The Categorical Basis of Combination: A Theory and Two Empirical Tests

Elizabeth G. Pontikes

University of Chicago Booth School of business

May 2018

Acknowledgements: I would like to thank Juan Alcacer, Bill Barnett, Diane Burton, Ian Cockburn, Michael Hannan, Greta Hsu, and Rosemarie Ziedonis for helpful comments and suggestions.

Abstract

Research in economics and sociology points to combinations as the source of novel developments. Most systematic studies of this thesis show technology combinations underlie high-impact inventions, and imply that results speak to market evolution. This study proposes limitations to the generalizability of this research, as it overlooks that every domain has multiple ways of assessing similarity and difference. The category system an evaluator uses partly determines what is a combination, and therefore what is novel. This idea is explored in two studies. Study 1 provides causal evidence of the cognitive mechanism with a laboratory experiment. Subjects randomly assigned to a category system rate as more *novel* firms that combine the assigned categories compared to firms that combine the alternate categories. Study 2 shows real-world implications. It investigates antecedents of *market novelty* in terms of venture capital (VC) investment in software organizations. Those that combine *market categories* are more likely to receive VC investment, but there is no effect from combining *technology categories*. This is not because market categories are more accurate. Combining technology categories (not market categories) is positively related to technology novelty (high-variance patent citations). Results suggest categories underlie assessments of novelty and associated rewards.

Schumpeter's insight that novelty emerges from "the carrying out of new combinations" spawned rich and diverse literatures (Schumpeter, 1934: 66). Recombination, or new combinations of existing elements, is widely accepted as the primary antecedent for novel developments. Vedres and Stark (2010:1157) "conceptualize entrepreneurship as recombination." Recombination is reified as "the 'holy grail' of innovation research" (Gruber, Harhoff and Hoisl, 2013: 837), the "ultimate source of novelty" (Fleming, 2001: 118; Schoenmakers and Duysters, 2010).

Researchers' prolonged interest in the subject arises from the idea that combination underlies new developments that transform the economy and society. But most systematic studies of this thesis narrowly focus on patents and publications. Inventions based on combinations of technologies have high-variance outcomes in terms of technology or scientific novelty, measured by citations (Fleming, 2001; Nerkar, 2003; Schoenmakers and Duysters, 2010; Uzzi, Mukherjee, Stringer and Jones, 2013; Foster, Rzhetsky and Evans, 2015; Leahey, Beckman and Stanko, 2016). Drawing on studies showing highly-cited patents are the basis for market developments (Trajtenberg, 1990), researchers suggest the link between technology combinations and novelty in markets. There is limited direct research into whether combinations are the basis for market novelty. Sociological studies show that combinations are the source of new cultural and organizational forms (Stark, 1996; Rao, 1998; Carroll and Swaminathan, 2000; Phillips, 2013), but these take an ex-post approach, tracing the roots of a successful category.

This study proposes the body of evidence that shows recombination is the antecedent of novel developments is incomplete, and perhaps misleading, because it overlooks that there are multiple ways of assessing similarity and difference in any domain. The key idea is that categories people hold are foundational to what is a combination and so underlie assessments of novelty. A "new combination" brings together elements that are *different*.

An important implication is the well-established relationship between technology combination and high-impact technologies does not necessarily explain transformative changes in the market. Many industries can be segmented by *market categorization*, based on how products are used, or *technology categorization*, based on how they are built. The more market categorization is decoupled from technology categorization, the less technology combinations will explain the major market disruptions that have motivated studies of market evolution (Schumpeter, 1934; Nelson and Winter, 1982; Weitzman, 1998; Stark, 2009).

This is especially important when there is a lot at stake in assessments of novelty. For example, Venture Capitalists (VCs) are intermediaries who shepherd radical new business ideas to become market realities. They invest in firms that have the potential to revolutionize how people live and work. VC backed organizations – from Microsoft to Netscape to Google to Facebook to Uber – have been sources of major changes in markets and society, generating large fortunes for their investors. After the fact, these successes may seem obvious, but a priori, it is unclear when a novel development will flourish (March, 1991). VCs aim to invest in non-consensus companies with the potential to revolutionize how people interact on the market (Pontikes, 2012; Pontikes and Barnett 2017). An open question is what types of firms show the promise of market-changing novelty and attract VC investment.

The literature on invention might suggest firms whose developments combine technologies are best poised to introduce revolutionary products. But many examples call into question this explanation. Facebook, Uber, and Airbnb did not create high-impact technologies. These firms introduced products that brought together unrelated aspects of the market: using a online interface to “flip” through a university face book, using your phone to hail a taxi, and accessing an online network to rent your apartment for short periods of time. These examples highlight that asserting that “novelty” is based in “combination” is not meaningful without sociological context. To define combination, we need to know how an evaluator assesses similarity and difference in a domain.

Sociological research shows where category boundaries are drawn creates meaning and value (DiMaggio, 1987; DiMaggio, 1997, Lamont and Molnar, 2002). This has implications not only for people, but also organizations and products (Hsu, 2006; Hsu, Hannan and Koçak, 2009; Hannan, 2010). Taking into account there are multiple ways any domain can be segmented (Durand and Paolella, 2013; Goldberg, Hannan and Kovács, 2016) raises the possibility that different categorizations lead to alternate conceptions of meaning and value. This will affect how combination, and therefore novelty, is assessed.

To fix ideas, we can consider each category system to define a "space" for a domain in which in which organizations (or other actors/objects) are located (Pontikes and Hannan, 2014). Figure 1 provides an example for athletic clothing. It shows two spaces, one defined by materials and one by design. There are two objects, a "merino daily shirt" and a "hit the trails skort." In design space, a shirt or skirt is located squarely in one category, while the skort—a blend of shorts and skirts—is a combination. But in materials space, it is the merino daily shirt – a combination of merino wool and polyester – that is the combination. Whether the object is novel depends on *both* its position *and* the category system used by the evaluator.

These ideas are investigated in two empirical studies. Study 1 provides causal evidence of the general cognitive mechanism. In a laboratory experiment, subjects put in a VC role evaluate an entrepreneur's business plan based on how *novel* it is. Findings support the hypothesis: when subjects are randomly assigned to a category system, they rate as more novel business plans that combine the assigned categories as compared to categories from the alternate system. Study 2 shows real-world implications of the theory. It investigates whether market or technology combination underlies investment choices of VCs in software organizations. Because VCs invest in companies with the potential to radically change the way people live and work, VC investment is a good measure of the kind of market-changing novelty that has interested scholars for a century (Schumpeter, 1934; Nelson and Winter, 1982; Weitzman, 1998; Stark, 2009; Vedres and Stark, 2010). Hypotheses are tested with unique data that locates entrepreneurial organizations in both a *technology space* (based on patent classes) and a *market space* (based on market categories from press releases). As expected, VCs invest in market novelty as captured by *market category combination*. Measures of technology combination do not have an effect. This is not simply because market categories are more accurate: *technology combination* (but not market category combination) has the expected positive effect on *technology novelty* (patent citations).

Findings contribute to sociological research on categorization. They support the idea that categories are defined with respect to an individual (or group), and that we must account for alternate category systems within a domain in order to understand effects of social categories. In the context studied, if we did not consider the categories relevant to VCs, researchers may have made the erroneous conclusion that combinations did not drive VC investment. Findings also add to the growing body of work showing systematic ways in which category *non*-conformity is valued (Pontikes, 2012; Goldberg et al, 2016; Paolella and Durand, 2016). Consistent with previous research, *novelty-seeking* evaluators like VCs reward organizations that cross category boundaries.

These studies also contribute to the literature on market evolution and innovation by showing that radical market changes are better predicted by market combinations than technology combinations. The large body of research on technology recombination does not provide a systematic explanation for VC investment – a seminal event in bringing transformative change to the market and greater society. More generally, the relationship between combination and novelty, central to this literature, must be defined with respect to sociological context.

Categories and Combination

Research on market evolution proposes that recombination is the ultimate source of novelty. But in the rich literatures that link new combinations to novelty, researchers have not defined recombination in a systematic way that can be applied across contexts.

Scholars studying a particular domain use institutional categories to demarcate what is similar and what is different, and a new combination blends elements across these boundaries. In research on invention, patent classes group technology components and recombination is when patents cross different classes (Fleming, 2001; Nerkar, 2003; Schoenmakers and Duysters, 2010; Gruber et al., 2013). Studies of scientific novelty use disciplines for boundaries over which "new combinations" are created (Uzzi, Mukherjee, Stringer and Jones, 2013; Foster, Rzhetsky and Evans, 2015; Rzhetsky, Foster, Foster and Evans, 2015; Leahey, Beckman and Stanko, 2016). Scholars of form emergence use rich histories of particular contexts to define boundaries for combination (DiMaggio, 1991; Stark 1996; Rao, 1998; Phillips and Owens, 2004; Carroll and Swaminathan, 2000).

This approach is defensible for a given study, but is problematic in creating a body of research that points to new combinations as the explanation for novelty. The idea that there can be a "new combination" of existing elements assumes an *a priori* state in which some elements are similar and some are different. But where lines of similarity and difference are drawn can substantially change what is a "combination." Similarity and difference are not inherent properties of objects. As philosopher Nelson Goodman argues in his famous critique of similarity, objects can be similar in multiple ways, so if we do not specify in what respects they are similar, the notion is meaningless. Similarity requires a frame of reference (Goodman, 1972).

One might counter that previous studies of recombination contain an implicit argument that the categories chosen to represent similarity and difference come from institutional knowledge that comprises the frame of reference for the study, echoing psychologists' response to Goodman (Medin, Goldstone and Gentner, 1993). But the problem remains if these assumptions are not made explicit, especially if there is more than one possible system of categorization. Returning to the example of clothing in figure 1: without additional information, we have no basis to assume that either material or design should take precedence in evaluating new combinations.

Claiming "new combinations" underlie novelty is vacuous without specifying the framework for comparison (Medin et al, 1993). It can result in misleading implications. For example, scholars who study technological recombination as the basis for high-impact inventions imply that these findings speak to breakthroughs that transform markets (Fleming, 2001; Schoenmakers and Duysters, 2010). The link presumes that novelty based on technology combinations will translate to market novelty, and overlooks how products based on existing technologies might be groundbreaking on the market.[1] This is not to say that recombination is meaningless. Rather, combination needs to be specified in conjunction with a frame of reference: the category system that defines the ways in which objects are similar or different.

In a hypothetical example, an inventor at a software organization might use an insight from an obscure algorithm to reduce processing times by orders of magnitude, so that images can be transmitted

[1] This is problematic if market categories are not entirely based on technology categories. In contexts where market categories are defined by technology categories, technology combination and market combination are one in the same, and the problem posed is trivial.

faster in photo sharing software. This technology combination would result in a foundational technology that would be highly cited by other inventors. It would allow the firm to release a much *better* product and perhaps increase prices. But it would not create an especially *novel* product with respect to how customers use it. On the other hand, if a software firm applied photo sharing technology to digital audio, and discovered that an artist choosing a sequence of photos would output pleasing music, this product would be groundbreaking on the market, combining elements of two different market categories. But the firm would not have created especially novel technologies that were highly cited.

