Paper to be presented at the

DRUID Society Conference 2014, CBS, Copenhagen, June 16-18

Why do category labels stick? Industry evolution and the battle for categorical dominance

Fernando F Suarez
Boston University School of Management
Strategy & Innovation
fsuarez@bu.edu

Stine Grodal
Boston University School of Management
Strategy & Innovation
grodal@bu.edu

Diego Zunino
Boston University
Strategy & Innovation
dzunino@bu.edu

Abstract

A growing body of literature has shown that categorical affiliation affects firm performance. However, little is known about what drives the adoption of category labels in emerging industries. Using data from the smartphone industry from 1998 to 2011 we track firms? adoption of category labels over time. We argue and find support for a trade-off between the degree of familiarity and originality when firms make affiliation choices. In contrast to the existing literature, we show that familiarity and originality are not two ends of a spectrum, but represent distinct dimensions. The labels that diffuse most widely are, therefore, the ones that balance both familiarity and originality simultaneously. In other words, successful category labels are familiar, but not too familiar to be uninteresting, and original, but not too original that stakeholders cannot relate to them. We end the paper by discussing the implications of our findings for category dynamics in emerging industries and the consequences our findings have when firms make affiliation choices.

Jelcodes:M10,M10

Why do category labels stick? Industry evolution and the battle for categorical dominance

Fernando F. Suarez
Boston University School of Management
595 Commonwealth Ave., Room 649-A
Boston, MA 02215, USA
+1 617-358-3572
fsuarez@bu.edu

Stine Grodal
Boston University School of Management
595 Commonwealth Ave., Room 639-A
Boston, MA 02215, USA
+1 617-353-5617
grodal@bu.edu

Diego Zunino
Boston University School of Management
595 Commonwealth Ave., Room 618
Boston, MA 02215, USA
dzunino@bu.edu

Why do category labels stick? Industry evolution and the battle for categorical dominance

ABSTRACT

A growing body of literature has shown that categorical affiliation affects firm performance. However, little is known about what drives the adoption of category labels in emerging industries. Using data from the smartphone industry from 1998 to 2011 we track firms' adoption of category labels over time. We argue and find support for a trade-off between the degree of familiarity and originality when firms make affiliation choices. In contrast to the existing literature, we show that familiarity and originality are not two ends of a spectrum, but represent distinct dimensions. The labels that diffuse most widely are, therefore, the ones that balance both familiarity and originality simultaneously. In other words, successful category labels are familiar, but not too familiar to be uninteresting, and original, but not too original that stakeholders cannot relate to them. We end the paper by discussing the implications of our findings for category dynamics in emerging industries and the consequences our findings have when firms make affiliation choices.

INTRODUCTION

In recent years, scholars have begun to examine the socio-cognitive aspects of industry emergence (Rosal et al., 1999; Lounsbury et al., 2003; Kaplan and Tripsas, 2008; Bingham and Kahl, 2013). This research complements a large literature focused predominantly on technology-driven aspects of industry evolution (Utterback, 1994; Anderson and Tushman, 1990). In particular, investigations focusing on the emergence and role of categories and their associated labels have gained particular traction (e.g. Navis and Glynn, 2010; Pontikes, 2012). Categories are defined as socially constructed partitions or taxonomies that divide the social space into groupings of objects perceived to be similar (Bowker and Star, 2000). Categories have been shown to shape firm performance (Zuckerman, 1999), de-diversification (Zuckerman, 2000) and to have effects on the performance of individual products (Hsu, 2005; Hsu, Hannan and Kocak, 2009). Consequently, a significant body of work has been devoted to determine possible performance penalties suffered by firms when stakeholders fail to understand firms' categorical affiliation. .

When categories are initially created they are a mere category labels, that is, words or phonemes that are used to reference objects that are perceived to be similar. Indeed, when producers are launching radical innovations they often lack terms for how to reference their new product (Kaplan and Tripsas, 2008). In order to communicate the meaning of their product to stakeholders they often invent new category labels, like "pocket PC", "camera phone" or "smartphone". When first created stakeholders have no prior experience with which kind of products might be grouped under a particular category label and the meaning of any given category label, therefore, tends to be shallow. Through an iterative and negotiated process by which labels are adopted and used by different stakeholders labels become infused with value, and gradually become meaningful categories (Peirce, 1931; Grodal et al., 2014). As the meaning of categories develop, they begin to form "rules or boundaries" that dictate which objects can claim membership to a given category (Hannan, Polos, and Carroll, 2007; Navis and Glynn, 2010). Category labels that do not get traction fail to get infused by meaning and are ultimately abandoned. As this process continues, industries often coalesce in a dominant category, or "the conceptual schema that most stakeholders adhere to when referring to products that address similar needs and compete for the same market space" (Suarez et al., 2013).

Despite the importance of these dynamics that leads some labels to become categories, and lead some category labels to be selected over others, we know relatively little about the process by which some of these labels "emerge and fall out of use" (Kennedy and Fiss, 2013, p. 1). This is an important gap in our understanding, since the performance of entrants into a new industry can depend on the categorical labels they choose when positioning their products (Rosa et al., 1999; Pontikes, 2012). In this paper, we tackle this gap by presenting an empirical study of the creation and selection of categorical labels in what became to be known as the "smartphone" industry. By collecting data from the official product launches of each smartphone in the period 1998 to 2011 in the U.S., we trace the categorical labels used by smartphone producers over time. Borrowing from linguistics and extant research on categories, we develop and test novel theory and hypotheses to explain why some labels tend to stick while others are abandoned.

