Fitting in or Starting New? An Analysis of Invention, Constraint, and the Emergence of New Categories in the Software Industry

Elizabeth George Pontikes

October 2008

Acknowledgements: I would like to thank Bill Barnett, Mike Hannan, Hayagreeva Rao, Jesper Sørensen, Glenn Carroll, Martin Ruef, Greta Hsu, and Jerker Denrell for helpful comments and suggestions, and the University of Chicago GSB for support.

ABSTRACT

How do new organizational categories come about? In this study, I suggest that both novel invention and constraints from existing categories affect whether new organizational categories will emerge. Novel invention is perceived differently depending on whether the existing category structure is constraining or lenient. Constraining labels and categories create clear expectations, making them less accommodating to novel inventions. I hypothesize that organizations that are *both* different in their knowledge creation *and* are affiliated with constraining (low leniency) category labels are likely to claim a new identity label. I study these ideas in the context of the software industry, for the years 1990-2002. I use patent data to identify novel invention, and create a data set using software press releases to identify market categorization. Results support the hypothesis. Firms that are unique in their knowledge development are more likely to create new market categories when they are affiliated with constraining (low leniency) categories or labels.

Between 1990 and 2002, software organizations used 394 different labels to describe themselves and they introduced 325 of these during that time period. Managers scrambled to provide the next “new thing” whether it be “CRM,” “e-busniess,” or “digital imaging.” Why did organizations use so many new labels to describe their identities? Were they so radically different from existing labels that they needed to create these new identities, or was trying to appear novel simply compelling? Organizations that present novel practices or products can potentially change their environments and have a profound impact on society. On the other hand, attempts to be novel often fail, and there are strong pressures for organizations to fit within existing systems. In light of this, what leads to novelty in an organizational environment?

This question is pertinent to another important question in organizations theory, namely the extent to which organizations create or conform to their surroundings. A large body of research in population ecology and institutional theory shows that organizations are shaped by their environments (Hannan and Freeman, 1977;DiMaggio and Powell, 1991;Meyer and Rowan, 1991). Political, social, and normative structures influence which structures and functions organizations build and incorporate (Edelman, 1992;Fligstein, 1996). Other literatures take a more active view of organizations, depicting managers as attempting to exploit an organization’s unique qualities to shape its environment (Thompson, 1967;Pfeffer and Salancik, 1978;Weick, 1979). An example of this view today comes within the literature on strategic management, where researchers identify organizational capabilities that differentiate them from their competitors in an effort to improve performance (Prahalad and Hammel, 1990;Teece, Pisano, and Shuen, 1997). At the heart of these discussions is the question of whether or when an organization will introduce novelty into its environment or conform to existing systems and structures.

Yet another literature looks at where novelty comes from for an agentic organization. In the tradition of Schumpeter (1947), researchers in organizational knowledge and technology conceive of novelty as resulting from external shocks to an environment in the form of new technical developments. Researchers distinguish between different types of inventions that are more or less likely to spur the creation of a new label or category in market space. For example, there are competence-destroying compared to competence-enhancing innovations (Tushman and Anderson, 1986), process versus product inventions (Levin and Reiss, 1988), or exploration as opposed to exploitation (March, 1991). In this perspective, new organizational knowledge is the seed for new categories.

Together, these perspectives suggest that there are pressures to both conform and differentiate, and that the source of differentiation may come from exogenous shocks, such as new technological inventions. The organizational world shows evidence for all of these outlooks. Organizations are not merely conformist; they introduced or facilitated the diffusion of society changing technologies such as electricity, the telephone, the personal computer and the internet. At the same time, much organizational activity is routine, and even when organizations try to introduce something new often these attempts are not considered to be especially novel. In light of these different views, how can we understand the creation of novelty within an organizational environment? Here I argue novelty is a result of both deliberate actions on behalf of the organization and the environment where these actions occur. It is a function of the entity created by the organization– be it a technology, process, or new idea – as well as the context into which it is introduced. Novelty depends on an intersection of the "new thing" that is created and the social structure in which it is embedded.

I investigate these ideas through an empirical examination of invention and categories in the software industry for the years 1990-2002. The software industry contains a myriad of "spaces," or labels that classify software companies in a way that is both meaningful and difficult to define. I investigate organizations in this industry with respect to two domains: knowledge space and market space. Knowledge space is the arena in which organizations invent and develop new knowledge. Market space is where organizations identify with labels or categories as companies that provide specific products or services (figure 1). I suggest that studying the interaction between organizational knowledge and category constraints in market space is necessary to understanding the evolution of the category structure. Below, I argue that constraints imposed by existing labels and categories affect whether novelty in knowledge space leads organizations to introduce new labels into market space.

Market Space Categorization

The existing category structure is a lens through which people view the world. As a result, an object that does not conform to the expected structure may be misunderstood. Different categories present different logics of competition for organizations in terms of how organizations compete and the criteria for success or failure (Barnett, 2008). Success in one context does not imply success in another; in fact success in one product market can contribute to failures in another (Barnett and Pontikes, 2008). In addition, organizations that simultaneously straddle categories tend to fare badly: organizations that do not have a focused identity in an established category are devalued on the stock market (Zuckerman, 1999), and films that get classified in multiple genres are not reviewed as highly (Hsu, 2006). Categories help to set expectations to which organizations often conform.

On the other hand, research in strategic management has focused on how firms can differentiate themselves or their products from competitors (Porter, 1985;Prahalad and Hammel, 1990;Teece, et al., 1997). A differentiated organization has the chance to be evaluated more favorably by prospective customers or employees. But in order to be favorably evaluated an organization must first be evaluated. If an organization is perceived to be exceptionally different from a group of competitors it may not be considered at all. It is important for organizations to be both comparable with and differentiated from others (Callon and Muniesa, 2005). Organizations tend to be evaluated in a two-stage process, where one that does not conform to expectations risks being ignored, but one that does not differentiate from its reference group risks not being selected (Zuckerman, Kim, Ukanwa, and von Rittman, 2003). How an organization is categorized can affect the resources it will attract.

Categorization is an important human process that extends beyond organizations. Categories allow people to access a large amount of information by grouping together objects using category schemata (Rosch and Lloyd, 1978). They also influence people's expectations and, as a result, their evaluations (Osherson and Smith, 1982). People perceive groups to be less variable when they are given a common label (Park and Hastie, 1987), and objects grouped under a common label are also graded in terms of their typicality (Rosch and Mervis, 1975;Rosch and Lloyd, 1978). For instance, "chair" and "table" are typical types of "furniture," whereas "rug" might be considered a partial, or atypical, member of that category.

An organization is categorized by many different audiences such as potential customers, investors, partners, or employees, and its identity is a function of both the organization and the relevant audience (Hsu and Hannan, 2005;Hannan, Pólos, and Carroll, 2007). For example, an organization that sells books online could be considered an "e-business" by technology experts, a "bookstore" by stock market analysts, and an "e-book retailer" by its employees and competitors.

If a company is looking to acquire an e-business, or if an investor wants to buy shares of a bookstore, or if a software engineer wants to work at an e-book retailer, then whether the organization is viewed as competent and legitimate in the relevant categories can determine its fate. Employees within organizations are one type of audience. They select a competitive reference group and position the organization against these competitors based on their perceptions of the organization's identity (Porac and Thomas, 1990) in a process that often replicates the existing market structure (White, 1981;Leifer and White, 1987).

Categorization systems seem to be fixed at any one point, but over time categories change and new categories emerge. Hannan, Pólos and Carroll (2007) formalize the emergence of organizational categories and provide a useful framework for analyzing classification systems. They suggest that in the first stage audience members cluster organizations by similarity and label these clusters. Next, an extensional consensus might develop, where audience members agree on which organizations belong to a label, but do not agree about the meaning of the label. Audiences agree on the meaning of a label in the next stage, which can be thought of as the label's schema that indicates the specific set of features that people expect affiliated organizations to possess. In this framework, schematized labels are called categories (Hannan, et al., 2007). This describes the longitudinal progression of category formation. But at any given point in time, an industry's classification system may include similarity clusters, labels, and categories.

Categorization: Constraint and Leniency

Not all labels and categories have the same impact on their members or on relevant audiences. One way they can differ is in the extent to which they create strong expectations about member organizations. This partly depends on how much audiences agree that members of a category or label should conform to a schema. For example, when the “minivan” category for automobiles was emerging, before there was a stable consensus about which product features a “minivan” should have, many different models were rated favorably. After a schema emerged for this category, the acceptability of particular models changed even though the automobiles did not change (Rosa, Porac, Runser-Spanjol, and Saxon, 1999). When categories of mutual funds became more ambiguous due to variability among products, new categories were more likely to emerge (Lounsbury and Rao, 2004). In the health care arena, the identities, aggregate numbers, and size of existing organizational forms influenced the emergence of new organizational forms (Ruef, 2000). Categories create expectations that influence people’s evaluations about organizations that claim membership, or organizations that claim to be something new.

Much previous research on organizational categorization investigates categories and forms; cases where there is widespread agreement among relevant audiences about the schema, or specific codes, for each category. In these situations, researchers find that there are penalties for boundary spanning and non-conformity (Zuckerman, 1999;Hsu, 2006;Hsu, Hannan, and Koçak, 2007). For example, if a person expects to see a comedy but instead is presented with a horror film, she is less likely to favorably rate the film. Boundary spanners attempt to meet expectations arising from multiple categories, but in doing so risk violating the expectations of audiences of any particular category. As a result, when there are widely held expectations that organizations should fit into established categories, organizations will be more likely to conform.

On the other hand, in early stages of category formation there is not as much agreement on how organizations should be classified. In these situations, individuals within and outside organizations are active in shaping the meanings of existing labels (DiMaggio, 1991) or in creating new labels (Fligstein, 1985;Santos and Eisenhardt, 2008). Taken together, these studies imply that early stage categories foster different types of activities with respect to boundary definition. These literatures suggest that established categories induce conformity, whereas nascent categories are malleable.

But when are categories early as opposed to late stage? Rather than evaluate categories according to a timeline, I propose that at any stage it is the amount of constraint imposed by categories or labels that affects how individuals navigate the classification structure. Labels that impose high constraints and that are well known tend to be categories. Labels that are low constraint, or lenient labels, may be early stage or unknown, but they need not be. In fact, an interesting case emerges when an organizational label gains widespread acceptance, but when audiences do not develop a clear meaning for the label. The label may be discussed by analysts, featured in the press, or adopted widely by other organizations, but it does not evoke specific expectations for member. This type of label is not nascent or emerging; it is established. How does it function differently from constraining categories?

An example is the label "e-business" that emerged in the late 1990s in the software industry. This label referred to organizations that sold products or services over the World Wide Web. It was widely recognized, promoted in the press, and adopted by producers. But it never developed a consensual meaning outside of its actual definition, which was doing business electronically. As the label became part of the public vernacular, businesses increasingly began to claim to be "e-businesses" and that affiliation provided some level of legitimacy from customers and financiers. However, no specific agreed-upon codes emerged to indicate what it meant for an organization to claim an affiliation with this label. Nevertheless, the label is not illegitimate or immature, nor is it a super-category in a nested hierarchy. It is a label that diffused broadly but that did not become meaningful. This type of label is neither a traditional organizational category that provides both legitimacy and constraint, nor is it an emerging category with opportunities for entrepreneurs to define the terrain. It is an established but lenient label. Whether labels or categories are constraining or lenient changes the expectations of audiences. Because of this, leniency or constraint affects whether member organizations will try to make a novel identity claim within market space.

New Category Creation

Existing institutions, forms, and categories are the platform on which new categories are constructed. New categories are created by either recombining of elements from existing categories (Rosa, et al., 1999;Lounsbury and Rao, 2004;Phillips and Owens, 2004), or by carving out a new space in opposition to existing categories (Carroll and Swaminathan, 2000;Rao, Monin, and Durand, 2003;Barnett, 2004). For example, the category "minivan" emerged through the combination of elements from the existing categories of "cars" and "trucks" (Rosa, et al., 1999). Incumbent record companies combined elements of traditional music with jazz music to create the emerging popular music category of jazz, a case where the hybrid music retained the original label "jazz," and the original form took an alternate name "hot jazz" (Phillips and Owens, 2004). In the health care arena, the number and size of forms with similar identities increased the likelihood that a new form would emerge (Ruef, 2000). On the other hand, the new category of "nouvelle" French cuisine rose in opposition to the existing "classical" cuisine, defying the schema for traditional French cuisine by allowing techniques and ingredients that were forbidden in the traditional cuisine (Rao, et al., 2003).

I study the creation of new organizational categories in terms of when an organization will use a new label to describe its identity. This label may later be picked up by other organizations and become an accepted category, or in the opposite extreme it may be ignored. I use press releases press releases that are created by employees within the organization as a source for an organization's identity claims to be affiliated with a label or category. These claims represent the insider audience's representation of the organization's identity to the public. The identity of an organization as seen by the audience of employees can be especially important, as it influences how employees will interpret external situations, and how they will direct the organization to respond (Dutton and Dukerich, 1991). The competitive group with which executives identify the organization will influence the dimensions managers will emphasize as they compete with others in their reference group. The identity claims will not necessarily be accepted by target audiences, but they are serious assertions that place the organization in an existing category or that create a new label with the intent to carve out a new organizational category. I focus on categorization at the organizational level, which in some instances may coincide with product categories. I include product categories only when the product category is claimed as an organizational identity.

In some industries, where gatekeepers play an active role in maintaining the classification structure, it might be the gatekeepers who first introduce a new label and producer organizations that adopt the identity. In other industries, including the software industry, producer organizations are on the forefront of category definition, and they even use categorization as a strategic weapon to try to define a niche that they can dominate (Santos and Eisenhardt, 2008). Although analysts can introduce labels to describe a group of organizations, these labels might or might not get picked up by software companies as identifiers of a new organizational category.

Constraint and New Categories

Part of the process of new category creation is differentiating the new label from existing organizational categories. Much previous research on new category creation presumes that these existing categories are constraining. Categories that are formed through recombination bring together elements from related categories when the combination of elements is incompatible with the initial categories. Categories that are formed through opposition use the opposite elements of a well-defined category. In both cases, agreed upon codes and boundaries for existing categories provide a structure against which new categories can be judged. Indeed, the disk array label never evolved into an organizational category, even though there were activists setting standards and building associations around the disk array product. Because organizations from many other organizational categories were selling disk arrays, they could not form a common identity that was set apart against a common standard (McKendrick and Carroll, 2001). The constraint presented by existing categories – which makes existing structures inflexible – is what facilitates the creation of new categories.

Whether labels or categories are constraining or lenient is fundamental to how categorization impacts organizations. Constraint should be understood independently of category maturity, size, or hierarchical nesting. Although it may be related to these constructs, it is not equivalent to them and is more fundamental. How does the amount of constraint imposed by existing labels and categories affect organizations? Specifically, how does it impact the ways that organizations use knowledge to differentiate themselves from their competitors?

Knowledge Space

In contrast to market space arguments, previous research in the knowledge space perspective argues that the creation of new knowledge, or technical invention, leads to new categories or labels in market space. Much of this literature builds on Schumpeter's (1947) assertion that existing structures are changed through the introduction of novel technologies, or "creative destruction." This perspective takes an active view of the organization, with entrepreneurs introducing new inventions into the existing flow of the market.

Research in this tradition focuses less on the link between technological novelty and market disruptions as on factors that contribute to the development of novel technologies. Firm concentration (Levin, Cohen, and Mowery, 1985;Gilbert, 2006) technological opportunities across industries (Scherer, 1967;Levin, et al., 1985;Jaffe, 1986), and firm age (Tushman and Anderson, 1986;Sørensen and Stuart, 2000) all affect the likelihood that an organization will create novel inventions. In addition, inventions that draw on different knowledge as compared to previous inventions tend to make contributions that are higher variance than those that are more similar to what has come before (Fleming, 2001). This means that inventions that end up being important do tend to have been different from what came before – more novel – but so were inventions that are very unimportant. But do these important inventions lead to market space disruptions? Although many studies in the knowledge space perspective indicate novel technological development will benefit an organization through creative destruction, few have directly tested this relationship.

