

Concept innovation in the software industry: 1990 - 2002

Abstract

This study investigates *concept innovation*, when firms create new market category labels to differentiate their products. Data suggest that concept innovation is frequent. But little is known about its antecedents. This study proposes that concept innovation is based both on recombinant processes and on constraints from existing classification. Firms that combine elements across market categories are more likely to engage in concept innovation – when categories are constraining. But when a firm’s categories are lenient, the relationship weakens and leniency leads to concept innovation as firms attempt to resolve ambiguity. Hypotheses are tested using three measures of combination, and results support hypotheses for all three measures in a longitudinal analysis of 4,566 firms and 456 market categories in the software industry between 1990 and 2002.

A central question in strategic management concerns how novel ideas emerge. Novelty is compelling: new gadgets, fashions, or technologies tempt people who are eager for something different. But there is a tension between whether it is best for organizations to be more similar to or different from their peers. Similarity makes organizations recognizable and legitimate, while differentiation provides a buffer from competition (Deephouse 1999; Porter 1980; Hannan and Freeman 1977; DiMaggio and Powell 1983). Firms with novel offerings are the extreme on this continuum: often so differentiated that they are categorically distinct. These are the first movers into – or perhaps creators of – new market categories.

Little is known about what spurs the introduction of new market categories into a domain. This is striking given the strategic importance of staking out a position in an industry. Classic strategy research emphasizes competitive positioning in mature markets: which to enter and how to differentiate (Porter 1980). The literature on market categorization studies mature market categories and emphasizes the value of conformity (Hsu 2006; Hsu, Hannan and Koçak 2009; Negro, Hannan and Rao 2010). But market positioning is not just about choosing to enter an existing market and whether to conform or differentiate within it. There is also the option for a firm to try and pioneer a new category. To do this, a firm can introduce a label, a “tag” that is used to communicate about a new concept. I call this *concept innovation*.

With concept innovation, producers signal that they are so differentiated that they should be evaluated by new categorical standards. Once a concept label is introduced, if it diffuses, it may become a new market category, leading to changes in the market’s structure. Both market evolution and the trade-off between conformity and differentiation are important topics in strategy, and changes to classification systems are fundamental to research on categories.

Concept innovation is central to these lines of inquiry. Despite this, scholars have not studied the antecedents of concept innovation.

Previous research on market emergence looks at how boundaries, standards, and definitions are set. One stream describes this process as primarily technological (Abernathy and Utterback 1978; Utterback and Suarez 1993). But cognitive factors are also important. Because new products are unfamiliar, people form new concepts in order to understand and evaluate them (Clark 1985; Kaplan and Tripsas 2008). Firms can introduce a product as the basis for an emerging market category, but this is not the end of the story. How people make sense of new offerings determines how an industry evolves (Suarez, Grodal and Gotsopoulos 2014). A number of studies focus on this sense making process in terms of how people define boundaries and create shared meanings for early-stage market categories (Rosa, Porac, Runser-Spanjol and Saxon 1999; Kennedy 2008; Ruef and Patterson 2009; Rindova, Ferrier and Wiltbank 2010; Navis and Glynn 2010; Bingham and Kahl 2013). But these studies do not investigate when market categories are first pioneered.

Data from the empirical context of this study show that concept innovation is frequent in the software industry during the 1990s. Software firms introduced 361 new market labels between 1990 and 2002. Some of these caught on and became established market categories; others petered out (see figure 1). This presents a much more dynamic process of market category evolution than is assumed in studies of competitive strategy, organizational categorization, and entrepreneurship. And it indicates that producers are on the forefront of this process.

Through concept innovation, producers frame their offerings using a new concept label, which they hope becomes the basis for a new market category. Cognitively, categories facilitate

perception and learning (Park and Hastie 1987; Murphy 2004; Tenenbaum, Griffiths and Kemp 2006). Market categories affect how people evaluate a product (Porac and Thomas 1990; Perretti, Negro and Lomi 2008; Hsu, Hannan and Koçak 2009; Smith 2011; Leung and Sharkey 2013), and labels are used to communicate about concepts and categories (Murphy 2004; Lupyan, Rakison and McClelland 2007; Burshell and Mitchell 2016; Hannan et al. 2017). Given this, it should come as no surprise that some producers engage in concept innovation, using labels to frame their offerings as new and different, seeking a favorable position in the market (Santos and Eisenhardt 2009; Pontikes and Barnett 2015).

This study proposes that whether a firm engages in concept innovation depends on both whether its offerings combine categorical elements and on existing classification. It draws on two research streams: one that asserts that novelty comes from new combinations, the other that takes a structural view and proposes that existing categories constrain new category emergence.

In the tradition of Schumpeter, the combinatory view asserts that the ultimate source of novelty is when existing elements are brought together in new ways. Schumpeter (1934) proposed that economic development was based on the “carrying out of new combinations” (p. 66). For him, new combinations fundamentally change the structure of the market; they underlie what he later terms “Creative Destruction ... the essential fact about capitalism,” (1942, p. 83). Scholars across a variety of areas have asserted (and found) support for his main idea associating novelty with combinations. But researchers have not systematically investigated whether combinations lead to the types of qualitative changes in the market that he describes. Technology researchers find that inventions based on combinations have high citation impact, but they do not directly investigate the introduction of discontinuous, novel developments (Fleming 2001; Rosenkopf and Nerkar 2001; Nerkar 2003). Research in entrepreneurship, strategy, and

organization theory provides evidence of the link between combinations and new products, markets, or organizational forms, but these are typically ex-post studies of the historical roots of a category that has succeeded (DiMaggio 1991; Utterback and Suarez 1993; Stark 1996; Hargadon and Sutton 1997; Powell and Sandholtz 2012).

This study directly investigates antecedents of concept innovation, a strategic attempt by a firm to make a discontinuous change in the market structure. Drawing on the Schumpeterian tradition, it proposes that firms that create products, processes, technologies, or other developments that bring together elements from different market categories are likely to engage in concept innovation. But the combination story is not the complete picture. In the structural perspective, category emergence is not simply defined by a firm's outputs. Rather, the institutional environment affects how offerings are understood (Rao 1998; Santos and Eisenhardt 2009; Navis and Glynn 2010). This study notes that many market categories are *lenient*: they have porous boundaries and exert few constraints on member organizations. It argues that category leniency affects whether producers engage in concept innovation, in two ways. First, when categories are lenient they are less useful, and so firms engage in concept innovation to clarify their categorical identities, regardless of how unique their offerings are. Second, leniency weakens the relationship between new combinations and concept innovation. When categories are lenient, new combinations are not categorical misfits and so can credibly be part of the existing category.

This means that it is not only new combinations that prompt concept innovation – and thus the seeds for market evolution. Rather, to understand concept innovation, we must take into account a firm's existing classification. And when we do this, it turns out categorical constraint fosters novelty. At first blush, this may seem at odds with research in new institutionalism, which

emphasizes that constraints promote conformity (DiMaggio and Powell 1983). This may lead to lay hypotheses that constraining categories stifle innovation. This study proposes the opposite: constraints facilitate concept innovation because they establish clear boundaries against which differentiated firms can set themselves apart.

Concept Innovation

Concept innovation is an important strategic tactic, especially in domains that prize novelty and innovation (PlayBigger 2015). Through concept innovation, producers can signal that their offerings are unique and should be evaluated by different categorical standards. Firms engage in concept innovation either by coming up with an entirely new market category label, or by claiming a sparsely used label that has not yet been socially defined.¹

In defining *concept innovation*, I draw on research that differentiates *concepts* from *categories*. Concepts are mental representations (Murphy 2004). A category is the set of objects to which a concept is applied (Medin and Rips 2005; Hannan et al, 2017). People communicate about concepts and categories using labels (Neisser 1989). In this study, the unit of analysis is a firm. With concept innovation, people within the firm collectively develop a new concept, and, on behalf of the firm, introduce a label that refers to the concept. If other actors (other firms, analysts, customers) adopt the concept, start applying the label to other firms, and come to a consensus on its definition, the concept associated with the label evolves into a socially meaningful market category.

¹ In interviews, a number of entrepreneurs and executives referenced instances where others began using the same label around the same time. In some instances, the firm knowingly adopted a label used by another firm because they liked the term and sought to shape its definition; in another case firms engaged in different activities adopted the label by coincidence. This study treats both the coining of entirely new term, or adoption of a very early-stage label, as concept innovation.

Exploratory interviews with executives and venture capitalists show that firms are actively engaged in concept innovation. Producers not only create innovative *products* but also create labels for new *concepts* that describe their offerings. Both are essential to success. One entrepreneur interviewed describes attempts to influence analysts to start reporting on a label they pioneered:

[Our CEO] went and had coffee with [Gartner and Forrester analysts] and said, ‘OK, here’s the deal, you guys are not covering this industry, [*web management*].’ And [the analyst said] ‘wait, what’s [*web management*]?’ And [our CEO] said, ‘well, these companies: [our company] and two or three [others]. And, to be clear, [*web management*] wasn’t super innovative, but it was labeling it, and it was specifically labeling it ... when we felt like [*web management*] was what we wanted to call it. So it was strategic the way we were thinking about it.’²

Another executive highlights the importance of coining labels to describe a firm’s offerings:

When you’re bringing something new into an industry, you need to find some lingo. Like when we first started ... I had to go educate investors and prospects on what this kind of outsourced model was. Because this terminology of ‘SaaS’ or ‘ASP’ didn’t exist. And so, I’d related it to down terminals or mainframes or client-server, all these things that hopefully would catch their interest. Like, ‘oh, yeah, I have seen that in the past.’ And then when that term ‘ASP’ came along, man, I grabbed onto that as fast as I could. And when that term ‘SaaS’ came along, I grabbed onto that as fast as I could. ... Every industry will eventually generate a tribal language. And that tribal language makes things a lot easier to explain. You know, [you] spend 10 minutes explaining the client-server model to get someone to try and understand ASP because the term doesn’t exist.

The reason new labels make things “a lot easier to explain” is that the label can evolve into a market category, where a consensus develops around the meaning of the term. When categories are first introduced, their meanings are “shallow,” but can deepen over time (Grodal, Gostopolous, and Suarez 2015). Collective schemas emerge as people communicate about the label, selecting analogies that define the category within the broader context (Bingham and Kahl 2013). Then, the category might diffuse. The more that similar peers claim a new label, the more likely an organization will also become an early claimant (Burshell and Mitchell 2016). Category definitions informally emerge as people infer category meanings based on the characteristics of affiliated members (Pontikes and Hannan 2014).

² The category label has been ghosted to preserve anonymity.

The processes by which a label evolves into a meaningful social category have been well studied (usually in a longitudinal analysis of a single market). But less attention has been paid to the earliest stages where firms affiliate with new labels to differentiate, signal novelty, and possibly plant the seed for a new market category. Little is known about what prompts firms to engage in this strategic tactic.

Perhaps one reason why concept innovation has not been studied is a lot of research on market categories considers intermediaries to be authors of industry classification (Zuckerman 1999; Hsu, Hannan and Koçak 2009; Ruef and Patterson 2009; Negro, Hannan and Rao 2010; Leung and Sharkey 2013). Often, the implication is that analyst categories reflect technical or “natural” divisions in the market. More recently, researchers have noted that even analyst categories are strategically defined, and that analysts adopt categories pioneered by producers (Pontikes and Kim 2017). This is highlighted in a quote from an ethnography of Gartner, the preeminent high technology analyst firm, where an analyst lets slip that one of the market categories they claim to have coined was actually introduced by a firm:

So we said [the name for the new category] is: ‘something Feedback Management’. And we noticed that there is a company up in Boston [who] started to use the term ‘EFM’—Enterprise Feedback Management—and we went ‘that’s the term we like’. So we basically stole it and started using it (Pollock and Williams 2011).

Data from press releases suggest that this situation is not an anomaly. Comparing data on market category labels from firm press releases to reports issued by Gartner shows that usually firms are first to introduce a category label. Of the market categories in both press releases and Gartner reports, 75% first appear in press releases, and only 9% first appear in Gartner reports (the remaining 16% appear in the same year).³ And concept innovation is frequent – in the empirical context for this study, software producers introduced 361 new market category labels

³ Gartner reports are available in electronic format from 1995 onward, this comparison is for 1995 – 2002.

between 1990 and 2002. Producers are on the forefront of shaping how people conceptualize new offerings, and their strategic behavior introducing labels affects how market categories evolve.

Figure 1 illustrates concept innovation in software industry classification, presenting network maps of market category labels claimed by producers, for selected years. Nodes are labels, and edges link two labels based on the number of organizations that are in both. Dark nodes indicate concept innovation: category labels that are first mentioned in the press release corpus in that year.⁴ These plots show the interconnected market category landscape against which firms assess their offerings as they decide whether to attempt to carve out a new area in the space. As I will argue below, one consideration around whether to engage in concept innovation is how much a firm is combining elements of categories in their offerings. But, I suggest another consideration takes into account this categorical context. These plots also show varying success of concept innovation. Some new labels, like “data mart,” and “fraud detection” have become stable market categories, while labels like “internet commerce” became too broad to be useful, and others, like “color management,” were specific, ending up as a feature but not a market category.

