The unclear and changing status of green chemistry makes it an advantageous setting in which to investigate measurement because it could result in a lack of correspondence between the *labeling* of green chemistry and the *practice* of green chemistry.

3.2. Data and methods

We measured the diffusion of green chemistry using keyword search, database indexing, and the manual sorting of records by practicing green chemists (hereafter, "domain expert assessment"). We applied each of these measurement approaches to identify green chemistry records, published in 2008 and earlier, in the SciFinder Scholar database. We selected SciFinder Scholar because practicing academic chemists identify it as the authoritative database of chemistry research. Our comparison of search results from SciFinder Scholar against results obtained from ISI's Web of Science (another leading database) confirmed that SciFinder Scholar provides a far more complete collection of records (nearly three times as many in our case).

We restricted our search to articles appearing in English-language academic journals. We limited records to academic journal articles in an effort to capture the diffusion of green chemistry as a practice per se, rather than diffusion of more general commentary on green chemistry such as book reviews and opinion pieces. We limited the search to records published in English for purely pragmatic reasons; however, since English is the "lingua franca" for much of the scientific world, our results still reflect considerable geographic diversity.

3.2.1. Keyword search dataset

We constructed the keyword search dataset by searching for the phrase "green chemistry" in SciFinder Scholar. This keyword search queried the title, abstract, and other metadata for each publication, but not the full text. This search for "green chemistry" yielded 5799 articles.

The authors of each article determine the keywords for that article. Thus, keyword searches reflect instances in which authors claimed the "green chemistry" label. These instances may reflect the actual practice of green chemistry or instances in which authors claimed the label but did not engage in the practice.

3.2.2. Database indexing dataset

We constructed the database indexing dataset by capturing those articles that received a "green chemistry" index term from SciFinder Scholar. This search yielded 5592 articles. These index terms are assigned by employees of the database company, not by the authors of individual articles. Informal interviews with reference librarians and with indexers indicated that employees vary, however, in their scientific knowledge and that they generally claim expertise, if at all, in a broad area such as "natural sciences" rather than a specific field such as "green chemistry." Moreover, indexers often draw upon keywords, thus deferring at times to article authors when making assignments.

3.2.3. Domain expert assessment

A key feature of our database and analysis is a "domain expert assessment" dataset. Our intention with this domain expert assessment dataset was to create a group of all articles that met the technical definition of green chemistry – based upon the principles in Appendix 1 – such that we could compare the keyword and index-term datasets against this group. Our first step in creating this dataset was to search the SciFinder Scholar database for a wide range of keywords (beyond just "green chemistry") in an attempt to capture all articles that might draw upon or reflect the practice of green chemistry. Our author team includes two practicing PhD chemists who specialize in green chemistry. Based upon their knowledge of the field, their conversations with a number of other practicing green chemists, and an examination of the search terms used in other analyses of green chemistry (e.g., Nameroff et al., 2004), they determined that the broad set of keywords listed in Appendix 2 would capture the full range of green chemistry articles.

This initial broad search yielded 10,231 articles. Of course, given our broad search terms, many of these articles were "false positives," meaning that they were not actually green chemistry articles. Our next step, therefore, was to filter the results so that we only retained articles that met the technical definition of green chemistry. The two PhD chemists on our author team designed and executed our filtering approach. They determined that an article should be excluded if it did not report on experiments or findings that employed one or more of the 12 principles of green chemistry (see Appendix 1) – regardless of whether or not the article had a green chemistry index term or keyword. They tested this approach by constructing an initial database of 100 articles and by then comparing their individual determinations of which articles were included and excluded. In this initial set, they (initially) disagreed on only one article (the disagreement centered not on whether this article met the technical definition of green chemistry, but rather on whether the article was a review article or a research article).

One of the two chemists then took the lead by reviewing the abstracts (and, when necessary, the full text) of all 10,231 articles identified in the broad search to determine if it met the green chemistry criteria. In weekly meetings with the other chemist and the full research team, he highlighted any ambiguous articles, which always accounted for less than 1% of the articles reviewed that week. He and the other chemist then reviewed these ambiguous articles along with a sample of all articles included and excluded. The two chemists' inter-rater reliability scores on these samples exceeded 98%. It is important to note that the 12 principles of green chemistry have little technical ambiguity. As a result, it is straightforward for a trained chemist to determine whether or not a particular reaction, process, or concept qualifies as green chemistry. Our approach is consistent, therefore, with Lee's (2009) study of organic food laws, in that someone with requisite technical knowledge in the area can easily distinguish records of interest. As a check on our process and assessments, we consulted with a third practicing chemist, a tenured professor who also specializes in green chemistry, who validated this approach.

Some examples may be helpful in illustrating how the chemists determined which articles to include in our domain expert dataset: A 2008 article in the journal *Oil and Gas Science and Technology* (Economou et al., 2008) focuses on ionic liquids, a core material for the practice of green chemistry, and describes their use in "environmentally benign processes." This article, however, does not use the term "green chemistry" and is not picked up in the keyword or index term dataset. Because use of ionic liquids in place of other solvents invokes green chemistry principle five (see Appendix 1), the domain experts included this article. By contrast, another 2008 article, published in the *Journal of Industrial Ecology* (Kapur et al., 2008) includes "green chemistry" as a keyword; but, because the article does not build upon, reference, or use any of the 12 principles of green chemistry, the domain experts did not include it.

We acknowledge that this lack of ambiguity and the existence of coherent, stable, and widely acknowledged criteria may mark green chemistry as an unusual case (indeed, those features mark it as attractive for our comparison of different measurement approaches). As Kaplan and Radin (2011), Kennedy et al. (2010), Granqvist et al. (2013) and others argue, the determination of "what" falls within in field is often subject to important political and sociological pressures. Our interest in this study, however, is less around the margins or boundaries of the field and more around the potentially different views of diffusion obtained via these three forms of measurement (keywords, index terms, and domain expert assessment).

Ultimately, our domain expert assessment dataset contained 4763 articles and our full dataset (including articles captured via any of the three approaches) contained 6394 articles. With these records identified, we then coded each article for the journal in which it appeared; the organizations represented (according to author affiliations); the organization types (firm, university, government, or other); the countries represented (by organization locations); and collaborative patterns on an individual and organizational basis. Because SciFinder Scholar provides organizational information for only the corresponding author, we turned to the front-page text of the actual articles to glean and code this information.

Our analysis follows previous studies on measurement of innovation (e.g., Hagedoorn and Cloodt, 2003; Nelson, 2009, 2012) in that we compare diffusion patterns over time and across these different dimensions. Our study, therefore, does not have independent and dependent variables, or the regression analyses that these permit, since our research interest is comparative and not causal or relational.