Studies do show that organizational positions within knowledge space are linked to categorization in market space. An organization's knowledge can help refine its market niche and, by extension, how it is classified (Podolny, Stuart, and Hannan, 1996;Stuart and Podolny, 1996;Carroll and Swaminathan, 2000). Further, organizational knowledge is an important competency that managers can use to differentiate from competitors in market space (Henderson and Clark, 1990;Teece, et al., 1997;Tripsas, 1997). In addition, studies of particular markets show that some technological inventions change existing structures and lead to new categories in market space (Tushman and Anderson, 1986), supporting the idea that novel knowledge development can lead to disruptions in market space categorization.

However, the relationship between positions knowledge space and in market space often is not a straightforward one-to-one mapping. In early telephony, whether organizations competed or had mutualistic effects on one another in market space depended on interactions between technological standardization, geographical location, and organizational form (Barnett and Carroll, 1987;Barnett, 1990). After aircraft engine control producers outsourced several aspects of production, some retained a broad footprint in knowledge space while others cultivated a narrower focus. This depended on the predictabilities and interdependencies of their components in market space (Brusoni, Prencipe, and Pavitt, 2001). In the biotech industry, knowledge ties between organizations in different market space categories has created a unique network-based categorization structure (Powell, Koput, and Smith-Doerr, 1996;Powell, Koput, White, and Owen-Smith, 2005). How organizational knowledge affects market space categorization partly depends on the existing classification structure.

Interaction: Novel Invention and Lenient Labels

I argue that the creation of a new label depends on both novel organizational knowledge and the levels of constraint or leniency within the existing classification structure. I propose that novelty is determined by perceptions of people who are steeped in an existing context and that novel technology cannot be measured independently of an environment's category structure. Rather than attempt to refine the definition of knowledge space novelty by creating more finegrained technical distinctions, I suggest that we consider both the technical newness of an invention as well as the constraints arising from the existing classification structure.

The extent to which existing categories provide or lack constraint will influence how a novel invention will be received. Classification systems help people navigate the organizational world. At the same time, they create expectations about what types of activities are appropriate for category members. In order to derive benefits associated with membership in a category, organizations must comply with expectations. When a category is clearly defined and schematized, there is a widespread consensus of what a member should or should not do. This type of category presents constraint; if an organization is to continue to be associated with such a category, its activities will be restricted. For an organization in a constraining category, the creation of very different types of knowledge could violate expectations and push an organization out of a category. In addition, constraint creates negative space within the environment. By clearly defining what a member ought to do, relevant observers can infer that other activities are distinct from existing structures, helping audiences recognize the difference between the old categories and the new category. Constraining categories both *push* and *pull* an organization with novel knowledge into a new category.

On the flip side, when there is not a widespread consensus about what affiliates of a label should or should not do, the label is lenient and members are less constrained. They will be able to engage in a wide range of activities, including knowledge development, and their affiliation with the label will not be questioned. In addition, because there is not a clearly accepted notion of what a member should do, there is also not as much negative space, which makes it less likely that a new label will be distinguishable from existing structures. Therefore I hypothesize:

Hypothesis: When organizations are both different from others in knowledge space, and are in constraining (low leniency) categories or labels, they are more likely to create a new market category or label.

Empirical Context: The Software Industry

I investigate this hypothesis within the context of the software industry. The software industry has generally been elusive to researchers. It developed under the public radar, was shaped by many independent organizations, and its products are intangible. This industry provides a lot of variation in how much constraint was imposed by different labels or categories. Software companies have typically taken the lead in creating and validating labels and categories within the industry. Analysts take on the nomenclature a little bit later, and financial markets have yet to catch up.[1] In addition, it is an environment where knowledge was important, where many different types of knowledge were relevant, and where organizations continually created new technologies.

The software industry was not referred to as such until the late 1960s but software has been around since just after computers were commercialized, in the 1950s. In 1957 the creation of FORTRAN, the first higher-level programming language, allowed programmers to code software to run on many different machines, setting the stage for software as a separate product from computers. In 1968, IBM announced that it was unbundling its hardware and software, providing an opportunity for independent vendors in the software industry, and in the 1970s initial industry classifications emerged. The main division was between "system" software and "application" software, and within applications, software was defined either by industry, or as a

[1] Software stocks are currently divided into only five sectors: application software, business software & services, internet software & services, security software and services, and technical and system software.

cross-industry application (Steinmueller, 1995;Campbell-Kelly, 2003).

In the 1980's the introduction of the Personal Computer led to growth in the software industry. Despite this, in the early part of the decade it was not discussed in the business press. From 1966 until 1980, *Businessweek* did not publish any articles about the software industry, and thereafter its next article on the industry was published in 1984 (Campbell-Kelly, 2003). Meanwhile, the software industry continued to develop. By the end of 1983, there were about 35,000 PC software products offered by about 3,000 vendors. VisiCalc, the "killer app" that is often attributed with unleashing the PC revolution, became available in 1979. In 1982 the term "productivity application" originated, which referred to the spreadsheet, word processor, and personal database (Campbell-Kelly, 2003). Around the same time, a consensus emerged that the best way for a computer to multitask was to develop a windowing system. Growing hardware markets for hard disks, display monitors, modems, and printers gave rise to software markets for utility software. Improvements in printing led to desktop publishing. By the end of the 1980s, software vendors offered thousands of programs for specialized applications and dozens more for general purpose applications (Steinmueller, 1995).

The software industry progressed through innovation, but this also developed under the official radar. Although copyright protection was always available, software was not always eligible for patent protection. In *Gordon v. Benson* in 1972, the courts upheld that software programs were merely algorithms that could not be patented. This decision was mostly overturned in *Diamond v. Diehr* in 1981, which decided that a software program could be patented if it was embedded within an apparatus. The debate about whether software should be patentable raged among officials and researchers. Meanwhile, software companies continued to patent in large numbers, receiving patent approval for inventions that were questionable at the time, such as pure data structures, methods for performing calculations in a data processor, data compression algorithms, and the like. After a series of cases that increasingly supported the patentability of software, in 1998 the last barrier to patenting pure software was overturned (Cohen and Lemley, 2001). Despite the official discussion however, as Cohen and Lemley (2001) note in the *California Law Review*, the approval of software patents was a routine practice long before the courts recognized it. Ironically, software patents may even have been granted too broadly, because the PTO did not hire software experts who could adequately evaluate the patents. The PTO did not see the expected surge of software patenting after the official ruling in 1998 (Cohen and Lemley, 2001).

By 1990, the software industry had matured. The mid-1990s saw a consolidation at the top of the software industry. According to *Software Magazine*, in 1994 the top ten of their software companies accounted for 63% of the revenues of the top hundred (not of the entire industry). Still, this did not slow the rest of the industry; the same review identified twenty-three companies achieve more than 50% growth over the previous year (Bucken, 1995). In the mid 1990's companies operated in a number of software labels, including relational databases, network management tools, ERP, security software, object management software, networking applications, middleware, financial applications, human resource management, CAD, and integrated voice response systems (Frye and Melewski, 1995).

In 1991, Tim Berners-Lee and Robert Calliau, researchers at CERN in Switzerland, released the HTML document format and the HTTP protocol that ran on the existing TCP/IP network infrastructure, which they dubbed as the World Wide Web. These advances created the internet as we know it today and brought another opportunity for software vendors (Fabrizio and Mowery, 2007). Existing companies shifted focus to creating client/server products that could be used over the World Wide Web (Geppert, 1998). The World Wide Web provided a scalable answer to client/server computing that software companies scrambled to provide. At MicroStrategy, Inc., a business intelligence software company, a recent hire out of college suggested creating a web version of their client/server product; this was the first the executives had even heard of the World Wide Web. They implemented his suggestion, and in 1996 released "DSS Web," which allowed their clients to access their software over the web.[2] The boom of the late 1990's fueled the growth of software companies. They began to focus on data mining, OLAP (On-Line Analytical Processing), and object-oriented programming (Comerford, 1998). Less constraining, or lenient labels emerged including Customer Relationship Management (CRM), e-commerce software, and e-business software; companies affiliated with these created a wide range of different types of software (Hayes, 2000).

The history of the software industry is unique and complex. Its technologies did not fit into standard ways of measuring innovation, and the importance of thousands of small independent vendors did not conform to standard ways of measuring industries. As a result, it was overlooked by mainstream business for many years. Much research on the industry focused on a few prominent firms and did not understand the industry in its entirety. Nevertheless, independent software firms continued to innovate and create their own organizational and product distinctions, supporting a technical infrastructure that became so important to society that it could no longer be ignored. Innovation in this industry was important and categorization was organic, providing a good context to study the relationship between invention, labeling, and category creation.

[2] From personal communications during my employment with MicroStrategy. Executives used to recount the company's history.

Data and Measures

Market Space

I study categorization in the software industry during 1990 – 2002 using press releases. Electronic documentation of press releases – one of the main forums where software companies described themselves and shaped the industry – are available after 1985, and press releases began to be used across the industry around 1990. Therefore I focus this research on the modern software industry, 1990 – 2002.

Software companies actively issue press releases to distribute news. This is true for large companies and small, independent vendors who are not documented in official data, but who contributed to shaping the industry. In the time period studied, software companies used press releases to announce new products, customers, partners, patent grants, and financing. Press releases were an important public face for software companies, and they are not especially costly to produce. So this data source captures a wide range of small and young organizations that are otherwise difficult to track. Because companies classify themselves within these documents, it captures the labeling and categorization system of the industry as it was emerging.

Therefore, I use press releases to create a data set of software companies and their claims to categories or labels. Within each press release, a company will claim an affiliation with a category or label. Figure 2a shows an example of a press release from Accrue Software. Figure 2b shows additional example identity claims from organizations in these data.

(Figure 2 about here)

I created a data set of software companies and labels they use to identify themselves from the 268,963 press releases issued during 1990 – 2002 that contain at least three mentions of the word “software.” Appendix A describes the data collection in detail. The final data include

3,365 distinct software organizations and 394 distinct categories and labels, over 14,357 organization-years and 3049 category-years. Figure 3 shows the number of software firms over time, and figure 4 shows the number of categories and labels by year.

(Figures 3 and 4 about here)

In this analysis I take into account whether organizations are dedicated members of a particular category or label or whether they are partial members to measure the extent to which a category or label constrains its members. I measure *leniency* for every category and label in these data, which is the category or label's *lack* of constraint. Lenient labels are those whose members are not constrained from identifying with many other labels within the software industry. I construct a measure for label or category leniency that takes into account the proportion of members that identify with *other* categories and labels, as well as the number of *unique* other categories with which they are identified. Appendix A details the construction of this measure:

$$leniency_C = fuzz_C \cdot N_{ocat} \qquad \text{(A5)}$$

A mapping of the categories and labels of the software industry for selected years is provided in figure 5. These figures show the extensive classification system of the software industry over time. The dimensions of these figures are not meaningful, but the distance between the categories or labels indicates how strongly they overlap. Each category or label is represented by a circle, the size of which is based on its leniency. There are a number of categories with high leniency that overlap with many other labels and categories, clustered in the center. Low leniency or high constraint categories lie both in the center and toward the edges. In many cases, two, three or four labels or categories form an isolated cluster, indicating high overlap, but high constraint. These figures illustrate the complexity of classification in the software industry over time. There are few categories where boundaries are absolute; most that have a substantial number of members also have a non-trivial amount of overlap.

(Figure 5 about here)

Knowledge Space

Building on previous studies (Podolny, et al., 1996;Pontikes, 2007;Pontikes and Barnett, 2008), I create knowledge space measures for the software industry using patent and patent citation data from the United States Patent Office, which is associated with the National Bureau of Economic Research patent data project (Hall, Jaffe, and Trajtenberg, 2001). I use data that have been updated to include patents through 2002, maintained on Hall's web site. The U.S. patent and trademark office issues patents for inventions that are new, unique and non-obvious. All patents must cite relevant 'prior art' on which the new invention is based, and this indicates the knowledge foundation for the patent at hand. Two patents that cite the same patent as prior art are more similar in knowledge space than two patents that have no common citations.

For patents to be defensible, they must be specific and accurate, and the patent office is active in requiring that inventors' claims be focused and narrow. Inventors include patents they are aware of in a patent's citations, and the patent examiner will also add citations to ensure that prior art is comprehensively cited (Alcácer and Gittleman, 2006). In some studies, it is important that the prior art accurately reflect the knowledge that the inventor is actively aware of, and so citations added by the examiner are problematic. Here, I am interested in locating software organizations in knowledge space in order to determine if their innovations are especially different, or if they are similar to other inventions created within the industry. Patent citations locate an invention within knowledge space, and so citations added by an examiner are not only unproblematic, but they actually help refine the patents' positions.

Patents are assigned to a class and a subclass when they are issued. There are about 400 classes, so to make the database more tractable, the National Bureau of Economic Research has created a higher level classification system of 36 subcategories and six categories: Chemical, Computers & Communications, Drugs and Medical, Electrical and Electronics, and Other (Hall, et al., 2001). I use all patents granted in the Computers & Communications category to construct knowledge space for the software industry. This is a broad classification that contains knowledge relevant to software. Note that this includes patents of software organizations, as well as patents issued to individual inventors, universities, and non-software organizations. Therefore this creates a knowledge space that is distinct from the software market space.

Appendix B details the construction of knowledge space, and the construction of measures for an organization's knowledge space difference. This measures is based on how "close" an organization is in knowledge space, to organizations that are in *different* market space categories:

$$kprox_{A,D_A} = \frac{\underset{i \in A}{sum}(sim_{i,D_A})}{npat_A}, \qquad \text{(B5)}$$

Empirical Test

Model

To test the hypothesis, I estimate the rate of new label creation for an organization as the probability that the organization introduces a new label during time period Δt in the limit where $\Delta t \to 0$. This rate can be operationalized in terms two random variables, $Y(t)$, or the number of new labels introduced by an organization at time t, and t_n, or the time of the organization's introduction of its n[th] label:

$$r\ (t-t_n)=\lim_{\Delta t\to 0}\frac{\Pr\{Y(t-t_n+\Delta t)=n+1\mid Y(t-t_n)=n\}}{\Delta t} \qquad (1)$$

I estimate this rate as a function of characteristics of the organization, the categories of which the organization is a member, and environmental variables:

$$r(t-t_n)=r_0(t-t_n)\cdot\exp(\alpha_f\cdot\mathbf{x}_f+\alpha_c\cdot\mathbf{x}_c+\alpha_e\cdot\mathbf{x}_e)+\varepsilon \qquad (2)$$

The hypothesis testing and control variables are described in detail below. I use piecewise continuous hazard rate models to estimate this model, using the stpiece routine in Stata written by Jesper Sørensen. In order to estimate repeated events in these models, organizations that introduce a new label exit from the data set and enter with a new id. Therefore I cluster the variance by original organization id. In addition, organizations can change categories either by claiming a new label or by joining an existing category. For comparison, I also estimate an organization's rate of joining an existing category. Both estimations are modeled as competing risks.

Dependent variables

I test this hypothesis on the dependent variable that an organization uses a new label for its identity claim in a press release. A label is "new" in the first or second year that it is used in press releases. The second dependent variable is whether an organization joins an existing category, or uses a label that has been used in press releases for three or more years.

Hypothesis testing variables

I use label leniency and knowledge space position to test hypothesis 1. Since there is not one clear point when a label becomes a category, I characterize all categories and labels by their leniency, defined in equation (A5). Labels that are low in leniency are high in constraint (and are more likely to be categories). Hypothesis 1 predicts that organizations that are different in knowledge space will be more likely to introduce a new label when they are already affiliated with high constraint – or low leniency – categories or labels.

I test the hypothesis in two ways. First, I use the interaction between an organization's knowledge space proximity to organizations that are in *different* software categories to measure knowledge space difference as defined in (B5), and the mean leniency of the labels in which organization is a non-zero member, as defined in (A5). I also use the organization's knowledge space difference as defined in (B5) and split it into four variables: for organizations in high leniency, high/medium leniency, medium/low leniency, and low leniency labels or categories. All variables start for a given organization-year segment, and they are lagged by one year. In other words, I include the leniency of an organization's labels at time *t-1*, for its effect on new label creation at time *t*.