It is useful to articulate two scope conditions for empirically testing these ideas. First, the theory relates to evaluations of *novelty*. Novel items are not necessarily better or more profitable on average. Rather, they are in the tails of a performance distribution, either big wins or failures (March 1991; Fleming 2001). In an empirical test, it is important that the outcome captures novelty: for example asking people's assessments of novelty (e.g. in a laboratory study), studying choices of a novelty-seeking evaluator (e.g. VC investment), or investigating high variance outcomes (e.g. a dispersion parameter). Second, we must identify the categories that comprise the frame of reference a person uses when evaluating whether something is novel. In a historical study this can be discovered by understanding the goals of the evaluator. In a laboratory study, it can be directly manipulated.

Below, the theory is studied in two empirical contexts. Study 1 provides a causal test of the general cognitive mechanism using a laboratory experiment with random assignment to two category systems. Study 2 investigates real-world consequences in the context of VC investment in software organizations. VCs are novelty-seeking evaluators, and there are two important category systems in software: market categories and technology categories. Together, results suggest that categorical systems are foundational to what is a combination – and therefore what is novel – with real-world implications.

Study 1: Causal Evidence of the Cognitive Mechanism – A Laboratory Experiment

Study 1 investigates these ideas with a laboratory experiment designed to provide causal evidence for the claim that categories held by an evaluator affect assessments of novelty. In the laboratory, participants can be randomly assigned to different category systems. The experiment draws on the example of athletic clothing presented in figure 1. Participants are put in the role of a VC who wants to invest in a company with new and different products. They are tasked with assessing the *novelty* of an article of clothing described in a business plan, which fulfils the first scope condition. The category system that segments the industry is manipulated across conditions, addressing the second scope condition.

The experimental design is a between-subjects 2 (category system: materials vs. design) by 2 (clothing combines unrelated materials vs. unrelated designs). It measures novelty assessments of the article of clothing. Hypothesis 1 predicts that participants will rate as more novel combinations that blend the category system presented, as compared to combinations that blend the alternate category system. Figure 2 shows the 2 x 2 schematic of the design and expected outcomes.

--- Insert figure 2 about here ---

Method. The sample comprised 163 adults (39% female, M_{age}=35) recruited through Mechanical Turk, who passed the attention checks (out of 172 total).[2] Participants were paid $0.75 for their study involvement. Participants are told that an investor hosted a business plan competition to find new and different women's clothing ideas. The winner will receive a $50,000 investment. They are told the industry is categorized based on the randomly assigned system. They are then given a one-sentence description of the clothing idea taken from the business plan, and asked to rate its novelty.

Dependent variable. The dependent variable is *novelty*, which is comprised of 4 items rated on Likert scales that range from 1 to 7: "How much do you think the proposal describes a new type of clothing" (1 - "Not at all new", 7 - "Very new"), "How much do you think the proposal describes a different type of clothing?" (1 - "Not at all different", 7 - "Very different"), "How much will this appeal to women who want to buy clothes that are new and different?" (1 – "Not at all", 7 – "Very much"), "If the investor only wants to

[2] Results are the same if all subjects are used.

invest in clothes that are new and different, do you think he should invest $50,000 to produce this article of clothing?" (1 – "Definitely no", 7 – "Definitely yes").

Procedure. Participants were told an investor is considering investing in a company that makes women's clothing apparel. They were randomly assigned to a category condition. In the materials category condition, participants read:

Clothes in this market are categorized based on materials:

For leisure clothes:

- Leisure clothes are made of casual materials, polyester or cotton.
- They are made in a variety of designs, including dresses, skirts, shorts, and pants.

For professional clothes:

- Professional clothes are made of formal materials, silk or cashmere.
- They are made in a variety of designs, including dresses, skirts, shorts, and pants.

In the design category condition, participants read:

Clothes in this market are categorized based on design:

For leisure clothes:

- Leisure clothes are shorts or cropped pants.
- They are made in a variety of materials, including polyester, cotton, silk, or cashmere.

For professional clothes:

- Professional clothes are skirts or dresses.
- They are made in a variety of materials, including polyester, cotton, silk, or cashmere.

Participants then read that the investor wants to invest in a company that makes clothes that are new and different types. The investor hosted a competition for clothing ideas, where the winner would receive a $50,000 initial investment. Students from the local business school submitted their ideas. After several multiple-choice questions confirming that participants knew the category system and that the investor was looking for a novel business, they read an excerpt from the submission. They were randomly assigned to evaluate a materials combination (“`This article of clothing is a combination of silk and polyester materials`”) or a design combination (“`This article of clothing is a combination of shorts and skirt designs`”). Participants then rated the clothing idea on the dependent variable for *novelty*.

Results

The scale reliability was sufficiently high (Cronbach’s $\alpha = .92$), so responses for the 4 items were averaged to create the dependent variable of *novelty*. Figure 3 presents results, which support hypothesis 1. Participants rate combinations of the categories they are presented as more novel, $\beta = 4.03$ (.157), as compared to combinations of the alternate categories $\beta = 3.27$ (.174). An ANOVA test shows these effects are different with high statistical confidence, $F(1, 162) = 10.56$, $p < .005$. The pattern is the same for either combination (materials or design). The effect is consistent in a regression that controls for combination type, categories presented, and the participant’s demographic characteristics ($\beta = .96$ (.25) , $p < .0001$).

--- Insert figure 3 about here ---

Discussion

Study 1 provides causal evidence of the general cognitive mechanism. A combination is rated as more or less novel depending on the category system randomly assigned to the participant. The category system is manipulated by simply telling subjects the relevant categories that segment the domain. Findings indicate that it does not necessarily take extensive education or indoctrination for a categorical frame of reference to have an effect. This suggests the influence of categories on novelty assessments may be widespread.

Study 2: Venture Capital Investment Predicted by Market Category Combination

Study 2 investigates real-world implications of the theory in an analysis of VC investment in software organizations. VCs are novelty-seeking evaluators who invest in organizations with the potential to transform how people interact in the marketplace (MacMillan, Zemann and Subbanarasimha, 1987; Hisrich and Jankowicz, 1990; Lewis, 1999). VCs are "market-makers" who aim to radically change the market in terms of how people live and work (Pontikes, 2012). Because large successes cannot be precisely predicted, VCs maintain a portfolio of firms, expecting that most will fail but a few will generate outsized returns so that the overall portfolio is profitable (Sahlman, 1990).[3] This is a high-variance strategy, where novel organizations are in the "tails" of the performance distribution (March, 1991). VCs look for non-consensus approaches: firms doing something that runs counter to common wisdom (Pontikes and Barnett, 2017). For instance, a VC interviewed for this paper describes Airbnb as having changed how people vacation:

> you want to invest in those things, where you look at it and are like – that'll never work. But if it *does* – you know, dot, dot, dot. Like Airbnb. That'll never work. People will never go sleep in a kid's bedroom in someone else's house and pay a nightly rate. But if it did …

VCs that are early investors in an iconic success build reputations they leverage for access to future promising investment opportunities (Podolny 2001; Hochberg, Ljungqvist, and Lu, 2007). Studying VC investment satisfies the first scope condition, that VCs invest in organizations that are novel.

For the second scope condition, we must define the frame of reference for VCs. In the software industry, two important frames are *market space*, defined by market categories that reflect how products are used ("enterprise resource planning," "digital audio"), or *technology space*, defined by technology categories based on how products are created, for example the coding language or algorithms for data processing ("java," "page-rank") (Pollock and Williams, 2009; Wang, 2010). Much previous research on

[3] VCs do not *want* any investments to be failures. Their investment strategy reflects the recognition that successful novelty cannot be predicted ahead of time.

recombination and novelty is situated in technology space. But a consideration of VC goals suggests that technology space is not the frame for a typical VC.

Academic and popular literature shows VC strategy in software is to invest in firms with a unique *market* insight (MacMillan, Zemann and Subbanarasimha, 1987; Lewis, 1999). VCs invest in companies with ambiguous market categorization because these firms are best positioned to change how people use products (Pontikes, 2012). As a prominent Silicon Valley consulting firm describes, the ideal VC investment is one that "define[s] a new space [market category] by conditioning the market to buy and consume new technology in a new way" (Ramadan, Lochhead, Peterson and Maney, 2013: 19).

The research streams that propose combinations underlie novelty suggest that VCs should find organizations that create "new combinations" as having the most potential. From the discussion above, the relevant frame of reference for VCs is market categorization, reflecting the use-based perspective of a customer. It is not that technologies are irrelevant to VCs – in software, all companies are technological. But a market breakthrough is not necessarily a groundbreaking technological or scientific advance, as the examples of Facebook, Airbnb, or Uber illustrate. In some cases, entrepreneurs develop a technology combination which is also a market category combination and therefore transformative on the market. But we should expect the *systematic* relationship between combination and VC investment to be with respect to combinations of market categories.

Hypothesis 2a: Venture capitalists are more likely to invest in organizations with <u>high market category combination</u>, as compared to those with <u>low market category combination</u>.

Hypothesis 2b: <u>Market category combination</u> is more predictive of VC investment as compared to <u>technology category combination</u>.

Data and Methods

Testing the hypotheses requires data on market categories and technology categories, organizations' positions within each category system, and VC investment. This study uses unique longitudinal data on software organizations that contain this information.

The risk set for the empirical analysis is from a dataset that provides structured data on software organizations and market categories, compiled using computational text analytics applied to press releases. These data were merged with data on VC investment and patents (see Pontikes (2012) for a detailed description of the data collection).[4] This analysis uses a subset of these data: (1) organizations that have patented, as patents locate organizations and categories in market space and technology space (described below), (2) only young and private firms that are at risk for receiving VC financing (private and less than 15 years old),[5] and (3) the years between 1995, after legal barriers to software patenting in the United States were removed, and 2002, the last year of the data.[6] After filtering on these parameters, the final data include 368 organizations over 1,057 organization-years.

Dependent variable. The primary dependent variable is whether the organization receives venture capital financing in the current year. There are 174 VC funding events (103 organizations funded).

Independent variables. The two independent variables are market category combination and technology category combination. To compute these variables, two spaces are constructed, market space and technology space, using data on market categories and technology categories in this domain.

Much previous research on combination and novelty focuses on technology space, primarily analyzing patents. Following these studies, this analysis uses patent classes for technology categories (Lerner, 1994; Fleming 2001; Schoenmakers, Wand G. Duysters, 2010; Gruber et al., 2013; Kaplan and Vakili 2015). Patent classification is a good source for technology categorization for a number of reasons:

[4] Most software companies issue press releases and so these data include small and young organizations that never received funding and are missing from many data sets. This mitigates issues of selection bias that are present in many studies of VC funding where only firms that have already received a round of funding are in the risk set (Guler, 2007; Mann and Sager, 2007; Hsu and Ziedonis, 2013).
[5] Founding dates could not be located for some organizations, likely to be small, young firms that were not successful, and so excluding these might bias the results. Therefore, I also include private organizations where the founding date is unknown. Firms that IPO or turn 15 during the study period are removed from the risk set as a censored observation.
[6] The USPTO did not officially recognize software patents until the courts removed legal barriers in 1994 – 1995 Hall and MacGarvie, 2010; Cockburn and MacGarvie, 2011).

it categorizes important developments created by many firms, it is carefully maintained by governmental patent offices, and it allows these analyses to be compared to previous research. There is not another dataset of which the author is aware that comprehensively documents technology categories in this context.