4. Results

4.1. Article publication trends

Different approaches to the identification of green chemistry articles yielded very different results: keyword search yielded 5799 articles; database indexing yielded 5592 articles; and domain expert assessment yielded 4763 articles. Fig. 1 displays these results over time. As the figure illustrates, the three approaches identify different starting points for green chemistry, with domain expert assessment, keyword search, and database indexing placing the first green chemistry publication in 1990, 1995, and 1999, respectively.

Through the year 2000, domain expert assessment yields more results in any given year than either of the other two approaches. The comparatively low number of keyword articles suggests that many authors were not labeling their articles as "green chemistry" even though the underlying science reflected green chemistry. Further, the lack of articles in the index term data set prior to 1999 reflects the fact that SciFinder did not create an index term for green chemistry until late in that year.

In 2001, as Fig. 1 indicates, there is a shift as keywords and index terms identify more articles than does domain expert assessment. This result indicates that authors and indexers alike may have grown overzealous in their application of the label.

Of course, multiple pressures for under- and over-labeling may be active simultaneously and it is difficult to disentangle their effects on the basis of raw yearly counts alone. Thus, we compared each measurement approach on an article-by-article basis. As indicated in Fig. 2, both keywords and index terms miss a substantial, but declining portion of the articles identified by domain experts (indicated by solid lines in Fig. 2). Specifically, keyword and index searches miss the vast majority of expert-identified green chemistry articles published in the early years of the field, and miss about 10% of articles published in the most recent years.

Fig. 2 also indicates that keywords and index terms pick up a number of false positives (indicated by dashed lines in Fig. 2). In other words, these articles are labeled with "green chemistry" even though our experts judged that the underlying science does not correspond to green chemistry. These differences decrease over time, though experts still judge a full one-quarter of articles in the most recent years as "false positives."

As a first cut, these results indicate that keywords and index terms exhibit considerable differences from domain expert assessments and that these differences are especially severe in the first decade of diffusion. Next, we examine these differences across the dimensions of publication outlets, organizational demography, and collaboration.

4.2. Publication outlets

The journal population data reflect these same general patterns. Green chemistry research appeared in a large number of journals: keyword searching indicates 864 unique journals, database indexing indicates 849 unique journals, and expert assessment indicates 682 unique journals. As with the assessment of overall publishing activity, the data are especially inconsistent through 2001 (see

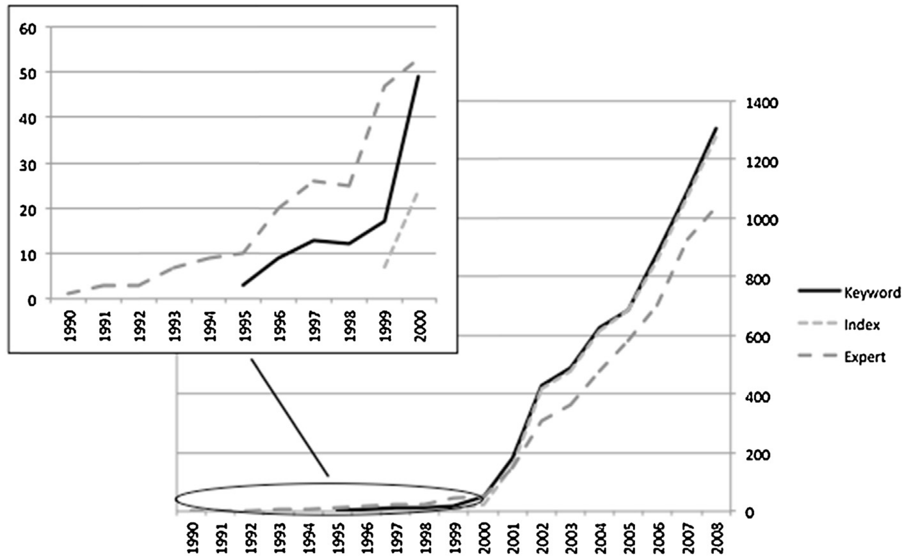

Fig. 1. Number of articles captured by each approach.

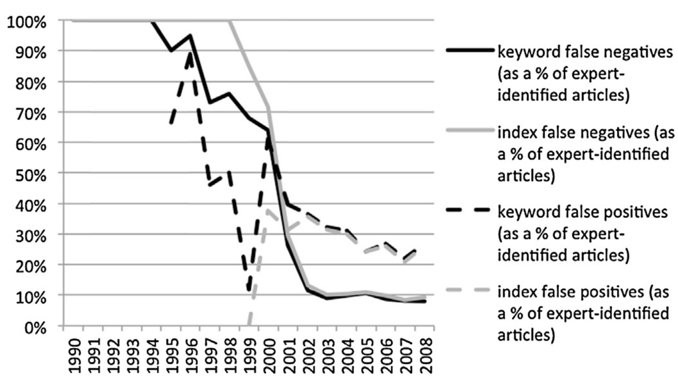

Fig. 2. Differences between keyword/index-term datasets and domain expert assessment (articles).

Fig. 3) and keywords and index terms appear to miss many of the journals that were among the first publishers of green chemistry research (according to domain experts). For example, *Chemical Engineering* published some of the earliest work in green chemistry (with two articles in 1991), but it is not included in the keyword and index datasets. *Chemistry for Sustainable Development* published a pair of 1993 articles, when there were less than 15 articles in the whole field, but keywords or index terms also miss this journal – perhaps because it published only three articles total (the last article appeared in 1996). Because keywords and index terms pick up later activity, they appear to miss publication outlets that were active early but that fizzled before field took off. More generally, these findings suggest that if keywords and index terms under-capture early activity in a field, they may miss entirely those participants that contribute then exit in this early period.

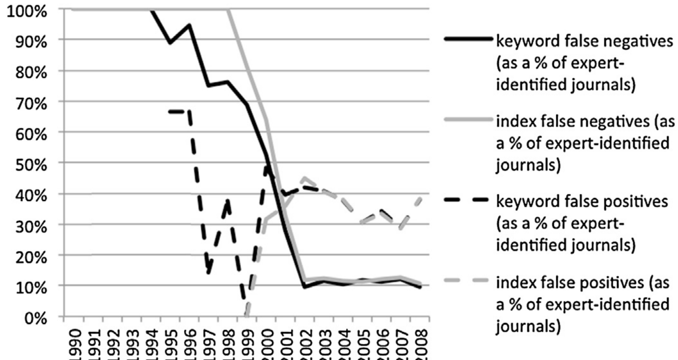

Fig. 3. Differences between keyword/index-term datasets and domain expert assessment (journals).

Fig. 3 also illustrates how, starting in 2001, keywords and index terms miss about 10% of the journals included in the expert assessment (see solid lines in Fig. 3). These errors, however, do not appear to affect the journals that publish the most green chemistry research: the top 10 journals according to expert assessment appear in the "top 12" of the index term dataset and the "top 13" of the keyword dataset.