Control variables

I control for the main effects of knowledge space proximity to organizations that are in different categories and mean leniency of the categories or labels of which the organization is a member. To test whether results are due to knowledge space *difference*, as opposed to general knowledge space activity, I include variables for knowledge space proximity to categories or labels in which the organization is already a member. Specifically, I control for the organization's knowledge space proximity to organizations that are in the same categories as the organization, defined in equation (B4), and for the interaction between knowledge space proximity to the labels or categories an organization is in and the mean leniency of the organization's categories. These are the counterparts to the knowledge space difference variables used to test the hypothesis.

In addition, I include the mean fuzziness of the organization's categories as defined in equation (A3), and the total number of patents the organization has released over its history.[3] I control for the number of press releases the organization has released in the last year, the number of categories in which it has non-zero membership, the number of acquisitions made by the organization, whether the organization is the only member of a category in which it is a member, and the number of software organizations in the industry.[4] I also include time period dummies. All variables start for a given organization-year segment, and they are lagged by one year.

Results

Descriptives are included in table 1, and correlations are provided in table 2. Inspection of the correlations among the variables shows high correlations between the knowledge space proximity variables and the respective interactions with category leniency, which could raise concerns about multicollinearity. As a result, I also include an estimation that separates knowledge space proximity for low, medium-low, medium-high, and high leniency. Some of the control variables also are highly correlated. To address this, I ran additional models (not included for the sake of brevity) that did not simultaneously include highly correlated variables. The effects of the coefficient estimates of the tested variables are similar and do not decrease in significance.

(Tables 1 and 2 about here)

[3] I include number of patents in the previous year and number of citations in models not shown, but the total number of patents provides a better model fit. The alternate measures provide similar results.
[4] I separate the total number of software organizations into those in the same category as the focal organization and the number of organizations in different categories, but these do not have different effects, and the total number of organizations provides the best model fit. Splitting this variable in the models yields similar results.

Table 3 provides estimation results of piecewise continuous models on an organization's rate of new label formation and its rate of joining an existing organizational category. Model 1 estimates the rate of new label creation, and includes control variables only. Both the mean fuzziness ($p<0.10$) and mean leniency ($p<0.05$) of an organization's labels increase the likelihood that the organization will start a new label. This supports the argument that the amount of leniency or constraint imposed by a label should be measured separately from its fuzziness (defined in Appendix A). The number of total patents an organization has been issued significantly decreases its propensity to start a new label, although it increases its propensity to join an existing category or label (model 3). An organization's patents in the previous year were included in models not shown, and did not have a significant effect, indicating that this effect is due to organizations with a history of knowledge space activity. The number of press releases issued by an organization has a positive and marginally significant effect indicating that organizations that are active in the public forum are also active in label creation. As expected, the number of categories in which an organization has non-zero membership, and the number of acquisitions it has completed in the past year both have a positive effect on new label creation. The number of distinct categories within the software industry also has a positive and significant effect, which may be an indicator of an increasing carrying capacity for categories in the industry.

When an organization is the only member of its category – so that it has already attempted to start a category – it is significantly more likely to claim a new label, but there is no significant effect on its likelihood to join an existing category (model 3). This picks up on organizations that are actively trying to start a new category, and are trying out new labels in the market. The number of software organizations has a negative and significant effect, indicating that the more crowded the industry becomes, the less likely organizations are to introduce new labels. In other models, I included both the density of the organization's own categories, and the density of other categories, but the density of an organization's own categories did not have a separate effect, and so model 1 includes the total number of organizations, which provides the best model fit.

(Table 3 about here)

Models 2 and 3 compare how main effects of knowledge space position affect an organization's likelihood to create a new label or to join an existing label or category. Model 2 estimates new label creation and includes main effects for an organization's knowledge space position. It builds on model 1, adding knowledge proximity to an organization's own categories and labels and proximity to different categories and labels. Model 3 estimates the likelihood that an organization joins an existing category. Results show that the main effect of knowledge space proximity to different categories is positive and significant ($p<0.05$) in both models. The effect is twice as large for new label creation. When an organization has high knowledge space proximity to other categories, it is likely to join an existing category or label, and is even more likely to create a new label. Model 2 shows a positive and significant effect ($p<0.01$) of an organization's knowledge space proximity to its own categories or labels on the rate of new label creation. This effect only holds for new label creation, and not for the rate of joining an existing category. I will re-visit this effect in the hypothesis tests.

Model 4 tests the hypothesis and includes the interaction between an organization's knowledge proximity to other categories and category leniency. The hypothesis predicts that there will be a positive relationship between knowledge space difference and new label creation when organizations are in constraining categories. Therefore I expect that the positive relationship between knowledge space proximity to different categories and new label creation will decrease as label leniency increases. Model 4 is an improvement in fit over model 2 at the $p<0.05$ level. Results show that when the interaction between knowledge difference and label leniency is included, the effect of knowledge space proximity to others doubles when evaluated at zero leniency, significant at the $p<0.05$ level. The interaction is negative and marginally significant at $p<0.10$ (two-tailed test). This supports the hypothesis: when indicates that when organizations are in constraining categories or labels, knowledge space difference has a positive and significant effect on new label creation. As labels become more lenient, the relationship disappears and even becomes negative in the range of these data. These results show that when an organization is *both* in constraining categories or labels *and* is different in knowledge space, it is more likely to create a new label.

Model 4 also includes the interaction between an organization's knowledge space proximity to its own labels and label leniency. Results show that the effects of knowledge space proximity to an organization's own categories and labels are positive and significant only for the interaction. An organization that is creating knowledge similar to its own labels or categories is *only* likely to start a new label when it is also in a lenient label. This result was unexpected but is consistent with the ideas presented above. We would not expect that an organization that is proximate in knowledge space to its own labels or categories would be pushed out due to constraints. Results show that this is not the case; an organization's knowledge space proximity to its own labels and categories only has a positive effect when label leniency is high. It is possible that organizations that are very active creating knowledge similar to their own labels may attempt to create related “spin-off” labels. This type of re-branding may be more acceptable to audiences when organizations are affiliated with lenient labels.

Model 5 estimates the same independent variables as model 4 on an organization’s rate of joining an existing category. Results show that when the interactions are included, none of the knowledge space variables have significant effects, indicating that label leniency does not affect the relationship between knowledge space proximity to different labels or categories and the rate of joining an existing label or category. These results support the assertion that constraints arising from categories affect novelty, not an organization’s tendency to change.

Finally, model 6 tests hypothesis 1 by including knowledge space proximity to different categories separated into four variables: for organizations that are in low, low/medium, medium/high, and high leniency categories. Results show that there is a positive and significant effect ($p<0.05$) of knowledge space proximity to different labels and categories only when organizations are in labels or categories that have low or medium/low leniency. The effect disappears when organizations are in labels or categories that have medium/high or high leniency.[5] In addition, the magnitude of the effect doubles when organizations are in low, as opposed to low/medium labels. This indicates that there is a positive relationship between knowledge space difference and new label creation only when organizations are in lenient labels or categories. As constraints increase, the strength of the relationship reduces and then disappears. In sum, the results of models 1- 6 provide strong support for hypothesis 1. An organization’s likelihood to create a new label is increased

[5] When the model is run to only include knowledge space proximity to different categories for high and medium/high constraint categories, there is an improvement in fit over model 2 at the $p<0.05$ level.

when it is both different in knowledge space *and* is in constraining labels or categories.

Figure 6 illustrates the effect, by plotting the multiplier of an organization's rate of new label creation by knowledge difference for organizations in constraining as opposed to lenient labels. When an organization is in a constraining label or category there is a positive relationship between an organization that is different in knowledge space, and the rate of new label creation. For organizations in a constraining label, one with knowledge difference of 100 is twice as likely to create a new label as compared to an organization that is not different from others in knowledge space. For organizations in lenient labels, there is not a strong relationship between knowledge difference and the rate of new label creation.

Figure 7 provides a three-way plot that shows the rate of new label creation for an organization's knowledge difference and label leniency. This figure shows that the positive relationship between knowledge difference and the rate of new label creation becomes more muted with increasing label leniency. For a leniency of zero, the multiplier of the rate of new label creation increases sharply with knowledge difference. As leniency increases, the rate does not increase as sharply. This figure also illustrates the positive main effect of leniency on new label creation: when knowledge difference is zero, organizations in lenient labels are more likely to create a new label. Because low leniency labels and categories constrain organizations from identifying elsewhere, we expect that those affiliated with lenient labels are more likely to create a new identity. Interestingly, for organizations with high knowledge difference, this trend reverses. At knowledge difference of 150, there is an inverse relationship between leniency and new label creation. This indicates that knowledge difference has a strong effect on organizations that are in constraining labels. When they are not creating technical novelty, these organizations are unlikely to try out a new identity on the market. But when they are developing new technologies, organizations in constraining categories and labels are especially likely to introduce a new label into the market.

Patenting Organizations Only

Next, I test whether the results reported above are due to differences between organizations that are active in knowledge space and those that are not. Table 4 contains models that estimate the rate of new label formation only for organizations that have already patented.

(Table 4 about here)

Model 7 is run on control variables only, and the effects are similar to those in model 1. Model 8 includes an organization's knowledge space proximity to different categories and labels and to its own categories and labels. In this subset of data statistical power is reduced, and so although the effect of knowledge space proximity to an organization's own categories and labels remains positive and significant at $p<0.01$, the effect of knowledge space proximity to different labels or categories loses significance.

Model 9 includes the same variables as model 4 on patenting organizations only, and results show that when the interaction is included, knowledge space difference for organizations that are in constraining labels or categories (low leniency) increases in magnitude and becomes marginally significant at the $p<0.10$ level (two-tailed test), providing support for hypothesis 1. The magnitude of the coefficient, 0.006, is slightly lower than the coefficient estimated in models of all organizations, where the effect is 0.0081, and the interaction is not significant in this model, partly due to a decrease in the size of the coefficient. Model 9 also shows that knowledge space proximity to an organization's own labels or categories only increases its propensity to create a new label when it is in lenient labels, consistent with the results for all organizations reported in model 4.

Model 10 includes the same variables as model 6, and results are consistent with those reported in model 6. When knowledge space proximity to other labels or categories is split into four variables depending on label leniency, knowledge space difference has a positive and marginally significant effect at the $p<0.10$ level (two-tailed test) on new label creation only when organizations are in low leniency or medium/low leniency labels. Again, the reduction in significance is partly due to an inflation of the standard errors, and partly due to a slight decrease in the magnitude of the coefficient. The coefficients in model 10 for patenting organizations are 85%-90% of the values of those reported for all organizations. In all, the models run on patenting organizations are consistent with the results obtained for all organizations, providing additional support for hypothesis 1.

Leniency or Fuzziness?

I use equation (A5) to measure how constraining or lenient labels and categories are, in order to test hypothesis 1. However, an alternative hypothesis might suggest that it is the category fuzziness, as defined in equation (A4) that is driving these effects, and that the leniency metric is unnecessary. I test against this alternate hypothesis using a model, similar to model 4, that substitutes mean fuzziness (from equation (A4)) for label leniency in the interactions. Results show that the interaction of knowledge space difference with fuzziness does not have a significant effect on the rate of new label formation, indicating that the results present above are due to strength of category or label constraint, as suggested.

Discussion and Conclusion

Above, I proposed that an organization's propensity to create a new label and disrupt market space would depend on both organizational knowledge and the extent to which labels and categories are constraining or lenient. Specifically, I hypothesized that there would be a positive relationship between knowledge space difference and new label creation when organizations were in constraining categories or labels. Results support this hypothesis. When organizations are proximate in knowledge space to different categories, they are only likely to introduce a new label into the market structure when they are in constraining labels or categories. These results indicate that market space disruptions are not simply a result of how existing categories are structured, nor do they only emerge from enterprising organizations that introduce new technologies into the market. Rather, to understand how the classification structure evolves, we must consider *both* the structure of market space and organizational contributions such as novel knowledge development.

These results address two literatures that have developed independently, each describing how new organizational categories emerge. The knowledge space perspective examines invention as the source of new category formation and the market space perspective looks at how existing structures lead to new categories. Both literatures provide important insights into how categories are formed. However, if novel invention is separated from the categorical context into which it is introduced, category creation cannot be fully understood. Novelty in knowledge space interacts with constraints in market space to affect when new categories are introduced in the market.

Here I focus on how novel knowledge interacts with market space constraints to affect market space outcomes. A related question might ask which organizations are more likely to create novel knowledge. Do market space constraints affect an organization's propensity to explore in knowledge space? Previous research shows that individuals borrow elements from rival categories even when categories are sharp and oppositional. The more actors borrow, the lower the penalty for borrowing (Rao, Monin, and Durand, 2005). Similarly, I would expect also that market space constraints would affect the propensity for organizations to create novel knowledge. Organizations that are affiliated with constraining labels or categories may be less likely to create different knowledge. Results here show that when they do, they are more likely to introduce a new label. The findings introduced here are consistent with Rao, Monin, and Durand (2005) in that as labels become more lenient organizations are less likely to introduce a new label when they have created novel knowledge. Perhaps they are also more likely to create novel knowledge.

More broadly, this dissertation may speak to the literature on network forms of organization. In networked forms of organization, the position of an organization within a network of related organizations is an important factor that contributes to its success (Powell, et al., 2005). It is possible that lenient labels facilitate network forms of organization. Members of lenient labels may be more likely to form ties with diverse sets of other organizations. If this is the case, networked forms of organization may be less likely to give rise to new market space categories.

This research also contributes to the literature on the role of technology in organizations. Findings indicate that how we evaluate technologies is fundamentally linked to the structure of market space. The determination of technological novelty depends partly on the market context. Looking back, the differences between a category that emerged and existing categories may seem rooted in obvious technical distinctions. But this is because – by definition –

categorization highlights differences. It is not as easy to predict the formation of a new category when looking forward; inventions that are novel in knowledge space do not always inspire the formation of a new category.

In addition, this study may speak to the literature on strategic change. Managers attempt to position their organizations to maximize both the amount of attention received by important audiences and the likelihood that they will favorably compare with competitors. Constraining labels may provide a more stable resource base since potential customers will be more certain of what to expect from member organizations. At the same time, constraining labels will facilitate comparisons among competitors and may favor incumbents over newcomers. Depending on the organizations' and competitors' competencies, managers may seek to identify with constraining or lenient labels. This study shows that one core competency – organizational knowledge – has different effects on how managers strategically attempt to identify their organizations, depending on existing category and label constraints. Extending this, it is plausible that other competencies also affect the types of labels or categories managers claim for their organizations' identities.

One reason that managers' identity claims for an organization are strategic is that these claims help to define the organization's competitive group. Different contexts present different logics of competition that organizations must navigate, and success in one context does not necessarily imply success in another (Barnett, 2008). Because of this, an organization may find that the same products and product strategies meet with more success when the organization competes in a new context. As a result, what the organization claims to be can be used strategically. Results here indicate that strategic claims to new labels are based on organizational competencies with respect not only to a pool of competitors but also to the logics of competition associated with the competitive environment.

These findings also speak to theories of entrepreneurship. Often, entrepreneurs are credited with disrupting existing market structures and creating novelty on the market. However, these data from the software industry indicate that new labels are introduced by both new and existing organizations. When existing organizations develop new knowledge and are also in constraining categories or labels, they tend to introduce new labels into market space. But to what extent can managers emphasize or downplay how much new knowledge violates existing category codes? It is possible that entrepreneurs and managers of young companies are more likely to actively try to differentiate new technologies by claiming a new label. If so, entrepreneurs may attempt to define these new labels by creating definitions that implicitly compare them with constraining as opposed to lenient categories and labels.