--- Figure 1 about here ---

For a firm that is trying to do something different, they do not simply create a novel product and try to sell it. They also present a concept that conveys what the product does. Before a customer can want to purchase something, she needs to understand *what it is* she is buying. This presents difficulties for firms that have unique offerings. They can position their offerings as part of an established category and try to convey that they provide a “better version.” But this

⁴ These maps only include the main component, so unconnected categories that only one firm is in are not represented. Therefore, these figures present a subset of concept innovation.

can backfire if customers see the firm's offering as non-standard and thus unnecessary.⁵ Concept innovation is an alternative.

Like any strategic tactic, there are risks associated with concept innovation. If all goes well, the nascent label will evolve into an established market category with stable demand. But, as with any novel development, the success of a new concept label is difficult to predict ahead of time (March 1991). If the label does not catch on and become a market category, the firm may find itself in an unrecognized market with no demand. But if the label becomes a popular market category defined around a firm's offerings, the firm will be well positioned to compete.

Cognitive underpinnings of concept innovation: the role of labels

Labels are central to the definition of concept innovation. One might question whether firms can pioneer new concepts and categories without introducing labels. Research in cognitive science indicates that it would be difficult. People use labels to communicate about abstract concepts (Vygotsky [1934] 2012; Neisser 1989). A unique label indicates that a set of objects can be treated equivalently (for some purposes). In laboratory studies, people perceive objects as more similar when the same label is applied to them, and more different if a different label is applied (Park and Hastie 1987; Sloutsky, Lo and Fisher 2001; Winawer et al. 2007; Lara, Hahn, Yu and Yamauchi 2012). Cognitive research also shows that labels are not simply tags that are attached to concepts and categories, but are important for cognition (Lupyan 2012). In laboratory studies,

⁵ Pollock and Williams (2011) describe a firm that was looking for a new "Customer relationship management (CRM)" product, who were initially interested in an innovative firm, Pico. They asked Gartner to recommend a set of solutions, and Gartner would not recognize Pico as a CRM firm. The buying firm ended up going with one of Gartner's recommendations. To some extent, Pico agreed with Gartner's position – they argued that Gartner did not have a category for their offering. They thought their solution was better. Gartner saw it as substandard. Pico failed to do the work of concept innovation to help customers (and analysts) understand *what they were*.

labels prompt people to think categorically (Gelman and Heyman 1999) and even alter mental representations (Lupyan 2008a; 2008b).

With concept innovation, producers introduce a new label to represent a new type, similar to what researchers do in laboratory experiments. Just as participants in the lab begin to form a new concept associated with a label, producers hope concept innovation triggers customers, investors, analysts, and other firms to start to form a new concept for their offering. Affiliating with a new label indicates that the producer believes their offering should be judged by new categorical standards. To succeed at pioneering a new category, more has to happen. Subsequent processes of meaning construction are well researched (Rosa, Porac, Runser-Spanjol and Saxon 1999; Colyvas 2007; Kennedy 2008; Navis and Glynn 2010; Granqvist, Grodal and Woolley 2013; Bingham and Kahl 2013). But before any of this takes place, a new concept label must be introduced.⁶ Researchers have not studied the initial underpinnings for concept innovation. Below, I explore combinatorial and contextual factors that prompt producers to engage in this tactic.

Novelty as New Combinations

In his groundbreaking work, Schumpeter proposes a theory of economic development in which “the carrying out of new combinations” (1934, p. 66) leads to “discontinuous change” in the economic structure that “forever alters and displaces the equilibrium state previously existing” (p. 64). His ideas inspired a rich literature about what leads to science and engineering inventions

⁶ When market categories are forming, often more than one label is used for offerings with similar functionality. Part of the process of category emergence is people coming to consensus on a label for a category (Grodal, Gostopoulou, and Suarez 2015). But even if different terms refer to similar activities, psychological research shows that the fact that a different label is used is significant because it cues people to look for differences. My interviews suggest that firm executives have an intuitive understanding of this and deliberately use new labels to differentiate their products. Other studies have investigated the important issues of coming to a social consensus the meanings of labels and categories. The objective of this study is to understand what prompts firms to introduce new labels in the first place.

(Fleming 2001; Nerkar 2003; Gruber, Harhoff and Hoisl 2013). But his writings indicate that, for him, innovation includes more than technological advances (Nelson and Winter 1982, p. 277). He defines five cases of development “(1) the introduction of a new good ... (2) ... a new method of production... (3) the opening of a new market, (4) ... a new source of supply of raw materials ... (5) the carrying out of the new organization of any industry” (Schumpeter 1934, p. 66). In Schumpeter’s account, novel development is a qualitative, discontinuous change in the market: categorical in nature. Schumpeter does not write about concepts and categories, but his descriptions of economic development are what producers aim to do when they engage in concept innovation. This suggests that concept innovation might arise from producers who develop “new combinations.”

A rich literature applies Schumpeter’s ideas to technological development, and shows empirical support for the link between new combinations and high impact inventions. Databases of patents and journal articles record citations, allowing for ex-ante studies that compare “radical” inventions based on new combinations to “incremental” ones that draw from the same knowledge pool. These studies investigate whether an invention or article is highly cited, under the assumption that high-impact technologies become the basis for major changes in the economic structure (Trajtenberg 1990). Research finds support for the idea that new combinations are the basis for high-impact developments, and also underscores the inherent uncertainty in how novel scientific and engineering developments are received. Patents and papers with distant or unusual citations have especially high or low citations (Fleming 2001; Uzzi, Mukherjee, Stringer and Jones, 2013; Foster, Rzhetsky, and Evans, 2015; Leahey, Beckman and Stanko, 2016).

Studies in technology development provide cohesive, systematic evidence that new combinations are a primary source of high-impact technologies. But it is a narrow assumption that technology advances capture the range of changes in the economy about which Schumpeter theorizes. Nelson and Winter (1982) acknowledge this. For them, new combinations are the basis for any type of novelty: “The creation of any sort of novelty in art, science, or practical life - consists to a substantial extent of a recombination of conceptual and physical materials that were previously in existence,” (p. 130).⁷ Studies in sociology, entrepreneurship, and strategy show that novelty, more broadly defined, has its basis in combination: market actors blend capitalist and socialist elements to create a new type of property ownership in post-Soviet Hungary (Stark 1996), musicians combine rare instruments to create novel music in jazz (Phillips 2013), and business groups bring together diverse and familiar knowledge to generate disruptive innovations (Vedres and Stark 2011).

In the technology literature the outcome studied is performance variability. Studies in this tradition do not directly investigate whether new combinations lead to an invention that is a discontinuous, qualitative change. Yet, this distinction is central to foundational research on market evolution (Schumpeter 1934; 1942; Nelson and Winter 1982; Tushman and Anderson 1986). Studies that do directly investigate the creation of new “types” usually investigate the historical roots of a successful category. These provide additional evidence for the view that novelty is based in combinatory processes, for example in the emergence of new organizational forms in non-market settings (DiMaggio 1991; Rao 1998; Powell and Sandholtz 2012), or in the

⁷ Nelson and Winter (1982) mention that Schumpeter’s treatment of innovation is broader than technological progress alone, and caveat against “overdrawing ... distinctions between technological and organizational considerations” in studying innovation, but their model focuses on the narrower issue of technological progress (pp. 277-278).

emergence of new products in markets (Utterback and Suarez 1993; Hargadon and Sutton 1997), but cannot speak to whether there is a systematic pattern.

There is a great deal of evidence from a wide range of literatures supporting the link between new combinations and novel developments. These studies are of two main types: (1) studies in technology that take an ex-ante approach to measuring new combinations, but do not directly look at the initiation of a discontinuous change, and (2) studies that investigate a discrete change but take an ex-post approach. Ex-ante studies that systematically investigate what leads to the introduction of a novel type are rare. In a notable exception, Ruef (2000) investigates the emergence of new forms in the health care industry, in a community ecology analysis of 48 organizational forms, using regulatory events to time new form emergence. But even this study does not capture the earliest stages of concept innovation. By the time the government recognizes a category, it likely has been part of the social fabric for some time.

Addressing this gap, this study investigates concept innovation. Affiliating with a new label is a discreet and discontinuous act that indicates a producer is attempting to pioneer a new market category. In the empirical analysis, antecedents of concept innovation can be systematically studied using unique data that track market categorization for hundreds of labels and thousands of firms over more than a decade.

Drawing on the above literatures, I propose that when firms combine elements across market categories they are especially likely to engage in concept innovation. A development that combines elements is not similar to any one category, and the less similar something is to existing categories, the more likely producers will introduce a new label to describe it. There are a number of ways firms can create new combinations. In technology, engineers combine sand and aluminum to create semiconductors (Fleming 2001). In cultural markets, chefs combine

ingredients across different cuisines (Rao, Monin, and Durand 2003), equipment across different vintner styles (Negro, Hannan and Rao 2011) or instruments across different music genres (Philips 2013). In product creation, entrepreneurs combine elements of technology, process, and design: for example, the typewriter draws from the piano swinging arms and hammers for imprinting letters (Utterback and Suarez 1993), and automobiles combine carriages with the internal combustion engine (Rao 1994). Firms may intentionally combine to produce something novel, or they may stumble on a new combination and decide it has potential. Either way, a development based on combination—in terms of technology, product, services, processes, or other pertinent effort—will appear to be categorically different from what already exists. This means:

Hypothesis 1: Firms that combine elements across market categories are more likely to engage in concept innovation, as compared to those that do not combine elements across market categories.

Category Leniency and Concept Innovation

New developments are not stand-alone entities, but are understood through the lens of existing classification (Smith 2011). This means characteristics of existing categories might influence how firms use category labels. I propose that the extent to which existing categories are *lenient* affects firms' propensity to engage in concept innovation.

The standard narrative in the category literature is that mature social categories develop strong boundaries and an agreed-upon social meaning, which makes them valuable for abstract reasoning (Hannan, Pólos and Carroll 2007; Negro, Hannan and Rao 2010; Kovács and Hannan 2010). But categories need to be flexible enough to describe a heterogeneous reality (Bowker

and Star 2000). Scholars increasingly recognize that mature categories may remain ambiguous in their meaning, for example if actors with multiple interests try to influence a category's definition (Grodal, forthcoming). Often, classifications contain categories that are *lenient*: with porous boundaries, a high degree of overlap, and ambiguous social meaning (Pontikes and Barnett 2015). This is illustrated in software market categories depicted in figure 1. Lenient market categories include “platform,” “customer relationship management (CRM),” and “business intelligence (BI),” which overlap with upwards of 30 other categories during the period of study.⁸ A study of market categorization in software describes the ambiguous nature of lenient categories:

These names [market categories MRP, ERP, and CRM] refer not to a specific homogeneous product but to a more or less heterogeneous collection of artefacts ... Such terminologies proposed a boundary that linked a group of (often quite various) artefacts while differentiating them from others (Pollock and Williams 2011:195).

It is difficult to articulate a specific definition for a lenient category. In a blog titled “What’s not BI?” a Forrester analyst struggles to define the lenient category “business intelligence.”

One of the favorite pastimes of BI analysts everywhere—not just at Forrester—is defining and redefining this uber-category known as BI. What exactly is it? Or rather—considering that almost every data management technology has been swept into BI’s gravitational orbit at one time or another by somebody somewhere—what is BI not? What’s analytics? What are decision support systems? How does this relate to knowledge management, content management, social media, and so forth? (Kobielus 2010).

Categories with broad overlap exert less constraint on members because overlap leads to (and reflects) uncertainty about category meanings. A diverse membership broadens how people typify a category. In the laboratory, people learn categories by observing patterns of similarity among objects that share a label (Sloutsky 2003; Lupyan 2012). A similar process takes place in markets, where people infer category meanings based on characteristics of members (Pontikes and Hannan 2014). When people encounter firms that claim the same label but that are

⁸ Lenient categories are not superordinate in a hierarchy. Superordinate categories are those that contain all members of subordinate categories. Lenient categories have a high degree of overlap but do not contain all members of the overlapping categories. This study only uses categories at the same hierarchical level.

dissimilar, they will not form strong expectations of exactly what membership in the category entails. For example, in the market category for “disk arrays,” member organizations also identified with a number of other categories. As a result, people came to perceive the “disk array” category in various ways (McKendrick and Carroll 2001).