On the other hand, about 35% of reported journals in the keyword and index term datasets do not appear in the domain expert dataset (dashed lines in Fig. 3). For example, *Environmental Science and Technology* is ranked as the sixth most frequent journal in the keyword dataset (and seventh in the index term dataset). Domain expert assessment, by contrast, ranked the journal 39th (it is worth noting that this journal is an American Chemical Society, or ACS, publication and that the ACS has an active interest in green chemistry). Similarly, the *International Journal of Life Cycle Assessment* appears seventh in the keyword dataset and sixth in the index dataset. According to expert assessments, however, this journal does not actually publish green chemistry research.

4.3. Organizational demography

We identified organizations adopting green chemistry by the appearance of an organization's name in affiliation with an author on an article. We counted only unique organizations in a given record to avoid assigning undue weight to records featuring numerous authors from the same organization. The lists of the most active overall organizations according to each measure are relatively consistent: the top 10 most active organizations according to expert assessment appear in the "top 11" in the index term dataset and the "top 12" in the keyword dataset.

In terms of dates of first engagement, however, the discrepancies are much greater. For example, the domain expert data show the US EPA – an important early advocate of green chemistry according to qualitative histories of the field (Linthorst, 2010) – publishing as early as 1992, whereas the keyword and index data

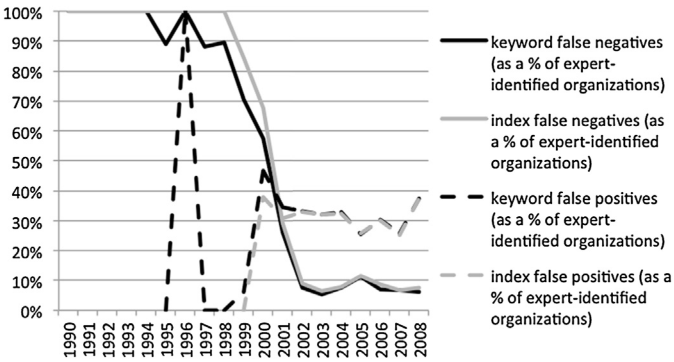

Fig. 4. Differences between keyword/index-term datasets and domain expert assessment (organizations).

do not pick up the US EPA until 1998 and 1999, respectively. Aston University in the UK, one of the earliest organizations in the field according to the expert assessment data (1990), does not enter the keyword and index term datasets until 2002. Similarly, Union Carbide chemical company is one of the earliest organizations in the expert data (1991), but enters the keyword and index datasets only in 2003. In fact, other early organizations are missed entirely since their publishing activities were limited to the early years of the field: Both 3 M and the Norway Water Technology Center published in 1991, but neither organization is included in the keyword or index term datasets.

The patterns of discrepancy between measures reflect these differences over time, as illustrated by Fig. 4. Thus, as with the assessment of articles, missed organizations (per the expert dataset) decrease dramatically after the first few years, ending around 10%. False positives, however, account for almost one-third of organizations.

These early differences also show up in the relative prevalence of organization types. Across all years, the distribution of universities, firms, and other organizations is relatively consistent. In the first five years, however, the measures show very different patterns of involved organizations. Specifically, the keyword dataset highlights a large percentage of unaffiliated authors and other types of organizations: 55% in the keyword dataset, versus 5% in the index dataset and 9% in the domain expert dataset. Conversely, the index-term dataset highlights a dominant role for universities: 70% of organizations, versus 27% per keywords and 41% per domain experts. The expert dataset, by contrast, shows a far greater role for firms: 32% of organizations, versus 10% per index terms and 2% per keywords (incidentally, many of our interviewees highlighted the critical role played by firms in green chemistry's emergence: Many green chemistry practices grew out of industry and one motivation for the Presidential Green Chemistry Challenge Awards was to encourage firms to share information about their practices with the public). Overall, therefore, this analysis highlights how different measures lead to different conclusions about the role of the public and private sectors in green chemistry's emergence.

4.4. Collaboration patterns

Our next assessment concerned the collaborative patterns reflected in the articles. Fig. 5 displays these patterns for both individual-level collaboration (lines in Fig. 5) and organizational-level collaboration (bars in Fig. 5). In both cases, the measures closely align starting in 2001. In the years beforehand, however, the discrepancies are quite pronounced (these differences are not an artifact of the small number of articles in the keyword and index term datasets in the 1990s since the percentages for each measure are displayed as a function of the articles captured by that same measure). The differences show that the articles picked up by keywords (the black line and bars in Fig. 5) in the 1990s are far less collaborative – on both an individual and organizational basis – than those picked up by expert assessment (the medium-gray line and bars); in other words, the articles picked up by keywords are more likely to be single author and single organization publications. Thus, the different measures suggest different roles for collaboration in the emergence of the field (and, in turn, different network structures), with expert assessment portraying a more collaborative field at the outset.

4.5. Cross-tabs between false positives/negatives and organization types, countries, and collaborative patterns

Finally, we examined whether the rate of false positives and false negatives varied across organization types, countries, and collaboration patterns. Table 3 displays the results of these cross-tabs. The dark-shaded cells in the table indicate error rates that are at least 25% higher than in the dataset as a whole (e.g., especially "high" error rates). The light-shaded cells indicate error rates that are at least 25% lower than in the dataset as a whole (e.g., relatively "low" error rates). Across the three categories of organizations – universities, firms, and other organization types

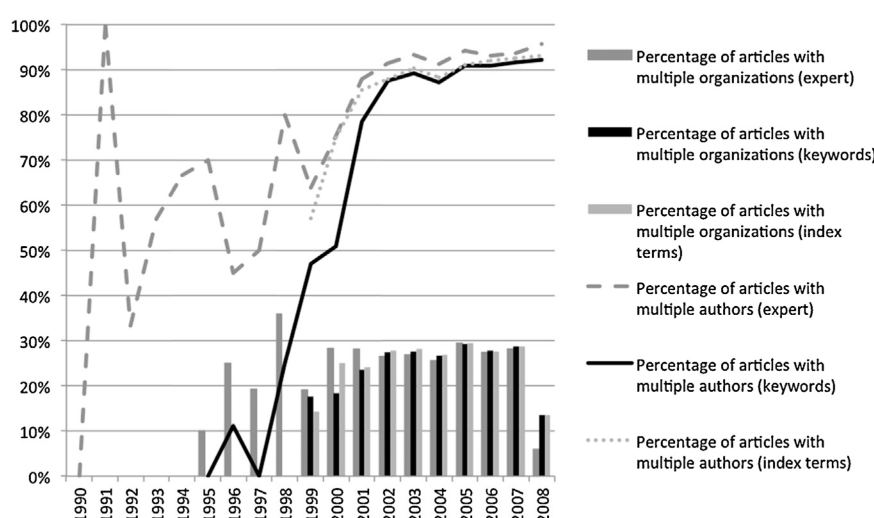

Fig. 5. Differences between keyword/index-term datasets and domain expert assessment (collaboration).