Finally, these ideas contribute to the literature on categorization by introducing leniency and constraint as a measurable characteristic of categories and labels. In this study, I investigate how low or high leniency labels and categories generally affect the evolution of the market structure. At any point in time, market space is populated by labels and categories at various stages of definition. Rather than either focusing on high constraint categories or investigating domains where categories are emerging, characterizing labels and categories by the extent to which they are constraining or lenient allows for the understanding of different dynamics within the same domain. Further, some highly visible labels confer legitimacy, but are not constraining. Shifting the focus from mature versus immature or nascent versus stable categories, to characterizing any label or category by its constraint allows for a flexible and inclusive analysis of category dynamics.

How organizations are classified has important implications for both organizations and society as a whole. Accordingly, understanding how categories emerge, and by implication how structures change, is an important question in organization theory. This paper explores these ideas in two domains, knowledge space and market space, and results show that both an organization's knowledge and existing categories powerfully shape the identities of organizations and the evolution of the category structure. Understanding interactions between these domains, both in terms of position and structure, provides a more comprehensive examination of the evolution of market categories.

References

Alcácer, J., and M. Gittleman
2006 "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations." The Review of Economics and Statistics, 88: 774-779.

Barnett, W.
1990 "The organizational ecology of a technological system." Administrative Science Quarterly, 35: 31-60.

Barnett, W., and G. Carroll
1987 "Competition and mutualism among early telephone companies." Administrative Science Quarterly, 32: 400-421.

Barnett, W.P.
2004 "From Red Vienna to the Anschluss: Ideological Competition among Viennese Newspapers During the Rise of National Socialism." American Journal of Sociology, 109: 1452-1499

2008 The Red Queen among Organizations: How Competitiveness Evolves. Princeton, NJ: Princeton University Press.

Barnett, W.P., and E.G. Pontikes
2008 "The Red Queen, Success bias, and Organizational Inertia." Management Science.

Bogaert, S., et al.
2006 "Contentious Legitimacy: Professional Association and Density Dependence in the Dutch Audit Industry 1884-1939." Stanford University Working paper.

Brusoni, S., et al.
2001 "Knowledge Specialization, Organizational Coupling, and the Boundaries of the Firm: Why Do Firms Know More Than They Make?" Administrative Science Quarterly, 46: 597-621.

Bucken, M.
1995 "State of the Industry (the 13th Annual Software Magazine top 100)." Software Magazine: 4.

Callon, M., and F. Muniesa
2005 "Peripheral Vision: Economic Markets as Calculative Collective Devices." Organization Studies, 26.

Campbell-Kelly, M.
2003 From Airline Reservations to Sonic the Hedgehog: A History of the Software Industry. Cambridge, Massachusetts: The MIT Press.

Carroll, G.R., and A. Swaminathan

2000 "Why the microbrewery movement? Organizational dynamics of resource partitioning in the U.S. brewing industry." American Journal of Sociology, 106: 715-762.

Cohen, J., and M. Lemley
2001 "Patent Scope and Innovation in the Software Industry." California Law Review, 89: 1-57.

Comerford, R.
1998 "Software Engineering." IEEE Spectrum.

DiMaggio, P.J.
1987 "Classification in art." Annual Review of Sociology, 52: 440-455.

1991 "Constructing an Organizational Field as a Professional Project: U.S. Art Museums, 1920-1940." In W. J. Powell, and P. J. DiMaggio (eds.), The New Institutionalism in Organizational Analysis. Chicago: The University of Chicago Press.

DiMaggio, P.J., and W. Powell
1991 "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." In P. D. W. Powell (ed.), The New Institutionalism In Organizational Analysis: University of Chicago Press.

Dutton, J., and J. Dukerich
1991 "Keeping an Eye on the MIrror: Image and Identity in Organizational Adaption." Academy of Management Journal, 34: 517-554.

Edelman, L.
1992 "Legal ambiguity and symbolic structures." American Journal of Sociology, 97: 1531-1576.

Fabrizio, K., and D. Mowery
2007 "The Federal Role in Financing Major Innovations: Information Technology During the Postwar Period." In N. Lamoreaux, and K. Sokoloff (eds.), Financing Innovation in the United States, 1870 to Present. Cambridge: MIT Press.

Fleming, L.
2001 "Recombinant Uncertainty in Technological Search." Management Science, 47: 117-132.

Fligstein, N.
1985 "The Spread of the Multidivisional Form Among Large Firms, 1919-1979." American Sociological Review, 50: 377-391.

1996 "Markets as Politics: A Political-Cultural Approach to Market Institutions." American Sociological Review, 61: 656-673.

Frye, C., and D. Melewski
1995 "Wide Window Beckons to Suppliers: Just Do It." Software Magazine.

Geppert, L.
1998 "Technology 1998: Analysis & Forecast." IEEE Spectrum.

Gilbert, R.
2006 "Looking for Mr. Schumpeter: Where are we in the Competition-Innovation Debate?" In A. Jaffe, et al. (eds.), Innovation Policy and the Economy. Cambridge: The MIT Press.

Hall, B., et al.
2001 "The NBER patent citations data file: Lessons, insights, and methodological tools." NBER Working Paper Series. Cambridge, MA: National Bureau of Economic Research.

Hannan, M.T., and J. Freeman
1977 "The population ecology of organizations." American Journal of Sociology, 82: 929-964.

Hannan, M.T., et al.
2007 Logics of Organization Theory. Princeton: Princeton University Press.

Hayes, I.
2000 "Application First, Delivery Second!", Software Magazine: 63-64.

Henderson, R., and K. Clark
1990 "Architectural innovation: The reconfiguration of existing product technologies and the failure of established forms." Administrative Science Quarterly, 35: 9-30.

Hsu, G.
2006 "Jacks of all Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production." Administrative Science Quarterly, 51.

Hsu, G., and M.T. Hannan
2005 "Identities, Genres and Organizational Forms." Organization Science, 16: 474-490.

Hsu, G., et al.
2007 "Multiple Category memberships in Markets: A Formal Theory and Two Empirical Tests." Working paper.

Jaffe, A.
1986 "Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value." The American Economic Review, 76: 984-1001.

Leifer, E., and H. White
1987 "A Structural Approach to Markets." In M. Mizruchi, and M. Schwartz (eds.), Intercorporate Relations: The Structural Analysis of Business. Cambridge: Cambridge University Press.

Levin, R., et al.

1985 "R&D Appropriability, Opportunity, and Market Structure: New Evidence on Some Schumpeterian Hypotheses." The American Economic Review, 75: 20-24.

Levin, R., and P.C. Reiss
1988 "Cost-Reducing and Demand-Creating R&D with Spillovers." The RAND Journal of Economics, 19: 538-556.

Lounsbury, M., and H. 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.

March, J.
1991 "Exploration and exploitation in organizational learning." Organization Science, 2: 71-87.

McKendrick, D., and G. Carroll
2001 "On the genesis of organizational forms: Evidence from the market for disk drive arrays." Organization Science, 12: 661-682.

Meyer, J., and B. Rowan
1991 "Institutionalized Organizations: Formal Structure as Myth and Ceremony." In P. D. W. Powell (ed.), The New Institutionalism In Organizational Analysis: University of Chicago Press.

Osherson, D., and E. Smith
1982 "Gradedness and conceptual combination." Cognition, 12: 299-318.

Park, B., and R. Hastie
1987 "Perception of Variability in Category Development: Instance - versus abstraction-based stereotypes." Journal of Personality and Social Psychology, 53: 621-636.

Pfeffer, J., and G. Salancik
1978 The External Control of Organizations. New York: Harper & Row.

Phillips, D., and D. Owens
2004 "Incumbents, Innovation, and Competence: the Emergence of Recorded Jazz, 1920 to 1929." Poetics, 32: 281-295.

Podolny, J., et al.
1996 "Networks, Knowledge, and Niches: Competition in the Worldwide Semiconductor Industry, 1984-1991." The American Journal of Sociology, 102: 659-689.

Pontikes, E.G.
2007 "Knowledge Sharing or Competition? How Knowledge Space Progress Affects Knowledge Progress and Market Survival." Working Paper.

Pontikes, E.G., and W.P. Barnett

2008 "Technical Change Among Organizational Communities." Working Paper.

Porac, J., and H. Thomas
1990 "Taxonomic Mental Models in Competitor Definition." The Academy of Management Review, 15: 224-240.

Porter, M.E.
1985 Competitive Advantage: Creating and Sustaining Superior Performance. New York: The Free Press.

Powell, W., et al.
1996 "Interorganizational collaboration and the locus of innovation." Administrative Science Quarterly, 41: 116-145.

2005 "Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences." The American Journal of Sociology, 110: 132-205.

Prahalad, C.K., and G. Hammel
1990 "The core competence of the corporation." Harvard Business Review, 68: 79-91.

Rao, H., et al.
2003 "Institutional Change in Toque Ville: Nouvelle Cuisine as an Identity Movement in French Gastronomy." American Journal of Sociology, 108: 795-843.

2005 "Border Crossing: Bricolage and the Erosion of Categorical Boundaries in French Gastronomy." American Sociological Review, 70: 968-991.

Rosa, J., et al.
1999 "Sociocognitive Dynamics in a Product Market." Journal of Marketing, 63: 64-77.

Rosch, E., and B. Lloyd
1978 Cognition and Categorizatoin: Lawrence Erlbaum Associates.

Rosch, E., and C. Mervis
1975 "Family Resemblances: Studies in the internal structure of categories." Cognitive Psychology, 7: 573-605.

Ruef, M.
2000 "The emergence of organizational forms: A community ecology approach." American Journal of Sociology, 106: 658-714.

Santos, F., and K. Eisenhardt
2008 "Constructing Niches and Shaping Boundaries: A Model of Entreprenurial Action in Nascent Markets." Working paper.

Scherer, F.M.

1967 "Market Structure and the Employment of Scientists and Engineers." American Economic Review, 57: 524-531.

Sørensen, J., and T. Stuart
2000 "Aging, Obsolescence, and Organizational Innovation." Administrative Science Quarterly, 45: 81-112.

Steinmueller, W.
1995 "The U.S. Software Industry: An Analysis and Interpretive History." In D. Mowery (ed.), The International Computer Software Industry: Oxford University Press.

Stuart, T., and J. Podolny
1996 "Local search and the evolution of technological capabilities." Strategic Management Journal, 17: 21-38.

Teece, D., et al.
1997 "Dynamic Capabilities and Strategic Management." Strategic Management Journal, 18: 509-533.

Thompson, J.
1967 Organizations in Action. New York: McGraw Hill.

Tripsas, M.
1997 "Unraveling the Process of Creative Destruction: Complementary Assets and Incumbent Survival in the Typesetter Industry." Strategic Management Journal, 18: 119-142.

Tushman, M., and P. Anderson
1986 "Technological Discontinuities and Organizational Environments." Administrative Science Quarterly, 31: 439-465.

Weick, K.
1979 The Social Psychology of Organizing. Menlo Park, CA: Addison-Wesley Publishing Company.

White, H.
1981 "Where do markets come from?" The American Journal of Sociology, 87: 517-547.

Zuckerman, E.
1999 "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount." The American Journal of Sociology, 104: 1398-1438.

Zuckerman, E.W., et al.
2003 "Robust Identities or Nonentities? Typecasting in the Feature-film labor market." American Journal of Sociology, 108: 1018-1074.

Tables

Table 1. Descriptives

	Mean	Stdev	Min	Max	Mean	Stdev	Min	Max
	All Organizations (N=14357)				Patenters only (N=2843)			
Org claims a new label	0.03	0.18	0	1	0.05	0.23	0	1
Year	1998	3	1990	2002	1998	3	1990	2002
Knowledge proximity to different labels/categories x mean leniency of organization’s labels/categories	60	375	0	11538	304	799	0	11538
Knowledge proximity different labels/categories, organization is in low leniency labels/categories (0-5]	0.49	6.14	0	315	2.47	14	0	315
Knowledge proximity different labels/categories, organization is in low/medium leniency labels/categories (5-20]	1.02	9.08	0	332	5.16	20	0	332
Knowledge proximity different labels/categories, organization is in medium/high leniency labels/categories (20-30]	0.43	5	0	196	2.15	12	0	196
Knowledge proximity different labels/categories, organization is in high leniency labels/categories (30+)	0.77	7	0	237	3.87	15	0	237
Knowledge proximity to different software labels/categories	2.70	14	0	332	14	29	0	332
Knowledge proximity to organization's labels/categories x mean leniency	42	348	0	10680	211	758	0	10680
Knowledge proximity to organization's labels/categories	1.09	8.24	0	219	5.48	17.85	0	219
Mean leniency of organization's labels/categories	25	28	0	131	31	26	0	131
Mean fuzziness of organization's labels/categories	0.38	0.24	0	0.8026	0.49	0.13	0	0.7694
Number of total patents over org's history	23	327	0	14707	116	728	1	14707
Number of press releases from organization	9.68	25.46	0	1148	22	50	0	1148
Number of categories of which organization is a non-zero member	1.72	1.90	0	29	2.67	2.51	1	29
Number of acquisitions made by organization	0.03	0.18	0	4	0.07	0.30	0	4
Number of software categories org is not in	218	136	0	355	280	75	66	355
Org is only member of its category	0.01	0.11	0	1	0.01	0.12	0	1
Number of software organizations	1440	641	103	2176	1470	623	103	2176

Table 2a. Correlations for all organizations (N=14357)

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Organization claims a new label	(1)									
Year	(2)	-0.04								
Knowledge proximity to different labels/categories x mean leniency of organization’s labels/categories	(3)	0.06	-0.01							
Knowledge proximity different labels/categories, organization is in low leniency labels/categories (0-5]	(4)	0.02	-0.03	0.02						
Knowledge proximity different labels/categories, organization is in low/medium leniency labels/categories (5-20]	(5)	0.03	-0.05	0.27	-0.01					
Knowledge proximity different labels/categories, organization is in medium/high leniency labels/categories (20-30]	(6)	0.01	-0.02	0.35	-0.01	-0.01				
Knowledge proximity different labels/categories, organization is in high leniency labels/categories (30+)	(7)	0.05	0.01	0.84	-0.01	-0.01	-0.01			
Knowledge proximity to different software labels/categories	(8)	0.06	-0.05	0.75	0.43	0.64	0.38	0.49		
Knowledge proximity to organization's labels/categories x mean leniency	(9)	0.09	0.02	0.64	-0.01	0.01	0.20	0.61	0.39	
Knowledge proximity to organization's labels/categories	(10)	0.09	0.00	0.62	-0.01	0.06	0.34	0.56	0.45	0.93
Mean leniency of organization's labels/categories	(11)	0.03	0.51	0.08	-0.06	-0.05	0.00	0.09	-0.02	0.09
Mean fuzziness of organization's labels/categories	(12)	0.09	0.34	0.09	-0.03	0.05	0.04	0.07	0.07	0.07
Number of total patents over org's history	(13)	0.02	0.01	0.04	0.01	0.02	0.01	0.04	0.04	0.03
Number of press releases from organization	(14)	0.10	0.11	0.08	0.00	0.03	0.04	0.06	0.07	0.11
Number of categories of which organization is a non-zero member	(15)	0.13	0.20	0.06	-0.02	0.03	0.06	0.04	0.06	0.11
Number of acquisitions made by organization	(16)	0.05	0.02	0.04	0.01	0.03	0.02	0.04	0.05	0.05
Number of software categories org is not in	(17)	0.07	0.56	0.07	0.02	0.02	0.02	0.06	0.06	0.07
Org is only member of its category	(18)	0.01	-0.08	-0.02	0.07	-0.01	-0.01	-0.01	0.01	-0.01
Number of software organizations	(19)	-0.03	0.98	-0.01	-0.04	-0.06	-0.03	0.01	-0.06	0.02

Table 2a (cont'd). Correlations for all organizations (N=14357)