In a lenient category, almost “anything goes.” This can pose a problem for category members, as categories are most valuable when they draw clear boundaries that accurately group objects (Negro et al 2010; Kovacs and Hannan 2010). Clear boundaries present constraints that facilitate the purpose of classification: organizing knowledge, simplifying cognition, and enabling communication and coordination (Murphy 2004; Bowker and Star 2000). As a result, I expect category leniency to trigger activist attempts to reduce categorical ambiguity. Previous research shows that ambiguous environments inspire actors to engage in processes of social construction (Weick 1995; Aldrich and Fiol 1994). In the mutual fund industry, the more variation within a category of funds, the more likely an analyst is to create a new category (Lounsbury and Rao 2004). Managers try to provide simple categorical frameworks in their public statements to reduce uncertainty (Lounsbury and Glynn 2001; Rindova, Ferrier and Wiltbank 2010). When category boundaries are muddied, producers mobilize to convey a focused identity (Negro, Hannan and Rao 2011), or disassociate from the category (Granqvist, Grodal and Woolley 2013).

One way to clarify a firm’s categorical position is to engage in concept innovation.⁹ An ethnographic study of five entrepreneurs shows that they spend considerable time claiming and demarcating a category they hope to control (Santos and Eisenhardt 2009). Quotes from my interviews, as referenced above, reinforce this point. For lenient categories, there is often

⁹ This is only one strategy for clarifying categories. Others include activist efforts to delineate boundaries of a muddied category: organizing industry associations, standard setting organizations, lobbying analysts, etc.

considerable uncertainty over what members provide. As a result, I expect producers in lenient categories to engage in concept innovation—regardless of whether they have created new combinations—in an attempt to clarify their identities and structure the ambiguous environment.

Hypothesis 2: Firms in high leniency market categories are more likely to engage in concept innovation, compared to firms in low-leniency categories.

I also suggest leniency weakens the relationship between new combinations and concept innovation. Low-leniency categories present constraints that reflect shared expectations that certain activities are expected of organizational members and that other pursuits do not belong. The more constraining a category is, the less likely a development that combines different elements will be accepted as part of the category, and the more likely firms will perceive it as distinct (and expect that important external constituencies will do the same). For example, NetApp, a software organization in the constraining “network storage” category, developed a new product for file serving. It was based on novel technology that combined disk storage improvements, a specialist operating system, and RAID. It was markedly different than the general-purpose workstations typical of “network storage” organizations, and NetApp pioneered the new “security appliance” category. Not only did they combine elements from different categories to create a novel product, but they were also in a constraining category. As a result, they could readily differentiate through concept innovation.

But what happens if we relax the assumption that categories are constraining? For lenient categories, a new combination will not as readily distinguish a new class of goods. For example, similar to NetApp, WebTrends developed products to report on Internet activity, combining knowledge from user profiling, client/server data access, caching, distributed computing, and IP address filtering. WebTrends was a member of the lenient “enterprise solutions” category.

Members of this category engaged in a wide range of activities, and WebTrends' combinations were easily subsumed under the lenient category. They did not engage in concept innovation.

In this way, the leniency of a firm's categories may affect whether new combinations are antecedents for concept innovation. A firm that blends diverse elements is not a credible member of a constraining category; as a result, there is more opportunity to define an adjacent category around the new combination. Conversely, firms that combine elements across categories can easily fit within lenient categories; as a result, new combinations do not benefit from a distinct categorical identity any more than any other product. Category constraints both *push* and *pull* differentiated producers to engage in concept innovation. This suggests:

Hypothesis 3: The positive relationship between firm combinations and concept innovation decreases with the leniency of the firm's categories.

Together, hypotheses 2 and 3 suggest that leniency fundamentally alters the process by which new market labels are introduced into a domain. When existing categories are constraining, firms engage in concept innovation because they have combined elements across category boundaries and so their offerings are categorical misfits. But when categories are lenient, the 'differentness' of the offering is not as relevant. In lenient contexts, firms use concept innovation for the purpose of clarifying muddied boundaries.

Empirical Test: The Software Industry

These hypotheses are tested in the software industry between the years 1990 and 2002. The software industry is a dynamic context where novelty is prized and market categories are important (Campbell-Kelly 2003; Tingling and Parent 2004; Pollock and Williams 2007). An informal classification based on product markets has emerged to help people make sense of the

industry, created through interactions among producers, industry analysts, the media, investors, and customers (Wang 2009).

Producers actively influence categorization. One way they do this is by simply claiming a particular category, which leads people to update their definition of the category (Negro, Hannan and Rao 2011; Granqvist, Grodal and Woolley 2013; Pontikes and Hannan 2014). Concept innovation is another important way they influence classification, and it is common in this setting: 361 new market category labels were introduced during the time period studied. The importance of novelty, categorization, and concept innovation makes this a good context to study the hypotheses introduced above.

Data and Methods

This study uses press releases to track market categorization. These data are used to measure concept innovation and leniency. Three different measures are used to capture combinations, which are computed using data from press releases and from patents.

Market Classification

The initial source of data to identify software firms and the market categories they are in are press releases issued between 1990 and 2002.¹⁰ Software firms actively issue press releases to announce news: a new product, executive team member, partnership, award, presence at a conference, etc. Press releases are not costly to produce, and so they track small and young firms that are missing from standard data sets. Within almost every press release, a firm will claim a market category label to describe what it does, providing a historical record of market categories

¹⁰ Data from 1989 were collected to determine concept innovation in 1990.

firms were in over time. This makes press releases a good source of data to study market categorization and concept innovation.

Press releases in *Businesswire*, *PR Newswire*, and *Computerwire* that contained at least 3 mentions of the word “software” were the initial source of data. There were 268,963 of these. Firms were extracted from press releases using custom text-matching programs, keying off firm identifiers such as “inc,” “corp,” etc. Results were inspected by hand to generate a list of 4,566 software firms in the press releases. Within press releases, firms claim category labels to describe what they do, such as “[*company*] is a leading provider of [*category label*].” Identity statements were extracted for each firm from press releases. Table 1 lists sample identity statements. As these identity statements convey, in this context producers are the objects classified, and classification is based on the features of their offerings.

--- Insert table 1 about here ---

An extensive list of market category labels was compiled from articles in *Software Magazine* and *Computerworld*, as well as from an inspection of the identity statements, to capture both “successful” categories as well as labels that do not catch on. Text matching programs searched identity statements for these labels, resulting in data on all category labels firms were in for each year. The final data contain 456 market category labels and 4,566 firms.

A number of checks were run to verify the validity of these data. One question may be whether press releases are relevant – it is likely that they are not widely read. But the media uses press releases to report on firms (Soltes 2009), leading to a wider dissemination than readership of the press releases would indicate. I also investigated whether market categories from press releases were reflected in other public statements from firms. I compared press release claims to

the archived web and to annual reports for a sample of firms, and found that categories in press releases were on Web sites 74% of the time, and in 10K statements 81% of the time.

Another question may be whether self-classification is valid. A series of studies find that self-classification (from annual reports) is a better predictor of financial outcomes, as compared to SIC or NAICS codes (Hoberg and Phillips 2010; 2012). Previous research uses self-classification to study venture capital financing (Pontikes 2012) and media coverage (Kennedy 2008). Self-claimed categories reflect patenting behavior (Pontikes and Hannan 2014). Finally, a question might be whether category labels from producers are used by anyone else. I compared market categories from press releases to categories covered in Gartner reports, and found that over half of the market categories from press releases were covered in Gartner reports.¹¹ This shows that there is a common language between Gartner reports and producer claims in press releases. It also shows that there are a number of categories that are not picked up by the analyst, indicating that these data capture a large breadth of category labels, which is required to study concept innovation.

I merged these data with other data sets in order to construct controls. I gathered data on public companies and IPOs from Thomson Financial and data maintained by Jay Ritter.¹² I included data on VC financing from the Venture Economics database from Thomson Financial. I also included information on size (based on revenue) from the Software 500, an annual listing of the largest public and private software firms, based on a survey conducted by *Software Magazine*. I searched for founding dates for all firms in press releases, on Hoovers,

¹¹ Gartner reports are available in electronic format from 1995 onward. Therefore the comparison is from 1995 – 2002. The title of the report typically references the market category covered. I searched for whether the category label was in the title of a Gartner report.

¹² <http://bear.warrington.ufl.edu/ritter/ipodata.htm>.

BusinessWeek, company Web sites, and Wikipedia. Founding dates were located for 3,705 of the 4,566 firms. Press releases were searched for mergers and acquisitions.

Model

The hypotheses are tested by modeling a firm's rate of concept innovation as a function of independent and control variables:

$$r(t - t_n) = r_o(t - t_n) \cdot \exp(\beta_{ind} \cdot \mathbf{x}_{ind} + \alpha_{control} \cdot \mathbf{x}_{control} + \varepsilon) \quad (1)$$

The firm-year is the unit of analysis. Firms are at risk of engaging in concept innovation for the years they are active in press releases: 4,566 firms over 18,192 firm-years. One measure of combination uses patent data, and so estimations including this covariate are run on patenting firms only, a risk set which includes 789 firms over 4,012 firm-years.

Dependent variable: Concept innovation

The dependent variable is whether a firm engages in concept innovation in the given year. Because these unique data track all market category labels used by software producers, they contain the earliest uses of labels (and include labels that did not diffuse). This allows for a systematic study of concept innovation. Of the 456 market category labels in these data, 361 were first introduced in the time period. My interviews and popular accounts indicate that the earliest uses of a label are often traced back to more than one actor. Firms may engage in concept innovation by borrowing and recasting a little-known term, or firms are unaware that another had recently started using the same label (sometimes leading to a prolonged debate about who used it first). This is reflected in these data: most new labels do not gain traction until after the second

year (if they gain traction at all).¹³ To capture concept innovation in these data, it is coded as a 0/1 indicator variable the first year a firm affiliates with one of the 361 new labels in the corpus, if it is *also* in the first or second year that the label *has ever been used* in the entire press release corpus. There are 787 events of concept innovation by 581 firms (302 events of concept innovation by 144 patenting firms).¹⁴ Supplementary analyses are run with alternate definitions of concept innovation: (1) excluding labels that grew quickly in the first two years (those with more than 5 members in the first two years), as all of the early affiliations with these labels may not represent concept innovation, (2) including only labels that diffused among firms (those that eventually had at least 5 members), and (3) including only category labels that were picked up by Gartner.

Independent Variables: Combinations and Leniency

Independent variables are constructed to measure whether a firm creates combinations (hypotheses 1 and 3) and whether its categories are lenient (hypotheses 2 and 3), for each firm-year.

There are various ways in which firms can create developments based on new combinations: in their inventions, through product design, by applying a technology to a different use, by incorporating a process, and so on. I construct three measures to capture these different aspects of combination:

- Firm combination measures whether the language in a firm's press releases is similar to language used by firms in different market categories (ones the firm is not in).

Firms that are very similar to many different categories score high on this measure.

¹³ 75% of new labels have only one or two members in the first two years that they appear in these data.

¹⁴ The dependent variable is set to zero for the first year a firm appears in press releases because in these cases we do not know for how long the firm had previously used the label.

This general measure of combination captures anything discussed in press releases, including descriptions of products and services, design, how offerings are marketed, strategic alliances, the executive team, and so on. A strength of this measure is that it includes these many facets of combination. Another strength is that language can capture subtle aspects of categorical blending. The down side to this measure is that it does not convey the specific ways in which a firm creates new combinations.

- Category combination, the simplest of the three, measures the number of market categories the firm is in (as evidenced by the categories claimed in press releases). It captures a breadth-based aspect of recombination. Firms can score high on this measure with one offering that is applicable to multiple market categories, or if they have multiple products in different categories. In either case, firms with high category combination are familiar with elements of multiple categories and so have more opportunities to combine them. A strength of this measure is that it captures combination among categories with which the firm is most familiar. A down side is that it does not capture when firms are combining the distant elements associated with categories they are not in.
- Technology combination captures whether a firm's technologies draw on elements of different market categories (ones the firm is not in). Using the firm's position in a patent citation network, this measure captures how close in knowledge space a firm's patents are to different categories. The strength of this measure is that it captures recombination knowledge space, a plane that is not defined by language used on the market. It is the least likely to be intentionally influenced by the firm for the purposes of creating a new market category. The down side is that it only captures a specific

type of combinatory behavior, and therefore is only defined for firms that patent (a subset of the firms in these data).

Individually, each measure has strengths and weakness. Each captures different aspects of combination. Firm combination is the most general, representing a number of ways in which firms combine elements. It reflects both deliberate attempts to craft language as well as unintentional combinations. Technology combination is the most specific and reflects the underlying technical components used by a firm. This measure is the least likely to be influenced by marketing (it is unlikely a firm would include patent citations for this purpose). Both firm and technology combination capture a firm drawing on elements from categories it is not in.

Category combination measures the opportunity to combine elements from categories the firm is in. Together, these three measures capture whether a firm is creating new combinations in its various activities. As such, using all three provides a robust test of hypotheses 1 and 3. Below, I detail how each measure is constructed.