Table 3
Cross-tabs between false positives/negatives and organization types, countries, and collaborative patterns.[a]

	Keywords		Index terms	
	% of keyword hits that are false positives	% of expert hits missed by keywords	% of index hits that are false positives	% of expert hits missed by index terms
Full dataset	28.1	12.5	26.7	13.9
By organization type				
Universities	26.5	11.9	25.8	13.1
Firms	46.0	20.8	46.1	23.5
Other	33.9	14.8	30.1	16.7
By country				
China	9.4	6.3	9.4	6.8
India	10.8	9.0	10.6	9.6
US	48.2	26.4	46.4	28.6
Iran	4.1	2.7	4.1	2.7
Japan	27.1	20.0	27.5	21.4
Italy	25.5	6.3	25.8	7.9
UK	41.4	14.1	38.9	17.0
France	20.3	7.3	20.0	9.9
Germany	42.5	20.2	40.8	22.5
Spain	32.2	5.6	33.1	9.6
By number of authors on a publication				
1 author	54.7	22.5	47.7	28.8
2 authors	23.5	9.2	23.0	10.2
3 authors	31.0	14.3	30.9	15.4
4 authors	24.5	11.7	23.9	12.4
5 or more	21.0	11.9	21.1	13.1

[a] Dark-shaded cells are at least 25% higher than the full dataset average (e.g., they have especially high error rates); light-shaded cells are at least 25% lower than the full dataset average (e.g., they have relatively low error rates).

such as government laboratories – firms have especially high error rates: Both the keyword and index term measures yield almost a 50% false positive rate, indicating that almost one-half of the firm-affiliated research labeled as "green chemistry" is not, in fact, green! At the same time, these measure miss more than one-fifth of all relevant publications with this organization type.

Table 3 also lists the 10 countries most active in publishing green chemistry research. China, India, Iran, and France each have relatively low error rates, indicating that keywords and index terms are fairly accurate measures of green chemistry research in these countries. Conversely, the United States and Germany have especially high error rates both for false positives and for false negatives – again, approaching 50% for false positives! (In Germany, especially, "green" carries political connotations due to the relatively high profile of the "Green Party" in the German political system. Such an association could lead to under-labeling of green chemistry in the German case – false negatives – though it does not explain the high rate of false positives).

Finally, single-author publications appear especially susceptible to both false positives and false negatives. It is possible that additional authors act as a "check" upon the accuracy of keyword assignments (that are subsequently picked up by indexers): Although a single author can add the "green chemistry" keyword even if it is connection to their research is tenuous, multiple authors may prevent one another from making such associations. These findings around single authors also could explain the patterns presented in Fig. 5 around collaboration patterns; if single authors are more likely to claim the "green chemistry" keyword even where it may not apply, then the total pool of articles captured using this keyword will appear less collaborative since this total pool will include more single-authored articles.

4.6. Matching mechanisms and findings

To summarize our results, we found that keywords, index terms and domain expert assessment yield very different conclusions about total publications, journals, organizations, and collaboration patterns. We found that these differences are present over the entire time period examined and are especially prevalent in the first decade, thus signaling special relevance to studies of emergence and early diffusion patterns. We also found that errors rates varied across organization types, countries, and collaboration patterns.

Although our intention was not to engage in hypothesis testing around the specific mechanisms that we described earlier – such tests would require uncovering each author's and indexer's specific motivations – our results are consistent with the operation of all four mechanisms (see Table 1). Moreover, the green chemists whom we interviewed offered multiple examples of each mechanism in action (unfortunately, we lack the extensive and detailed interview data that would be necessary to link specific authors, indexers and labeling behaviors in a systematic fashion). Specifically, our study demonstrates how pre-labeled emergence can shape researchers' perceptions through a high rate of false negatives in the early period. Before a term enjoys widespread acceptance – and, in the case of index terms, before it is logged as a unique index term – relying on instances of that term to identify instances of a practice will likely result in dramatic undercounting, as we demonstrate.

False negatives also are consistent with strategic avoidance, or the intentional lack of labeling despite awareness of the term. Again, we lack systematic interview data to link the lack of labeling to this specific mechanism. Strategic avoidance, however, would result in underreporting of activity throughout a diffusion period, which our data do demonstrate.

Symbolic adoption, by contrast, results in false positives and may continue to be active in later stages as an innovation has increased in apparent legitimacy. Again, this pattern matches our data, which show false positive rates around 30% throughout the 2000s (Figs. 2–4). We also found that journals may engage in symbolic adoption, leading to similar patterns for index terms. In fact, the far greater number of journals picked up by keywords and index terms (versus experts) suggest that some journals may publish green chemistry research only in "label" but not in practice.

Finally, obliteration by incorporation – the assumption that a well-accepted practice need not be labeled – would result in false negatives in later time periods. Once more, our data are consistent with this mechanism, with false negatives constituting approximately 10% of articles – and, by extension, about 10% of journals and organizations – in the mid- and late-2000s. Though we intend our study to illustrate differences between measurement approaches – not to disentangle theoretical mechanisms – our results suggest that the four identified mechanisms hold explanatory promise.

5. Discussion

Innovation diffusion is a central topic in strategic management and entrepreneurship. Nevertheless, measuring innovation diffusion remains challenging and we have little data on how different measures may affect our assessments of emergence and patterns over time. This paper contributes to a growing literature concerned with the correspondence between measures and the activities or effects that these measures are attempting to capture (Aguinis and Edwards, 2014; Boyd et al., 2013; Chatterji et al., 2009; Colyvas and Powell, 2009; Freeman and Soete, 2009). Specifically, we provide empirical evidence that measurement of innovative practices is, indeed, a problem. We document the timing, magnitude and direction of potential biases for the case of green chemistry, and we present potential solutions in the form of domain expert assessment and comparative measures. In the sections below, we discuss where innovation measurement may be most problematic, what role domain experts might play, and how researchers might leverage multiple measures, along with the policy implications that follow from our study.

5.1. Contextualizing measurement concerns

A central question for innovation scholars concerns the extent to which these results may or may not challenge the theoretical mechanisms and conclusions of prior studies. To this point, our data suggest that the answer is largely dependent upon the specific research question. For example, one of our most persistent findings concerns the undercounting and overcounting of diffusion over time, including the widely varying data in the first decade. If a diffusion study is focused on field emergence and associated constructs, such as determining which organizations are "early adopters" of a particular practice, then our results suggests that current keyword and index-term approaches may present significant challenges and could call conclusions into question. Similarly, our results suggest that keywords and indexing may challenge research that relies upon the start and end times of various diffusion cycles. Research that investigates more general patterns, however, may be less challenged.