In particular, we posit that the success of a given categorical label in being adopted and used by other stakeholders depends largely on resolving an inherent tension between two drivers of adoption: originality and familiarity. Categorical labels that stick are more conducive than others to facilitate information exchange between disparate parties; they do this by simultaneously being distinctive enough so that they convey the originality of the underlying product, and familiar enough to be easily comprehensible (Bingham and Kahl, 2013). Managing the inherent tension between originality and familiarity lead us to hypothesize a non-linear relationship between familiarity and label adoption, and between originality and label adoption. Our results largely support our hypotheses, and thus our results not only represent one of the first empirical studies of category label adoption, but also open interesting avenues for future research.

Understanding what makes categorical labels stick is important because of the direct performance implications that these findings can have for firms that enter new industries. Indeed, the categorical labels that firms choose for their products can shape how consumers receive these products (Hsu, Hannan and Kocak, 2009). A firm's choice of category label might shape the consumers' understanding of the firm's products and penalize the company if it subsequently tries to reposition its products into another category that is achieving dominance in the industry (Zuckerman, 1999). Moreover, category labels that gain traction can give a firm the opportunity to shape the emerging industry and gain competitive advantage (Santos and Eisenhardt, 2009).

THEORY AND HYPOTHESES

Categories are arguably the most widely-researched construct among several that have been proposed to capture the socio-cognitive dimension of industry emergence, such as "field frames" (Lounsbury 2001; Lounsbury, Ventresca and Hirsch, 2003); "technological frames" (Kaplan and Tripsas, 2008; Orlikowski and Gash, 1994); and "schemas" (Bingham and Kahl, 2013). Categories, often seen as partitions that divide the social space in groups of objects perceived to be similar (Bowker and Star, 2000; Negro et al., 2011), can have important implications for firm strategy (Suarez et al., 2013), and for the design and evolution of industry products (Seidel and O'Mahony, 2013). Categories can indeed determine the set of characteristics that their object members are expected to possess and that differentiates them (or not) from members of other categories (Vygotsky, 1986).

Category labels are the first instantiation of categories in an emergent industry. Once created, labels begin to develop semantic links to other categories and their associated labels, thus deepening their meaning until some eventually become established categories (Peirce, 1931; Bingham and Kahl, 2013). When stakeholders observe a category label, they construct the group of objects that they perceive as being associated with it (Yamauchi and Markman, 1998). Therefore, a category can also be defined as the set of objects to which a particular label applies (Bowker and Star, 1999). While producers are the main creators of labels in the process of introducing their products to an emergent industry, being socially constructed, labels can also be created by other stakeholders such as users, industry analysts or observers, bloggers, etc. For instance, the "mountain bike" label was created by users who tweaked their standard bikes for racing down hills, the "robot" label was created by a writer of a science fiction book, and the label "impressionism" was created by art critics (originally with a negative connotation) to refer to the work of Monet and other painters of the time.

Most previous studies on categories have placed emphasis on the relevance of one or a few specific categories within an industry and the process by which they become legitimized (Kennedy, 2008; Navis and Glynn, 2012), or the costs or "categorical discount" of spanning multiple categories (Zuckerman, 1999). More generally, researchers have sought to understand the extent to which firms spanning categorical boundaries face a performance penalty (Hsu et al. 2009) and also the conditions under which spanning categorical boundaries might be permissible (Ruef and Patterson, 2009; Fleischer, 2009; Pontikes, 2012; Granqvist et al. 2013).

In contrast, little research has been conducted to unveil the dynamics of competing categories within a nascent industry; i.e. the contestation process that leads to the adoption of some categories and the abandonment of others. In order to tackle this challenge, in this article we build upon recent theoretical developments (Bingham and Kahl, 2013; Grodal et al., 2014) to propose that category labels have to resolve the paradox of being simultaneously (1) distinctive enough that they convey the novelty of the underlying product and attract the attention of stakeholders, and (2) familiar enough to be easily comprehensible. By establishing semantic links to existing categories and labels, a new category label can appear more familiar to stakeholders and thus increase its likelihood of being adopted. Bingham and Kahl (2013), for instance, discuss how in the late 1940s the new "computer" category label became associated with the "machine" category, which helped stakeholders assimilate the new category to something they already knew. However, "computer" also became gradually associated with the "brain" category, which conveyed novelty and new possibilities. Hargadon and Douglas (2001) also allude to this tension: "Without invoking existing understandings, innovations may never be understood and adopted in the first place. Yet, by hewing closely to existing institutions, innovators risk losing the valued details, representing the innovation's true novelty, that ultimately change those institutions. Success, then, requires entrepreneurs to locate their ideas within the set of understandings and patterns of action that constitute the institutional environment in order to gain initial acceptance, yet somehow retain the inherent differences in the new technology that ultimately will be needed to change those institutions" (p. 478).

While most of the existing literature has posited familiarity and originality as opposite points of a spectrum, we develop a more nuanced account. Indeed, "originality" is more than simply being "unfamiliar", because originality can also arise from novel combinations (Pieters, Warlop, and Wedel, 2002). An established literature has shown that most innovations are formed through the recombination of existing elements (Schumpeter, 1939), and that the originality of these recombinations influence both the success of the innovation (Fleming, 2001; Fleming, Mingo, and Chen, 2007) and how widely they diffuse (Grodal and Thoma, 2014).

A category label can be original while still using very familiar words of phonemes. For example, when John Burton Carpenter in the 1970s created the compound label "snowboard" it was at the same time original, because it combined two words "snow" and "board" that had never been used together, but at the same time it was familiar, because even stakeholder who were exposed to the compound for the first time were able to associated it with some initial meaning. Indeed, compounds, defined as "the simple concatenation of any two or more nouns [or other words] functioning as a third nominal" (Downing, 1977: 810), are an important way for stakeholders to create category labels that build links to existing categories, thus invoking familiarity and, at the same time, allow for new and unique recombinations (Lieber, 1983; Wry, Lounsbury and Jennings, forthcoming). We propose that compounds are a most effective ways by which label creators can achieve a balance between novelty and familiarity that will make their labels successful (Berger et al., 2012), therefore:

H1: The use of compounds in the creation of category labels in an emerging industry is positively associated with its degree of adoption.