Firm combination

Firm combination is computed using a natural language processing bag of words algorithm. The text of a firm's press releases issued in each year are each bag of words, and all firm-year "bags" comprise the corpus. I programmed an algorithm in python that defines the language space for this corpus, and then located each firm-year in the space. Words and phrases of up to three words are the features that define the space. Stop words are removed, and each word is stemmed using snowball stemmer from the nltk.stem package. I tokenize the corpus using Tfidfvectorizer, which uses term-frequency-document-frequency (tf.idf) weights that reflect how important the word or

phrases are in the corpus (Leskovec, Rajaraman, and Ullman 2014). I use the top 20,000 features (tf.idf weighted words and phrases). The algorithm outputs a feature vector that locates each firm in each year in the space. Market categories are then located in this space by summing the vector locations of the firms in the category. Cosine similarity between the firm vector (\mathbf{f}) and category vector (\mathbf{c}) measures how similar the firm's language is to each category, for that year:

$$\cos_sim_{f,c,y} = \frac{\langle \mathbf{f}, \mathbf{c} \rangle}{\|\mathbf{f}\| \|\mathbf{c}\|} \quad (2)$$

This ranges from 0 – 1. Firm combination measures how similar a firm is in this language space to all categories it is not in, which is the sum of the firm's cosine similarity to categories it is not in. The natural log is taken as the measure exhibits skew:

$$firm_combination_{f,y} = \ln\left(\sum_{f \notin c} \cos_sim_{f,c,y}\right) \quad (3)$$

A similar metric is constructed for supplementary analyses that computes the firm's language similarity to its own categories.

Category combination

Category combination counts the number of market categories the firm is in. The natural log is taken as the measure exhibits skew:

$$category_comb_{f,y} = \ln(\#categories_{f,y}) \quad (4)$$

Technology combination

Technology combination is computed using the location of a firm's patents in a knowledge space that is based on patent citation overlap.¹⁵ Patent data are obtained from the NBER US Patent

¹⁵ One concern about using patents in a study of software firms is that historically the U.S. patent office did not allow software to be patented. This decision was effectively overturned in 1994 – 1995, but restrictions were loose

Citations data file (Hall, Jaffe and Trajtenberg 2001), which has been extended to include patents through 2006.¹⁶

An n-dimensional “knowledge space” is created using all Computers & Communications patents (not just patents issued to firms in these data). Knowledge space is a citation network using a five-year window of all patents applied for in the current year and four years prior.¹⁷

Citation overlap is used to locate patents close to or far from one another in knowledge space, dividing the number of overlapping citations by the number of citations made by the focal patent:

$\alpha_{mn} = \frac{s_{mn}}{s_m}$, $0 \leq \alpha_{mn} \leq 1$ (Podolny, Stuart and Hannan 1996). s_{mn} is the number of shared

citations between patent m and n , and s_m is the total number of citations by patent m . Second degree similarity is also computed, so that two patents that are similar to a third can have non-zero similarity to each other. Similarity between any patent m and n is measured by multiplying their similarities to a third patent, k : $\sigma_{mn} = \max_{\alpha_{mk}>0 \wedge \alpha_{kn}>0} (\alpha_{mk} \cdot \alpha_{kn})$. The patent k that yields the maximum similarity between patents m and n is used in the computation ($\sigma_{mn} = \alpha_{mn}$ if there is nonzero first degree similarity).

Technology combination is computed by measuring how close a firm’s patents are in knowledge space to market categories it is not in. Areas of knowledge space are mapped to market categories based on the category membership of the firm that is assigned the patent.¹⁸ All measures are weighted by a firm’s “grade of membership” in each category (so firms in multiple categories do not have outside influence). Grade of membership of firm A in market category

after *Diamond vs. Diehr* in 1981. The approval of software patents was routine even before the courts officially recognized it (Cohen and Lemley 2001; Mann 2005).

¹⁶ This study uses the 2002 update, which was posted on Bronwyn Hall’s Web site.

¹⁷ Patents are timed based on application dates. Only patents that were eventually granted are used.

¹⁸ Previous research shows that this mapping is meaningful: firms are more likely to enter market categories to which they are proximate in knowledge space, and are more likely to exit categories distant in the space.

k , $\mu_A(k)$, is computed as: (the number of press releases in which organization A claims category k) divided by (the number of press releases in which it claims any category) in a given year.

Technology combination for firm A is calculated using the knowledge space proximity of each of A 's patents (m) to market categories that A is not in, $D \in \{D_A: \mu_A(D) = 0\}$. A patent n is affiliated with D if it was issued to an organization B that is a member of D , weighted by B 's grade of membership in D :

$$prox_{(m \in P_A), D_A} = \sum_{n: (n \in P_B) \wedge (B: \mu_B(D) > 0) \wedge (D \in D_A)} [\mu_B(D) \times \sigma_{mn}] \quad (5)$$

P_A is the patent portfolio of organization A in the given year, and P_B is the patent portfolio of an organization B that is a member of category $D \in D_A$. Patent level proximities are then averaged over the number of patents issued to organization A that year. Because the distribution is skewed, the natural log is used:

$$Tech_comb = \ln \left(\frac{\sum_{(m \in P_A)} [prox_{m, D_A}]}{npat_A} \right) \quad (6)$$

A subset of firms from the press release data patent (789 firms). Therefore, for hypothesis tests that use technology combination, the primary estimations use firms that have patented as the risk set. Additional models test effects on the risk set of all firms. Firms that do not patent in a year are assigned a value of zero for technology combination.

Leniency

A measure for category leniency is constructed based on how much a category overlaps with other categories in the domain. Leniency is built on *fuzziness*, an established construct in cognitive science and the literature on market categorization (Rosch and Mervis 1975; Hampton 1998; Hannan, Pólos and Carroll 2007). A category has fuzzy boundaries if it has high overlap with at least one other category. Lenient categories are not only fuzzy, but also overlap broadly

with many other categories in the domain. This broad overlap reflects that members are not highly constrained from adopting a variety of other identities. Leniency is calculated by multiplying fuzziness by the number of distinct categories with which the focal category overlaps. Fuzziness is calculated as the inverse of *contrast*, which is the average grade of membership of firms in the category (Hannan 2007):

$$fuzz_k = 1 - \frac{1}{N_k} \sum_{i:\mu_i(k)>0} \mu_i(k) \quad (7)$$

Where N_k is the number of firms in category k and $\mu_i(k)$ is the grade of membership of firm i in category k , as defined above. Then leniency is:

$$Leniency_k = fuzz_k \times \ln(N_{ocat}) \quad (8)$$

N_{ocat} is the number of other categories with which category k overlaps.

Hypothesis 1 is tested by including covariates for firm combination (equation 3), category combination (equation 4), and technology combination (equation 6). Hypothesis 2 is tested by including leniency (equation 8). Hypothesis 3 is tested by including the interaction between each of the combination variables and leniency. A secondary measure is used to test hypothesis 3 for technology combination, where technology combination is included in pieces in the model, for organizations in low/medium and high leniency categories. High leniency is defined as organizations in the upper 35% of the maximum level of leniency in the given year.¹⁹

Controls

A number of control variables are included that might influence concept innovation. The number of members in a firm's categories (weighted by grade of membership and logged) is included to

¹⁹ Results are not sensitive to the threshold.

control for competitive effects. To control for firm size, a categorical variable is included that indicates the firm's ranking in the Software 500 in the previous year,²⁰ and a 0/1 variable is included for whether the firm is public. An indicator variable is also included for whether the firm received venture capital funding in the previous year. Controls are included for the firm's number of patents and citations in the previous year.²¹ The tenure of the firm since the inception of the data is included to control for age. Year fixed effects are included in all models to control for temporal effects. In supplementary analyses, controls also include the number of acquisitions made by the organization, number of press releases issued by the organization, the tenure of the organization's categories measured since the first mention of the category in these data, and the age of the organization (for the subset of organizations for which founding dates could be located). All independent and control variables are lagged by one year.

Estimation

The instantaneous hazard rate of an organization engaging in concept innovation as a function of independent and control variables (equation 1) is estimated with a piecewise exponential model using the `stpiece` routine in STATA. Duration pieces are included for whether an organization waits 0—1 year, 1—2 years, 2—5 years, 5—10 years, or more than 10 years before creating a new category label, measured from when the organization last entered an existing category or engaged in concept innovation.²² In order to estimate repeated events in these models, organizations exit the data enter with a new id. Therefore, robust variance estimators are used.

²⁰ This is computed as 0 if the firm is unranked, 1 for $500 \leq \text{rank} < 400$, 2 for $400 \leq \text{rank} < 300$, 3 for $300 \leq \text{rank} < 200$, 4 for $200 \leq \text{rank} < 100$, 5 for $100 \leq \text{rank} < 50$, 6 for $50 \leq \text{rank} < 10$, 7 for $\text{rank} \leq 10$. Results are robust to including a 0/1 indicator variable for whether the firm ranked in the Software 500, or including dummy variables for rank bands.

²¹ Results are robust to including the natural log of patents and citations, and the total number of patents and citations over time. The number in the previous year is most predictive of concept innovation.

²² In this way concept innovation and entering an existing category are modeled as competing risks.

Analyses are primarily run on the risk set of all organizations; patenting organizations only are used for tests of technology combination. Data for all organizations contain 4,566 organizations over 18,192 organization-years, with 787 events of concept innovation. Data for patenting organizations contain 789 organizations over 4,012 years, with 302 events of concept innovation. Tables 2 - 3 contain descriptive statistics and correlations for all firms and for patenting firms.

--- Insert tables 2 - 3 about here ---

There are a number of highly correlated variables, notably between category leniency and other category metrics. This is partly because there are more categories that become more lenient over time. This suggests it is important to control for year fixed effects and for other category variables. At the same time, it can raise concerns about multicollinearity and therefore the stability of the results. To address this, I checked the variance inflation factor for the covariates and all values were less than ten, indicating that multicollinearity is not a substantial issue. In addition, models without highly correlated controls yield consistent results.

Results

Table 4 includes tests of hypotheses 1 and 2. Column 1 contains controls only. Column 2 includes leniency of a firm's categories, which has a positive effect ($\beta = 0.447$ (0.0802), $p < 0.001$). A one standard deviation increase in category leniency is associated with a 66% increase in the rate of concept innovation. This effect remains stable when combination measures are also included in columns 3 – 5, indicating that the effect is net of whether a firm creates new combinations. This provides support for hypothesis 2.

Columns 3 – 6 test hypothesis 1. Column 3 includes category combination, which yields a positive effect with $\beta = 1.056$ (0.103), $p < 0.001$. A one standard deviation increase in category combination is associated with an 85% increase in the rate of concept innovation. Column 4 adds firm combination, which has a positive effect with $\beta = 0.239$ (0.0371), $p < 0.001$. A one standard deviation increase in firm combination is associated with a 50% increase in the rate of concept innovation. Column 5 adds technology combination in models run on all organizations, and the effect is positive, $\beta = 0.155$ (0.0477), $p < 0.01$. Column 6 runs the same estimation on a risk set of firms that have patented, and results are similar ($\beta = 0.203$ (0.0678), $p < 0.01$). A one standard deviation increase in technology combination for a patenting firm is associated with a 21% increase in the rate of concept innovation.²³ Results show support for hypothesis 1 across all three measures of combination.²⁴

--- Insert table 4 about here ---

Table 5 presents tests of hypothesis 3. Column 1 includes an interaction between category combination and the leniency of the firm's categories, and the effect is negative, as expected ($\beta = -0.483$ (0.0811), $p < 0.001$). Column 2 includes an interaction between firm combination and category leniency, and again the effect is negative ($\beta = -0.204$ (0.0311), $p < 0.001$). Effects of technology combination are again included in models run on all organizations and patenters only. When compared to all organizations, the effect is negative and significant ($\beta = -0.246$ (0.0694), $p < 0.001$), but in models run on patenters only the effect of the interaction is negative but not significant at conventional levels ($\beta = -0.127$ (0.0936), $p = 0.18$). It is possible that the noise is due to the functional form of the interaction. Column 5 includes technology combination in pieces: for firms in low/medium leniency categories, and for firms in high leniency categories.

²³ All combination effects are similar in models that exclude leniency.

²⁴ The effect of firm combination is much reduced and not significant at standard levels in models of patenters only. This may indicate that for firms that patent, technology is more relevant for new combinations.

For firms in lower leniency categories, there is a positive relationship between technology combination and concept innovation ($\beta = 0.365$ (0.0914), $p < 0.001$), but the effect weakens substantially and becomes nonsignificant for firms in high leniency categories ($\beta = 0.063$ (0.0831), $p = 0.45$). We can reject the null that these coefficients are equal at $p < 0.01$. For all three measures of combination, there is a positive relationship between combination and concept innovation when the firm is in constraining categories, and this relationship weakens substantially when categories are lenient.²⁵ For category and firm combination, the effect is apparent using the continuous interaction. But results indicate that the functional form of the effect of leniency on the relationship between technology combinations and new concept innovation is not best modeled by a log-linear function; rather, the relationship between technology combinations and new category formation decreases sharply for firms in high leniency categories. Together, these effects show support for hypothesis 3.