Our results also suggest that measurement concerns may be most acute when a label is ambiguous and/or when the diffusion subject is difficult to observe. For these reasons, assessing the diffusion of innovative practices or fields, as with the examples given in Table 2, may be more difficult than assessing the diffusion of specific artifacts, like therapeutic drugs (Coleman et al., 1957), hybrid corn (Griliches, 1957), mechanical reapers (David, 1975), stretched aircraft airframes (Rosenberg, 1982), automated teller machines (Saloner and Shepard, 1995), or other exemplars of diffusion research. Even when the meaning of a label, however, is clear – as with these examples – diffusion measurement can be difficult if the artifact is controversial, illegal, or competitive since these attributes could lead to concealed adoption. Conversely, if a study explores a long-established practice, then obliteration by incorporation suggests that false negatives in later years are the main issue. If the practice is ambiguous, like "going green" or "adopting TQM," then false positives, by contrast, may be the primary concern. In short, then, our results suggest that in order to determine whether keyword- and index-based approaches may pose a challenge, researchers should carefully consider the possible operation of each of the four mechanisms that we identify alongside the particular characteristics of their focal product or practice. The fact that false positives and false negatives vary over time and appear simultaneously, however, means that these issues probably demand consideration in most cases and cannot be subsumed under a simple error term.

5.2. Incorporating domain experts

One of our central recommendations is to incorporate domain experts into the identification and execution of search strategies. The use of experts also responds to calls for increased interdisciplinary research and for grounded studies that draw deeply upon participants embedded in phenomena of interest (e.g., Barley and Kunda, 2001). We find that domain experts can perform multiple roles. First, experts can help to refine the keywords and index terms themselves. In our case, expert input allowed us to identity the specific keywords lists in Table 2. In their study of solar PV technology, Wu and Mathews (2012) employed technology experts to verify the patent classifications (a form of index) that they used. Similarly, in their study of embedded software in automobiles, Xie and Miyazaki (2013) used experts to help them identify keywords and to identify "where" in the patent record to search for these keywords (e.g., title, abstract, or claims). In fact, Islam and Miyazaki (2010), Mina et al. (2007), Nameroff et al. (2004) and many other studies all employ variants of this approach, using expert-guided searches to reduce the incidence of false negatives.

Our study, however, demonstrates that experts can play two other roles, too. First, experts can identify evidence and markers that may extend beyond the "usual suspects." In our case, experts pointed us toward using the SciFinder Scholars database rather than the ISI Web of Science database. In a different domain, Martin (2011) relies on Wall Street industry analysts – another type of "expert" – to develop his set of comparator firms in the software industry, noting that this approach provides a more accurate picture of the focal firms' competition than would simply identifying those firms that share an SIC code. In pointing toward alternative data sources that may be more thorough, reliable, or accurate, the use of experts can thus reduce false positives and/or false negatives.

Finally, experts can filter out false positives that almost inevitably result from more expansive attempts to capture diffusion. In our case, for example, experts trimmed the initial expansive search from 10,231 articles to 4763 articles. This use of experts, therefore, reduces the appearance of false positives in the final database, serving as an important counterweight to biases associated with the use of broad search terms alone (e.g., Islam and Miyazaki, 2010; Nameroff et al. (2004)).

In turn, combining these multiple uses of experts, as we do here, can reduce both false negatives and false positives. The result is a dramatically improvement in a dataset's accuracy, as we demonstrate.

Of course, we acknowledge the difficulty of finding, selecting, and involving such experts. Our own work benefited from close collaboration between management scholars and chemists at the same institution, an advantage of the diverse university environment. Here again, the nature of the research question and setting should inform the potential benefit from incorporating expert or practitioner assessments, which can then be weighed against the cost in terms of research time, expense, and other factors. Moreover, different studies may need to employ different numbers and/or types of domain experts. Green chemistry, for example, is relatively unambiguous and our experts displayed high agreement in their judgments. But in other cases – for example, where a field is both broad and specialized, or where it lacks coherence or agreement – a larger team of experts may be necessary to refine and validate measures.

5.3. Leveraging multiple measures

More generally, and regardless of the incorporation of experts into a research design, our results encourage tighter linkages between the innovations investigated and the measures employed to study these innovations – along with a greater discussion of how any particular study's results may be tied to or influenced by the measure. Thus, index terms, to use one example, can be a powerful tool for quickly gathering a set of relevant documents, individuals or organizations. The blind use of index terms absent consideration of our four identified mechanisms, however, cannot be expected to yield an accurate or unbiased assessment of diffusion.

In making this call, our belief is not that one approach is necessarily or inherently superior to another one. In fact, our results may be interpreted as *rejecting* a purely "objective" view of diffusion and instead providing strong evidence that perspective differ and that diffusion assessments must be taken in context. In turn, our results suggest that researchers may improve the measurement of their focal phenomena by drawing upon multiple sources and by comparing across them (e.g., Järvenpää et al., 2011; Kleinknecht et al., 2002; Lee et al., 2011; Nelson, 2009, 2012). Lee et al. (2011), for example, combine patent classifications (a type of indexing) with keyword searches in order to improve their search results. Nelson (2009) demonstrates how a combination of patents, publications and licenses can provide better results than any measure in isolation.

Beyond attempting to improve the accuracy of measurement, the use of multiple measures also can be informative in the differences that may be revealed. For example, where there is disagreement between different measures of the same phenomenon, researchers might leverage these disagreements themselves to focus their attention on potential patterns or variables of interest (these disagreements, for instance, might be revealing of the social and/or political construction of innovations and emerging fields). For this reason, measurement differences are not necessarily something for researchers to minimize, but rather something to investigate as potentially informative about where and how individuals and organizations interpret and contest diffusion. In turn, uncovering unexpected relationships in a dataset can spur new investigations and collaborations. Such activities, however, require multiple measures in the first place.

5.4. Policy implications

Finally, our work holds important policy implications. Specifically, our work suggests that government support may both spur and respond to mislabeling and measurement errors. For example, government support for a field may lead some researchers to adopt a particular label – either disingenuously or as a new label for work that they already were performing (in fact, government agencies can provide some of the strongest incentives for mislabeling of research when they emphasize certain program areas as "valued"). In turn, such increases in labeled research may lead to misguided conclusions about the role and effectiveness of government support.

Conversely, a failure to track innovation (as evident in false negatives) may suggest to policymakers that a "jump start" is needed when, in fact, it may not be. Or, false negatives may adversely affect research funding and government support for an emerging field: To the extent that government agencies rely upon broad assessments of activity in a field, strategic avoidance and pre-labeled emergence may lead to a vicious cycle in which agencies perceive a lack of activity and, therefore, do not offer support, while researchers witness a lack of support and, therefore, continue not to claim a label. Funders of all sorts, therefore, would be wise to be careful in their research assessments, especially when our four identified mechanisms may be active.

The differences that we witness among organization types and countries also hold important policy implications. Specifically, the results in Table 3 suggest that firms may not be as active in green chemistry as keyword- and index-based measures may suggest. Similarly, keyword- and index-based measures would appear to over-represent the activities of the US and Germany relative to China, India and Iran. In turn, policies based upon perceptions of significant firm involvement or of US or German research dominance may be built upon false premises.