Market space is defined by market categories drawn from organizations' self-classifications in press releases. Press releases provide a historical record of market categories used by the thousands of software firms active during the time period. Categories in press releases describe the organizations' offerings to customers and investors. They can be detected from natural language due to advances in computational analytics. Previous research shows that these self-classifications are meaningful measures of market categorization (Pontikes, 2012; Pontikes and Hannan, 2014), and they provide a better map of the competitive space as compared to institutional classification like NAICS or SIC that researchers have previously relied upon (Hoberg and Phillips, 2015).

The next question is how to locate organizations and categories within each space. For technology space, following previous research, patents are used (Podolny, Stuart and Hannan, 1996; Fleming 2001). For market space, one approach might be to define market combination using only market category claims, and compare its effect on VC financing to that of technology combination. But this comparison leaves open a plausible alternative hypothesis that effects arise because positions in market space and technology space are based on different types of firm attributes, not because of the different category systems.

A better approach also uses patents as the basis for market space. Patents track detailed information about important components underlying firms' offerings, and can be linked to market categories based on the category membership of the assignee. This approach can directly capture whether firms' offerings are built on elements that bring together market categories (described below).[7] Importantly, a patent-based measure is comparable to technology category combination. Differences between the measures arise from differences in category systems, not because evaluators attend to different firm attributes, providing a conservative test of the hypotheses. The trade-off is that rich data on firms' underlying components are only

[7] Market category claims can capture combination, as an ambiguous position on the market reflects whether an organization combines. But using patent-based locations for firms and categories provides a direct measure.

available for organizations that patent. For the present analysis, this is not a problematic limitation, since the objective is to compare effects of market category combination to technology category combination, and technology categories are only available for patenting firms.[8]

A question may arise regarding the validity of using software patents because intellectual property (IP) protection is not especially important in this industry. In this study, patents are used as a historical record of the knowledge development of organizations, so whether patents provide adequate IP protection is not relevant. Software companies patent for reasons beyond IP protection, including to have an institutional record of their inventions (Lemley, 2000). Previous research shows that software patents influence venture capital financing, IPOs, acquisition, and market entry (Mann, 2005; Mann and Sager, 2007; Cockburn and MacGarvie, 2011). There is a meaningful connection between the patenting behavior and market category affiliations in software (Pontikes and Hannan, 2014). Patents provide rich information on firm attributes that is not otherwise available, which can be used to measure market and technology combination. Patent portfolios were gathered from the NBER U.S. Patent Citations data file (Hall, Jaffe and Trajtenberg, 2001), which has been extended through 2006, and the Patent Network Dataverse (Li et al., 2014).[9]

Market category combination is computed based on how proximate an organization is to different market categories, using positions in a patent network. First, a citation network of software patents is built using a five-year window for every year in the data. Organizations are located in the network by their patents. Market space is created by projecting market categories onto this network based on the areas of the network where organizations in the respective categories patent. Market categories for software

[8] Studying patenting organizations makes this analysis comparable to previous research on innovation, which uses patents as the risk set. Effects of combination should be interpreted with respect to other patenting firms. The analysis does not shed light on how combination based on organizational features not captured by patents affects VC funding. Results from previous research speak to this question, and show that firms that take ambiguous market space positions have a higher likelihood of receiving funding (Pontikes, 2012). This provides further evidence for the general idea presented: combination, ambiguity, and non-conformity underlie novelty (when defined with respect to the relevant category system).

[9] NBER data contain original and current class and subclass assignment for the patent's primary class. The Patent Network Dataverse contains multiple class and subclass assignments for the patent's current class only. All measures are timed by the application year of patents (only patents that were later granted are used).

organizations come from press releases. Proximities between patents are calculated using citation overlap (Podolny, Stuart and Hannan, 1996; Pontikes and Hannan, 2014). Market category combination is computed as the proximity in market space of an organization's patents to market categories it is not in. Computational details are included in the study 2 methodological appendix (Appendix A).

Technology category combination is based on how proximate an organization is to different technology categories. Previous research uses different measures of combination based on patent classes (see Gruber *et al* (2013) for a review). This study uses the following (details in Appendix A):

1. Originality is derived from the Herfinindahl index and captures a patent's similarity to different technology categories based on its citations' patent classes (Henderson, Jaffe, and Trajtenberg, 1998). It is perhaps the most widely used measure of recombination (Gruber et al, 2013), and so is the metric used in the main analysis. Technology categories used in previous research are either patent classes maintained by the US Patent Office (USPTO) or International Patent Classification (IPC) maintained by the World International Patent Organization (WIPO). USPTO patent classes have been updated, so original and current classes can be different. To provide a conservative test, originality is computed using all three category systems: USPTO original class, IPC, and USPTO current class.

Variants of two additional measures are used in supplementary analyses:

2. Breadth counts the number of technology categories to which a patent is assigned. Such measures have previously been used in research on combination and novelty (Lerner, 1994; Leahey, 2006).
3. Network Proximity is a measure computed identically to market category combination (Appendix A, equations A1 and A2), but measures proximity to *technology categories* the organization is not in. This measure is based on the same patent citation network, to test against a concern that effects might result from the different way market category combination is computed.

Controls

Within-category development tests whether effects of market category combination arise from blending components *across market categories*, or if it captures something else about positions in the patent network. It is computed identically to market category combination, but measures an organization's knowledge space proximity to patents in its own categories.

Additional controls that influence VC financing are included.[10] Because some firms do not patent every year, an important control is whether the firm has patented last year (all IVs are lagged by one year).[11] The number of other organizations in the same market categories and the number of firms in the category that received VC funding (and its square) are included to control for the popularity and competitiveness of the organization's market category (Pontikes and Barnett, 2017).[12] The number of acquisitions made by the organization, its tenure in the press release data, and a 0/1 indicator variable for whether it was ranked in *Software Magazine*'s Software 500 are included to control for firm size and quality. The number of previous rounds of financing is included to capture quality or a tendency for VCs to try to salvage an investment (Guler, 2007). All independent and control variables are lagged by one year.

--- Insert tables 1-2 about here ---

Table 1 contains descriptive statistics and correlations for all organization-years in the data, and table 2 only years where at least one patent was issued. Comparing the two tables reveals that the apparently high correlations between market category and technology category combination is an artifact of organizations not patenting every year. Table 2 shows that for patenting years, the correlation between market category and technology category combination is low. Correlations among technology combination

[10] The base model includes controls that influence (or are expected to influence) funding for patenting firms. Some important factors influencing VC funding for all firms in the full data, like category fuzziness (Pontikes, 2012), are not significant for a risk set of patenting firms only. This is likely because investors use patent-based combination as a more relevant measure of potential market novelty and do not rely on categorical framing. All effects reported are robust to the inclusion of these variables, which are reported in supplementary analyses in Appendix B.
[11] Results are robust to including: number of patents, number of citations, cumulative patents, cumulative citations, (and the natural log of these variables), average citations per patent, and number of claims.
[12] Organizations in multiple categories are weighted by their grade of membership in the category, as defined in Appendix A.

measures remain high. This suggests that market categories and technology categories represent different categorization systems.

This also points to the importance of testing whether effects are sensitive to how combination is treated in years where organizations do not patent. The main analyses control for whether an organization has patented that year, as a 0/1 indicator variable, and assigns market and technology combination a value of 0 in years where no patent was issued. The control captures the effect of patenting, and the combination variables the effect of combining across the respective category systems. Additional analyses use a number of different specifications, including assigning market combination to the last observed value with a time discount, and only including years when firms patent in the risk set. Effects are not sensitive to these alternate specifications, which are described in detail in the appendices.

Model and Estimation

The hazard rate of receiving VC funding in a given year is modeled a function of the independent and control variables:

$$r(t-t_n) = r_o(t-t_n) \cdot \exp(\beta_{ind} \cdot \mathbf{x}_{\boldsymbol{ind}} + \alpha_{control} \cdot \mathbf{x}_{\boldsymbol{control}} + \varepsilon) \qquad (7)$$

The rate is estimated using a piecewise exponential model with the stpiece routine in STATA. Receiving VC financing is a repeated event, and organizations that receive funding exit and enter with a new ID. Therefore, standard errors are clustered by firm. Duration is the time since the organization was last funded, or when it entered the data.

Results

Evidence for hypotheses 2a and 2b is can be seen figures 4 and 5, which plot the mean number of venture capital funding events by market category combination and technology category combination. These figures present trends without any controls. They show a positive relationship with market category combination, but a fairly flat relationship for technology category combination, in line with the hypotheses.

--- Insert figures 4 - 5 about here ---

Table 3 reports statistical tests of hypothesis 2a, piecewise continuous hazard rate estimations on the likelihood that an organization receives VC funding. Column (1) contains controls only. Column (2) tests hypothesis 2a by including market category combination. The effect is positive and significant ($p < 0.05$), providing support for the hypothesis. Column (3) includes the control for within-category development, to test if the effect is capturing combination or something about position in the patent network. This control does not affect funding, and the positive effect of market category combination remains ($p < 0.01$). A one standard deviation increase in market category combination is associated with a 50% increased likelihood of receiving VC funding.[13] This is substantial: twice as large as the increased likelihood of being funded after having received a previous round (25%),[14] an effect that previous literature has shown to result in a strong bias in favor of investment (Guler, 2007).

Market category combination is only defined for years when the organization patents, and so should also be considered in conjunction with the effect of having patented. This effect is null in the base model (column 1). When market category combination is included, the effect of having patented is *negative* and significant ($p < 0.01$), indicating that organizations that develop non-combinatory knowledge are less likely to receive VC investment. Considering the two variables together, being in the top 45% of market category combination results in a net positive effect on VC investment. Appendix B reports results alternate specifications to account for years when organizations do not patent. Results are not sensitive to these tests.

--- Insert table 3 about here ---

Table 4 contains tests of hypothesis 2b. Technology category combination does not have a positive effect on VC investment, either when included in models alone (top rows) or with market category combination (bottom rows).[15] The effect of market category combination is robust. This provides support for hypothesis 1b, that market category combination is more predictive of VC investment than technology

---

[13] This is computed using the estimate from table 3 column 3, and a standard deviation for market category combination from table 1: exp(0.374 * 1.09) = 1.50.
[14] Computed from estimates reported in table 3 column 3: exp(0.220*1) = 1.25.
[15] Estimations show when controlling for other factors the discrepancy between effects on funding for market and technology combination are even more pronounced than figures 2 and 3 show. Net of controls, there is a slight (albeit non-significant) negative relationship between technology combination and VC funding.

category combination. Additional analyses testing alternate measures of technology combination and the robustness of effects are detailed in Appendix B.

--- Insert table 4 about here ---

One question might be whether the results above are due to VCs using market categories as an alternate category system, or whether market categories are a simply a better way of segmenting the industry. To address this, I investigate whether technology category combination is associated with technology novelty, as measured using future patent citations. Previous studies that link technology or scientific combination with novelty use this outcome (Fleming, 2001; Nerkar, 2003; Schoenmakers and Duysters, 2010; Uzzi, Mukherjee, Stringer and Jones, 2013; Foster, Rzhetsky and Evans, 2015).