--- Insert table 5 about here ---

Effects are illustrated in figure 2, which plots the relationship between concept innovation and the three measures of combination, for firms in low leniency categories (one standard deviation below mean or as indicated in the model) and high leniency categories (one standard deviation above the mean or as indicated in the model). For each measure of combination, the positive relationship with concept innovation is much stronger for firms in low-lenieny categories as compared to those in high-lenieny categories. There is the most pronounced relationship for category combination, where one standard deviation above mean category combination is associated with a fourteen-fold increase in concept innovation for firms in low-lenieny categories, but this drops to a three-fold increase for firms in high-lenieny

²⁵ Effects are similar when all three interactions are included in the model.

categories (compared to firms with zero combination).²⁶ For firm combination, one standard deviation above the mean is associated with a seven-fold increase for low-leniency categories, and only a 17% increase for high-leniency categories.²⁷ For technology combination, one standard deviation above the mean for a patenting firm in low/medium leniency categories shows a 89% increase in concept innovation, compared to only a 12% increase for those in high-leniency categories.²⁸

--- Insert figure 2 about here ---

Effects of control variables are reported from table 4, column 5. The number of other organizations in a firm's categories has a negative effect on concept innovation ($\beta = -0.185$ (0.055), $p < 0.01$), perhaps indicating that popular market categories are resource rich. Larger and more successful firms are more likely to engage in concept innovation: rank in the Software 500 ($\beta = 0.104$ (0.0196), $p < 0.001$), being public ($\beta = 0.178$ (0.083), $p < 0.05$), having many patents ($\beta = 0.162$ (0.086), $p = 0.059$), or having received venture capital funding in the previous year ($\beta = 0.469$ (0.109), $p < 0.001$) all have positive effects. There is not a strong effect of time pieces until a firm has been in a category over 10 years, when it is very unlikely to engage in concept innovation, likely picking up heterogeneity in firm strategies.

Additional tests

One concern related to the firm combination and technology combination measures is whether effects are due to firms combining elements across categories, or if they simply are picking up on general language or inventive activity. Table 6 includes models that address this alternative.

Column 1 includes a covariate for whether a firm in press releases uses *language similar to its*

²⁶ For low leniency, $\exp(2*1.48-.48*1.48*0.46)=14$; for high leniency, $\exp(2*1.48-.48*1.48*2.74)=2.76$

²⁷ For low leniency, $\exp(0.6*3.74-0.204*3.74*0.46)=6.6$; for high leniency, $\exp(0.6*3.74-0.204*3.74*2.74)=1.17$.

²⁸ For low/medium leniency, $\exp(0.365*1.75)=1.89$; for high leniency, $\exp(0.063*1.75)=1.12$.

own categories, computed as in equation 3 but summing over cosine similarity to market categories the firm is in. The effect of the covariate is negative and there is a high probability the null cannot be rejected ($\beta = -0.258$ (0.241); $p = 0.283$). Column 2 includes a measure for whether a firm is developing *technologies similar to its own categories*, like equation 6 but measuring the firm's knowledge space proximity to categories it is in. The effect of the covariate is near zero and there is a high probability the null cannot be rejected ($\beta = -0.044$ (0.225); $p = 0.844$). Effects of leniency, combination, and the interaction, are robust. This suggests that it is proximity to *different* categories that drives the effects reported above. This supports the idea that it is the combination of different elements that prompts concept innovation, providing additional evidence in support of hypothesis 1.

--- Insert table 6 about here ---

Next, I test for whether effects are sensitive to the definition of concept innovation used in the dependent variable. I test against three alternative definitions in estimations presented in table 7. For each alternate dependent variable, separate estimations are run for the three measures of combination, including the same covariates as table 5 columns 1, 2, and 5, respectively. For the sake of parsimony, table 7 includes effects of the independent variables only for these nine models.

--- Insert table 7 about here ---

One concern may surround the cut-off employed for defining concept innovation: affiliating with a category label in its first or second year that it has ever been used. This is used because most labels do not gain traction until after their second year. However, some labels do see early growth; in these cases some early claims might not represent concept innovation. In the first column in table 7, the dependent variable is defined as when a firm claims a category label

in the label's first two years of existence, *if the label had five or fewer members in its first two years*. This excludes categories with early traction. Results are similar.

Another question may be whether results are due to idiosyncratic labels that never catch on. This calls into question whether effects of concept innovation speak to category creation more broadly. I address this question in estimations reported in columns 2 and 3 in table 7. Column 2 tests if effects persist for category labels that catch on with other producers. The dependent variable defines a concept innovation event only if the category label eventually had at least 5 members. Results continue to support the hypotheses. Column 3 tests if effects persist for category labels that catch on with industry analysts. The dependent variable defines a concept innovation event only if the category label is used in Gartner reports (these are available from 1995 onward). Effects persist.

Results are consistent with the inclusion of additional controls: the number of times the firm previously engaged in concept innovation, number of acquisitions made by the firm, the number of press releases issued, category age, whether the firm is the only member of its category, and the age of the firm.

Discussion

An important topic in strategy research concerns how a firm should position on the market. Studies have focused on the trade-off between being similar to or differentiating from competitors (Porter 1980; Deephouse 1999), mostly presuming that market categories are fixed. Yet, an alternative is to engage in *concept innovation*, where firms introduce new market category labels to describe what they do. Concept innovation is not only an important strategic tactic, it is also a first stage of market evolution. Once introduced, labels may coalesce into a

category and change the market structure. There are many studies on product and category emergence from a variety of perspectives (Utterback and Suarez 1993; Hargadon and Sutton 1997; Ruef 2000; Fleming 2001; Verdes and Stark 2012; Powell and Sandholtz 2012), but little is known about this tactic.

Previous research on changes in market structures have found that novel developments are based on combinations of existing elements, an idea that traces back to Schumpeter (1934). He proposed that “the carrying out of new combinations” (1934, p. 66) is the primary source of economic development. But subsequent research has not systematically studied whether combinations lead to the types of discontinuous changes depicted in his works. In the technology literature, systematic studies of patents find that combinations are associated with high impact inventions (Fleming 2001; Rosenkopf and Nerkar 2001; Nerkar 2003), but this literature does not look at whether these inventions are a qualitative change in the market. In entrepreneurship, strategy, and organization theory, researchers find evidence that new products, market categories, and organizational forms are based on combinations, but these study the formation of a (by definition) successful category (DiMaggio 1991; Utterback and Suarez 1993; Hargadon and Sutton 1997; Powell and Sandholtz 2012).

This is one of the first studies to investigate antecedents of the earliest stages of market category formation, by studying concept innovation. Results show that, consistent with the view that novelty arises from new combinations, firms that blend elements across market categories engage in this tactic. But contextual factors mitigate this effect. There is a strong relationship between combinations and concept innovation when existing categories are constraining, but it weakens considerably when categories are lenient. In these contexts, leniency itself gives rise to

concept innovation, presumably as firms introduce new labels to clarify their categorical identity in an ambiguous environment.

A systematic study of this topic is enabled by unique data that contain details of market classification for hundreds of market categories and thousands of firms over thirteen years. Computational advances in text analytics have created opportunities for creating such datasets, where early stage labels can be extracted from free text and assembled into structured data. Whereas previous studies on market creation look at the history of one successful market, these data contain hundreds of instances of concept innovation, allowing for a study of whether (and when) combinations are associated with the tactic. These data also contain far more variation in new market categories than is typical: they include labels that took off and became important categories as well as labels that never gained traction.

Results support the combinatory view when existing categories are constraining; combinations across categories are associated with concept innovation. This holds for three different measures of combination that capture different facets of how firms blend elements across categories: firm combination, category combination, and technology combination. Support across all three measures indicates that many types of combination can underlie novel developments, consistent with both Schumpeter's expansive definition (1934, p. 66), and also with previous studies that use different measures of combinations depending on the empirical context (DiMaggio 1991; Utterback and Suarez 1993; Stark 1996; Hargadon and Sutton 1997; Rao 1998; Philips 2013; Powell and Sandholtz 2012).

Findings also show that the relationship between combination and concept innovation depends on the categorical environment. When market categories are constraining and well bounded— as they have traditionally been conceptualized in strategy, economics, and sociology —

there is a pronounced relationship between combining elements across categories and concept innovation. But leniency weakens the effect. This supports the idea that combinations do not differentiate offerings when category constraints are weak and boundaries are porous. It implies that contexts with constraining as opposed to lenient categories give rise to very different processes when it comes to concept innovation, and thus market category emergence. When categories are constraining, the traditional view that new categories result from new combinations holds. But when categories are lenient, seeds for new market categories arise from actors negotiating categorical boundaries, regardless of whether they combine elements across those boundaries.

Novel developments are not stand-alone entities, but are interpreted by actors whose views are shaped by existing categories. Context affects how firms position themselves within the market. There are echoes of this process even in the physical world, where perception has a material effect. The Heisenberg Uncertainty Principle in quantum mechanics states that an experimenter cannot know both the position and momentum of a particle with certainty. As Heisenberg (1927) noted, “the path [of a particle] comes into being only because we observe it.” Comparably, in markets, firms engage in concept innovation because of how they interpret their developments through the lens of existing classification.

This study contributes to the growing literature that builds on cognitive science to understand macro-level effects of how markets are structured. Scholars have drawn on cognitive research to understand strategic behavior in competitive positioning (Porac and Thomas 1989; 1990; Durand and Paolella 2013; Cattani and Porac 2017). Researchers from the sociological tradition are increasingly interested in the cognitive foundations of cultural and market categories (DiMaggio 1997; Pontikes and Kim 2017; Hannan et al 2017). Findings here suggest that market

actors strategically use labels to change how people conceptualize their (and their competitors') products, and that this behavior has macro-level implications for the evolution of markets.

The software industry is fast-paced, innovative, and relatively young. One might think that these characteristics make the industry particularly accommodating to lenient categories. But lenient categories are prevalent in other domains. For example, "nanotechnology" has become an important area of scientific research that has become quite lenient (Grodal forthcoming). "Social entrepreneurship" is a lenient category that is important in business: it applies to any person who uses entrepreneurial principles to address a social problem. In academia, "strategy" is a lenient category that refers to a wide range of research. Many systems of classification result in overlapping, networked structures (Bowker and Star 2000), indicating that findings regarding lenient categories may be widely applicable.

Results may have implications for other literatures for which blending, bridging, or bricolage is salient. For instance, in network research, brokers are advantaged because they have access to a diverse base of information through bridging ties (Burt 2005). Social movements become prominent through issue bricolage and tactical blending (Jung, King and Soule 2014). If blending processes are altered by environmental constraint, it may be useful for these literatures to also consider categorical constraints where blending occurs.

Studying concept innovation complements research on category formation and boundary construction. A rich literature investigates the processes of meaning formation in category emergence: the boundary work that takes place formally and informally as people use new terms to communicate about new ideas, middlemen define emerging categories in reports, and activists lobby to establish legal codes that regulate who can affiliate with a term (Carroll and Swaminathan, 2000; Navis and Glynn 2010; Bingham and Kahl 2013, Grodal forthcoming).

Eventually, a category may become a taken for granted social fact (Berger and Luckmann 1967; Carroll and Hannan 2000). Missing in this literature is the crucial first step to category formation: when a label for a new concept is introduced. This study helps close that gap by studying factors associated with concept innovation.

This study also speaks to the question of whether firms create or conform to their surroundings. Where organization theory has focused on environmental constraints that induce homogeneity, leading to the replication of existing categories (Meyer and Rowan 1977; DiMaggio and Powell 1983; Fligstein 1996; Hannan and Freeman 1977; Zuckerman 1999), the literature in strategic management emphasizes how firms act to differentiate (Deepphouse 1999; Porter 1980) or change their environments (Abernathy and Utterback 1978; Utterback and Suarez 1993; Suarez, Grodal and Gotsopoulos 2014). Findings here present a unique take on constraints: rather than simply inducing conformity, category constraints create contrast within the environment that is necessary to enable the strategic tactic of concept innovation. Both firm outputs and context interact to influence concept innovation, a primary way through which a firm can evolve the market structure.