		(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Mean leniency of organization's labels/categories	(11)	0.06								
Mean fuzziness of organization's labels/categories	(12)	0.08	0.64							
Number of total patents over org's history	(13)	0.03	0.03	0.04						
Number of press releases from organization	(14)	0.13	0.18	0.24	0.46					
Number of categories of which organization is a non-zero member	(15)	0.13	0.35	0.58	0.21	0.58				
Number of acquisitions made by organization	(16)	0.06	0.04	0.09	0.09	0.22	0.20			
Number of software categories org is not in	(17)	0.07	0.65	0.85	0.03	0.21	0.48	0.07		
Org is only member of its category	(18)	-0.02	-0.10	-0.18	-0.01	-0.02	-0.04	-0.02	0.00	
Number of software organizations	(19)	0.00	0.51	0.32	0.01	0.10	0.19	0.02	0.56	-0.07

Table 2b. Correlations for patenting organizations only (N=2843)

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Organization claims a new label	(1)									
Year	(2)	-0.05								
Knowledge proximity to different labels/categories x mean leniency of organization's labels/categories	(3)	0.08	-0.04							
Knowledge proximity different labels/categories, organization is in low leniency labels/categories (0-5]	(4)	0.02	-0.09	-0.04						
Knowledge proximity different labels/categories, organization is in low/medium leniency labels/categories (5-20]	(5)	0.04	-0.13	0.22	-0.05					
Knowledge proximity different labels/categories, organization is in medium/high leniency labels/categories (20-30]	(6)	0.01	-0.06	0.32	-0.03	-0.05				
Knowledge proximity different labels/categories, organization is in high leniency labels/categories (30+)	(7)	0.07	0.01	0.84	-0.05	-0.07	-0.04			
Knowledge proximity to different software labels/categories	(8)	0.08	-0.15	0.72	0.40	0.61	0.35	0.45		
Knowledge proximity to organization's labels/categories x mean leniency	(9)	0.15	0.02	0.61	-0.05	-0.05	0.17	0.58	0.33	
Knowledge proximity to organization's labels/categories	(10)	0.14	-0.01	0.59	-0.05	0.00	0.31	0.53	0.39	0.92
Mean leniency of organization's labels/categories	(11)	-0.02	0.58	0.12	-0.20	-0.18	-0.04	0.16	-0.15	0.17
Mean fuzziness of organization's labels/categories	(12)	0.02	0.43	0.09	-0.27	-0.01	0.03	0.08	-0.07	0.08
Number of total patents over org's history	(13)	0.02	0.02	0.00	-0.01	-0.02	-0.01	0.01	-0.01	-0.01
Number of press releases from organization	(14)	0.10	0.07	0.00	-0.04	-0.03	0.00	0.01	-0.04	0.06
Number of categories of which organization is a non-zero member	(15)	0.12	0.17	-0.03	-0.11	-0.04	0.03	-0.02	-0.08	0.08
Number of acquisitions made by organization	(16)	0.07	0.02	0.01	-0.01	0.00	0.00	0.02	0.00	0.03
Number of software categories org is not in	(17)	-0.04	0.97	0.00	-0.07	-0.12	-0.05	0.05	-0.11	0.07
Org is only member of its category	(18)	0.02	-0.07	-0.05	0.14	-0.03	-0.02	-0.03	0.02	-0.03
Number of software organizations	(19)	-0.04	0.98	-0.05	-0.09	-0.16	-0.08	0.01	-0.18	0.04

Table 2b (cont'd). Correlations for patenting organizations only (N=2843)

Patenting orgs N=2843		(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Mean leniency of organization's labels/categories	(11)	0.09								
Mean fuzziness of organization's labels/categories	(12)	0.07	0.52							
Number of total patents over org's history	(13)	0.00	0.03	0.04						
Number of press releases from organization	(14)	0.08	0.11	0.15	0.49					
Number of categories of which organization is a non-zero member	(15)	0.12	0.18	0.36	0.30	0.66				
Number of acquisitions made by organization	(16)	0.05	0.02	0.06	0.10	0.23	0.22			
Number of software categories org is not in	(17)	0.03	0.54	0.39	0.01	0.02	0.11	0.01		
Org is only member of its category	(18)	-0.04	-0.14	-0.44	-0.02	-0.04	-0.08	-0.03	-0.07	
Number of software organizations	(19)	0.00	0.58	0.40	0.02	0.06	0.16	0.02	0.98	-0.06

Table 3. Piecewise continuous hazard rate models on an organization's likelihood to create a new label or join an existing label/category.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Dependent Variable:	**New label**	**New label**	**Existing category**	**New label**	**Existing category**	**New label**
Knowledge proximity to different labels/categories x leniency of organization's labels/categories				-0.0002 +	0.00002	
				(0.0001)	(0.0001)	
Knowledge proximity to different labels/categories		0.0040 *	0.0026 *	0.0081 **	0.0023	
		(0.0020)	(0.0010)	(0.0028)	(0.0016)	
K- prox to different labels/categories, organization in low leniency labels/categories (0-5]						0.0093 *
						(0.0037)
K- prox to different labels/categories, organization is in low/med leniency labels/categories (5-20]						0.0048 *
						(0.0021)
K- prox to different labels/categories, organization is in med/high leniency labels/categories (20-30]						-0.0089
						(0.0066)
K- prox to different labels/categories, organization is in high leniency labels/categories (30+)						-0.0012
						(0.0042)
Knowledge proximity to organization's labels/categories x mean leniency				0.0003 **	-0.00003	
				(0.0001)	(0.0001)	
Knowledge proximity to organization's labels/categories		0.0077 **	0.0022	-0.0032	0.0033	0.0118 **
		(0.0024)	(0.0017)	(0.0046)	(0.0040)	(0.0029)
Mean leniency	0.0056 *	0.0052 *	-0.0015	0.0050 *	-0.0015	0.0055 *
	(0.0025)	(0.0025)	(0.0009)	(0.0026)	(0.0010)	(0.0025)
Mean fuzziness	0.9359 +	0.9990 *	0.7819 **	1.060 *	0.7784 **	1.066 *
	(0.4886)	(0.4945)	(0.1968)	(0.4945)	(0.1972)	(0.4963)
Number of total patents over organization's history	-0.00010 **	-0.00009 **	0.00003 +	-0.00009 **	0.00001	-0.0001 **
	(0.00004)	(0.00004)	(0.00002)	(0.00004)	(0.00004)	(0.0000)

** p<0.01 * p<0.05 + p<0.10

Table 3 (cont'd). Piecewise continuous models on an organization's likelihood to create a new label or join an existing category.

Dependent Variable:	Model 1 New Label		Model 2 New label		Model 3 Existing category		Model 4 New label		Model 5 Existing category		Model 6 New label	
Number of press releases from organization last year	0.0015	+	0.0014	+	0.0017	*	0.0014	+	0.0018	*	0.0013	+
	(0.0008)		(0.0008)		(0.0007)		(0.0008)		(0.0008)		(0.0008)	
Number of categories organization is a member of	0.1544	**	0.1502	**	0.0608	**	0.1509	**	0.0601	**	0.1492	**
	(0.0220)		(0.0228)		(0.0160)		(0.0227)		(0.0160)		(0.0231)	
Number of acquisitions last year	0.2910	*	0.2654	*	0.1435	+	0.2733	*	0.1455	*	0.2636	*
	(0.1258)		(0.1261)		(0.0749)		(0.1259)		(0.0747)		(0.1263)	
Number of software categories org is not in	0.0255	**	0.0249	**	0.0329	**	0.0246	**	0.0329	**	0.0248	**
	(0.0083)		(0.0080)		(0.0114)		(0.0079)		(0.0114)		(0.0081)	
Org is only member of its category	0.6992	+	0.7191	+	0.2220		0.7162	+	0.2217		0.7107	+
	(0.3856)		(0.3874)		(0.1983)		(0.3884)		(0.1982)		(0.3888)	
Number of software organizations	-0.0026	*	-0.0025	*	-0.0027	*	-0.0024	*	-0.0027	*	-0.0025	*
	(0.0010)		(0.0010)		(0.0013)		(0.0010)		(0.0013)		(0.0010)	
Year 1993 - 1994 dummy	-0.9849	**	-0.9752	**	-0.8299	**	-0.9737	**	-0.8314	**	-0.9680	**
	(0.3253)		(0.3205)		(0.3195)		(0.3185)		(0.3203)		(0.3220)	
Year 1995 - 1996 dummy	-1.706	**	-1.707	**	-1.906	**	-1.705	**	-1.908	**	-1.674	**
	(0.5311)		(0.5202)		(0.5894)		(0.5165)		(0.5911)		(0.5236)	
Year 1997 - 1998 dummy	-2.305	**	-2.365	**	-2.649	**	-2.359	**	-2.652	**	-2.302	**
	(0.6832)		(0.6715)		(0.6924)		(0.6687)		(0.6943)		(0.6776)	
Year 1999 - 2000 dummy	-2.311	**	-2.307	**	-2.945	**	-2.333	**	-2.948	**	-2.223	**
	(0.8129)		(0.8024)		(0.6602)		(0.8028)		(0.6616)		(0.8081)	
Year 2001 - 2002 dummy	-3.982	**	-3.927	**	-3.424	**	-3.956	**	-3.430	**	-3.828	**
	(0.8308)		(0.8268)		(0.5209)		(0.8283)		(0.5217)		(0.8317)	

** $p<0.01$ * $p<0.05$ + $p<0.10$

Table 3 (cont'd). Piecewise continuous models on an organization's likelihood to create a new label or join an existing category.

Dependent Variable:	Model 1 New Label		Model 2 New label		Model 3 Existing category		Model 4 New label		Model 5 Existing category		Model 6 New label	
0-1 year since org has joined/started category/label	-4.962	**	-4.951	**	-4.500	**	-4.962	**	-4.502	**	-4.961	**
	(0.5422)		(0.5236)		(0.7975)		(0.5187)		(0.8001)		(0.5322)	
1-2 year since org has joined/started category/label	-5.152	**	-5.120	**	-4.551	**	-5.129	**	-4.554	**	-5.128	**
	(0.5545)		(0.5360)		(0.8042)		(0.5308)		(0.8068)		(0.5441)	
2-5 year since org has joined/started category/label	-5.272	**	-5.236	**	-4.504	**	-5.244	**	-4.508	**	-5.245	**
	(0.5661)		(0.5475)		(0.8068)		(0.5424)		(0.8094)		(0.5555)	
5-10 year since org has joined/started category/label	-5.736	**	-5.695	**	-4.399	**	-5.698	**	-4.403	**	-5.696	**
	(0.6461)		(0.6302)		(0.8048)		(0.6258)		(0.8075)		(0.6374)	
10+ year since org has joined/started category/label	-4.344	**	-4.301	**	-3.760	**	-4.303	**	-3.764	**	-4.305	**
	(1.049)		(1.041)		(0.953)		(1.040)		(0.9558)		(1.045)	
Log Likelihood	-1725.39		-1716.49		-4856.62		- 1713.44		-4857.19		-1713.46	
Degrees of freedom	19		21		21		23		23		24	
Number of observations	14357		14357		14357		14357		14357		14357	

** p<0.01 * p<0.05 + p<0.10

Table 4.. Piecewise continuous hazard rate models on patenting organizations only.

Dependent variable: New label	**Model 7**		**Model 8**		**Model 9**		**Model 10**	
Knowledge proximity to different labels/categories x leniency of organization's labels/categories					-0.0001			
					(0.0001)			
Knowledge proximity to different labels/categories			0.0033		0.0060	+		
			(0.0023)		(0.0033)			
Knowledge proximity different labels/categories, organization is in low leniency labels/categories (0-5]							0.0081	+
							(0.0046)	
Knowledge proximity to different labels/categories, organization is in low/med leniency labels/categories (5-20]							0.0041	+
							(0.0023)	
Knowledge proximity to different labels/categories, organization is in med/high leniency labels/categories (20-30]							-0.0107	
							(0.0067)	
Knowledge proximity to different labels/categories, organization is in high leniency labels/categories (30+)							-0.0011	
							(0.0042)	
Knowledge proximity to organization's labels/categories x mean leniency					0.0004	**		
					(0.0001)			
Knowledge proximity to organization's labels/categories			0.0096	**	-0.0038		0.0136	**
			(0.0024)		(0.0044)		(0.0029)	
Mean leniency	0.0004		-0.0020		-0.0050		-0.0013	
	(0.0044)		(0.0045)		(0.0048)		(0.0046)	
Mean fuzziness	1.317		1.404	+	1.662	*	1.670	*
	(0.8222)		(0.8258)		(0.8188)		(0.8296)	
Number of total patents over organization's history	-0.00007	**	-0.00005	*	-0.00005	*	-0.00005	*
	(0.0000)		(0.0000)		(0.0000)		(0.0000)	

** p<0.01 * p<0.05 + p<0.10

Table 4 (cont'd). Piecewise continuous hazard rate models on patenting organizations only.

Dependent variable: New label	**Model 7**		**Model 8**		**Model 9**		**Model 10**	
Number of press releases from organization last year	0.0008		0.0008		0.0007		0.0008	
	(0.0008)		(0.0007)		(0.0007)		(0.0008)	
Number of categories organization is a member of	0.1845	**	0.1693	**	0.1739	**	0.1705	**
	(0.0401)		(0.0406)		(0.0397)		(0.0407)	
Number of acquisitions last year	0.3578	*	0.3558	*	0.3662	*	0.3518	*
	(0.1512)		(0.1520)		(0.1524)		(0.1527)	
Number of software categories org is not in	0.0740	*	0.0688	*	0.0665	*	0.0709	*
	(0.0298)		(0.0296)		(0.0296)		(0.0296)	
Org is only member of its category	1.561	*	1.576	*	1.576	*	1.556	*
	(0.6498)		(0.6596)		(0.6625)		(0.6678)	
Number of software organizations	-0.0084	*	-0.0079	*	-0.0075	*	-0.0083	*
	(0.0034)		(0.0034)		(0.0034)		(0.0034)	
Year 1993 - 1994 dummy	-2.392	**	-2.244	*	-2.219	*	-2.282	*
	(0.8926)		(0.8811)		(0.8741)		(0.8805)	
Year 1995 - 1996 dummy	-4.049	*	-3.817	*	-3.769	*	-3.827	*
	(1.583)		(1.564)		(1.556)		(1.561)	
Year 1997 - 1998 dummy	-4.649	*	-4.501	*	-4.460	*	-4.427	*
	(1.950)		(1.930)		(1.918)		(1.932)	
Year 1999 - 2000 dummy	-4.447	*	-4.056	*	-4.185	*	-3.922	+
	(2.071)		(2.051)		(2.035)		(2.052)	
Year 2001 - 2002 dummy	-4.344	*	-3.776	*	-3.923	*	-3.543	+
	(1.865)		(1.866)		(1.853)		(1.873)	

** p<0.01 * p<0.05 + p<0.10

Table 4 (cont'd). Piecewise continuous hazard rate models on patenting organizations only.

Dependent variable: New label	Model 7		Model 8		Model 9		Model 10	
0-1 year since org has joined/started category/label	-8.208	**	-7.837	**	-7.807	**	-8.063	**
	(2.292)		(2.268)		(2.265)		(2.280)	
1-2 year since org has joined/started category/label	-8.532	**	-8.191	**	-8.153	**	-8.396	**
	(2.316)		(2.290)		(2.285)		(2.299)	
2-5 year since org has joined/started category/label	-8.794	**	-8.364	**	-8.323	**	-8.583	**
	(2.311)		(2.288)		(2.285)		(2.298)	
5-10 year since org has joined/started category/label	-8.881	**	-8.412	**	-8.362	**	-8.607	**
	(2.384)		(2.358)		(2.353)		(2.367)	
10+ year since org has joined/started category/label	-20.11	**	-19.19	**	-18.14	**	-19.62	**
	(2.441)		(2.420)		(2.419)		(2.428)	
Log Likelihood	-530.20		-521.28		-517.61		-518.21	
Degrees of freedom	19		21		23		24	
Number of observations	2843		2843		2843		2843	

** p<0.01 * p<0.05 + p<0.10

Figures

Figure 1. Knowledge space and market space for the software industry

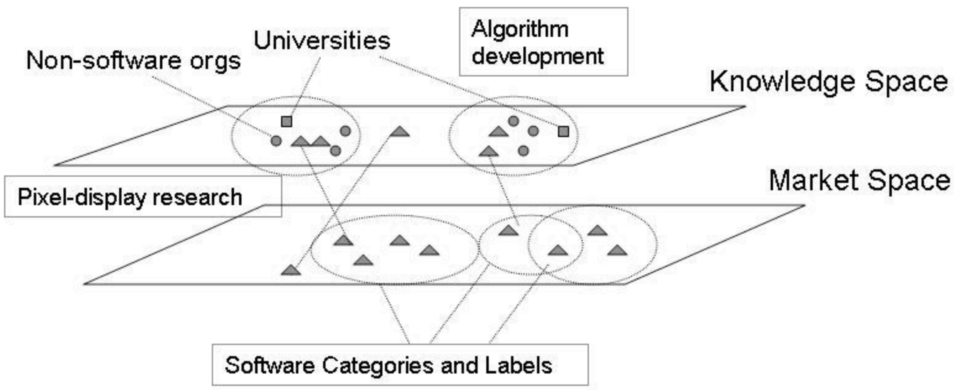


Figure 2a. Example press release from Accrue. Accrue claims to be "a leading provider of eBusiness analysis software and services."