--- Insert table 5 about here ---

Negative binomial estimations are run on patent citations eight years after the grant date. Returning to the first scope condition, the outcome of interest is how *novel* a patent is, which is predicted to result in high-variance outcomes (Fleming, 2001). So this analysis focuses on effects of the dispersion parameter (alpha). Results are reported in table 5. They show the expected positive effect when computed using IPC or current USPTO classes ($p < 0.05$).[16] The patent-level measure of market-category combination (the natural log of equation (2)) is also included, and it has a negative effect on the dispersion parameter ($p < 0.001$). Combination across *technology categories* (but not *market categories*) has a positive effect on *technology novelty*.[17]

Discussion

Results show real-world implications of the theory. Consistent with findings from study 1, the relationship between combination and novelty depends on the category system relevant to the evaluator. Here, combinations of *market categories* (but not *technology categories*) predict VC investment. This is not just

[16] That the effect is not apparent for USPTO original class gives credence to the idea that the USPTO classification was not tailored to software in the pre-1995 era, before software patenting was officially recognized.
[17] Technology combination also has a positive effect in models that exclude market combination.

because market categories are more relevant in this setting. In line with previous literature, technology novelty is predicted by combinations of *technology categories*. Results should not be taken to imply that inventive developments are unimportant. Novel inventions may underlie product advances. But technology combination is not the primary antecedent of the types of groundbreaking market developments that VCs seek.

General Discussion

The idea that new combinations are the ultimate source of novelty underlies research on sociological, economic, and technological change. This paper suggests that combination—and therefore novelty—depends not only on characteristics of the object, but also on sociological context. The category system that characterizes the evaluator's frame of reference partly determines what is a combination. Findings from two empirical studies support this view.

Study 1 presents causal evidence of the general cognitive mechanism. In the laboratory experiment, categories used by evaluators are manipulated through random assignment. The same combination is assessed as more or less novel depending on the categories assigned. Study 2 shows real-world consequences. Venture capitalists are more likely to invest in organizations that combine *market* categories. Combination across technology categories does not significantly correlate with investment. Results show combination is a basis for novelty, but there is an important caveat: it depends on the categories used by the evaluator.

Results have implications for long-standing research on market evolution that proposes new combinations underlie novel developments. The large body of evidence supporting this claim is incomplete, and possibly misleading. Previous research on technological combination suggests the link between recombination and high-impact technologies explains the emergence of radical market changes. This study calls into question that claim, showing that sociological context is foundational to this relationship. Combination must be defined in the same frame of reference used to evaluate novelty. Researchers need to account for multiple category systems in any domain, and identify the relevant one for the evaluator(s) studied. If one did not consider the frame of reference for VCs, he might come to the incorrect conclusion that recombination is not associated with VC investment. But VCs do value firms that combine in their frame of reference – across market categories.

These findings are relevant to other contexts. Every domain has different systems of categorization that capture multiple ways of assessing similarity and difference. In economic contexts, these often include *market categorization* and *technology categorization*, two systems that are not unique to software. For example, automobiles can be categorized by their target market ("economy", "luxury"), or engine type ("four cylinder," "V6"). Pharmaceutical drugs can be classified by the conditions they treat ("high blood pressure," "hair loss"), or by their molecular make-up ("minoxidil," "sildenafil citrate"). Multiple systems also apply to cultural settings. Fine art categories might be based on technique ("broken brush strokes," "intense shadows") or the subject depicted ("pastoral," "urban scene").

It is possible that in a particular context one category system is a direct reflection of another. For example, in some industries market categories might be defined by technologies, in which case technology combination and market combination are one in the same. But scholars of market evolution should be cautious assuming technology categories are defaults. The software industry is a high-technology setting, and findings shows no evidence that technology category combination underlies VC investment. It is also possible that in some industries VC strategies would directly focus on technology novelty. For example, if IP protection is important, VCs may invest in firms with high-impact technologies, in which case we might expect both market category and technology category combination to affect investment. It will be important in future research to explicitly specify the frame of reference of the evaluator, and to study outcomes that measure the type of novelty theorized.

The theory proposes that assessments of novelty depend on the category systems used by an evaluator. Study 1 provides causal evidence that supports this mechanism, by manipulating the evaluator's category system through random assignment. Results show that people can easily be primed to use a particular category system: simply using different categorizations to describe the setting results in measurable effects. This suggests in a practical application, VCs (or other evaluators) may quickly adopt the category system used by their colleagues and other entrepreneurs. This means the frame of reference for an evaluator's current role is likely influential. It will be interesting in future research to study whether an evaluator's history (for example education or previous career) also influences novelty assessments.

Study 2 complements study 1 as an example of how advances in computational text analytics allow scholars to study measurable, economic effects of different category systems. Researchers can mine descriptions from historical documents for structured data on category systems in a domain. In this case, market categories are documented in press releases, and texts from hundreds of thousands of press releases were processed with computational analytics to extract market categories. These complement technology categories already documented in patent classification.

In study 2, patents are used to measure the types of developments an organization has created, an example of how patents can be used in research, not only as measures of IP protection, but also to track firm attributes over time. Patents do matter, even in contexts like software where IP protection is not strong, because they provide data to measure the knowledge a firm develops. Previous research shows that in industries where IP protection is strong, like biotechnology and semiconductors, simply having a patent is a predictor of VC investment (Baum and Silverman, 2004; Hsu and Ziedonis, 2013). To the extent that VCs also seek novel organizations in these contexts, results suggest there may be positive effects of combination in addition to effects from simply having a patent.

Findings speak to research in organizational learning that conceptualize novelty as a multidimensional construct (Rosenkopf and McGrath, 2011). For example, an organization can engage in exploration by building alliances along a "functional" dimension, but be more exploitive in alliances along a "structural" dimension (Lavie and Rosenkopf, 2006). Researchers find optimal outcomes may balance exploration along one dimension and exploitation along another (Rosenkopf and Nerkar, 2001). This study expands these ideas by considering the perspective of the audience, showing that different types of novelty are valued depending on the frame of reference of the evaluator.

This study also may help reconcile recent research that questions the link between recombination and novelty. Kaplan and Vakili (2015) analyze patent texts to identify when a patent originates a new topic.

They show that both recombination across patent classes (using citations) and new topic formation (using text) result in high patent citations. But these effects are independent: recombination does not lead to new topic formation. They interpret these results as calling into question a whether recombination is the primary antecedent of novelty. Findings here suggest an alternative: perhaps text-based novelty and citation-based recombination rely on different category systems.

Findings have implications for sociological research on categorization (DiMaggio, 1987; DiMaggio, 1997, Lamont and Molnar, 2002). They add to the growing body of work that documents ways certain evaluators value categorical *non*-conformity in markets (Pontikes, 2012; Goldberg et al., 2016; Paolella and Durand, 2016). Consistent with previous research, a novelty-seeking evaluator prizes combinations across categories. This helps reconcile findings in the category literature that documents the appeal of category conformity (Hsu 2006; Kovács and Hannan, 2010) with the conclusions of the innovation literature on the value of combining categories (Fleming, 2001; Uzzi, Mukherjee, Stringer and Jones, 2013; Foster, Rzhetsky and Evans, 2015; Leahey, Beckman and Stanko, 2016). Results also support the idea that categories should be defined and studied with respect to an audience (Hannan, Pólos and Carroll, 2007; Hannan, 2010). Different actors (or the same actor in different roles) use different categorical systems. Most contexts yield multiple classifications: for example how academics and practitioners classify research or the critic versus layman's categorization of music. This study shows that combination, a concept important to many literatures, is fundamentally dependent on these category systems.

References

Baum, J., and B. Silverman. 2004. "Picking Winners or Building Them? Alliance, Intellectual, and Human Capital As Selection Criteria in Venture Financing and Performance of Biotechnology Startups." *Journal of Business Venturing* 19:411-436.

Carroll, G., 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.

Cockburn, I M, and M. J. MacGarvie. 2011. "Entry and Patenting in the Software Industry." *Management Science* 57:915-933.

DiMaggio, P. J. (1987). Classification in Art. *American Sociological Review*, *52*, 440-455.

DiMaggio, P. J. 1991. "Constructing An Organizational Field As a Professional Project: U.S. Art Museums, 1920-1940." Pp. 267-292 in *The New Institutionalism in Organizational Analysis*, edited by Walter J Powell and Paul J DiMaggio. Chicago: The University of Chicago Press.

Dimaggio, P. J. (1997). Culture and cognition. *Annual Review of Sociology*, *23*, 263-287.

Durand, R., and L. Paolella. 2013. "Category Stretching: Reorienting Research on Categories in Strategy, Entrepreneurship, and Organization Theory." *Journal of Management Studies* 50:1100-1123.

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

Foster, J. G., A. Rzhetsky, and J. A. Evans. 2015. "Tradition and Innovation in Scientists' Research Strategies." *American Sociological Review* 80:875-908.

Goldberg, A., M. Hannan, and B. Kovács. 2016. "What Does It Mean to Span Cultural Boundaries? Variety and Atypicality in Cultural Consumption." *American Sociological Review.*

Goodman, N. 1972. "Seven Strictures on Similarity." *Problems and Projects*. Indianapolis: Bobbs-Merril, 437-447.

Gruber, M, D. Harhoff, and K. Hoisl. 2013. "Knowledge Recombination Across Technological Boundaries: Scientists Vs. Engineers." *Management Science* 59:837-851.

Guler, I. 2007. "Throwing Good Money After Bad? Political and Institutional Influences on Sequential Decision Making in the Venture Capital Industry." *Administrative Science Quarterly* 52:248-285.

Hall, B., A. Jaffe, and M. Trajtenberg. 2001. "The NBER Patent Citations Data File: Lessons, Insights, and Methodological Tools." in *NBER Working Paper Series.* Cambridge, MA: National Bureau of Economic Research.

Hall, B. H, and M. MacGarvie. 2010. "The Private Value of Software Patents." *Research Policy* 39:994-1009.

Henderson, R., A. Jaffe, and M. Trajtenberg. 1998. 'Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting 1965-1988," *Review of Economics and Statistics* 80: 119-127.

Hannan, M. T. 2010. "Partiality of Membership in Categories and Audiences." *Annual Review of Sociology* 36:159-181.

Hannan, M. T, L. Pólos, and G. Carroll. 2007. *Logics of Organization Theory: Audiences, Codes and Ecologies.* Princeton, NJ: Princeton University Press.

Hisrich, R D, and A. D. Jankowicz. 1990. "Intuition in Venture Capital Decisions: An Exploratory Study Using a New Technique." *Journal of Business Venturing* 5:49-62.

Hoberg, G., & Phillips, G. (2015). Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*.

Hochberg, Y. V., Ljungqvist, A., & Lu, Y. (2007). Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance*, *62*(1), 251-301.

Hsu, D. H., and R. H. Ziedonis. 2013. "Resources As Dual Sources of Advantage: Implications for Valuing Entrepreneurial-firm Patents." *Strategic Management Journal* 34:761-781.

Hsu, G. 2006. "Evaluative Schemas and the Attention of Critics in the US Film Industry." *Industrial and Corporate Change* 15:467-496.