Together, these possibilities highlight why governments should take steps to ensure that innovations measures do, in fact, measure innovation. In turn, preliminary efforts – such as the "STAR Metrics" program jointly organized by the US National Science Foundation, National Institutes of Health, and Office of Science and Technology Policy (Lane, 2010) – might be usefully expanded by attending specifically to issues of labeling, including the ways in which labels may be manipulated and the many motivations that individuals and organizations may have to engage, knowingly or not, in such manipulations.

5.5. Limitations and future research

An important boundary condition for our investigation is that we focus on the diffusion and measurement of innovations and, specifically, of innovative practices, strategies, and other phenomena that are not particularly amenable to discrete measurement. Of course, a wide range of other things have diffused, including civil service reforms (Tolbert and Zucker, 1983), the ordination of women in churches (Chaves, 1996), same-sex partner benefits (Creed et al., 2002), and AACSB accreditation among business schools (Casile and Davis-Blake, 2002). Although it is likely that our identified mechanisms may apply to these settings, too, further research could better delineate the generalizability and boundary conditions of our findings.

Moreover, we limit our attention to attempts to capture diffusion through keywords and indexing. Thus, we do not extend our theory or findings to the prevalent approach of citation-based diffusion measurement. Of course, the two approaches are not wholly distinct and some of our identified mechanisms could apply to citation-based studies, too. For example, Lampe (2012) notes that patent applicants often fail to make relevant citations for competitive reasons, a variant of strategic avoidance. Similarly, a number of patent citation studies employ patent "classes," which are conceptually similar to index terms in that they attempt to establish a category for records of common interest. In turn, patent classes likely suffer from similar limitations as index terms (for example, the patent class for nanotechnology, "977," was not introduced until August 2004. Thus, reliance upon this class to capture nanotechnology patenting activity would suffer from pre-labeled emergence biases, failing to capture activity prior to 2004). In the realm of publication citations, MacRoberts and MacRoberts (1986) discuss how authors, for "political reasons," often cite publications that did not inform their work – leading to a similar dynamic as symbolic adoption – while Lederberg (1977) focuses on the phenomenon of obliteration by incorporation in publication citations. Together, these examples suggest that our identified mechanisms – and, possibly, our empirical results – may be broadly applicable to an even wider set of diffusion and emergence studies than we originally theorized; however, future research is necessary to test these suggestions directly.

6. Conclusion

Although innovation is widely celebrated as a driver of economic growth and firm competitiveness, measuring innovation remains a difficult challenge. This study contributes to a growing body of work on "measurement" by identifying four mechanisms that can account for divergence – e.g., both overstatement and understatement – between popular innovation measures and "actual" innovation patterns. Through an in-depth exploration of the emergence and diffusion of "green chemistry," we then empirically demonstrate that the common measurement approaches of keywords and index terms can dramatically misstate the magnitude, timing, organizational demography, and collaborative basis of diffusion patterns, and we highlight error rates approaching 50% for some organization types (e.g., firms) and for research conducted in certain countries (e.g., the United States). Given the centrality of these variables across a wide array of innovation studies, our study, therefore, raises serious concerns about the empirical basis of some innovation research. Indeed, scholars have long noted the interplay between theory, data, and policy: data are the basis on which we craft and test our theories, and data are the basis on which we design and assess our policies. In turn, demonstrations of extreme data biases and weaknesses – as we demonstrate here – suggest that both theories and policies may be built upon a shaky foundation.

Our intention in presenting this research, however, is not to call into question previous efforts. In fact, our research design does not permit us to comment on the veracity of *any* particular study. Instead, by pointing to potential challenges – and to potential solutions – our hope is to mobilize innovation researchers to address measurement issues. Given the importance of innovation, and the central role that data play in shaping our assessment of it, further research into measurement issues is sure to bear much fruit.

Acknowledgements

For comments, suggestions and insights, we thank Anne Parmigiani, Stine Grodal, and participants in the 2011 West Coast Research Symposium on Technology Entrepreneurship and the 2012 Academy of Management Annual Conference. The T.J. Meyer Fund for Sustainability and the Environment supported this research. A shorter and preliminary version of this paper appeared in the 2012 Academy of Management Best Paper Proceedings.

Appendix 1. The 12 principles of green chemistry

1. It is better to prevent waste than to treat or clean up waste after it is formed.
2. Synthetic methods should be designed to maximize the incorporation of all materials used in the process into the final product.
3. Wherever practicable, synthetic methodologies should be designed to use and generate substances that possess little or no toxicity to human health and the environment.
4. Chemical products should be designed to preserve efficacy of function while reducing toxicity.
5. The use of auxiliary substances (e.g., solvents, separation agents, etc.) should be made unnecessary wherever possible and innocuous when used.
6. Energy requirements should be recognized for their environmental and economic impacts and should be minimized. Synthetic methods should be conducted at ambient temperature and pressure.
7. A raw material of feedstock should be renewable rather than depleting wherever technically and economically practicable.
8. Unnecessary derivatization (blocking group, protection/deprotection, temporary modification of physical/chemical processes) should be avoided whenever possible.
9. Catalytic reagents (as selective as possible) are superior to stoichiometric reagents.
10. Chemical products should be designed so that at the end of their function they do not persist in the environment and break down into innocuous degradation products.
11. Analytical methodologies need to be further developed to allow for real-time, in process monitoring and control prior to the formation of hazardous substances.
12. Substances and the form of a substance used in a chemical process should be chosen so as to minimize the potential for chemical accidents, including releases, explosions, and fires.

Source: Anastas and Warner (1998).

Appendix 2. Initial keyword list for expert assessment dataset

Green Chemistry
Benign by Design
Sustainability Metrics*
Benign Chemistry
Benign Synthesis
Environmental Pollution Control*
Protection of Environment*
Technology of Sustainable Environment*
Green Technology*
Clean Technology*
(* = refined by the term "chemistry")