An important way in which novel category labels become familiar is by establishing links to existing categories and their associated labels. Category labels that fail to make these connections will most likely be abandoned because stakeholders will be confused about the labels' meaning. For example, an early label for what later became the "snowboard" industry was the "snurfer," a label that failed to make enough connections to existing labels and categories and therefore was abandoned. While the label drew from the two words "snow" and "surfer" these association were disguised making it hard for stakeholders encountering the category label "snurfer" for the first time to make sense of the label. While being unfamiliar might be problematic being too familiar has its disadvantages as well, because when category labels increase in familiarity they become taken-for-granted and thereby fail to elicit scrutiny by stakeholders (Hsu and Grodal, 2014). Indeed, too much familiarity may render the label obvious or uninteresting and thus fail to capture the attention of stakeholders. We therefore hypothesize,

H2: There is an inverted U-shape relationship between the familiarity of a category label and its degree of adoption in an emerging industry.

A similar dynamic occurs with originality.). While label creators should strive to have their labels be original as to make them attractive to other stakeholders, as we have argued above, the recombination of words in the creation of category label like the recombination of novel technologies might be taken to an extreme: "The set of potential combinations and, a fortiori, the possible ways that each set of potential combinations can be combined has become essentially infinite" (Fleming, 2001). In other words, label creators have limitless possibilities to create original, novel labels by recombining words and phonemes that have never been recombined before. However, when labels are too novel and lack semantic connections to existing labels or categories, they risk being difficult to comprehend by stakeholders. Indeed, Fleming (2001) finds recombinations based in local search tend to be more successful than those based on more distant search. It follows that,

> *H3: There is an inverted U-shape relationship between the originality of a category label and its degree of adoption in an emerging industry.*

The tension that labels creators need to manage between originality and familiarity implies that, other things being equal, the effect of an increase in originality on the adoption level of a label depends on the label's position in the familiarity score – and vice versa. Label creators need to strike the right balance between originality and familiarity, which implies that,

> *H4: There is a positive interaction effect between the originality and familiarity of a category label and its degree of adoption in an emerging industry.*

DATA AND METHODS

Data

To test our hypotheses, we constructed a unique dataset of category labels and their adoption in what we know today as the smartphone industry, which emerged in the late 1990s. More than 280 category labels were introduced in the U.S. smartphone industry over our 14 years study period 1998 to 2011, for a total of 3,920 label-year observations for our analysis. We identified category labels by examining the press releases that smartphone manufacturers used to introduce their new products to the market. Press releases are a reliable record of the category labels claimed by different companies over time, since they represent the firms' best efforts to communicate the position of their products within market categories. Indeed, press releases have already been used in existing research on categories; Pontikes (2012), for instance, finds that more than 1 unique category label (mean 1.5, median 1.3) per producer are issued every year.

Our dataset tracks category labels from 52 companies used for 1,642 devices identified as belonging to the "smartphone" category, launched from 1998 to 2011 in the United States. Figure 1 shows the number of models defined as smartphones produced by each company over the 14 years observed. The figure illustrates the different categorical strategies followed by the different competitors; for instance while Apple uses very few categorical labels, Samsung does the opposite, using more than one hundred category labels in their product press releases. The figure also suggests that the number of category labels used by a producer is not necessarily correlated with the producer's performance in the market. For example, Alcatel and Siemens, the companies that used the largest number of category labels, had already exited the industry before the end of our data period. Figure 2 shows that the total number of labels used by all producers increases over time up to a point, and then begin to decrease, exhibiting a pattern consistent with existing theory (e.g. Suarez et al, 2013). Due to the historical nature of our study we located press release in multiple locations. We collected the majority of the press releases from the manufacturers' websites (about 60%); we found close to 20% of the press releases on handset distributors and technology websites; less than 5% came from carriers' websites; and 9% were found searching on Lexis-Nexis (see Table 1 for a summary of sources). For about 12% of the devices, we could not find information about their press releases, so these were left out of our sample. Irrespective of the location where we identified the press releases they were all press releases issued by the producing firm.

*** INSERT FIGURE 1 HERE ***

*** INSERT TABLE 1 HERE ***

*** INSERT FIGURE 2 HERE ***

Once we had identified all 280 category labels used across all of the producers we coded whether a label was a compound or not. If the label was a compound we decomposed it into its component parts. For example, we decomposed the compound "camera phone" into its two component words "camera" and "phone". In total the smartphone manufacturers used 166 words in their category labels.

We constructed a measure of a category label's familiarity and originality by drawing on data from three major U.S. newspapers namely, *The New York Times*, *The Wall Street Journal*, and *USA Today*. We randomly selected one day per month for each year in our data collection period, excluding weekends because weekend editions tend to differ among these newspapers, and collected all the articles that were published on that particular day. On average, each random day provided 400 to 450 articles, with some exceptions for holidays where the articles count dropped from 190 to 320. On average we collected about 4,500 articles per year over 14 years, for a total of 63,055 articles as shown in Table 2.

We used a new tool for relational content analysis, MemeStat, to obtain both counts of words and counts of their co-occurrence with other words both at the story (entire article) and paragraph level. To alleviate concerns of external validity derived from using the sample of 63,055 articles from four newspapers we conducted a separate data collection in Factiva, where we searched all articles published on all days of the year in the 69 newspapers included in the database (the list of newspapers is in the appendix). Below we explain how we developed each of our variables in more detail.