Kaplan, S., and K. Vakili. 2015. "The Double-edged Sword of Recombination in Breakthrough Innovation." *Strategic Management Journal* 36: 1435 – 1457.

Kovács, B., & Hannan, M. T. (2010). The Consequences of Category Spanning Depend on Contrast. In G. Hsu, G. Negro, & O. Kocak (Eds.), *Categories in Markets: Origins and Evolution* (pp. 175-201). Bingley, UK: Emerald Group Publishing, Ltd.

Lamont, M., & Molnar, V. (2002). The Study of Boundaries in the Social Sciences. *Annual Review of Sociology*, *28*, 167-195.

Lavie, D., and L. Rosenkopf. 2006. Balancing exploration and exploitation in alliance formation. *Academy of management journal, 49*(4), 797-818.

Leahey, E., C. M. Beckman, and T. L. Stanko. 2016. "Prominent but Less Productive: The Impact of Interdisciplinarity on Scientists Research." *Administrative Science Quarterly.*

Leahey, E. 2006. "Gender Differences in Productivity Research Specialization As a Missing Link." *Gender & Society* 20:754-780.

Lemley, M A. 2000. "Reconceiving Patents in the Age of Venture Capital." *J. Small & Emerging Bus. L* 4:137.

Lerner, J. 1994. "The Importance of Patent Scope: An Empirical Analysis." *The RAND Journal of Economics* 25:319.

Lewis, M. 1999. *The New New Thing: A Silicon Valley Story.* WW Norton & Company.

Li, Guan-Cheng et al. 2014. "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975–2010)." *Research Policy* 43:941-955.

MacMillan, I. C,. L. Zemann, and P. N. Subbanarasimha. 1987. "Criteria Distinguishing Successful from Unsuccessful Ventures in the Venture Screening Process." *Journal of Business Venturing* 2:123-137.

Mann, R. J. 2005. "Do Patents Facilitate Financing in the Software Industry?" *Texas Law Review* 83:1-71.

Mann, R. J., and T. W. Sager. 2007. "Patents, Venture Capital, and Software Start-ups." *Research Policy* 36:193-208.

March, J. G. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science* 2:71-87.

Medin, D. L., R. L. Goldstone, and D. Gentner. 1993. Respects for similarity. *Psychological Review*, *100*(2), 254.

Nelson, R., and S. Winter. 1982. *An Evolutionary Theory of Economic Change.* Cambridge: Harvard University Press.

Nerkar, A. 2003. "Old Is Gold? The Value of Temporal Exploration in the Creation of New Knowledge." *Management Science* 49:211-229.

Paolella, L., and R. Durand. 2016. Category Spanning, Evaluation, and Performance: Revised Theory and Test on the Corporate Law Market. *Academy of Management Journal*, *59*, 330-351.

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

Phillips, D. J. (2013). *Shaping jazz: Cities, labels, and the global emergence of an art form*. Princeton University Press.

Podolny, J. 2001. "Networks as the Pipes and Prisms of the Market." *American Journal of Sociology* 107:33-60.

Podolny, J,. T. Stuart, and M. T. Hannan. 1996. "Networks, Knowledge, and Niches: Competition in the Worldwide Semiconductor Industry, 1984-1991." *American Journal of Sociology* 102:659-689.

Pollock, N., and R. Williams. 2009. "The Sociology of a Market Analysis Tool: How Industry Analysts Sort Vendors and Organize Markets." *Information and Organization* 19:129-151.

Pontikes, E. G., and W. P. Barnett. 2017. The Non-consensus Entrepreneur: Organizational Responses to Vital Events. *Administrative Science Quarterly*, *62*(1), 140 - 178.

Pontikes, E.G. 2012. "Two Sides of the Same Coin: How Ambiguous Classification Affects Multiple Audiences' Evaluations." *Administrative Science Quarterly* 57:81-118.

Pontikes, E. G, and M. T. Hannan. 2014. "An Ecology of Social Categories." *Sociological Science* 1:311-343.

Ramadan, A., C. Lochhead, D. Peterson, and K. Maney. 2013. "Time to Market Cap: The New Metric That Matters." *Play Bigger Advisors, LLC.*
Rao, H. 1998. *"Caveat Emptor:* The Construction of Nonprofit Consumer Watchdog Organizations." *American Journal of Sociology* 103:912-961.
Rosenkopf, L., and A Nerkar. 2001. "Beyond Local Search: Boundary-spanning, Exploration, and Impact in the Optical Disk Industry." *Strategic Management Journal* 22:287-306.
Rosenkopf, L., and P. McGrath. 2011. Advancing the Conceptualization and Operationalization of Novelty in Organizational Research. *Organization Science*, *22*(5), 1297-1311.
Rzhetsky, A., J. G. Foster, I. T. Foster, and J. A. Evans. 2015. "Choosing Experiments to Accelerate Collective Discovery." *Proc Natl Acad Sci U S A* 112:14569-14574.
Sahlman, W. 1990. "The Structure and Governance of Venture-Capital Organizations." *Journal of Financial Economics* 27:473-521.
Schoenmakers, W., and G. Duysters. 2010. "The Technological Origins of Radical Inventions." *Research Policy* 39:1051-1059.
Schumpeter, J. 1934. *The Theory of Economic Development.* Cambridge, MA: Harvard University Press.
Stark, D. 1996. Recombinant Property in East European Capitalism. *American Journal of Sociology*, *101*(4), 993-1027.
Stark, D. 2009. *The Sense of Dissonance: Accounts of Worth in Economic Life*. Princeton, N.J.: Princeton University Press.
Trajtenberg, M. 1990. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *The RAND Journal of Economics* 21:172.
Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones. 2013. "Atypical Combinations and Scientific Impact." *Science* 342:468-72.
Vedres, B., and D. Stark. (2010). Structural Folds: Generative Disruption in Overlapping Groups. *American Journal of Sociology*, *114*, 1150-1190.
Wang, P. 2010. "Chasing the Hottest IT: Effects of Information Technology Fashion on Organizations." *MIS Quarterly* 34:63-85.
------. 2009. "Popular Concepts Beyond Organizations: Exploring New Dimensions of Information Technology Innovations." *Journal of the Association for Information Systems* 10:1-30.
Weitzman, M. L. 1998. "Recombinant Growth." *The Quarterly Journal of Economics* 113:331-360.

Figures

Figure 1. Example of two types of combination for clothes.

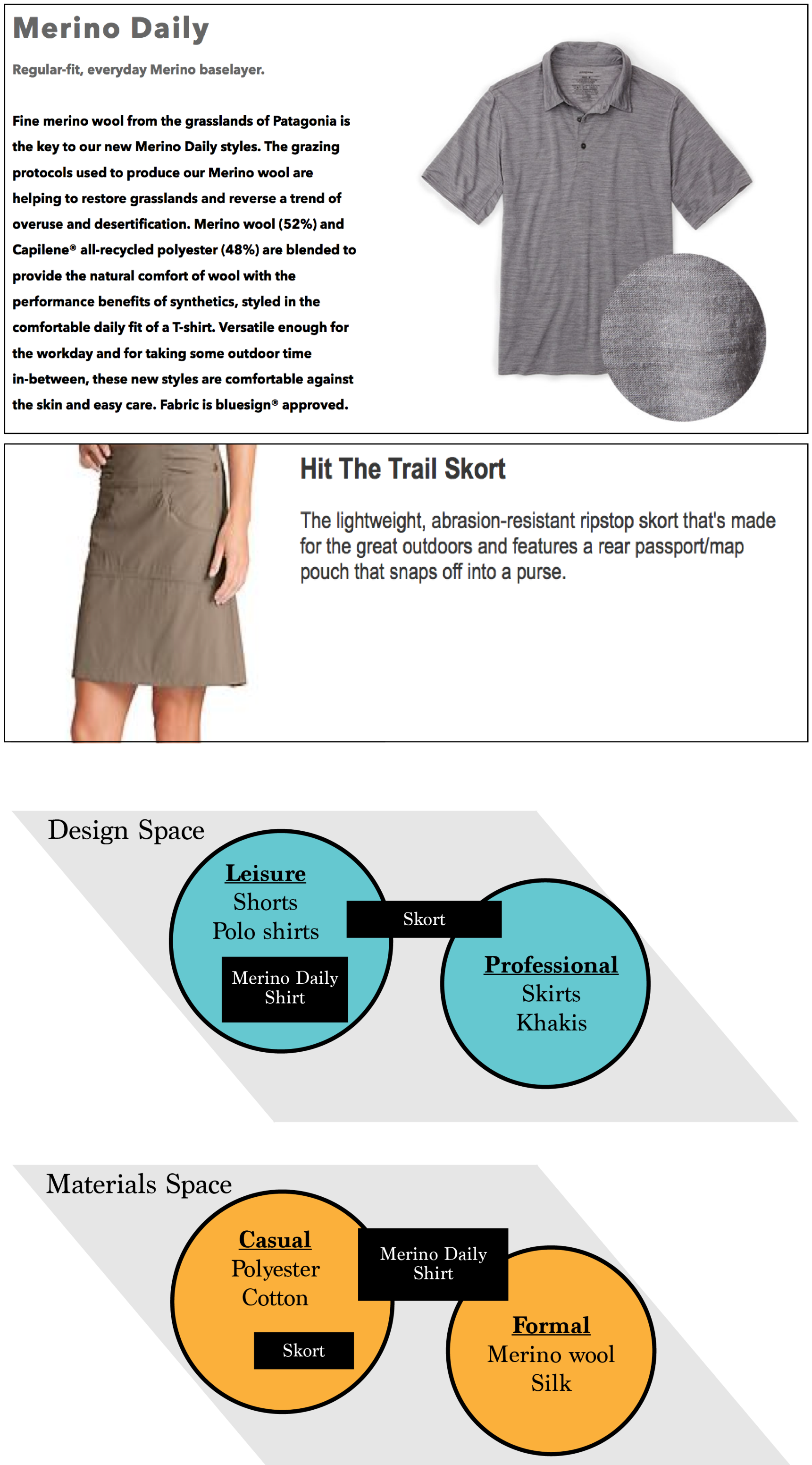


Figure 2. Hypothesis 1: Experiment Design and Expected Effects.

	Materials combination	**Design** combination
Categories based on: **Materials**	+ Novelty	- Novelty
Categories based on: **Design**	- Novelty	+ Novelty

Figure 3. Tests of hypothesis 1. Participants' assessments of novelty for an article of clothing. The same object is appraised as more novel when it combines the category system to which the participant is randomly assigned.