References

Abrahamson, E., Rosenkopf, L., 1993. Institutional and competitive bandwagons: using mathematical modeling as a tool to explore innovation diffusion. Academy of Management Review 18, 487–517.
Abrahamson, E., Fairchild, G., 1999. Management fashion: lifecycles, triggers, and collective learning processes. Administrative Science Quarterly 44 (4), 708–740.
Achilladelis, B., Schwarzkopf, A., Cines, M., 1990. The dynamics of technological innovation: the case of the chemical industry. Research Policy 19 (1), 1–34.
Aguinis, H., Edwards, J.R., 2014. Methodological wishes for the next decade and how to make wishes come true. Journal of Management Studies 51 (1), 143–174.
Anastas, P.T., Kirchhoff, M.M., 2002. Origins, current status, and future challenges of green chemistry. Accounts of Chemical Research 35 (9), 686–694.
Anastas, P.T., Warner, J.C., 1998. Green Chemistry: Theory and Practice. Oxford University Press, New York.
Arora, A., 1997. Patents, licensing, and market structure in the chemical industry. Research Policy 26 (4), 391–403.
Arora, A., Landau, R., Rosenberg, N., 1998. Chemicals and Long Term Economic Growth. John Wiley &Sons, New York.
Barley, S.R., Kunda, G., 2001. Bringing work back in. Organization Science 12 (1), 76–95.
Beaudry, C., Allaoui, S., 2012. Impact of public and private research funding on scientific production: the case of nanotechnology. Research Policy 41 (9), 1589–1606.
Boyd, B.K., Bergh, D.D., Ireland, R.D., Ketchen, D.J., 2013. Constructs in strategic management. Organizational Research Methods 16, 3–14.
Bromley, P., Powell, W.W., 2012. From smoke and mirrors to walking the talk: Decoupling in the contemporary world. The Academy of Management Annals 6 (1), 483–530.
Carpenter, V.L., Feroz, E.H., 2001. Institutional theory and accounting rule choice: an analysis of four US state governments' decisions to adopt generally accepted accounting principles. Accounting, Organizations and Society 26, 565–596.
Casile, M., Davis-Blake, A., 2002. When accreditation standards change: factors affecting differential responsiveness of public and private organizations. Academy of Management Journal 45 (1), 180–195.
Chatterji, A.K., Levine, D.I., Toffel, M.W., 2009. How well do social ratings actually measure corporate social responsibility? Journal of Economics and Management Strategy 18 (1), 125–169.