*** INSERT TABLE 2 HERE ***

Dependent Variable

Count of mentions in press releases. The dependent variable in our analysis is a measure of the adoption of a given category label. This measure provides greater granularity over alternative measures of category label success, such as survival. We compute the level of adoption as the number of times a category label has been used by any smartphone producer within the first year from the label introduction in the manufacturer's press release. The most-used label in any year had 101 mentions, in 2010, while most category labels are mentioned just once.

Explanatory Variables

Two-word compound is a dummy variable that assumes the value of one if the category label is a recombination of two words (as for example in "pocket computer"), and zero otherwise. We created another dummy variable to control for the rather rare occurrence of compounds with three or more words (as for example in “electronic messaging device”).

Familiarity. We constructed a familiarity index by searching our corpus of articles for the frequency of each word within any given year. We then measured how common each word was relative to the words used in all the category labels for that year. More precisely, we normalized each label word by dividing its count by that of the most cited word across all labels in a given year. For category labels consisting of more than one word, we constructed the familiarity score by generating the arithmetic average of the count of the compound words. Given our hypotheses, familiarity enters the regressions both in linear and quadratic terms. For robustness, we develop three different measures of familiarity: at the article level and paragraph level using our sample derived from three newspapers, and than a third measure at the article level considering the larger set of 69 newspapers.

Originality. This variable represents for any given compounded category label how novel the recombination of words within the compound are. We measured this by identifying how common it was for two or more words to appear together in the same press release or story for any given year. For example, within this excerpt from *The Wall Street Journal,* July 10, 2007 part of our sample: “The colorful computers and the ads are part of an effort by Dell to redefine its brand, which had lost its focus in recent years as the PC business changed. Dell became famous with the direct-distribution sales model pioneered by founder Michael Dell, in which it sold computers over the phone and on the Internet. Instead of aiming to create an image for the brand, Dell's ads used to focus on the technical specifications of its computers, such as the speed of its processors.” we would for example code the word “PC”, and ”phone” as co-occurrences in a given paragraph . We measure originality as the inverse of the share of the number of co-occurrences of two words over the number of times the two words are mentioned individually. For 2007 the category label “PC phone” would, thus, have a low originality score if the two words together only appeared in that paragraph. We subtracted the intersection from the sum of the two words in order to avoid counting the intersection twice. Logically, the originality index is illustrated in Figure 3 and may be represented as following:

$$Originality_{it} = 1 - A \cap B_{it} A \cup B_{it} = 1 - A \cap B_{it} A_{it} + B_{it} - A \cap B_{it}$$

*** INSERT FIGURE 3 HERE ***

Category labels with more than two words are only 7% of the category label-year observations and just 17 of them receive between one and three mentions. For simplicity, given the minimal loss, we limit the analysis of originality to two-word category labels. Theoretically, a value of originality equal to zero means that there is no occurrence where word A is mentioned without word B, while a value of originality equal to one means that the two words have not co-ocurred together in that year. The originality index enters the regression both in linear and quadratic terms. As we did with familiarity, we performed the analysis of originality at the article and paragraph level in the 3-newspaper sample, and also at the article level using the larger set of 69 newspapers.

<u>Control Variables</u>

We controlled for several other characteristics of category labels, in addition to information about the smartphone producers and including year dummies. We created a variable that counts the number of words used in a category label, in order to explore whether having more elaborate compounds has an effect on label adoption. Notably, this variable shows only a 0.27 correlation with the compound dummy variable. We also control for specific characteristics of the compounds and the information they convey. We created a dummy variable which assumes the value of one if the category label contains the name of a specific technological features such as “Bluetooth” and “color screen”, and zero otherwise, because such references may initially increase adoption but may also become more quickly abandoned. We also created a dummy variable for a category label mentioning a particular generation of technology, such as “2G” or “3G.” Finally, we created a dummy variable for the specific case of category labels that contain references to an operating system (without specifying generations), such as “Android” or “Windows”. Similarly, we created a dummy to check for the impact of using a trademarked word in the label, such as “Linux”, or “Blackberry”. These variables are not highly correlated, with correlation scores ranges from 0.21 to 0.30. We also control for the effect that the size of a producer has on the adoption of a label it uses. We operationalize this construct by creating a dummy variable “large creator” that assumes the value of one if a firm that uses a particular label is among the five largest firms by revenue at the year of introduction, zero otherwise. As a final check, we added a control for labels that reference specific types of usage; to this effect, we created a variable that assumes the value of one if the category label contains words such as “fashion” or “business.”

Methods

Because we deal with count data, we used three specifications to look for associations, moving from a linear model and progressively relaxing the assumptions. To facilitate the interpretation of results we first ran a log-link model that keeps the linear structure but uses the logarithm of the count dependent variable, as shown below:

$$\log count_{it} = \alpha + XB_i + \gamma_1\, familiarity_{it} + \gamma_2\, familiarity_{it}^2 + \delta_1\, creativity_{it} + \delta_2\, creativity_{it}^2 + \xi\, creativity_{it} * familiarity_{it} + \tau_t + \varepsilon_{it}$$

Where X is the vector of time-invariant characteristics of the category label, and τ is a set of dummy for each year observed. We test the sign of the coefficients of interest using a Poisson model, i.e. a maximum-likelihood model that is the standard in the analysis of the count data. The model distribution is the following:

$$f(y, \lambda) = \Pr(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

The peculiarity of the Poisson model is the assumption of expected value and variance of the data equal to λ, known as the dispersion index. This assumption can be too strict, so we relax the assumption and use the Negative Binomial distribution, which does not rely on the identity of variance and expected value to the dispersion index, but assumes overdispersion—that is variance larger than the expected value.

$$f y,\ \phi,\sigma 2=\Pr Y=y=i=1n\Gamma\phi\sigma 2-1+yi\, yi!\Gamma\phi\sigma 2-1\sigma 2-1\sigma 2yi\sigma 2-\phi\sigma 2-1$$

RESULTS

Table 3 reports the descriptive statistics of the variables. The table shows that category labels use on average three words. Almost 40% of labels reference technological features, 20% reference an operating system, 15% reference a technology generation, 11% of the category labels contained a trademark term, and about 4% of labels reference particular usages. As expected, our measures of familiarity and originality show that these two variables are not simply opposites of a continuum. Table 4 we present examples of category labels that fall in different cells in a familiarity-originality grid. It is interesting to note that, consistent with our theorizing the label "smartphone" that became the dominant category (see Suarez et al., 2013) in this industry, scores "medium" in both familiarity and originality.