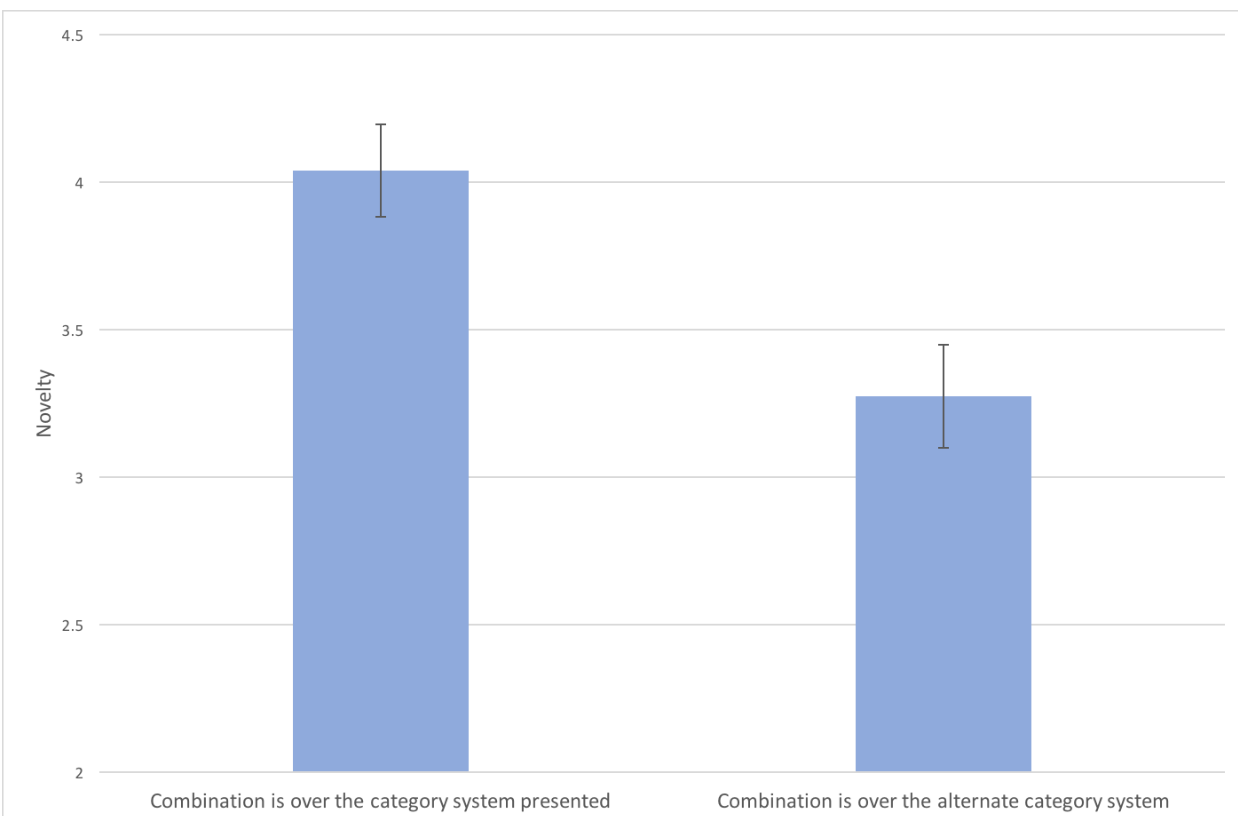


Figure 4. Evidence for hypothesis 2a: Mean VC funding events by market category combination.

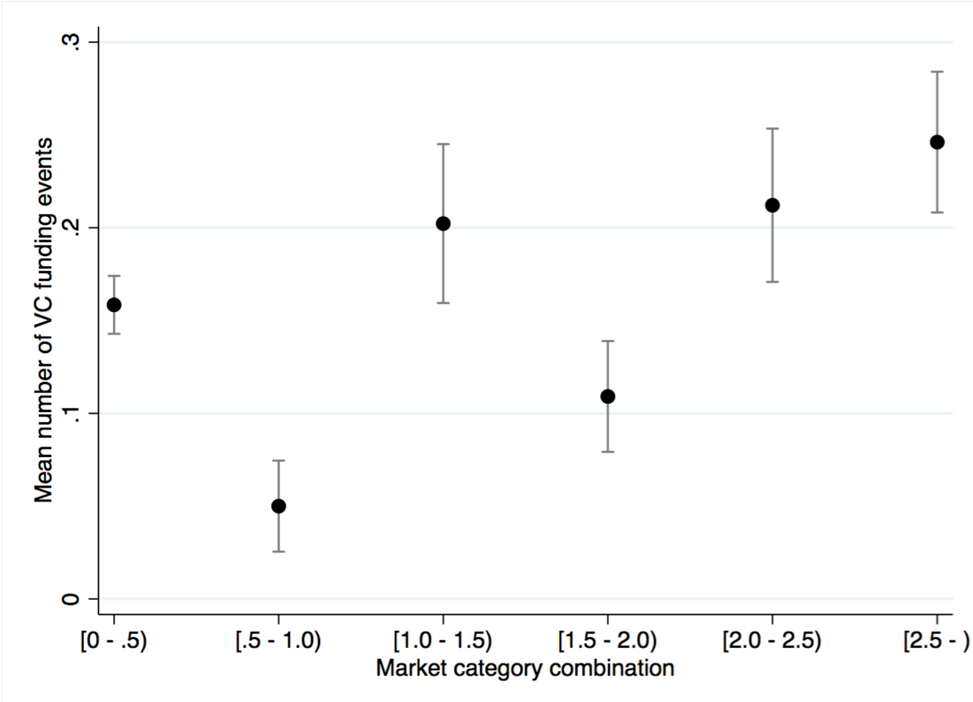
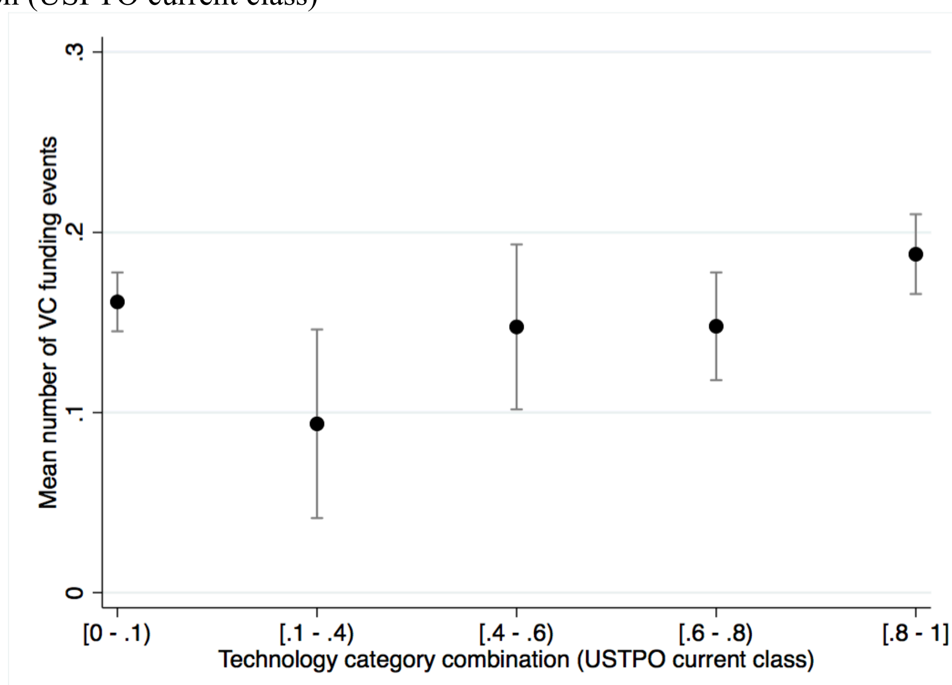


Figure 5. Evidence for hypothesis 2b: Mean VC funding events by technology category combination (USPTO current class)



Tables

Table 1. Study 2 Descriptive Statistics and Correlations

	Mean	Standard Deviation	Min	Max		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Received VC funding	0.1646	0.3710	0	1	(1)														
Market category combination	0.9292	1.091	0	3.962	(2)	0.07													
Within-category development	0.1061	0.2454	0	1.631	(3)	0.04	0.55												
Originality (USPTO original class)	0.3944	0.4024	0	1	(4)	0.01	0.65	0.32											
Originality (IPC class)	0.4610	0.4385	0	1	(5)	0.02	0.74	0.38	0.88										
Originality (USPTO current class)	0.4016	0.4092	0	1	(6)	0.02	0.71	0.35	0.90	0.90									
Patented last year	0.5572	0.4969	0	1	(7)	0.02	0.76	0.39	0.87	0.94	0.88								
No. members of categories (weighted; logged)	3.452	1.275	0	6.121	(8)	0.12	-0.08	0.28	-0.07	-0.09	-0.07	-0.08							
No. category members received VC funding (weighted)	2.693	3.703	0	25.07	(9)	0.08	-0.18	-0.01	-0.11	-0.14	-0.11	-0.16	0.36						
No. category members received VC funding squared (weighted)	20.96	59.12	0	628.8	(10)	0.02	-0.16	-0.05	-0.10	-0.12	-0.10	-0.13	0.22	0.91					
Number of acquisitions	0.0255	0.1905	0	4	(11)	0.01	0.04	0.08	0.00	0.02	0.01	0.04	0.02	-0.03	-0.03				
Tenure in data	3.547	2.245	0	13	(12)	-0.14	-0.35	-0.12	-0.33	-0.38	-0.35	-0.37	0.05	0.04	0.03	0.00			
Number of previous rounds of financing	1.105	2.124	0	15	(13)	0.41	-0.04	0.00	-0.05	-0.06	-0.06	-0.07	0.13	0.09	0.04	-0.02	0.07		
Ranked in Software 500	0.1003	0.3005	0	1	(14)	-0.07	0.06	0.04	0.01	0.00	0.00	0.04	0.06	-0.02	-0.03	0.22	0.09	-0.04	
Year	1999	1.947	1995	2002	(15)	0.08	-0.43	-0.21	-0.28	-0.34	-0.31	-0.36	0.28	0.40	0.30	0.04	0.23	0.16	0.04

N = 1,057

Table 2. Market category and Technology Category Correlations, Years where Organization Patents

	Mean	Standard Deviation		(1)	(2)	(3)
Market category combination	1.668	0.9507	(1)			
Originality (USPTO original class)	0.7077	0.2620	(2)	-0.06		
Originality (IPC class)	0.8273	0.2042	(3)	0.14	0.36	
Originality (USPTO current class)	0.7206	0.2654	(4)	0.15	0.58	0.48

N = 589

Table 3. Tests of hypothesis 2a: piecewise continuous hazard rate models on receiving VC funding.[1]

	(1)	(2)	(3)
Market category combination		0.317*	0.374**
		(0.127)	(0.132)
Within-category development			-0.477
			(0.363)
Patented last year	-0.248	-0.690**	-0.688**
	(0.184)	(0.268)	(0.267)
No. members of categories (weighted; logged)	0.0884	0.0889	0.127+
	(0.0642)	(0.0631)	(0.0690)
No. category members received VC funding (weighted)	0.126*	0.114*	0.118*
	(0.0578)	(0.0571)	(0.0579)
No. category members received VC funding squared (weighted)	-0.0096*	-0.0088*	-0.0089*
	(0.0042)	(0.0041)	(0.0041)
Number of acquisitions	0.509+	0.490+	0.516+
	(0.282)	(0.281)	(0.286)
Tenure in data	-0.0922+	-0.103*	-0.102*
	(0.0516)	(0.0521)	(0.0519)
Number of previous rounds of financing	0.212***	0.218***	0.220***
	(0.0361)	(0.0343)	(0.0339)
Ranked in Software 500	-0.617+	-0.649+	-0.643+
	(0.337)	(0.333)	(0.336)
Time piece: 0-1 year since last received funding	-2.698***	-2.699***	-2.847***
	(0.529)	(0.532)	(0.540)
Time piece: 1-3 years since last received funding	-3.695***	-3.604***	-3.738***
	(0.560)	(0.548)	(0.556)
Time piece: 3-5 years since last received funding	-6.085***	-5.965***	-6.089***
	(1.183)	(1.173)	(1.180)
Time piece: 5+ years since last received funding	-16.81***	-16.70***	-16.82***
	(0.692)	(0.673)	(0.677)
Log pseudo likelihood	-426.0	-422.8	-422.1
Degrees of freedom	19	20	21

+p<.10 *p<.05 **p<.01 *** p < 0.001

[1.] There are 174 events for 368 organizations over 1,057 organization-years. Risk set restricted to organizations < 15 years old (or founding date unknown) that have previously patented. All independent variables are lagged by one year (for some variables the lag is specified for clarity). Year dummies are included in all models.