References

- Abernathy, William J, and James Utterback. 1978. "Patterns of Industrial Innovation." *Technology Review* 80:41-47.
- Alcácer, Juan, and Michelle Gittleman. 2006. "Patent Citations As a Measure of Knowledge Flows: The Influence of Examiner Citations." *The Review of Economics and Statistics* 88:774-779.
- Aldrich, Howard, and Marlene Fiol. 1994. "Fools Rush In? The Institutional Context of Industry Creation." *Academy of Management Review* 19:645-670.
- Berger, Peter and Thomas Luckmann. 1967. *The Social Construction of Reality*. New York: Anchor Books.
- Bingham, Christopher B, and Steven S. J. Kahl. 2013. "The Process of Schema Emergence: Assimilation, Deconstruction, Unitization and the Plurality of Analogies." *Academy of Management Journal* 56:14-34.
- Bowker, Geoffrey C, and Susan S. L. Star. 2000. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.
- Burshell, Jessica and W. Mitchell. 2016. "The Social Construction of Market Categories: How Proximate Social Space Creates Strategic Incentives to be Early Claimants of the Fiscal Sponsor Label." in *Research in the Sociology of Organizations*, edited by Henrich Greve and Seidel, Marc-David. 50: 171-221.
- Burt, Ron. 2005. *Brokerage and closure: An introduction to social capital*. Oxford: Oxford University Press.
- Campbell-Kelly, Martin. 2003. *From Airline Reservations to Sonic the Hedgehog: A History of the Software Industry*. Cambridge, Massachusetts: The MIT Press.
- Carroll, Glenn and Michael Hannan. 2000. *The Demography of Corporations and Industries*. Princeton: Princeton University Press.
- Carroll, Glenn and Anand Swaminathan. 2000. "Why the Microbrewery Movement? Organizational Dynamics of Resource Partitioning in the U.S. Brewing Industry." *American Journal of Sociology* 106:715-762.
- Cattani, Gino and Joe Porac. 2017. "Categories and Competition." *Strategic Management Journal* 38: 64-92.
- Clark, Kim B. 1985. "The Interaction of Design Hierarchies and Market Concepts in Technological Evolution." *Research Policy* 14:235-251.
- Cohen, Julie, and Mark Lemley. 2001. "Patent Scope and Innovation in the Software Industry." *California Law Review* 89:1-57.
- Colyvas, Jeannette A. 2007. "From divergent meanings to common practices: The early institutionalization of technology transfer in the life sciences Stanford University." *Research Policy* 36:456-476.
- Deephouse, David L. 1999. "To Be Different, or to Be the Same? Itsa Question (and Theory) of Strategic Balance." *Strategic Management Journal* 20:147-166.
- DiMaggio, Paul J. 1991. "Constructing An Organizational Field As a Professional Project: U.S. Art Museums, 1920-1940." Pp. 267-292 in *The New Institutionalism in Organizational Analysis*, edited by Walter J Powell and Paul J DiMaggio. Chicago: The University of Chicago Press.
- DiMaggio, Paul J. 1997. "Culture and Cognition." *Annual Review of Sociology* 23: 263 – 287.

- DiMaggio, Paul J, and Walter W. W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48:147-160.
- Durand, Rodolphe and Lionel Paoletta. 2013. "Category Stretching: Reorienting Research on Categories in Strategy, Entrepreneurship, and Organization Theory." *Journal of Management Studies* 50: 1100-1123.
- Fleming, Lee. 2001. "Recombinant Uncertainty in Technological Search." *Management Science* 47:117-132.
- Fligstein, Neil. 1996. "Markets As Politics: A Political-Cultural Approach to Market Institutions." *American Sociological Review* 61:656-673.
- Foster, J. G., A. Rzhetsky, and J. A. Evans. 2015. "Tradition and Innovation in Scientists' Research Strategies." *American Sociological Review* 80:875-908.
- Gelman, S A, and . D. Heyman. 1999. "Carrot-Eaters and Creature-Believers: The Effects of Lexicalization on Children's Inferences About Social Categories." *Psychological Science* 10:489-493.
- Granqvist, N, S Grodal, and . L. Woolley. 2013. "Hedging Your Bets: Explaining Executives' Market Labeling Strategies in Nanotechnology." *Organization Science* 24:395-413.
- Grodal, Stine. Forthcoming. "Field Expansion and Contraction. How Communities Shape Social and Symbolic Boundaries." *Administrative Science Quarterly*.
- Grodal, Stine, Aleksios Gostopoulos, and Fernando Suarez. 2015. "The Coevolution of Technologies and Categories During Industry Emergence." *Academy of Management Review* 40: 423-445.
- Gruber, M, D Harhoff, and K Hoisl. 2013. "Knowledge Recombination Across Technological Boundaries: Scientists Vs. Engineers." *Management Science* 59:837-851.
- Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg. 2001. "The NBER Patent Citations Data File: Lessons, Insights, and Methodological Tools." in *NBER Working Paper Series*. Cambridge, MA: National Bureau of Economic Research.
- Hampton, James A. 1998. "Similarity-based Categorization and Fuzziness of Natural Categories." *Cognition* 65:137-165.
- Hannan, Michael T, and John Freeman. 1977. "The Population Ecology of Organizations." *American Journal of Sociology* 82:929-64.
- Hannan, Michael T., Gael LeMens, Greta Hus, Balazs Kovacs, Giacomo Negro, Laszlo Polos, Elizabeth Pontikes, and Amanda Sharkey. 2017. Concepts and Categories: Foundations for a Sociological Analysis. Working Manuscript.
- Hannan, Michael T, László Pólos, and Glenn Carroll. 2007. *Logics of Organization Theory: Audiences, Codes and Ecologies*. Princeton, NJ: Princeton University Press.
- Hargadon, Andrew and Robert Sutton. 1997. "Technology Brokering and Innovation in a Product Development Firm." *Administrative Science Quarterly* 42: 716-749.
- Heisenberg, Werner. 1927. "Über Den Anschaulichen Inhalt Der Quantentheoretischen Kinematik Und Mechanic." *Zeitschrift Fur Physik* 43:172-98; reprinted as "The Physical Content of Quantum Kine- matics and Mechanics", translation by J.A. Wheeler and W.H. Zurek, in *Quantum Theory and Measurement*.
- Hoberg, G, and G Phillips. 2010. "Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis." *Review of Financial Studies* 23:3773-3811.
- Hoberg, Gerald, and Gordon Phillips. 2012. "Industry Choice and Product Language." *Working Paper*.

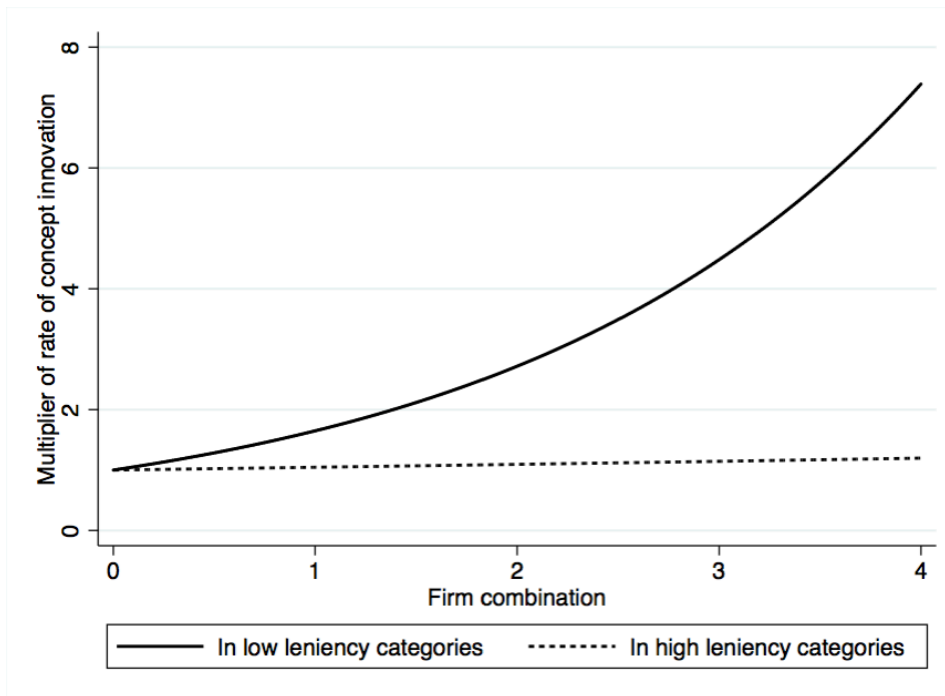
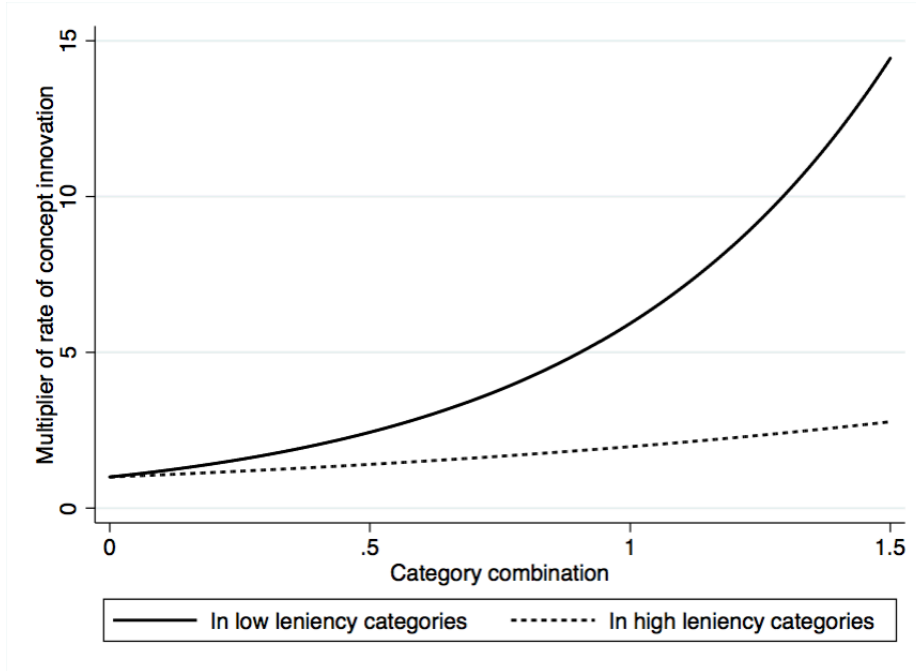
- Hsu, Greta. 2006. "Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production." *Administrative Science Quarterly* 51:420-450.
- Hsu, Greta, Michael Hannan, and Ozgecan Koçak. 2009. "Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests." *American Sociological Review* 74:150-69.
- Jung, Wooseok, Brayden King and Sarah Soule. 2014. "Issue Bricolage: Explaining the Configuration of the Social Movement Sector, 1960—1995." *American Journal of Sociology* 120: 187-225.
- Kaplan, S, and M Tripsas. 2008. "Thinking About Technology: Applying a Cognitive Lens to Technical Change." *Research Policy* 37:790-805.
- Kennedy, Mark T. 2008. "Getting Counted: Markets, Media and Reality." *American Sociological Review* 73:270-295.
- King, Brayden G, and Nicholas N. A. Pearce. 2010. "The Contentiousness of Markets: Politics, Social Movements, and Institutional Change in Markets." *Annu. Rev. Sociol* 36:249-267.
- Kobielus, James. 2010. "What's Not BI? Oh, Don't Get Me Started... Oops Too Late... Here Goes." *Forrester Blogs*.
- Kovács, Balázs, and Michael Hannan. 2010. "The Consequences of Category Spanning Depend on Contrast." Pp. 175-201 in *Categories in Markets: Origins and Evolution*, edited by Greta Hsu, Giacomo Negro and Ozgegan Kocak. Bingley, UK: Emerald Group Publishing, Ltd.
- Lara, Frankie, Amanda Hahn, Na-Yung Yu, and Takashi Yamauchi. 2012. "Arbitrary Category Labels Can Change Similarity Judgments of Human Faces." Pp. 1179-1184 in *Proceedings of 34th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Leahey, E., C. M. Beckman, and T. L. Stanko. 2016. "Prominent but Less Productive: The Impact of Interdisciplinarity on Scientists Research." *Administrative Science Quarterly* 62: 105 – 139.
- Leskovec, J., A. Rajaraman, and J.D. Ullman. 2014. "Data Mining." in *Mining of Massive Datasets*, Cambridge: Cambridge University Press, 1-17.
- Leung, Ming, and Amanda Sharkey. 2013. "Out of Sight, Out of Mind? Evidence of Perceptual Factors in the Multiple-Category Discount." *Organization Science* 25:171-184.
- Lounsbury, Michael, and Hayagreeva Rao. 2004. "Sources of Durability and Change in Market Classifications: A Study of the Reconstitution of Product Categories in the American Mutual Fund Industry, 1944-1985." *Social Forces* 82:969-999.
- Lounsbury, Michael, and Mary M. A. Glynn. 2001. "Cultural Entrepreneurship: Stories, Legitimacy, and the Acquisition of Resources." *Strategic Management Journal* 22:545-564.
- Lupyan, Gary. 2008a. "The Conceptual Grouping Effect: Categories Matter (and Named Categories Matter More)." *Cognition* 108:566-577.
- , 2008b. "From Chair to "chair": A Representational Shift Account of Object Labeling Effects on Memory." *Journal of Experimental Psychology: General* 137:348-69.
- , 2012. "Linguistically Modulated Perception and Cognition: The Label-feedback Hypothesis." *Front Psychol* 3:54.
- Lupyan, Gary, David D. H. Rakison, and James J. L. McClelland. 2007. "Language Is Not Just for Talking: Redundant Labels Facilitate Learning of Novel Categories." *Psychol Sci* 18:1077-83.