BusinessWire

Accrue **Announces BuyPath Offering Unmatched Merchandising Analysis of the Visitor Path From Entry to Purchase**

1,098 words
4 October 1999
13:21 GMT
Business Wire
English
(c) 1999 Business Wire

FREMONT, Calif.--(BUSINESS WIRE)--Oct. 4, 1999--

New Feature of Accrue Insight(TM) eBusiness Analysis

Application Provides Powerful Analysis of Web Site Navigation

by Customer Segment

Accrue Software, Inc. (NASDAQ: ACRU), a leading provider of eBusiness analysis software and services, today announced BuyPath(TM), a new feature of Accrue Insight(TM) that enables eBusiness marketers to analyze and compare site navigation for customer segments and to gain insights into visits that involve transactions or that touch high-value content.

Using BuyPath, the marketer can determine which visitor segments are the most valuable and which referrers and content are most effective in accomplishing eBusiness goals. A key goal of ecommerce is converting visitors into customers. BuyPath enables the comparison of the navigation patterns of precisely-defined customer segments.

For example, comparing new visitors from Yahoo! against returning visitors from Excite, eCommerce

Figure 2b. Example identity claims to labels and categories for organizations in these data.

Company	**Date**	**Description**
Citrix Systems	February 2000	Citrix Systems, Inc. is a global leader in application server software and services
Plasmon	August 2000	Plasmon, a leading manufacturer of automated data storage solutions, today announced its Diamond® storage management software.
Watson General	May 1994	Watson General currently provides remote software monitoring systems.
Comergent Technologies	Sept 2002	Comergent Technologies® Inc., the leading provider of sell-side e-business software solutions
Accrue Software	October 1999	Accrue Software, a leading provider of e-business analysis software and services
ACP	July 2001	ACP provides enterprise web publishing and e-business solution
Alliance	March 2001	Alliance offers the technical and business advantages of the Sybase Enterprise Portal with a wide range of e-business solutions, including content, e-commerce, and business process automation and analysis

Figure 3. Number of software producers by year.

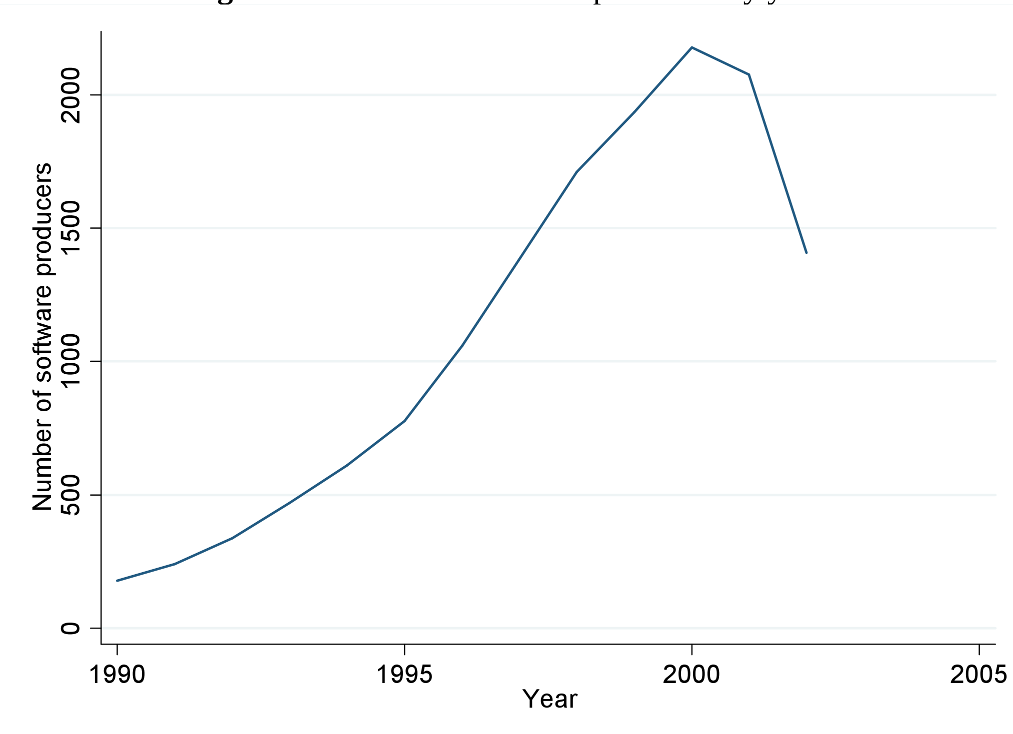


Figure 4. Number of software labels by year.

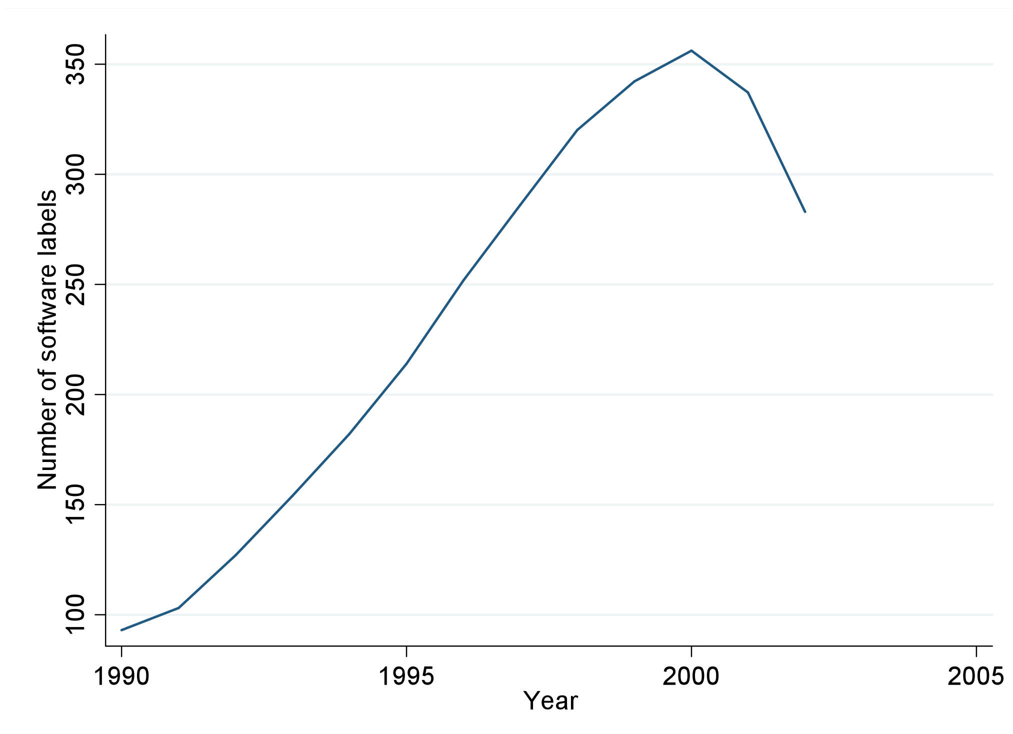
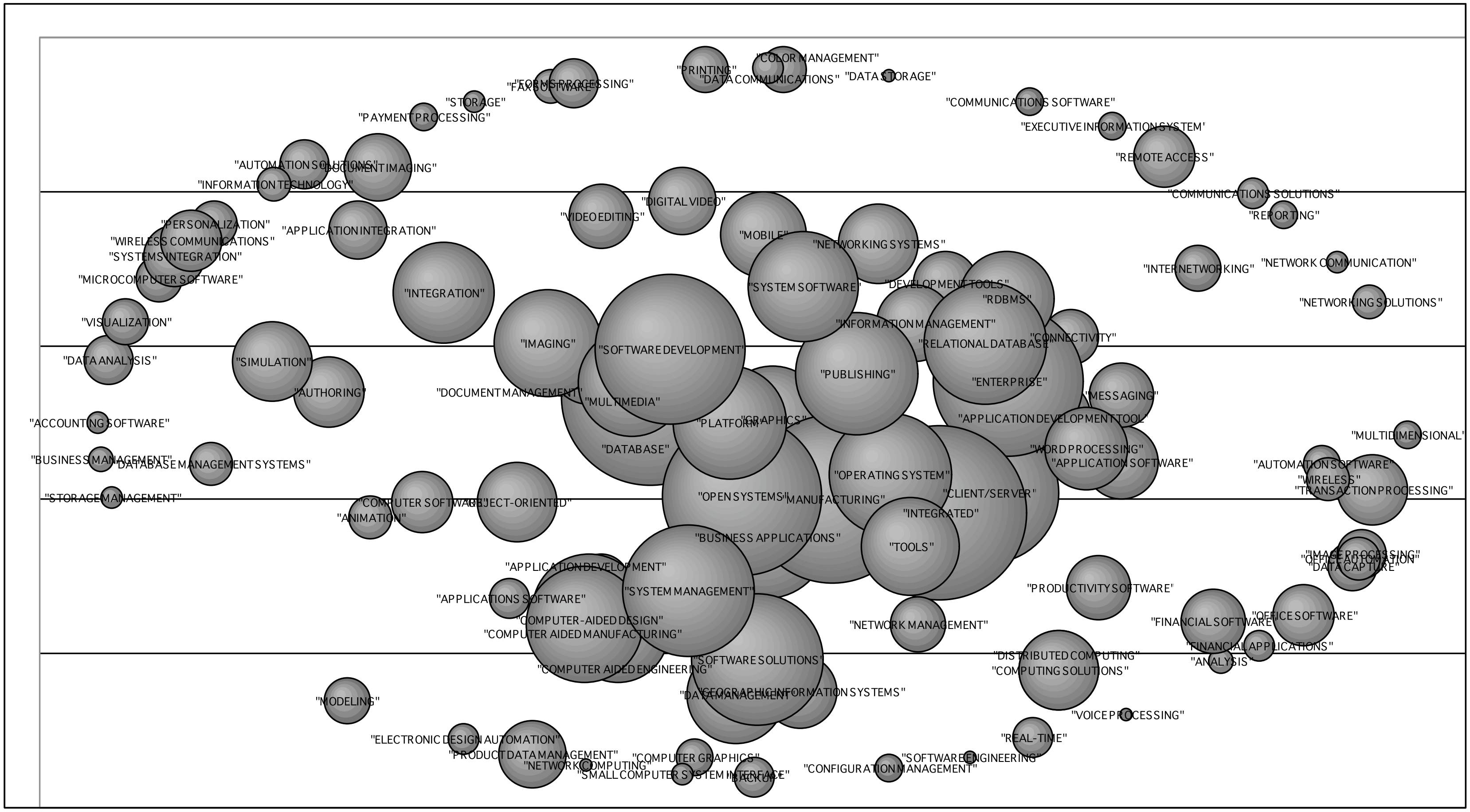
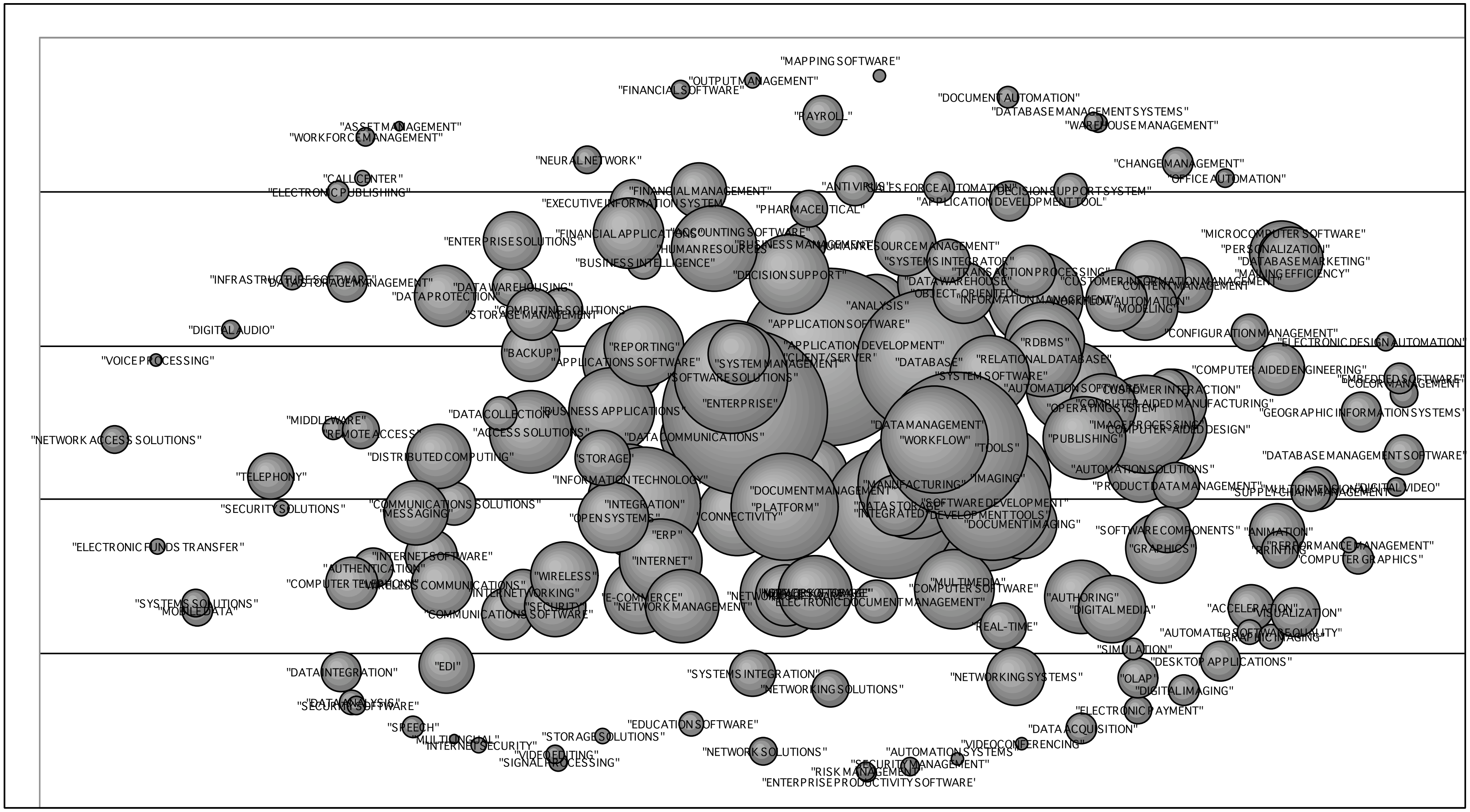


Figure 5. Labels in the software industry for selected years.