Table 4. Tests of hypothesis 2b: piecewise continuous hazard rate models on receiving VC funding.[1]

	Originality (USPTO original)	Originality (IPC)	Originality (USPTO current)
Technology category combination	-0.675+	-0.432	-0.115
	(0.391)	(0.452)	(0.378)
Patented last year	0.249	0.111	-0.164
	(0.361)	(0.413)	(0.331)
Log pseudo likelihood	-424.5	-425.7	-426.0
Degrees of freedom	20	20	20
Market category combination	0.349**	0.382**	0.385**
	(0.131)	(0.131)	(0.129)
Within-category development	-0.415	-0.455	-0.474
	(0.357)	(0.359)	(0.359)
Technology category combination	-0.579	-0.531	-0.263
	(0.405)	(0.461)	(0.370)
Patented last year	-0.235	-0.265	-0.512
	(0.432)	(0.463)	(0.391)
Log pseudo likelihood	-420.9	-421.5	-421.8
Degrees of freedom	22	22	22

+p<.10 *p<.05 **p<.01 *** p < 0.001

[1]Estimations reported for measures of technology combination, as indicated, included alone and with market category combination. There are 174 events for 368 organizations over 1,057 organization-years. Risk set restricted to organizations < 15 years old (or founding date unknown) that have previously patented. All models include controls for: number of category members (weighted), number of category members who receive VC funding and its square (both weighted), number of acquisitions, tenure in data, number of previous rounds of funding, whether the firm was ranked in the Software 500, and year dummies. Time pieces are included for 0–1, 1-3, 3-5, and 5+ years since the firm last received VC funding. All independent variables are lagged by one year (for some variables the lag is specified for clarity).

Table 5. Study 2 supplementary analysis: negative binomial models on 8-year citation counts for patents from the analysis. Effects on the dispersion parameter (alpha) and on the mean are reported: effects on the dispersion parameter are of primary interest to the hypotheses.[1]

	Originality: USPTO original class	Originality: IPC class	Originality: USPTO current class
Potential novelty: Effects on the dispersion parameter (alpha)			
Market category combination	-0.111***	-0.119***	-0.121***
	(0.0216)	(0.0222)	(0.0221)
Technology category combination	-0.0152	0.159*	0.115*
	(0.0416)	(0.0649)	(0.0466)
No. subclasses (logged)	-0.0546	-0.0558	-0.0455
	(0.0365)	(0.0365)	(0.0356)
No. classes (logged)	-0.0129	-0.0282	-0.0565
	(0.0618)	(0.0622)	(0.0602)
No. prior art citations	-0.00110	-0.00133	-0.00139
	(0.00087)	(0.00087)	(0.00088)
Variance in citations for class / 1000	0.0732*	0.0745*	0.0746*
	(0.0303)	(0.0305)	(0.0303)
Constant	-0.419	-0.510	-0.470
	(0.347)	(0.355)	(0.362)
Incremental benefits: Effects on the mean			
Market category combination	0.255***	0.249***	0.247***
	(0.0496)	(0.0505)	(0.0494)
Technology category combination	0.0233	0.0820	0.114+
	(0.0659)	(0.0547)	(0.0586)
No. subclasses (logged)	0.200***	0.200***	0.210***
	(0.0558)	(0.0529)	(0.0556)
No. classes (logged)	0.147*	0.145*	0.108
	(0.0671)	(0.0618)	(0.0677)
No. prior art citations	0.00884***	0.00869***	0.00858***
	(0.00084)	(0.00084)	(0.00080)
Mean citations for class	0.0331***	0.0333***	0.0331***
	(0.0041)	(0.0041)	(0.0040)
No. claims	0.0105***	0.0104***	0.0104***
	(0.0011)	(0.0011)	(0.0011)
Constant	1.085***	1.041***	1.062***
	(0.231)	(0.225)	(0.220)
Log pseudo likelihood	-44251.2	-44244.1	-44240.1
Degrees of freedom	14	14	14

+p<.10 *p<.05 **p<.01 *** p < 0.001

[1] Models run on 11,319 patents, estimated using the gnbreg estimator in STATA. Year dummies included in all models for both mean and dispersion effects.

Appendix A: Study 2 Methodological Appendix

Market category combination. Market space is constructed based on a citation network of software patents, using all patents relevant to software in a five-year window, for every year in the data (Podolny, Stuart and Hannan, 1996; Pontikes and Hannan, 2014).[18] Proximities between two patents (*m* and *n*) are calculated by dividing the common citations by the total citations for the focal patent (σ_{mn}).[19] Software organizations are located in the network using their patents in the focal year. Market categories are projected onto this network as the areas where organizations in the market category patent. Market categories for software organizations come from claims from press releases, which provide a historical record of market categories firms were in (Pontikes, 2012).

Market category combination is computed using the knowledge space similarity (σ_{mn}) between a patent *m* issued to organization *A* (*m* is in *A*'s patent portfolio P_A in the given year), and each patent *n* that is affiliated with a market category *A* is not in (*D*). A patent *n* is affiliated with market category *D* if it was issued to an organization *B* that is a member of *D* (*n* is in *B*'s patent portfolio P_B in the given year). The proximity is weighted by *B*'s grade of membership in category *D* ($\mu_B(D)$), so that organizations in multiple categories do not have outsized influence.[20] The proximity to different categories (D) for each patent *m* issued to organization *A* is computed as:

$$prox_{m,D} = \sum_{(n \in P_B)} [\mu_B(D) \times \sigma_{mn}] \qquad \text{(A1)}$$

Market category combination is an average of patent level proximities. Because the distribution is skewed, the natural log is taken:

$$(\text{Market category combination})_A = ln\left(\frac{\sum_{m \epsilon P_A} prox_{m,D}}{npat_A}\right) \qquad \text{(A2)}$$

Measures are computed yearly (for notational simplicity, time subscripts are not explicitly denoted).

[18] The patent network includes all patents in Computers & Communications (Hall, Jaffe, and Trajtenberg, 2001).
[19] Second degree similarity is also calculated between two patents with shared similarity to a third, by multiplying these similarities.
[20] Grade of membership is (the number of press releases in which organization *A* claims category *k*) divided by (the number of press releases in which it claims any category) in the given year.

Some organizations do not patent every year. A number of robustness checks and alternate specifications were tested to ensure that results are not sensitive to how this is handled. In the main analysis, market category combination is defined as 0 in years that organizations do not patent. A 0/1 flag is included as a control to indicate whether the firm patented that year. The patenting flag captures effects of having patented, and market category combination captures how combinatory are the (patented) developments the firm created. Including number of patents and patent citations, both in the respective year or cumulatively, do not change the effect. Tests are run using alternate measures of market category combination that account for historical combination in years when the organization does not patent (rather than assigning a 0 value), reported in Appendix B. The first alternate measure uses the value of market category combination in the last year the organization patented, discounted by the passage of time $\left(\frac{1}{t}\right)$. A second measure includes the cumulative measure of market category combination, discounted by the passage of time: $\sum_{t=t_o}^{t_c} \frac{(market\ category\ combination)_t}{t_c - t}$, where t_c is the current year, and t_o is the first year the firm entered the data.

<u>Technology Combination</u>

1. <u>Originality</u> is derived from a Herfindahl index, based on the class assignments of patent *j*'s cited patents (Henderson, Jaffe, and Trajtenberg, 1998). The bias-corrected modification proposed by Hall (2005) is used:

$$Orig_p = \frac{N_p}{N_p - 1}\left(1 - \sum_{k=1}^{K}\left(\frac{cite_{p,k}}{J}\right)^2\right) \tag{A3}$$

Patent p has J backward citations to k = 1 … K patent classes, and $cite_{p,k}$ is the number of p's citations in class k. N_p is the number of total classes patent p's citations are in, and multiplying the

Herfindahl index by $\frac{N_p}{N_p - 1}$ corrects for bias when the total number of citations are small (Hall 2005). This study investigates firm-level outcomes, so the patent-level measure is averaged:[21]

$$mean\_orig_A = \frac{\sum_{p=1}^{P} Orig_p}{P} \qquad \text{(A4)}$$

2. Breadth counts the technology categories to which a patent is assigned. Both the (natural log of the) number of classes and subclasses assigned to a patent are used. Multiple class and subclass assignments are available from the Patent Network Dataverse for current class and subclass.
3. Network Proximity uses the patent network created for market category combination, and computes a patent's proximity to different *patent classes* as in equation (A1), aggregated to the firm level as in equation (A2).

References

Henderson, R., A. Jaffe, and M. Trajtenberg. 1998. 'Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting 1965-1988," *Review of Economics and Statistics* 80: 119-127.

Hall, B. H. 2005. "A Note on the Bias in Herfindahl-type Measures Based on Count Data."

Podolny, J,. T. Stuart, and M. T. Hannan. 1996. "Networks, Knowledge, and Niches: Competition in the Worldwide Semiconductor Industry, 1984-1991." *American Journal of Sociology* 102:659-689.

Pontikes, E.G. 2012. "Two Sides of the Same Coin: How Ambiguous Classification Affects Multiple Audiences' Evaluations." *Administrative Science Quarterly* 57:81-118.

Pontikes, E. G, and M. T. Hannan. 2014. "An Ecology of Social Categories." *Sociological Science* 1:311-343.

[21] VCs might look at the originality of one promising patent. Using the patent with maximum originality ($\max_{p \in P} Orig_p$) yields the same results (available upon request).

Appendix B: Study 2 Supplementary Analyses

Alternate measures of technology combination

One question might be whether effects are due to the specific metric used for technology combination To address this, estimations were run using alternate measures, reported in table B1. Breadth, used in previous literature on innovation (Lerner, 1994; Leahey, 2006), is included in columns 1 – 2 as the number of classes or number of subclasses to which a patent is assigned.[22] This does not have an effect on VC financing, and does not change the market combination results.

There may be the concern that market category combination results arise because of the way it was computed using the patent citation network. To test against this, an alternate measure, network proximity, was created in the same manner as market category combination. Using the same patent citation network, proximity to organizations in different *technology categories* was calculated (details in Appendix A), for all three technology category systems. None of these measures are associated with VC investment, and effects market category combination measures remain.