Table 5 shows the results of the regression analysis for the familiarity variable. Model 1 is log link and shows the effects of the control variables and year dummies, before adding our independent variables familiarity and originality. The model shows that the use of a two-word compound is positively associated with adoption at the 10% significance level, while the dummy for compounds of more than two words fail to achieve significance. Moreover, the variable for label word count is negatively associated with adoption and statistically significant, but it has to be kept in mind that there are few labels with 3 or more words in the sample. The strongest effect on adoption among the controls is the size of the firm who introduces the label; this dummy is positive and significant at one percent level. The dummies for category labels that reference technology generations and technological features are negative and significant, but seem to have a relatively small effect on adoption. The dummy that captures category labels that reference an operating system is, in contrast, significant and positively associated with label adoption, while using a trademarked word in the label appears to have a negative association with adoption. Adding new words to the existing pool seems to have no effect. Finally, labels that reference usage types, and one-word labels that are derived from altering existing words, are negatively associated with adoption.

Model 2 we tested the effect of familiarity on adoption in a model with no controls and only year dummies, but familiarity fails to achieve significance. Model 3 adds a quadratic term; while familiarity is still not significant, its sign changes to positive and the quadratic term is negative and significant, which suggests support for Hypothesis 2. Model 4 adds the control variables to the model with the main and quadratic effects. The coefficient for familiarity is positive and achieves significance at the 5% level, while its quadratic term is still negative and more strongly significant than in Model 3. Models 5 and 6 test the robustness of the results by running a Poisson and Negative Binomial specification, respectively. The inverse u-shape results for familiarity still holds in these two models, and the significance of the coefficients is even stronger in the Poisson model. Notably, the coefficient for the two-word dummy increases its significant to the 1% level.

So far our analyses have been done at the paragraph level; i.e. by considering the number of times the words appear within the same paragraphs in a random sample of articles per year taken from the three major U.S. newspapers. As a further robustness test of the association between familiarity and category label adoption, we run the models again but at the article, not paragraph, level. Moreover, we run a model using not the sample from the three newspapers mentioned above, but all the articles published by the 69 major newspapers in a given year contained in the Factiva database. Table 6 shows the results for log link (models 1 to 3) and negative binomial specifications (models 4 to 6) run at different levels of analysis. Models 1 and 4 use the baseline paragraph level from the 3 newspaper sample, while Models 2 and 5 also use the 3 newspaper sample but uses the data from the full article. Models 3 and 6 consider the larger set of 69 newspapers, at the article level. Overall, the coefficients of the control variables are largely consistent with the previous regressions. The main and quadratic effects of familiarity on category label adoption are robust when using data at the paragraph level from the 3-newspaper sample, but they loose significance when using the data at the article level. However, the inverted U-shaped association between familiarity and category label adoption is again supported when we run the model using the larger 69 newspapers set at the article level.

Table 7 presents the results for the association between originality and label adoption. As in the previous tables, the baseline level of analysis is the paragraph level using the 3-newspaper sample; for simplicity, we only consider two words compounds. Models 1 to 3 are log link specifications, while the models 4 and 5 are respectively Poisson and Negative Binomial. Models 1 and 2 are run without adding control variables. In Model 1, originality fails to achieve significance. When we add the square term, the coefficient for the linear term is positive and significant at the 5% level, and the coefficient for the quadratic term is negative and also significant at the 5% level. This provides initial support for our Hypothesis 3. When we add controls and change the specifications in Models 3 to 5, the coefficients for originality remain consistent with our hypothesis about an inverted U-shape relationship between originality and category label adoption and become even more significant, at the 1% level. These models show that the main and quadratic effect of familiarity continues to suggest an inverted U-shape relationship between originality and category label adoption, and their coefficients are even stronger *vis-à-vis* the models that had not added the originality variables. In short, these results are consistent with Hypothesis 2 and Hypothesis 3.

Table 8 provides another robustness check, including the main and quadratic effects of both originality and familiarity on category label adoption, and considering different levels of analysis and specifications. Models 1 to 3 use log link specification, while models 4 to 6 use a negative binomial. The inclusion of both originality and familiarity in our models produces an overall improvement in the results; the inverse U-shapes relationship between originality and familiarity and category label adoption are retained, and the coefficients are now significant at all levels of analysis (including that at article level for the 3-newspaper sample, which was not significant in models 2 and 5 of Table 6). In Table 8, the effect of familiarity on label adoption measured using articles in the set of 69 newspapers is less than when the analysis is done at the paragraph level.. The association between originality and category label adoption also weakens a bit moving from paragraph level co-occurrence to the article level, although it remains significant throughout.

We finally test Hypothesis 4 by looking at the interaction between novelty and familiarity on category label adoption. Table 9 presents this analysis. Given the multicollinearity, or micronumerosity, some results lose significance when the interaction between these two variables is tested. Models 1 to 3 are log link specifications, models 3 to 6 are Poisson, and models 4 to 9 are negative binomial. The coefficient of the interaction between familiarity and creativity is positive and significant when the analysis is run at the paragraph level. The significant is diminished or lost, however, when we run the models at the article level using the 3-newspaper sample or the larger set of 69 newspapers. We note that when the coefficients are statistically significant, they retain the inverse U-shaped relationship with category label adoption predicted by the Hypotheses 2 and 3. Considering the level of multicollinearity, and the smaller power of the variables when measured at the article level in the population of newspapers, we consider Hypothesis 4 to be partially supported by our results.