1992



1995


"MAPPING SOFTWARE"
"OUTPUT MANAGEMENT"
"FINANCIAL SOFTWARE"
"DOCUMENT AUTOMATION"
"DATABASE MANAGEMENT SYSTEMS"
"WAREHOUSE MANAGEMENT"
"PAYROLL"
"ASSET MANAGEMENT"
"WORKFORCE MANAGEMENT"
"NEURAL NETWORK"
"CHANGE MANAGEMENT"
"OFFICE AUTOMATION"
"CALL CENTER"
"ELECTRONIC PUBLISHING"
"FINANCIAL MANAGEMENT"
"EXECUTIVE INFORMATION SYSTEM"
"ANTIVIRUS"
"SALES FORCE AUTOMATION"
"DECISION SUPPORT SYSTEM"
"APPLICATION DEVELOPMENT TOOL"
"PHARMACEUTICAL"
"ENTERPRISE SOLUTIONS"
"FINANCIAL APPLICATIONS"
"ACCOUNTING SOFTWARE"
"HUMAN RESOURCES"
"BUSINESS MANAGEMENT"
"HUMAN RESOURCE MANAGEMENT"
"MICROCOMPUTER SOFTWARE"
"PERSONALIZATION"
"DATABASE MARKETING"
"MAILING EFFICIENCY"
"BUSINESS INTELLIGENCE"
"SYSTEMS INTEGRATOR"
"INFRASTRUCTURE SOFTWARE"
"DATA STORAGE MANAGEMENT"
"DECISION SUPPORT"
"TRANSACTION PROCESSING"
"CUSTOMER INFORMATION MANAGEMENT"
"CONTENT MANAGEMENT"
"DATA WAREHOUSING"
"DATA PROTECTION"
"DATA WAREHOUSE"
"OBJECT-ORIENTED"
"INFORMATION MANAGEMENT"
"WORKFLOW AUTOMATION"
"MODELING"
"COMPUTING SOLUTIONS"
"STORAGE MANAGEMENT"
"ANALYSIS"
"DIGITAL AUDIO"
"APPLICATION SOFTWARE"
"CONFIGURATION MANAGEMENT"
"ELECTRONIC DESIGN AUTOMATION"
"REPORTING"
"APPLICATION DEVELOPMENT"
"RDBMS"
"VOICE PROCESSING"
"BACKUP"
"APPLICATIONS SOFTWARE"
"SYSTEM MANAGEMENT"
"CLIENT/SERVER"
"DATABASE"
"RELATIONAL DATABASE"
"COMPUTER AIDED ENGINEERING"
"SOFTWARE SOLUTIONS"
"SYSTEM SOFTWARE"
"EMBEDDED SOFTWARE"
"COLOR MANAGEMENT"
"AUTOMATION SOFTWARE"
"CUSTOMER INTERACTION"
"DATA COLLECTION"
"BUSINESS APPLICATIONS"
"ENTERPRISE"
"COMPUTER AIDED MANUFACTURING"
"OPERATING SYSTEM"
"GEOGRAPHIC INFORMATION SYSTEMS"
"MIDDLEWARE"
"REMOTE ACCESS"
"ACCESS SOLUTIONS"
"DATA MANAGEMENT"
"IMAGE PROCESSING"
"COMPUTER AIDED DESIGN"
"NETWORK ACCESS SOLUTIONS"
"DATA COMMUNICATIONS"
"WORKFLOW"
"PUBLISHING"
"TOOLS"
"DISTRIBUTED COMPUTING"
"STORAGE"
"DATABASE MANAGEMENT SOFTWARE"
"AUTOMATION SOLUTIONS"
"TELEPHONY"
"INFORMATION TECHNOLOGY"
"IMAGING"
"MANUFACTURING"
"DOCUMENT MANAGEMENT"
"PRODUCT DATA MANAGEMENT"
"SUPPLY CHAIN MANAGEMENT"
"MULTIDIMENSIONAL"
"DIGITAL VIDEO"
"COMMUNICATIONS SOLUTIONS"
"SECURITY SOLUTIONS"
"MESSAGING"
"INTEGRATION"
"PLATFORM"
"DATA STORAGE"
"SOFTWARE DEVELOPMENT"
"INTEGRATED DEVELOPMENT TOOLS"
"OPEN SYSTEMS"
"CONNECTIVITY"
"DOCUMENT IMAGING"
"SOFTWARE COMPONENTS"
"ANIMATION"
"ERP"
"PERFORMANCE MANAGEMENT"
"PRINTING"
"ELECTRONIC FUNDS TRANSFER"
"GRAPHICS"
"COMPUTER GRAPHICS"
"INTERNET SOFTWARE"
"INTERNET"
"AUTHENTICATION"
"WIRELESS"
"MULTIMEDIA"
"COMPUTER TELEPHONY"
"WIRELESS COMMUNICATIONS"
"INTERNETWORKING"
"COMPUTER SOFTWARE"
"AUTHORING"
"SYSTEMS SOLUTIONS"
"MOBILE DATA"
"E-COMMERCE"
"SECURITY"
"NETWORK MANAGEMENT"
"ELECTRONIC DOCUMENT MANAGEMENT"
"DIGITAL MEDIA"
"ACCELERATION"
"VISUALIZATION"
"COMMUNICATIONS SOFTWARE"
"REAL-TIME"
"AUTOMATED SOFTWARE QUALITY"
"GRAPHIC IMAGING"
"SIMULATION"
"DESKTOP APPLICATIONS"
"EDI"
"DATA INTEGRATION"
"SYSTEMS INTEGRATION"
"NETWORKING SYSTEMS"
"OLAP"
"NETWORKING SOLUTIONS"
"DIGITAL IMAGING"
"DATA ANALYSIS"
"SECURITY SOFTWARE"
"ELECTRONIC PAYMENT"
"EDUCATION SOFTWARE"
"DATA ACQUISITION"
"SPEECH"
"MULTILINGUAL"
"STORAGE SOLUTIONS"
"INTERNET SECURITY"
"VIDEOCONFERENCING"
"AUTOMATION SYSTEMS"
"VIDEO EDITING"
"NETWORK SOLUTIONS"
"SECURITY MANAGEMENT"
"SIGNAL PROCESSING"
"RISK MANAGEMENT"
"ENTERPRISE PRODUCTIVITY SOFTWARE"

1998

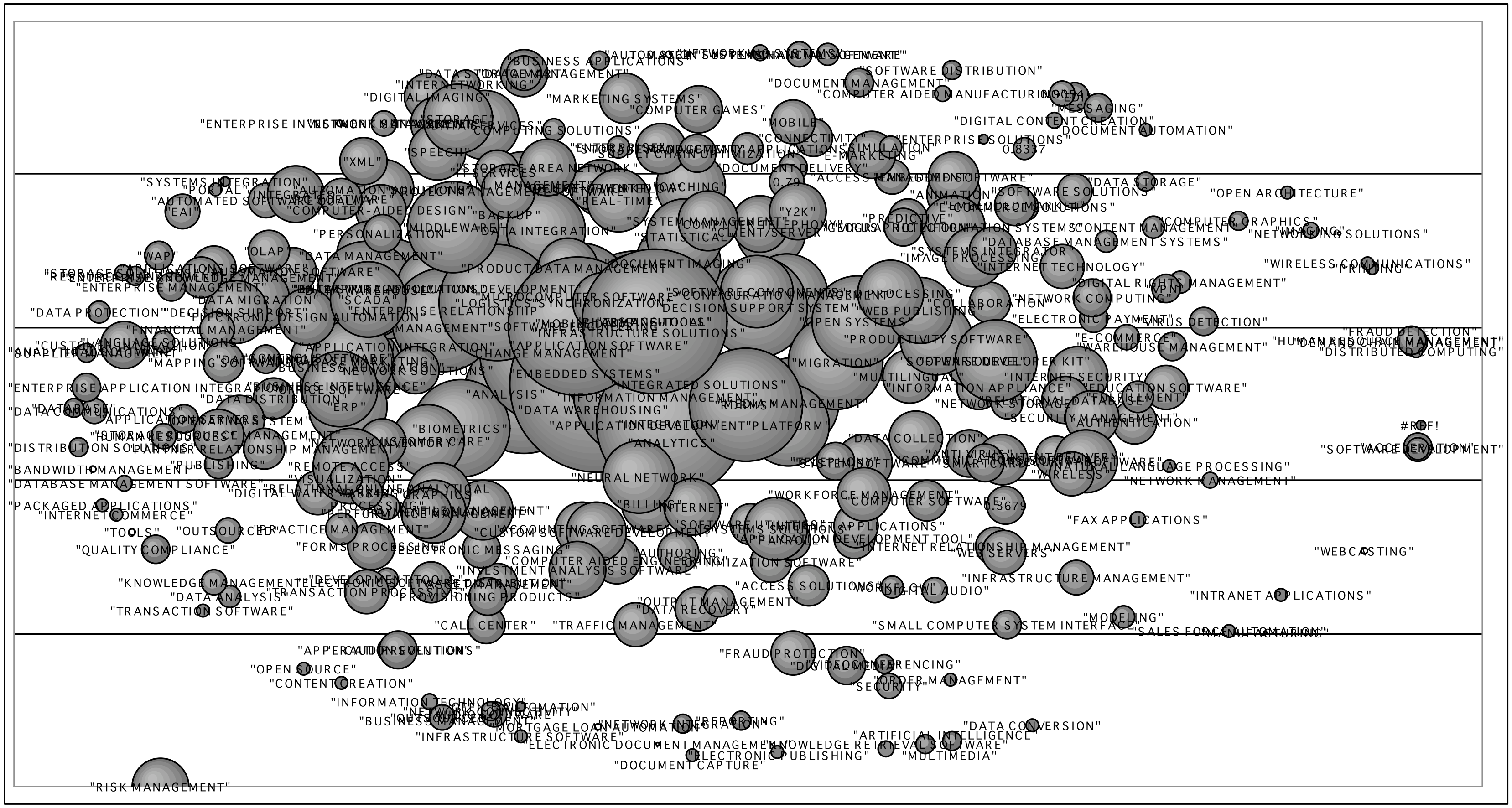
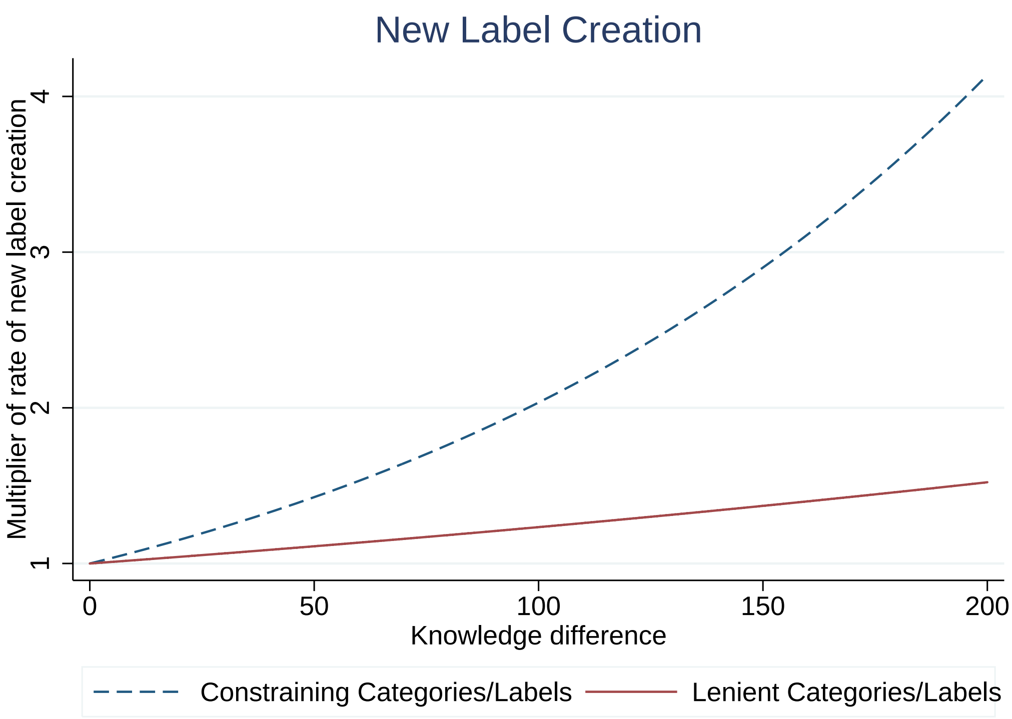
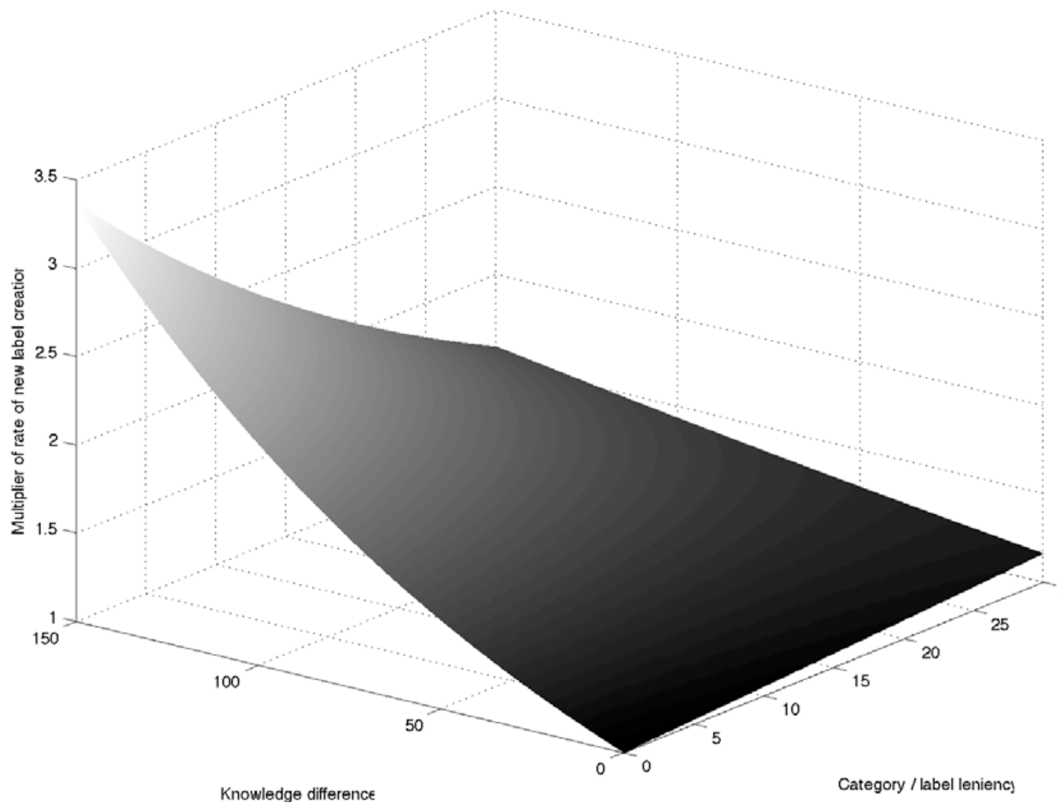


Figure 6. New label creation for organizations in constraining versus lenient categories and labels.



* Constraining categories/labels evaluated at leniency = 5
** Lenient categories/labels evaluated at leniency = 30

Figure 7. New label creation by knowledge difference and category / label leniency



Appendix A: Software Industry data collection and Market Space Measures

Data on the Software industry organizations and categories was collected from Press Releases issued by software companies during 1990 – 2002. *Businesswire*, *PR Newswire*, and *Computerwire* are the three publications that software companies generally use to release news, and they are available in electronic format for the time period of this study. Organizations in other industries besides software also use these publications, and so I filter on press releases that mention “software” at least three times. For my raw data, I collected every press release in these three publications with at lease three mentions of “software” dated 1990 – 2002. There are 268,963 of these. At least once in a press release, companies will refer to themselves by their full name, such as Oracle, Corp. To scrape the names of software organizations from the press releases, I wrote a program to automatically pull out words before Inc, Corp, Co, LLP, or capitalized Software. These rules cast a very broad net and returned both the names of software companies and extra “junk,” like sentences or phrases. The initial output contained over 300,000 rows of potential Software firms. To filter out the “junk,” I ran a series of cleaning steps, resulting in a list of 11,390 phrases that were potential software company names. I then ran manual searches through the raw press release data for each of these to determine if the name represented a software company. This produced 5063 potential software firms.