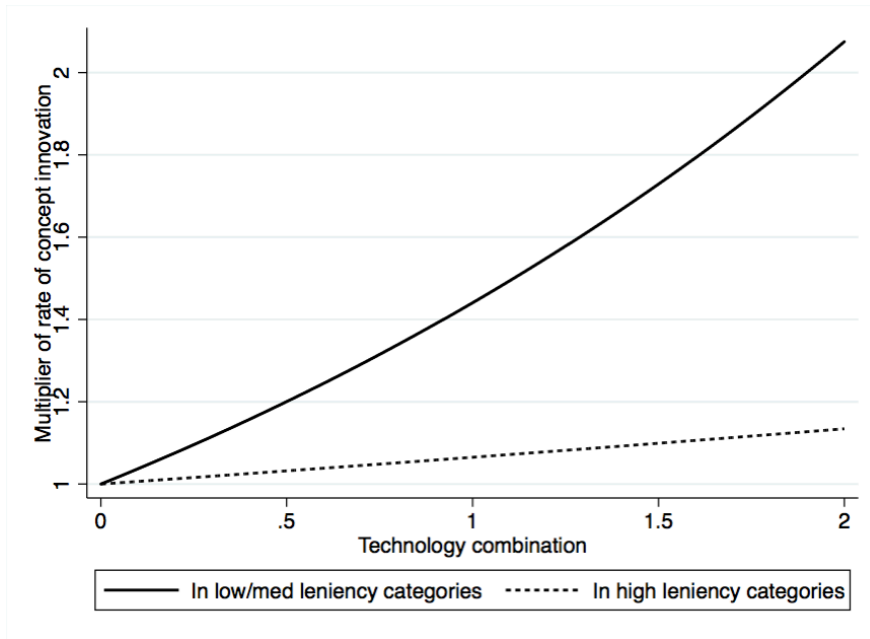
- Mann, Ronald J. 2005. "Do Patents Facilitate Financing in the Software Industry?" *Texas Law Review* 83:1-71.
- March, James G. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science* 2:71-87.
- McKendrick, David, and Glenn G. R. Carroll. 2001. "On the Genesis of Organizational Forms: Evidence From the Market for Disk Drive Arrays." *Organization Science* 12:661-682.
- Medin, Douglas L. and Lance J. Rips. 2005. "Concepts and Categories: Memory, Meaning, and Metaphysics." In *The Cambridge Handbook of Thinking and Reasoning*, edited by K. J. Holyoak and R. G. Morrison, Cambridge: Cambridge University Press 37-72.
- Meyer, John, and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure As Myth and Ceremony." *American Journal of Sociology* 83:340-63.
- Murphy, Gregory. 2004. *The Big Book of Concepts*. Cambridge, MA: MIT Press.
- Navis, Chad, and Mary M. A. Glynn. 2010. "How New Market Categories Emerge: Temporal Dynamics of Legitimacy, Identity, and Entrepreneurship in Satellite Radio, 1990-2005." *Administrative Science Quarterly* 55:439-471.
- Negro, Giacomo, Michael M. T. Hannan, and Hayagreeva Rao. 2010. "Categorical Contrast and Audience Appeal: Niche Width and Critical Success in Winemaking." *Industrial and Corporate Change* 19:1397-1425.
- Negro, Giacomo, Michael T. Hannan, and Hayagreeva Rao. 2011. "Category Reinterpretation and Defection: Modernism and Tradition in Italian Winemaking." *Organization Science* 22:1449-1463.
- Neisser, Ulric. 1989. *Concepts and Conceptual Development*. Cambridge: Cambridge University Press.
- Nelson, Richard, and Sidney Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge: Harvard University Press.
- Nerkar, Atul. 2003. "Old Is Gold? The Value of Temporal Exploration in the Creation of New Knowledge." *Management Science* 49:211-229.
- Park, Bernadette, and Reid Hastie. 1987. "Perception of Variability in Category Development: Instance-versus Abstraction-based Stereotypes." *Journal of Personality and Social Psychology* 53:621.
- Perretti, Fabrizio, Giacomo Negro, and Alessandro Lomi. 2008. "E Pluribus Unum: Framing, Matching, and Form Emergence in U.S. Television Broadcasting, 1940-1960." *Organization Science* 19:533-547.
- Phillips, Damon. 2013. *Shaping Jazz: Shaping jazz: Cities, labels, and the global emergence of an art form*. Princeton: Princeton University Press.
- PlayBigger. 2015. It's good to be king. Retrieved from <https://playbigger.com/category-kings>. Accessed on July 16, 2016.
- Podolny, Joel, Toby Stuart, and Michael M. T. Hannan. 1996. "Networks, Knowledge, and Niches: Competition in the Worldwide Semiconductor Industry, 1984-1991." *American Journal of Sociology* 102:659-689.
- Pollock, Neil, and Robin Williams. 2007. "Technology Choice and Its Performance: Towards a Sociology of Software Package Procurement." *Information and Organization* 17:131-161.
- , 2011. "Who Decides the Shape of Product Markets? The Knowledge Institutions That Name and Categorise New Technologies." *Information and Organization* 21:194-217.

- Pontikes, Elizabeth G. 2012. "Two Sides of the Same Coin: How Ambiguous Classification Affects Multiple Audiences' Evaluations." *Administrative Science Quarterly* 57:81-118.
- Pontikes, Elizabeth G, and Michael Hannan. 2014. "An Ecology of Social Categories." *Sociological Science* 1:311-343.
- Pontikes, Elizabeth G, and Ruben Kim. 2017. "Strategic Categorization." in *Research in the Sociology of Organizations. From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads*, edited by Rodolphe Durand, Granqvist, Nina and Anna Tyllstrom. 51: 71-111.
- Pontikes, Elizabeth G., and William W. P. Barnett. 2015. "The Persistence of Lenient Market Categories." *Organization Science* 26:1415 - 1431.
- Porac, Joseph F, and Howard Thomas. 1989. "Competitive Groups as Cognitive Communities: The Case of Scottish Knitwear Manufacturers." *Journal of Management Studies* 26: 397-416.
- Porac, Joseph F, and Howard Thomas. 1990. "Taxonomic Mental Models in Competitor Definition." *Academy of Management Review* 224-240.
- Porac, Joseph F, Howard Thomas, F Wilson, D Paton, and A Kanfer. 1995. "Rivalry and the Industry Model of Scottish Knitwear Producers." *Administrative Science Quarterly* 20:203-227.
- Porter, Michael E. 1980. *Competitive Strategy*. New York: Free Press.
- Powell, Walter W, and Kurt Sandholtz. 2012. "Chance, Necessite, Et Naivete: Ingredients to Create a New Organizational Form." in *The Emergence of Organizations and Markets*, edited by John Padgett and Walter W Powell. Princeton, New Jersey: Princeton University Press.
- Rao, Hayagreeva. 1994. "The Social Construction of Reputation: Certification Contests, Legitimation, and the Survival of Organizations in the American Automobile Industry: 1895-1912." *Strategic Management Journal* 15: 29-44.
- Rao, Hayagreeva. 1998. "*Caveat Emptor*: The Construction of Nonprofit Consumer Watchdog Organizations." *American Journal of Sociology* 103:912-961.
- Rao, Hayagreeva, Phillipe Monin and Rodolphe Durand. 2003. "Institutional Change in Toque Ville: Nouvelle Cuisine as an Identity Movement in French Gastronomy." *American Journal of Sociology* 108: 795-843.
- Rindova, Violina, Walter W. J. Ferrier, and Robert Wiltbank. 2010. "Value From Gestalt: How Sequences of Competitive Actions Create Advantage for Firms in Nascent Markets." *Strategic Management Journal* 31:1474-1497.
- Rosa, José Antonio, Joseph J. F. Porac, Jelena Runser-Spanjol, and Michael M. S. Saxon. 1999. "Sociocognitive Dynamics in a Product Market." *Journal of Marketing* 63:64-77.
- Rosch, Eleanor, and Carolyn Mervis. 1975. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7:573-605.
- Rosenkopf, L, and A Nerkar. 2001. "Beyond Local Search: Boundary-spanning, Exploration, and Impact in the Optical Disk Industry." *Strategic Management Journal* 22:287-306.
- Ruef, Martin. 2000. "The Emergence of Organizational Forms: A Community Ecology Approach." *American Journal of Sociology* 106:658-714.
- Ruef, Martin, and Kelly Patterson. 2009. "Credit and Classification: Defining Industry Boundaries in Nineteenth Century America." *Administrative Science Quarterly* 54:486-520.

- Santos, Filipe, and Kathleen Eisenhardt. 2009. "Constructing Markets and Shaping Boundaries: Entrepreneurial Power in Nascent Fields." *Academy of Management Journal* 52:643-671.
- Schumpeter, Joseph. 1934. *The Theory of Economic Development*. Cambridge, MA: Harvard University Press.
- Schumpeter, Joseph. 1942. *Capitalism, Socialism, and Democracy*. New York: Harper & Brothers Publishers.
- Sloutsky, Vladimir M. 2003. "The Role of Similarity in the Development of Categorization." *Trends in Cognitive Sciences* 7:246-251.
- Sloutsky, Vladimir M, Ya-Fen Lo, and Anna A. V. Fisher. 2001. "How Much Does a Shared Name Make Things Similar? Linguistic Labels, Similarity, and the Development of Inductive Inference." *Child Development* 72:1695-1709.
- Smith, Edward Bishop. 2011. "Identities As Lenses: How Organizational Identity Affects Audiences' Evaluation of Organizational Performance." *Administrative Science Quarterly* 56:61-94.
- Soltes, Eugene. 2009. "News Dissemination and the Impact of the Business Press." *Dissertation, the University of Chicago*.
- Stark, David. 1996. "Recombinant Property in East European Capitalism." *American Journal of Sociology* 101: 993-1027.
- Suarez, Fernando F., Stine Grodal, and Aleksios Gotsopoulos. 2014. "Perfect Timing? Dominant Category, Dominant Design, and the Window of Opportunity for Firm Entry." *Strategic Management Journal*.
- Tenenbaum, Joshua B, Thomas T. L. Griffiths, and Charles Kemp. 2006. "Theory-based Bayesian Models of Inductive Learning and Reasoning." *Trends Cogn Sci* 10:309-18.
- Tingling, Peter, and Michael Parent. 2004. "An Exploration of Enterprise Technology Selection and Evaluation." *The Journal of Strategic Information Systems* 13:329-354.
- Trajtenberg, Manuel. 1990. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *The RAND Journal of Economics* 21:172.
- Tushman, Michael and Phillip Anderson. 1986. "Technological Discontinuities and Organizational Environments." *Administrative Science Quarterly* 31: 439-465.
- Utterback, James M, and Fernando F. F. Suarez. 1993. "Innovation, Competition, and Industry Structure." *Research Policy* 22:1-21.
- Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. "Atypical Combinations and Scientific Impact." *Science* 342:468-72.
- Wang, Ping. 2009. "Popular Concepts Beyond Organizations: Exploring New Dimensions of Information Technology Innovations." *Journal of the Association for Information Systems* 10:1-30.
- Weick, Karl. 1995. *Sensemaking in Organizations*. London: Sage.
- Winawer, Jonathan, Nathan Witthoft, Michael M. C. Frank, Lisa Wu, Alex A. R. Wade, and Lera Boroditsky. 2007. "Russian Blues Reveal Effects of Language on Color Discrimination." *Proceedings of the National Academy of Sciences* 104:7780-7785.
- Vedres, Balazs and David Stark. 2010. "Structural Folds: Generative Disruption in Overlapping Groups." *American Journal of Sociology* 114: 1150-1190.
- Vygotsky, Lev. [1934] 2012. *Thought and Language*: MIT Press.
- Zuckerman, Ezra W. 1999. "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount." *American Journal of Sociology* 104:1398-1438.

Figure 2. Effects of combination across market categories, by category leniency.





¹ Plot based on estimates from table 5, columns 1, 2, and 5. Plots display x-axis range until one standard deviation above the mean.

Tables

Table 1. Sample press release identity statements with market category affiliations.