DISCUSSION AND FINAL REMARKS

Our research augments our understanding of the socio-cognitive dimension of industry emergence. Drawing from the literatures on categories and industry evolution we present an in-depth looks and empirical test of the socio-cognitive dynamics that takes place as an industry develops. Our study shows that, as theorized by recent research (Suarez et al., 2013), an initial period of categorical divergence where the number of categories in use increases over time, is followed by a period of convergence where the number of categories in use is reduced, a pattern illustrated in Figure 4. In doing so our study answers the call by Kennedy and Fiss (2013) to study the process through which categories are created, adopted and fall out of use.

*** INSERT FIGURE 4 HERE ***

We focus on category labels, because they represent the first instantiation of categories. During the early period of industry evolution producers introduce new category labels as they struggle to find the right term to describe their innovative products. As the industry evolves producers slowly ease to introduce new category labels and instead converge on using the same category label(s). In this study we set out to understand why some category labels are adopted while others abandoned. We argue that, in order to be successful, category labels have to overcome an inherent tension between familiarity and originality. Drawing from recent literature (Kennedy and Fiss, 2013; Grodal et al., 2014), we hypothesized a non-linear, inverse U-shape relationship between adoption and both familiarity and originality. In other words, successful category labels are familiar, but not too familiar to be uninteresting, and original, but not too original that stakeholders cannot relate to them. Our results, using data we collected on what is today called the smartphone industry, largely support our hypotheses. Indeed, smartphone, the "dominant category" (Suarez et al., 2013) in this industry, out competed more than 200 different labels that firms used to refer to their products and, as Table 4 shows, falls in that sweet spot of familiarity and originality.

The arguments and results of this article have direct and important implications for firms' strategies. It is likely that firms, when introducing their products, are not fully aware of the socio-cognitive dynamics that take place in their emergent industry, let alone its consequences. As existing research has shown (Zuckerman, 1999; Pontikes, 2012), choosing a categorical positioning that is not consistent with what customers and other stakeholders begin to accept as the major categories in an industry can have important consequences for the success of the firm's products and the firm's overall performance. While it is always possible for a firm to reposition its products using a different category label than that used when introducing the product, making such changes are costly. At the end of the day, even the most powerful firms have to conform to the dominant category once this has emerged. In the smartphone industry, for instance, Apple resisted for years the use of "smartphone" in their communications and advertising, emphasizing the power and customer awareness of their "iPhone,", which was first launched in 2007. However, by 2012 Apple began to use the "smartphone" category label to refer to the iPhone. Firms have to conform to the dominant category label because if they don't, they run the risks of not being in the preference set of customers when they look for a product in that category. When meaningful categories form from category labels, they create "rules of membership" or "boundaries" (Navis and Glynn, 2010)—that is, rules by which stakeholders determine which products belong to the category and which don't—that can be delineated and unforgiving.

The implication of our study is, therefore, that firms should pay more attention at how they create category labels for their products, and simultaneously follow closely the evolution of categorical labels in the industry. Balancing familiarity and originality is not trivial, and there may be different strategies open to firms, such as hedging by trying to position an early product in more than one category while learning and collecting information about which label works best. It is also clear from our results that firms are better off using two-word compounds when creating category labels. Compounds make it easier for firms to overcome the tension between originality and familiarity than single words, as one word can provide the familiar link while the other an original twist. Moreover, two familiar words that are not used together often may provide the originality needed for a label to succeed.

Despite our contributions, there are still several limitations to our study. First, we have only studied what drives the adoption of category labels, but have not explored in our regressions the link to product or firm performance. Given that many of the firms competing in the smartphone industry are large firms that produce many different products for different industries, it has not been possible yet to obtain a reliable measure of smartphone performance for all firms in the sample. However, it would be beneficial for future research to address this relationship. The increased understanding of the dynamics of category creation and adoption that our study provides is an important issue in its own right.

Second, one may question the external validity of our study, because we have studied one industry only. However, the dynamics of categorical evolution have been investigated in other industries (e.g. Rosa et al., 1999; Pontikes, 2012) and, while they do not present the kind of empirical analysis that we do here, the basic pattern of category creation and adoption seem to be similar to what we observed in the smartphone industry.

Figure 1. Number of category labels used by the different firms in the study, 1988-2011