Next, I coded a program to automatically search through the press releases for the sentences where organizations make claims to be affiliated with specific categories and labels. Software companies use fairly standardized language when they describe themselves, stating that the company “is a leader in” a category or label, “is a provider of” the category or label, “develops” the category or label, etc. I created a program to search through the press release data for the each software company followed by these keywords and to extract the press release date and the descriptive sentence for each company.

In the next step, I created a set of terms that identify categories and labels used to classify organizations during this time period, using both external sources and the descriptive sentences extracted from the press releases. First, I compiled a list of categories and labels that are used to classify companies in the industry publications *Software Magazine* and *Computerworld*. Then, I read the sentences extracted from the press releases that contained identity claims to compile a list of popular labels. In total, I created a list of 479 terms identifying categories or diffuse labels. I then matched each firm with a set of categories and/or labels over time. Due to the nature of these data, there is a possibility that an organization might mention a category or diffuse label in a different context, and so it is possible that this first match would cast too wide a net in terms of including organizations in categories. To minimize this, I only include organization-category entries that are mentioned in multiple years. This means that organizations that exist for less than one year are not included in these data. Nevertheless, this data set contains a sample of organizations and categories within the software industry that is much more inclusive than any alternative data set of which I am aware.

Acquisitions were an important part of the software industry, and they likely contributed to changes in categorization. Acquisitions were also publicly announced in press releases, making these data a good source for identifying acquisitions. Using perl, I automatically scanned all press release headlines for the name of each software organization in these data and a term related to acquisition, such as "merge" or "acquire." I then manually scanned through the results to identify acquisitions and the year of acquisition during this time period.

Next, I measure the constraint of each category or label in these data in a given year. Recall that organizational categories emerge when audiences develop a schema for the meaning of a label.

The extent to which members identify with emerging labels is critical to whether a category develops and is recognized by outsiders (DiMaggio, 1987;Rao, Monin, and Durand, 2003;Hannan, Pólos, and Carroll, 2007). Previous research shows that when organizations are members of multiple categories, they suffer from lower performance (Hsu, 2006). When members of an emerging categories also identify with other groups, the focal category is less likely to become taken for granted (McKendrick and Carroll, 2001;Bogaert, Boone, and Carroll, 2006).

To create a measure for label or category leniency, I build on Hannan, Pólos, and Carroll's (2007) formalization of category emergence, where organizations vary in the degree to which they identify as members of a category or label. I assign organizations a full or partial grade of membership depending on the number of times they self-identify with the respective category in their press releases, divided by the number of times they identify with any label or category:

$$w_{i,C_j} = \frac{I_{Cj}}{\sum_{j=1}^{N} I_{C_j}} \qquad 0 < w_{i,C_j} \le 1 \qquad \text{(A1)}$$

Here, I_{C_j} is the number of times organization i claims to be affiliated with category C_j in its press releases. In this study, I compute w_{i,C_j} yearly. This graded membership weights the number of times an organizations claims to affiliate with a focal category, divided by the number of claims it makes to affiliate with any category. If an organization makes ten claims, and nine are "business process management" and one is "enterprise," then w_{i,C_j} =0.9 for $j =$"business process management" and w_{i,C_j} =0.1 for $j =$ "enterprise." I then compute a category or label's "fuzzy" density by summing these partial memberships:

$$N_C = \sum_{i \in C} w_{i,C} \qquad \text{(A2)}$$

The fuzzy density takes into account whether members derive their primary identity from the label or category, or whether the label or category is a smaller part of a member's identity. This fuzzy density can be compared to the support of the label or category, N ,which sums the number of organizations that have non-zero grades of membership in the label or category. These measures allow researchers to distinguish between labels or categories with all dedicated members or with many partial members. Labels or categories that have all dedicated members (where $w_{i,C_j} \sim 1$) are more sharply defined than those with all partial members. This is captured by the contrast, which divides the fuzzy density by the number of non-zero category members:

$$contrast_C = \frac{N_C}{N} \qquad \text{(A3)}$$

A high level of contrast indicates that most members of the label or category are dedicated members. A label is more likely to represent a category if it has high contrast . Conversely, the extent to which a category does not have dedicated members, or its "fuzziness," is one minus the contrast:

$$fuzz_C = 1 - contrast_C \qquad \text{(A4)}$$

Fuzziness represents the extent to which members identify with at least one other label or category. The distribution of fuzziness for categories in these data is shown in figure A1.

(Figure A1 about here)

In the software industry, there are many labels and categories with a moderate level of contrast and fuzziness –that substantially overlap, but that also have a large proportion of dedicated members. When there is a moderate or high level of fuzziness, this may indicate that it is not clear what it means to be a classified as a member of the label. However, it also may mean that there are overlapping labels or categories. If most of the members of one label also identify with one other label, the label will have a medium level of contrast, but it might be highly restrictive. For instance, the labels “mobile” and “wireless” show moderate levels of contrast (between 0.3 and 0.6 over the date range of this study), but this is mainly due to the overlap between these two categories. On the other hand, the label “enterprise” has a contrast of around 0.5, but this is due to overlap with 284 other categories. The “enterprise” classification is an example of lenient label (has low constraint); it was used from 1991 through 2002 and has many organizational members, with a good portion of dedicated members, and is widely recognized by customers and analysts as a type of software. At the same time, claiming to be affiliated with this label does not strongly constrain an organization. As a result, software companies can plausibly claim an affiliation with every other label or category.

I refer to these low constraint labels as lenient. I create a measure for how lenient[1] a label is using fuzziness, defined in equation A3, as well as the number of distinct labels or categories with which affiliated organizations also identify. High fuzziness might simply indicate that a meaningful category overlaps with one other meaningful category. However, if a label has both high fuzziness, and if members identify with many different categories, the label also lacks constraint, or is lenient. Therefore I multiply a label’s fuzziness by the number of distinct other labels or categories to which members of the focal category or label also belong:

$$leniency_C = fuzz_C \cdot N_{ocat} \qquad \text{(A5)}$$

(Figure A2 about here)

Figure A2 illustrates the relationship between the fuzziness of a label or category and its leniency. At low levels of fuzziness leniency is also low. As fuzziness increases, the range of

[1] Leniency of a label is its lack of constraint. Labels that are low on leniency are high on constraint.

leniency also increases, and its highest levels are found for labels or categories with medium levels of fuzziness (or contrast). In the software industry, there are many labels with about half "dedicated" organizational members and half partial members. Of these, some have partial members that are scattered in terms of their alternate identifications, and are also identified with a large number of other categories and labels. Others are more restrictive, where partial members are only identified with a handful of other groups.

Bogaert, S., et al.
2006 "Contentious Legitimacy: Professional Association and Density Dependence in the Dutch Audit Industry 1884-1939." Stanford University Working paper.

DiMaggio, P.J.
1987 "Classification in art." Annual Review of Sociology, 52: 440-455.

Hannan, M.T., et al.
2007 Logics of Organization Theory. Princeton: Princeton University Press.

Hsu, G.
2006 "Jacks of all Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production." Administrative Science Quarterly, 51.

McKendrick, D., and G. Carroll
2001 "On the genesis of organizational forms: Evidence from the market for disk drive arrays." Organization Science, 12: 661-682.

Rao, H., et al.
2003 "Institutional Change in Toque Ville: Nouvelle Cuisine as an Identity Movement in French Gastronomy." American Journal of Sociology, 108: 795-843.

Appendix A Figures:

Figure A1. Distribution of fuzziness for labels in the software industry.

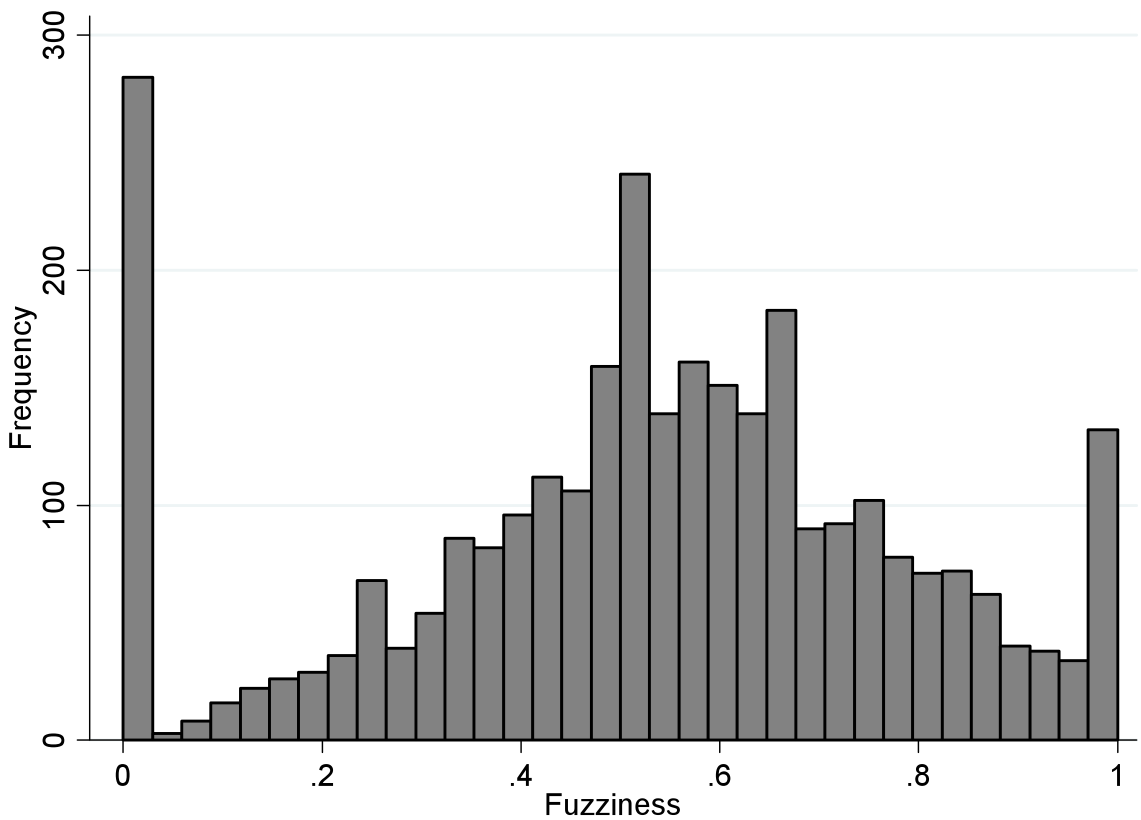
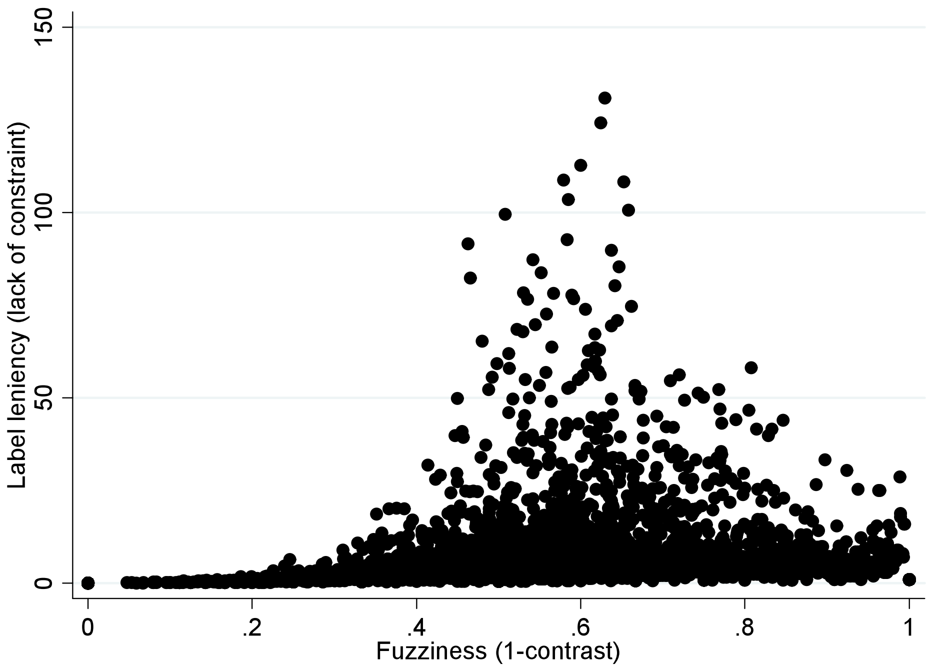


Figure A2. Relationship between label leniency and fuzziness.



Appendix B: Knowledge Space Construction and Measures

I create knowledge space measures for each year from 1990 through 2002 using a five-year window of all patents applied for in current year and four years prior. Using citation overlap, I measure the similarities between all pairs of patents applied for in the five-year window, and link them together, as illustrated in figure B1, which depicts a simplified knowledge space comprised of ten patents. The dots represent patents in the relevant categories, and the lines indicate that two patents are similar, marked by their similarity coefficients. In this example, patent 1 has a first order similarity to patents 2, 3, and 4, and second order similarities to patents 5, 6, 7, 9, and 10. Once knowledge space is constructed, I look for patents that were created by software organizations as identified from the press releases to locate organizations in knowledge space.

To measure similarity between patents, I take every patent in the five-year window, and compare it to every other patent in these data. Following Podolny et al (1996), I measure the number of overlapping citations between these patents, and divide by the number of citations made by the focal patent:

$$\alpha_{ij} = \frac{s_{ij}}{s_i} \qquad \text{(B1)}$$

Where s_{ij} is the number of shared citations between patent i and j, and s_i is the total number of citations by patent i. $\alpha_{ij} \in [0,1]$, and is the first-order similarity between patent i and all j's with which patent i shares at least one citation. Using this measure, I construct knowledge space measures at the patent level for each year from 1990 – 2002.

With knowledge space measures constructed, I compute second order similarities by looking at all patent k's that each patent j is similar to (for each j such that $\alpha_{ij} > 0$). The similarity measurement between any patent i and patent k is constructed by:

$$\sigma_{ik} = \max_{j,k}\{\alpha_{ij} \cdot \alpha_{jk}\} \tag{B2}$$

Using second order similarities allows for a more refined similarity measure. If there is a first degree similarity, or if α_{ik} >0, then $\sigma_{ik} = \alpha_{ik}$. However, even if α_{ik} =0, σ_{ik} may be non-zero, if patents i and k are both similar to a patent j. For example, in figure 2 patent 1 has a similarity of 0.5 to patent 2, and has a second order similarity of 0.9*0.6=0.54 to patent 7.

Knowledge space proximity

Next, I use first and second order similarities to measure where an organization is patenting in knowledge space. I do this by measuring how proximate an organization's patents are to patents associated with different market space categories. Patents are associated with a market space category if they are issued to an organization that claims a non-zero membership in that category. For each organization, for each year, I take every patent the organization applied for (that was eventually issued), and compute its similarity to every other patent in knowledge space within the five-year window. This creates a set of all patents to which the focal patent has non-zero similarity. Next I determine whether these other patents were issued to other software companies, and with which categories the companies (and hence the patents) are identified. I then compute knowledge space proximity metrics between the organization and its own categories, between the organization and different software categories, and between the organization and every other category in the software industry in the given year. I first compute how similar each *patent* belonging to organization A is to patents belonging to other organizations in the respective groups (G). I use organization A's patents from the current year, and compare them to all patents from members of the respective groups applied for[1] in that or the previous five years:

$$sim_{i,G} = \underset{i \in A, l \in G}{sum}(\sigma_{il}), \qquad \text{(B3)}$$

Next, I average these patent-level similarity measures over organization A's patents to create knowledge proximity metrics between organization A and its own categories, C_A per year:

$$kprox_{A,C_A} = \frac{\underset{i \in A}{sum}(sim_{i,C_A})}{npat_A}, \qquad \text{(B4)}$$

Between organization A and all different categories D_A, per year:

$$kprox_{A,D_A} = \frac{\underset{i \in A}{sum}(sim_{i,D_A})}{npat_A}, \qquad \text{(B5)}$$

[1] Only patents that were eventually granted are used.

Appendix B Figures:

Figure 3.1. A simplified knowledge space with ten patents.