Chaves, M., 1996. Ordaining women: the diffusion of an organizational innovation. American Journal of Sociology 101 (4), 840–873.
Cockburn, I., Henderson, R., 1998. Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. Journal of Industrial Economics 46 (2), 157–182.
Coleman, J., Katz, E., Menzel, H., 1957. The diffusion of an innovation among physicians. Sociometry 20 (4), 253–270.
Colyvas, J., Jonsson, S., 2011. Ubiquity and legitimacy: disentangling diffusion and institutionalization. Sociological Theory 29 (2), 27–53.
Colyvas, J.A., Powell, W.W., 2009. Measures, metrics, and myopia: the challenges and ramifications of sustaining academic entrepreneurship. Advances in the Study of Entrepreneurship, Innovation and Economic Growth 19, 79–111.
Creed, W.D., Scully, M.A., Austin, J.R., 2002. Clothes make the person? The tailoring of legitimating accounts and the social construction of identity. Organization Science 13 (5), 475–496.
Damanpour, F., 1991. Organizational innovation: a meta-analysis of effects of determinants and moderators. Academy of Management Journal 34 (3), 555–590.
David, P.A., 1975. Technical Choice, Innovation, and Economic Growth. Cambridge University Press, Cambridge.
David, P.A., 1986. Technology diffusion, public policy, and industrial competitiveness. In: Landau, R., Rosenberg, N. (Eds.), The Positive Sum Strategy: Harnessing Technology for Economic Growth. National Academy Press, Washington, DC, pp. 373–391.
David, R.J., Strang, D., 2006. When fashion is fleeting: transitory collective beliefs and the dynamics of TQM consulting. Academy of Management Journal 49 (2), 215–233.
Economou, I.G., Karatsani, E.K., Logotheti, G.E., Ramos, J., Vanin, A., 2008. Multi-scale modeling of structure, dynamic and thermodynamic properties of imidazolium-based ionic liquids: ab initio DFT calculations, molecular simulation and equation of state predictions. Oil and Gas Science and Technology 63 (3), 283–293.
Elsbach, K.D., Sutton, R.I., 1992. Acquiring organizational legitimacy through illegitimate actions: a marriage of institutional and impression management theories. Academy of Management Journal 35 (4), 699–738.
Fiss, P.C., 2008. Institutions and corporate governance. In: Greenwood, R., Oliver, C., Sahlin, K., Suddaby, R. (Eds.), The Sage Handbook of Organizational Institutionalism. Sage Publications, London, UK, pp. 389–410.
Fiss, P.C., Zajac, E.J., 2004. The diffusion of ideas over contested terrain: the (non)adoption of a shareholder value orientation among German firms. Administrative Science Quarterly 49 (4), 501–534.
Fiss, P.C., Zajac, E.J., 2006. The symbolic management of strategic change: sensegiving via framing and decoupling. Academy of Management Journal 49 (6), 1173–1193.
Freeman, C., Soete, L., 2009. Developing science, technology and innovation indicators: what we can learn from the past. Research Policy 38 (4), 583–589.
Gay, B., Dousset, B., 2005. Innovation and network structural dynamics: study of the alliance network of a major sector of the biotechnology industry. Research Policy 34 (10), 1457–1475.
Geroski, P.A., 2000. Models of technology diffusion. Research Policy 29 (4), 603–625.
Giroux, H., 2006. It was such a handy term: management fashions and pragmatic ambiguity. Journal of Management Studies 43 (6), 1227–1260.
Granqvist, N., Grodal, S., Woolley, J., 2013. Hedging your bets: explaining executives' market labeling strategies in nanotechnology. Organization Science 24 (2), 395–413.
Greve, H.R., 1995. Jumping ship: the diffusion of strategy abandonment. Administrative Science Quarterly 40 (3), 444–473.
Greve, H.R., 1996. Patterns of competition: the diffusion of a market position in radio broadcasting. Administrative Science Quarterly 41 (1), 29–60.
Griliches, Z., 1957. Hybrid corn: an exploration in the economics of technological change. Econometrica 25, 501–522.
Hagedoorn, J., Cloodt, M., 2003. Measuring innovative performance: is there an advantage in using multiple indicators? Research Policy 32 (8), 1365–1379.
Hendricks, K.B., Singhal, V.R., 1997. Does implementing an effective TQM program actually improve operating performance? Empirical evidence from firms that have won quality awards. Management Science 43 (9), 1258–1274.
Hu, A.G., Jaffe, A.B., 2003. Patent citations and international knowledge flow: the cases of Korea and Taiwan. International Journal of Industrial Organization 21 (6), 849–880.
Hudson, B., Okhuysen, G., 2009. Not with a 10-foot pole: core stigma, stigma transfer, and improbable persistence of men's bathhouses. Organization Science 20 (1), 134–153.
Islam, N., Miyazaki, K., 2010. An empirical analysis of nanotechnology research domains. Technovation 30 (4), 229–237.
Jaffe, A.B., 1989. Real effects of academic research. American Economic Review 79 (5), 957–970.
Jamasb, T., Pollitt, M.G., 2011. Electricity sector liberalisation and innovation: an analysis of the UK's patenting activities. Research Policy 40 (2), 309–324.
Järvenpää, H.M., Mäkinen, S.J., Seppänen, M., 2011. Patent and publishing activity sequence over a technology's life cycle. Technological Forecasting and Social Change 78 (2), 283–293.
Jonsson, S., 2009. Refraining from imitation: professional resistance and limited diffusion in a financial market. Organization Science 20 (1), 172–186.
Kalev, A., Shenhav, Y., De Vries, D., 2008. The state, the labor process, and the diffusion of managerial models. Administrative Science Quarterly 53 (1), 1–28.
Kaplan, S., Radin, J., 2011. Bounding an emerging technology: para-scientific media and the Drexler–Smalley debate about nanotechnology. Social Studies of Science 41 (4), 457–486.
Kapur, A., Keoleian, G., Kendall, A., Kesler, S., 2008. Dynamic modeling of in-use cement stocks in the United States. Journal of Industrial Ecology 12 (4), 539–556.
Katz, E., 1999. Theorizing diffusion: Tarde and Sorokin revisited. Annals of the American Academy of Political and Social Science 566 (1), 144–155.
Kennedy, M., Lo, J., Lounsbury, M., 2010. Category currency: the changing value of conformity as a function of ongoing meaning construction. Research in the Sociology of Organizations 31, 369–397.
Kleinknecht, A., Van Montfort, K., Brouwer, E., 2002. The non-trivial choice between innovation indicators. Economics of Innovation and New Technology 11, 109–121.
Lampe, R., 2012. Strategic citation. Review of Economics and Statistics 94 (1), 320–333.
Lane, J., 2010. Let's make science metrics more scientific. Nature 464, 488.
Lederberg, J., 1977. Forward. In: Garfield, E. (Ed.), Essays of an Information Scientist. ISI Press, Philadelphia, PA.
Lee, B.H., 2009. The infrastructure of collective action and policy content diffusion in the organic food industry. Academy of Management Journal 52 (6), 1247–1269.
Lee, J., Veloso, F.M., Hounshell, D.A., 2011. Linking induced technological change, and environmental regulation: evidence from patenting in the US auto industry. Research Policy 40 (9), 1240–1252.
Linthorst, J., 2010. An overview: origins and development of green chemistry. Foundations of Chemistry 12 (1), 55–68.
MacRoberts, M.H., MacRoberts, B.R., 1986. Problems of citation analysis: a critical review. Journal of the American Society for Information Science 40 (5), 342–349.
Martin, J.A., 2011. Dynamic managerial capabilities and the multibusiness team: the role of episodic teams in executive leadership groups. Organization Science 22 (1), 118–140.
Merton, R.K., 1968. Social Theory and Social Structure. Free Press, New York.
Merton, R.K., 1979. Foreword. In: Garfield, E. (Ed.), Citation Indexing: Its Theory and Application in Science, Technology, and Humanities. ISI Press, Philadelphia, PA.
Mina, A., Ramlogan, R., Tampubolon, G., Metcalfe, J.S., 2007. Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. Research Policy 36 (5), 789–806.
Mogoutov, A., Kahane, B., 2007. Data search strategy for science and technology emergence: a scalable and evolutionary query for nanotechnology tracking. Research Policy 36 (6), 893–903.
Nameroff, T.J., Garant, R.J., Albert, M.B., 2004. Adoption of green chemistry: an analysis based on US patents. Research Policy 33 (6), 959–974.
Nelson, A.J., 2009. Measuring knowledge spillovers: what patents, licenses and publications reveal about innovation diffusion. Research Policy 38 (6), 994–1005.
Nelson, A.J., 2012. Putting university research in context: assessing alternative measures of production and diffusion at Stanford. Research Policy 41 (4), 678–691.
Oliver, C., 1991. Strategic responses to institutional processes. Academy of Management Review 16 (1), 145–179.
Powell, W.W., DiMaggio, P., 1991. The New Institutionalism in Organizational Analysis. University of Chicago Press, Chicago, IL.
Powell, W.W., Koput, K.W., Smith-Doerr, L., 1996. Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology. Administrative Science Quarterly 41 (1), 116–145.
Rogers, E., 2003. Diffusion of Innovation, 5th ed. Simon and Schuster, New York.
Rosenberg, N., 1982. Inside the Black Box. Cambridge University Press, Cambridge.
Rosenberg, N., 1986. Factors affecting the diffusion of technology. Explorations in Economic History 10 (1), 3–33.
Saloner, G., Shepard, A., 1995. Adoption of technologies with network effects: an empirical examination of the adoption of automated teller machines. Rand Journal of Economics 26 (3), 479–501.
Scott, W.R., 2000. Institutions and Organizations, 2nd ed. Sage Publications, Thousand Oaks, CA.
Stephan, D.G., Atcheson, J., 1989. The EPA's approach to pollution prevention. Chemical Engineering Progress 85 (6), 53–58.
Stoneman, P., Diederen, P., 1994. Technology diffusion and public policy. Economic Journal 104 (425), 918–930.
Strang, D., Soule, S.A., 1998. Diffusion in organizations and social movements: from hybrid corn to poison pills. Annual Review of Sociology 24, 265–290.
Terlaak, A., Gong, Y., 2008. Vicarious learning and inferential accuracy in adoption processes. Academy of Management Review 33 (4), 848–868.
Tolbert, P.S., Zucker, L.G., 1983. Institutional sources of change in the formal structure of organizations: the diffusion of civil service reform, 1880–1935. Administrative Science Quarterly 28 (1), 22–39.
Walsh, V., 1984. Invention and innovation in the chemical industry: demand-pull or discovery-push? Research Policy 13 (4), 211–234.
Wang, T., Bansal, P., 2012. Social responsibility in new ventures: profiting from a long-term orientation. Strategic Management Journal 33 (10), 1135–1153.
Westphal, J.D., Gulati, R., Shortell, S.M., 1997. Customization or conformity? An institutional and network perspective on the content and consequences of TQM adoption. Administrative Science Quarterly 42 (2), 366–394.
Westphal, J.D., Zajac, E.J., 1994. Substance and symbolism in CEOs' long-term incentive plans. Administrative Science Quarterly 39 (3), 367–390.
Woodhouse, E.J., Breyman, S., 2005. Green chemistry as social movement? Science, Technology and Human Values 30 (2), 199–222.
Wu, C.Y., Mathews, J.A., 2012. Knowledge flows in the solar photovoltaic industry: insights from patenting by Taiwan, Korea, and China. Research Policy 41 (3), 524–540.

Xie, Z., Miyazaki, K., 2013. Evaluating the effectiveness of keyword search strategy for patent identification. World Patent Information 35, 20–30.

Zaheer, S., Mosakowski, E., 1997. The dynamics of the liability of foreignness: a global study of survival in financial services. Strategic Management Journal 18 (6), 439–463.

Zbaracki, M.J., 1998. The rhetoric and reality of total quality management. Administrative Science Quarterly 43 (3), 602–636.

Zuckerman, E.W., 1999. The categorical imperative: securities analysts and the illegitimacy discount. American Journal of Sociology 104 (5), 1398–1438.