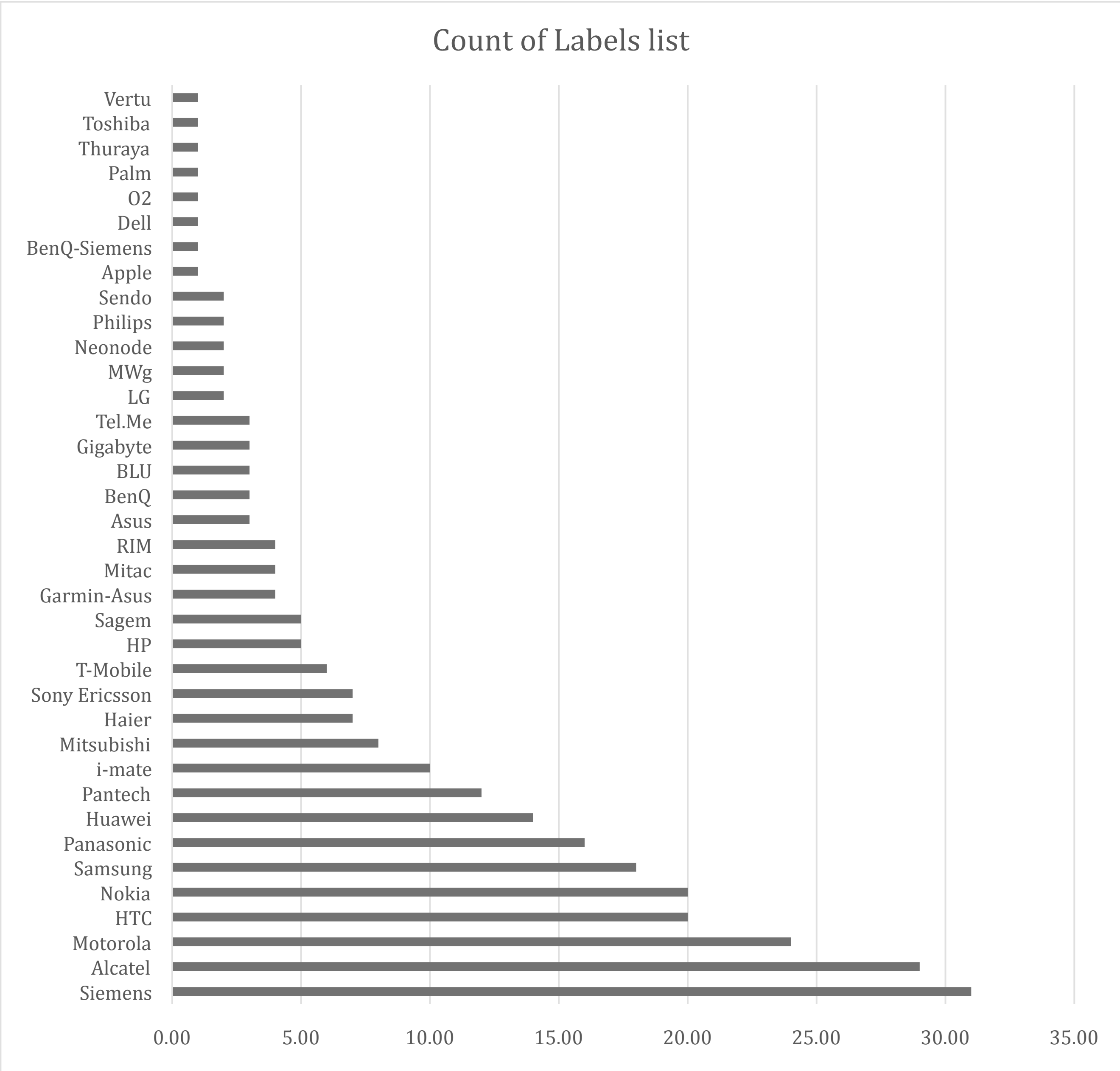

Figure 2. Number of labels used by all smartphone producers each year, 1998-2011

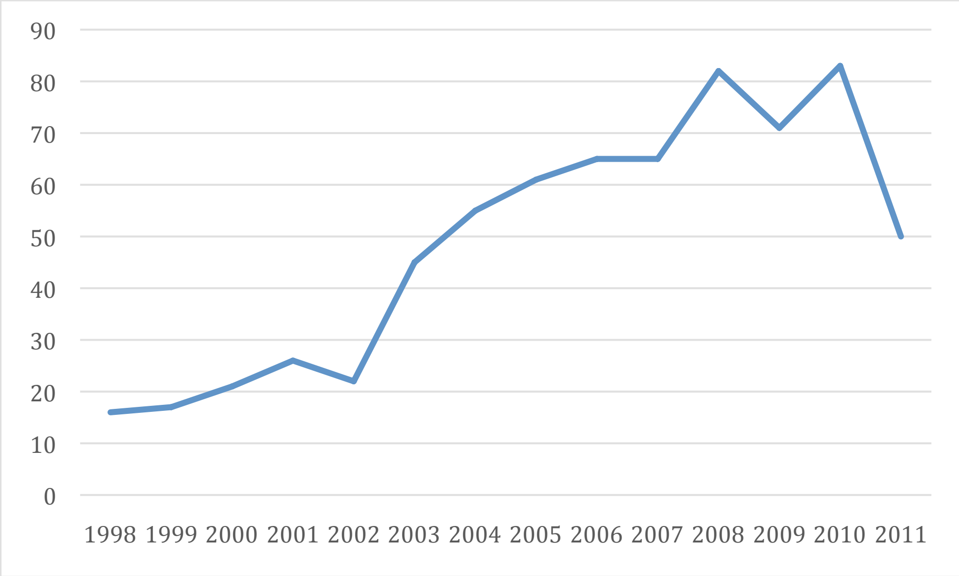

Figure 3. The Originality Measure

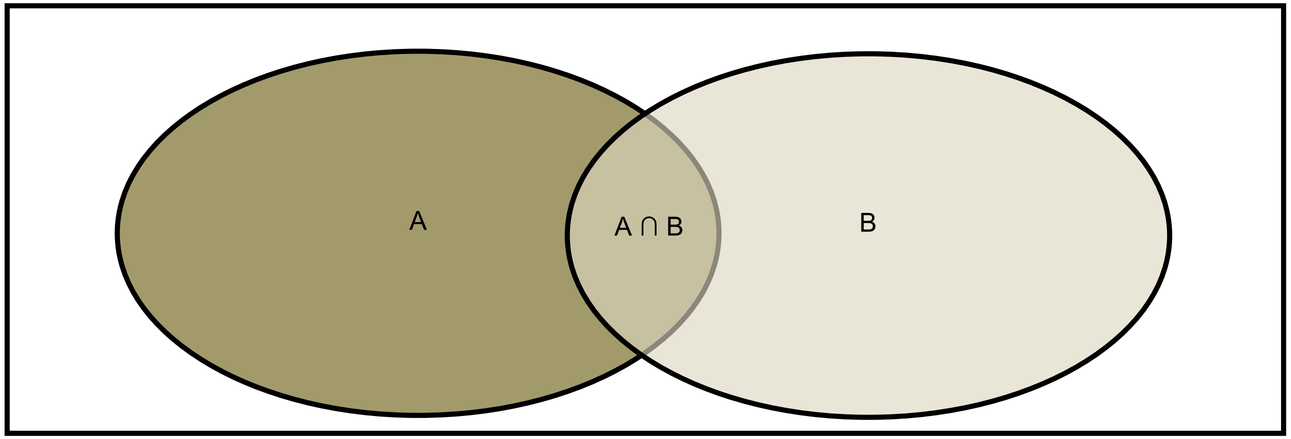

Figure 4. Label Adoption over time – Six Labels (percent of total adoption per year)