Organization	Date	Description
3D Systems	January 1994	3D systems develops, builds and markets solid imaging systems using a process known as stereolithography.
Manguistics	October 1994	Manguistics is the leading provider of software and services for supply chain management .
MicroStrategy	January 1996	MicroStrategy is the leading provider of relational OLAP (ROLAP) products and services for developing and accessing enterprise data warehouses .
VCON	October 1996	VCON is one of the leading manufacturers and marketers of desktop videoconferencing hardware and software products in the industry.
ACI	June 1997	Applied Communications, Inc (ACI), a leading provider of electronic payment software around the world, ...
Carnegie Group	October 1998	Carnegie Group is a provider of customer relationship management and advanced decision support solutions.
AVT	March 1999	AVT Corporation, a world leader in managed communications solutions .
Plasmon	November 2000	Plasmon, the leading provider of removable data storage products, plans to release ...
CardSoft	November 2000	CardSoft Inc., a leading provider of open platform software on small electronic devices.
Veridicom	November 2001	Veridicom, Inc. is a leader in fingerprint-based biometrics solutions.
Adonix	November 2002	Adonix delivers sophisticated ERP software to the underserved Fortune 50,000 of mid-market companies.

Table 2. Descriptive statistics and correlations, all firms (N=18,192 firm-years)

	Mean	Std. Dev	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Concept innovation	0.0433	0.2034	0	1												
(2) Firm combination	1.994	1.739	0	4.630	0.10											
(3) Category combination	0.8519	0.6285	0	3.738	0.15	0.62										
(4) Technology combination	0.1752	0.5602	0	4.153	0.10	0.13	0.17									
(5) Leniency	1.592	1.146	0	4.012	0.08	0.63	0.72	0.08								
(6) No. other firms in categories (weighted; logged)	2.629	1.861	0	6.827	0.10	0.64	0.86	0.14	0.88							
(7) Rank in Software 500	0.5261	1.465	0	7	0.14	0.23	0.35	0.15	0.21	0.26						
(8) Public	0.2706	0.4443	0	1	0.08	0.18	0.25	0.19	0.14	0.20	0.29					
(9) Received VC funding in previous year	0.0773	0.2671	0	1	0.04	0.19	0.16	0.03	0.19	0.17	-0.03	-0.08				
(10) Number of patents issued to organization / 100	0.0334	0.4629	0	24.3	0.06	0.04	0.08	0.15	0.02	0.05	0.08	0.08	-0.02			
(11) Number of citations to organizations' patents / 100	0.2346	3.898	0	248.2	0.05	0.04	0.08	0.13	0.03	0.05	0.09	0.06	-0.01	0.92		
(12) Firm tenure	2.731	2.791	0	13	0.06	0.34	0.54	0.16	0.50	0.53	0.30	0.34	-0.02	0.09	0.11	
(13) Year	1998	3.001	1990	2002	-0.07	0.28	0.18	-0.08	0.39	0.30	-0.02	-0.03	0.10	-0.03	0.00	0.23

Table 3. Descriptive statistics and correlations, patenting firms (N=4,012 firm-years)

	Mean	Std. Dev	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Concept innovation	0.0753	0.2639	0	1												
(2) Firm combination	2.710	1.438	0	4.578	0.04											
(3) Category combination	1.210	0.5276	0	3.738	0.12	0.31										
(4) Technology combination	0.7946	0.9648	0	4.153	0.11	0.01	-0.01									
(5) Leniency	2.024	0.793	0	3.876	-0.01	0.26	0.35	-0.15								
(6) No. other firms in categories (weighted; logged)	3.523	1.312	0	6.827	0.03	0.32	0.71	-0.05	0.69							
(7) Rank in Software 500	1.051	2.047	0	7	0.14	0.20	0.41	0.08	0.19	0.30						
(8) Public	0.5399	0.4985	0	1	0.07	0.15	0.21	0.01	0.08	0.15	0.27					
(9) Received VC funding in previous year	0.0815	0.2736	0	1	0.02	0.08	0.01	0.07	0.05	0.04	-0.05	-0.13				
(10) Number of patents issued to organization / 100	0.1514	0.9767	0	24.3	0.08	0.03	0.10	0.08	-0.02	0.05	0.08	0.06	-0.04			
(11) Number of citations to organizations' patents / 100	1.064	8.248	0	248.2	0.07	0.04	0.11	0.08	0.01	0.07	0.11	0.05	-0.03	0.92		
(12) Firm tenure	4.888	2.942	0	13	-0.03	0.07	0.29	-0.21	0.29	0.29	0.25	0.22	-0.15	0.08	0.12	
(13) Year	1998	2.867	1990	2002	-0.12	0.29	0.18	-0.27	0.56	0.41	0.04	0.00	0.04	-0.07	-0.01	0.42

Table 4. Tests of hypotheses 1-2. Piecewise continuous hazard rate models on firm rates of concept innovation. Models run on all firms and patenting firms.¹

	(1)	(2)	(3)	(4)	(5)	(6)
	All firms	All firms	All firms	All firms	All firms	Patenters
Category leniency		0.447 (0.0802)	0.579 (0.0790)	0.485 (0.0811)	0.498 (0.0812)	0.323 (0.121)
<u>Market category combination</u>						
Category combination			1.056 (0.103)	0.975 (0.104)	0.950 (0.103)	0.757 (0.158)
Firm combination				0.239 (0.0371)	0.235 (0.0373)	0.0722 (0.0608)
Technology combination					0.155 (0.0477)	0.203 (0.0680)
<u>Controls</u>						
No. other firms in categories (weighted; logged)	0.371 (0.0272)	0.193 (0.0429)	-0.171 (0.0567)	-0.184 (0.0554)	-0.185 (0.0553)	-0.200 (0.0898)
Rank in Software 500	0.162 (0.0210)	0.158 (0.0209)	0.113 (0.0200)	0.109 (0.0195)	0.104 (0.0196)	0.114 (0.0284)
Public	0.259 (0.0870)	0.285 (0.0879)	0.270 (0.0831)	0.215 (0.0825)	0.178 (0.0834)	0.225 (0.130)
Received VC funding in previous year	0.678 (0.111)	0.631 (0.110)	0.542 (0.110)	0.492 (0.109)	0.469 (0.109)	0.224 (0.184)
Number of patents issued to firm / 100	0.235 (0.0980)	0.239 (0.0914)	0.205 (0.0905)	0.187 (0.0818)	0.162 (0.0856)	0.149 (0.0700)
Number of citations to firm's patents / 100	-0.0215 (0.0099)	-0.0206 (0.0092)	-0.0183 (0.0089)	-0.0171 (0.0081)	-0.0151 (0.0084)	-0.0112 (0.0070)
Firm tenure	0.0458 (0.0166)	0.0412 (0.0167)	0.0121 (0.0169)	0.0155 (0.0170)	0.0107 (0.0171)	-0.0264 (0.0253)
Duration: 0-1 years	-3.450 (0.227)	-3.453 (0.227)	-3.935 (0.230)	-4.099 (0.229)	-4.054 (0.228)	-2.501 (0.425)
Duration: 1-2 years	-3.256 (0.224)	-3.368 (0.228)	-3.738 (0.223)	-3.917 (0.222)	-3.877 (0.222)	-2.482 (0.359)
Duration: 2-5 years	-3.708 (0.233)	-3.812 (0.236)	-4.064 (0.235)	-4.056 (0.234)	-4.018 (0.234)	-3.011 (0.403)
Duration: 5-10 years	-4.305 (0.387)	-4.413 (0.387)	-4.548 (0.387)	-4.436 (0.388)	-4.380 (0.388)	-2.838 (0.555)
Duration: 10+ years	-16.31 (0.461)	-16.42 (0.462)	-16.19 (0.465)	-15.64 (0.539)	-16.07 (0.544)	-13.56 (1.063)
Log pseudolikelihood	-2662.2	-2644.8	-2594.0	-2568.5	-2563.4	-886.8
Degrees of freedom	24	25	26	27	28	28
N	18192	18192	18192	18192	18192	4012

¹ Models run on 4,566 firms over 18,192 firm-years with 787 concept innovation events for all organizations, 789 firms over 4,012 firm-years with 302 events for patenters. Two-tailed tests. Year dummies are included in all models.

Table 5. Tests of hypothesis 3. Piecewise continuous hazard rate models on firm rates of concept innovation. Models run on all firms and patenting firms. Controls included in all models.¹

	(1)	(2)	(3)	(4)	(5)
	All firms	All firms	All firms	Patenters	Patenters
Category leniency	0.909 (0.102)	0.978 (0.102)	0.574 (0.0844)	0.454 (0.150)	
Organization is in high-leniency categories					0.589 (0.191)
<u>Market category combination</u>					
Category combination	1.991 (0.187)	0.896 (0.105)	0.957 (0.103)	0.762 (0.159)	0.718 (0.159)
Category combination x category leniency	-0.483 (0.0811)				
Firm combination	0.181 (0.0371)	0.610 (0.0676)	0.221 (0.0371)	0.0706 (0.0606)	0.0716 (0.0607)
Firm combination x category leniency		-0.204 (0.0311)			
Technology combination	0.152 (0.0468)	0.154 (0.0469)	0.666 (0.152)	0.468 (0.208)	
Technology combination x category leniency			-0.246 (0.0694)	-0.127 (0.0936)	
Technology combination, firm in low/med leniency categories					0.365 (0.0914)
Technology combination, firm in high leniency categories					0.0630 (0.0831)
Log pseudolikelihood	-2546.3	-2545.0	-2557.3	-885.9	-885.2
Degrees of freedom	29	29	29	29	29

¹ Models run on 4,566 firms over 18,192 firm-years with 787 concept innovation events for all organizations, 789 firms over 4,012 firm-years with 302 events for patenters. Two-tailed tests. Controls for number of other firms in categories, rank in Software 500, public, received VC funding in the previous year, number of patents issued to the firm, number of citations to firm's patents, firm tenure, year dummies, and duration pieces for 0-1, 1-2, 2-5, 5-10, and 10+ years since a firm joined or created a category included in all models.

Table 6. Supplementary analyses. Piecewise continuous hazard rate models on firm rates of concept innovation. Models run on all firms and patenting firms. Controls included in all models.¹

	(1)	(2)
	All firms	Patenters
Category leniency	0.960 (0.107)	
Organization is in high-leniency categories		0.595 (0.195)
<u>Market category combination</u>		
Category combination	1.061 (0.190)	0.711 (0.158)
Firm combination	0.647 (0.0693)	0.0716 (0.0607)
Firm combination x category leniency	-0.201 (0.0321)	
Language similarity to firm's categories	-0.258 (0.241)	
Technology combination	0.155 (0.0470)	
Technology combination, firm in low/med leniency categories		0.360 (0.0940)
Technology combination, firm in high leniency categories		0.0534 (0.103)
Technology similarity to firm's categories		0.0444 (0.225)
Log pseudolikelihood	-2544.4	-885.2
Degrees of freedom	30	30

¹ Models run on 4,566 firms over 18,192 firm-years with 787 concept innovation events for all organizations, 789 firms over 4,012 firm-years with 302 events for patenters. Two-tailed tests. Controls for number of other firms in categories, rank in Software 500, public, received VC funding in the previous year, number of patents issued to the firm, number of citations to firm's patents, firm tenure, year dummies, and duration pieces for 0–1, 1–2, 2–5, 5–10, and 10+ years since a firm joined or created a category included in all models.

Table 7. Supplementary analyses. Piecewise continuous hazard rate models on firm rates of concept innovation using alternate dependent variables. Results reported for combination covariates. Models run on all firms and patenting firms.¹

	<i>DV: Excludes fast growth categories</i>	<i>DV: Categories that diffuse</i>	<i>DV: Categories reported by Gartner</i>
<i>All firms; same covariates as table 5 column (1)</i>			
Leniency	0.876 (0.136)	0.969 (0.121)	1.080 (0.153)
Category combination	2.192 (0.244)	1.949 (0.215)	1.955 (0.312)
Category combination x category leniency	-0.617 (0.117)	-0.453 (0.0950)	-0.440 (0.135)
Number of events:	459	613	426
<i>All firms; same covariates as table 5 column (2)</i>			
Leniency	0.892 (0.125)	1.066 (0.118)	1.215 (0.135)
Firm combination	0.662 (0.0868)	0.609 (0.0785)	0.628 (0.102)
Firm combination x category leniency	-0.232 (0.0407)	-0.210 (0.0369)	-0.214 (0.0430)
Number of events:	459	613	426
<i>Patenters 1995 - 2002; same covariates as table 5 column (5)</i>			
Organization is in high-leniency categories	0.579 (0.250)	0.472 (0.221)	0.620 (0.255)
Technology combination, firm in low/med leniency categories	0.378 (0.114)	0.382 (0.107)	0.462 (0.120)
Technology combination, firm in high leniency categories	0.0464 (0.124)	0.137 (0.0948)	0.205 (0.105)
Number of events:	179	238	169

¹ Models run on 4,566 firms over 18,192 firm-years for all organizations, 789 firms over 4,012 firm-years for patenters. Two-tailed tests. Models include covariates as reported in table 5, as indicated.