--- Insert table B1 about here ---

Alternate measures of market category combination: accounting for years where firms do not patent

Tests were run to ensure results are not sensitive to how estimations treat years when organizations do not patent. In the first test, market category combination was assigned to the value in the last patenting year, divided by a $\left(\frac{1}{t}\right)$ time discount for years the organization did not patent. The effect remains (table B2, column 1). Effects are also robust when historical combination is taken into account using the cumulative measure of combination (computations described in appendix A, results available upon request).[23] Results are robust to including in the risk set only years when an organization patents.[24] Finally, the effect remains when non-patenting organizations are included in the risk set.[25]

---

[22] Multiple class assignments are available for USPTO current class only (details in Appendix A).
[23] Results are not sensitive to the functional form of the time discount (available upon request).
[24] In this case hazard rate models cannot be used, since organizations enter and exit the risk set. Instead, a logit and probit model is used to estimate the likelihood of VC funding in a year, and effects are robust in both specifications.
[25] We do not have information on market combination for non-patenting firms, so it is set to 0 in these models, and controls are included for whether the organization has previously patented or patented in the previous year.

--- Insert table B2 about here ---

Different VC Strategies

The first scope condition specifies this theory applies to a novelty-seeking evaluator, which characterizes the prototypical VC. But it is possible that some VCs do not adhere as strongly to espoused the paradigm. Specifically, VCs with lower reputations will not have access to the most promising novel investments, and so might be more likely to deviate from this model and invest in firms who are less risky, with a more predictable (but lower) payout. To explore this question, estimations are run that compares effects for receiving investment from high-reputation VCs, compared to lower-reputation VCs, presented in table B3. VC reputation comes from the LPJ reputation index (Lee, Pollock and Jin, 2011).[26] VCs invest in syndicates, with multiple firms investing in a round. Investment from a high-reputation VC is defined as when an organization receives funding in a round that includes a high-reputation VC firm, compared to rounds that do not include a high-reputation VC.[27] As expected, results show that the effect is strongest for the prototypical, high-reputation VCs. In fact, for lower reputation VC investment, the effect drops below conventional levels of statistical significance ($p < .15$).

--- Insert table B3 about here ---

Additional tests

Other questions are explored, reported in table B4. The first tests whether effects of market category combination are driven by VCs accurately predicting the technologies that will become important. Models include the *future citations* of the organizations' patents, reported column 1. Patent citations are the typical measure used to represent the impact of a patent (Trajtenberg, 1990).[28] It is unusual to include future events in a statistical model, but in this case the control provides a conservative test of the hypothesis. If causality is reversed and companies that are funded are more likely to have their patents cited, this biases results in

[26] Reputation scores are computed based on funds under management, number of start-ups invested in and amount invested, number of companies taken public, and the firm's age.
[27] High reputation is defined ranking between 1- 24 that year. Results are not sensitive to the cut-off.
[28] Importance can reliably be measured *ex post*, but there are not reliable present-time indicators of the future importance of patents (Fleming, 2001). Most studies use recombination as the ex-ante measure.

favor of the control. Effects indicate that results are not accounted for by VCs predicting which patents will become important.

--- Insert table B4 about here ---

Effects of controls

The number of category members does not have a significant effect on venture capital investment, and the number of market category members that received VC funding has a quadratic effect. Together, these results indicate that market category crowding does not have a competitive effect, perhaps because it indicates that demand is growing. Controlling for other effects (like previous funding rounds), older and larger firms are less likely to receive financing. As expected, the number of previous funding rounds is positively associated with receiving a subsequent round, as is having acquired another firm, which indicates growth potential.

Table B4 reports estimations that include additional controls. Column 2 includes the fuzziness of an organization's categories[29] and the number of categories the organization is in. Results are robust to these controls. Interestingly, although these controls predict VC funding in the full data (Pontikes 2012), neither are significant at conventional levels in the risk set of patenting organizations only. This suggests that for firms that patent, the technological basis of a firm's offering may be a stronger signal than its claimed position on the market. This reinforces the decision to use comparable, patent-based measures for both market category combination and technology category combination. Column 3 includes a non-monotonic effect for the number of competitors in a market category, and again results are robust.

References

Leahey, E. 2006. "Gender Differences in Productivity Research Specialization As a Missing Link." *Gender & Society* 20:754-780.

Lee, P. M., T. G. Pollock, and K. Jin. 2011. "The Contingent Value of Venture Capitalist Reputation." *Strategic Organization* 9:33-69.

Lerner, J. 1994. "The Importance of Patent Scope: An Empirical Analysis." *The RAND Journal of Economics* 25:319.

Pontikes, E.G. 2012. "Two Sides of the Same Coin: How Ambiguous Classification Affects Multiple Audiences' Evaluations." *Administrative Science Quarterly* 57:81-118.

Trajtenberg, M. 1990. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *The RAND Journal of Economics* 21:172.

[29] Measured as one minus the average grade of membership of organizations in the category.

Appendix B Tables

Table B1. Supplementary analysis: piecewise continuous hazard rate models on receiving VC funding. Alternate measures of technology combination.[1]

	No. subclasses (logged)	No. classes (logged)	Network proximity (USPTO original)	Network proximity (IPC)	Network proximity (USPTO current)
Market category combination	0.369**	0.374**	0.583*	0.521*	0.531*
	(0.133)	(0.132)	(0.228)	(0.247)	(0.223)
Within-category development	-0.470	-0.477	-0.450	-0.453	-0.460
	(0.363)	(0.363)	(0.359)	(0.361)	(0.360)
Technology category combination	-0.109	0.0143	-0.238	-0.173	-0.169
	(0.232)	(0.317)	(0.191)	(0.217)	(0.187)
Patented last year	-0.516	-0.703	-0.161	-0.274	-0.321
	(0.445)	(0.428)	(0.476)	(0.560)	(0.480)
Log pseudo likelihood	-421.9	-422.1	-421.2	-421.7	-421.7
Degrees of freedom	22	22	22	22	22

+p<.10 *p<.05 **p<.01 *** p < 0.001

[1]Estimations reported for measures of technology combination, as indicated, included alone and with market category combination. There are 174 events for 368 organizations over 1,057 organization-years. Risk set restricted to organizations < 15 years old (or founding date unknown) that have previously patented. All models include controls for: number of category members (weighted), number of category members who receive VC funding and its square (both weighted), number of acquisitions, tenure in data, number of previous rounds of funding, whether the firm was ranked in the Software 500, and year dummies. Time pieces are included for 0–1, 1-3, 3-5, and 5+ years since the firm last received VC funding. All independent variables are lagged by one year (for some variables the lag is specified for clarity).

Table B2. Supplementary analysis: piecewise continuous hazard rate models on receiving VC funding, using an alternate measure of market category combination. In years when the organization does not patent, the measure is assigned the previous value discounted for time.[1]

	(1)
Market category combination (alternate measure)	0.316*
	(0.136)
Within-category development (alternate measure)	-0.484
	(0.328)
Patented last year	-0.368+
	(0.202)
Log pseudo likelihood	-422.8
Degrees of freedom	21

+p<.10 *p<.05 **p<.01 *** p < 0.001

[1] There are 368 organizations over 1,057 organization-years. Risk set restricted to organizations < 15 years old (or founding date unknown) that have previously patented. Estimates are modeled as competing risks. All models include controls for: number of category members (weighted), number of category members who receive VC funding and its square (both weighted), number of acquisitions, tenure in data, number of previous rounds of funding, whether the firm was ranked in the Software 500, and year dummies. Time pieces are included for 0–1, 1-3, 3-5, and 5+ years since the firm last received VC funding. All independent variables are lagged by one year (for some variables the lag is specified for clarity).

Table B3. Supplementary analyses: piecewise continuous hazard rate models receiving funding in rounds that include or do not include high reputation VCs.[1]

	Round with high reputation VC	Round without high reputation VC
Market category combination	0.597**	0.268
	(0.189)	(0.177)
Within-category development	-0.430	-0.509
	(0.540)	(0.514)
Patented last year	-1.282**	-0.438
	(0.424)	(0.348)
Log pseudo likelihood	-194.5	-331.3
Degrees of freedom	21	21
Number of events	63	111

+p<.10 *p<.05 **p<.01 *** p < 0.001

[1] There are 368 organizations over 1,057 organization-years. Risk set restricted to organizations < 15 years old (or founding date unknown) that have previously patented. Estimates are modeled as competing risks. All models include controls for: number of category members (weighted), number of category members who receive VC funding and its square (both weighted), number of acquisitions, tenure in data, number of previous rounds of funding, whether the firm was ranked in the Software 500, and year dummies. Time pieces are included for 0–1, 1-3, 3-5, and 5+ years since the firm last received VC funding. All independent variables are lagged by one year (for some variables the lag is specified for clarity).

Table B4. Supplementary analysis: piecewise continuous hazard rate models on receiving VC funding.[1]

	(1)	(2)	(3)
Market category combination	0.377**	0.357**	0.374**
	(0.133)	(0.133)	(0.133)
Within-category development	-0.478	-0.470	-0.479
	(0.364)	(0.364)	(0.354)
Patented last year	-0.659+	-0.662*	-0.696**
	(0.365)	(0.269)	(0.268)
No. future citations (8 years, logged)	-0.0010		
	(0.0822)		
Fuzziness of organizations' categories		0.197	
		(0.937)	
No. categories organization is in (logged)		0.283	
		(0.243)	
No. members of categories (weighted; logged)	0.127+	0.0502	
	(0.0689)	(0.0879)	
No. members of categories (weighted)			0.0047+
			(0.0029)
No. members of categories squared (weighted)			-0.00001
			(0.00001)
No. category members received VC funding (weighted)	0.118*	0.132*	0.117*
	(0.0579)	(0.0600)	(0.0592)
No. category members received VC funding squared (weighted)	-0.00891*	-0.00941*	-0.00883*
	(0.00411)	(0.00426)	(0.00413)
Number of acquisitions	0.519+	0.443	0.510+
	(0.288)	(0.293)	(0.285)
Tenure in data	-0.103*	-0.111*	-0.101*
	(0.0519)	(0.0524)	(0.0504)
Number of previous rounds of financing	0.220***	0.222***	0.219***
	(0.0340)	(0.0334)	(0.0333)
Ranked in Software 500	-0.643+	-0.654*	-0.642+
	(0.336)	(0.330)	(0.338)
Time piece: 0-1 year	-2.844***	-3.000***	-2.584***
	(0.541)	(0.658)	(0.532)
Time piece: 1-3 years	-3.736***	-3.859***	-3.478***
	(0.556)	(0.678)	(0.552)
Time piece: 3-5 years	-6.086***	-6.192***	-5.835***
	(1.180)	(1.227)	(1.174)
Time piece: 5+ years	-16.81***	-16.90***	-16.58***
	(0.677)	(0.742)	(0.681)
Log pseudo likelihood	-422.0	-421.3	-422.1
Degrees of freedom	22	23	22

+p<.10 *p<.05 **p<.01 *** p < 0.001

[1] There are 174 events for 368 organizations over 1,057 organization-years. Risk set restricted to organizations < 15 years old (or founding date unknown) that have previously patented. All independent variables are lagged by one year (for some variables the lag is specified for clarity). Year dummies are included in all models.