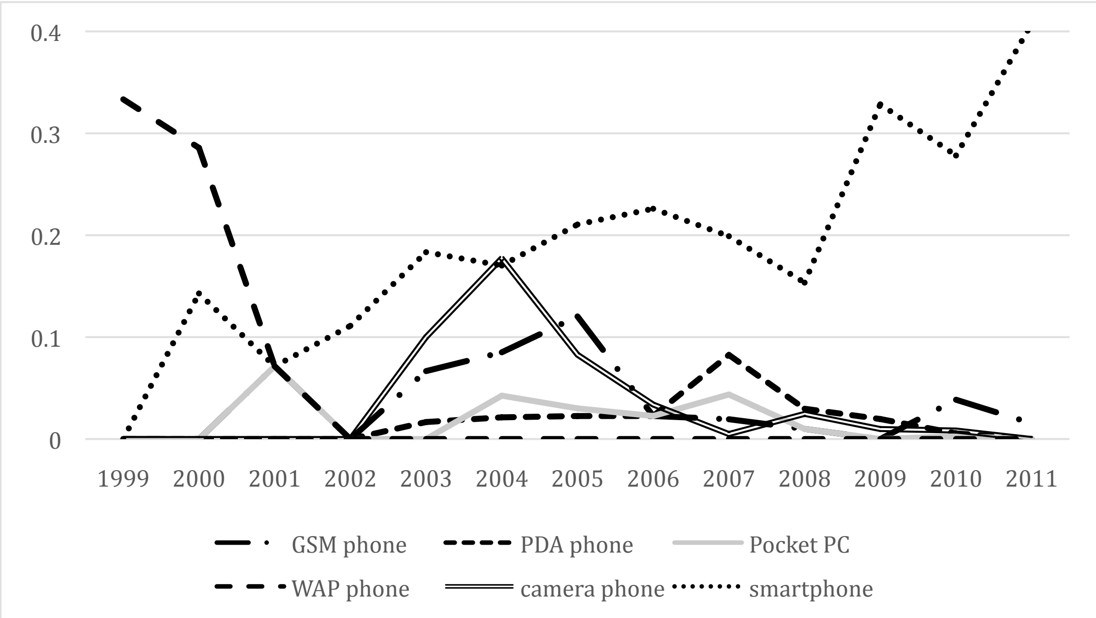

Table 1. Sources of Smartphone Press Releases in US and UK, 1998-2011

Source	Count	Share
Carrier PR	15	1%
Carrier Other	13	1%
LexisNexis	152	9%
Manufacturer Other	189	12%
Manufacturer PR	381	23%
Mobile Phone Website	397	24%
Sales Outlet	35	2%
Tech Website	268	16%
Other PR	1	0%
Blank	191	12%
Total	1642	100%

Table 2. Number of random articles considered each year*

Year	Count of Articles	Share
1998	4734	7%
1999	4709	7%
2000	4279	7%
2001	4861	8%
2002	4555	7%
2003	4866	8%
2004	4319	7%
2005	4492	7%
2006	4275	7%
2007	4285	7%
2008	4975	8%
2009	4376	7%
2010	4382	7%
2011	4496	7%
Total	6,3604	100%

* Drawn randomly from Factiva.

Table 3. Descriptive Statistics and Correlations

		Mean	S.D.	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Count of words	.45	3.00	0	101	1																	
2	Words per label	2.83	.96	1	5	-0.113	1																
3	Perishable Technology	.15	.36	0	1	-0.045	0.241	1															
4	Mentions a technology	.38	.48	0	1	-0.055	0.199	0.212	1														
5	Mentions an OS	.19	.39	0	1	-0.048	0.323	0.237	-0.313	1													
6	-powered, -enabled suffixes	.11	.31	0	1	-0.039	0.349	-0.016	-0.054	0.428	1												
7	Refers to a category of users	.04	.19	0	1	-0.014	-0.106	-0.081	-0.149	-0.092	-0.067	1											
8	Two-words compound	.36	.48	0	1	0.017	0.28	0.125	0.1	0.104	0.127	-0.023	1										
9	>two-words compound	.06	.24	0	1	-0.029	0.278	-0.065	-0.042	-0.006	0.009	-0.049	0.341	1									
10	Derivation label (e.g. communicator)	.05	.21	0	1	-0.019	-0.19	-0.093	-0.136	-0.105	-0.076	-0.042	-0.058	-0.056	1								
11	Adds new words to the pool	.43	.50	0	1	0.013	0.029	-0.1	-0.065	-0.208	-0.019	0.068	0.118	0.114	0.016	1							
12	Introduced by a large company	.04	.20	0	1	0.098	-0.111	-0.003	-0.039	-0.07	-0.052	0.057	-0.03	-0.024	0.028	0.016	1						
13	Familiarity (in-sample paragraph level)	.060	.12	0	1.22	-0.001	0.225	-0.073	-0.053	-0.078	-0.012	0.063	0.061	0.277	-0.043	0.077	-0.011	1					
14	Familiarity (in-sample article level)	.060	.12	0	1.22	-0.004	0.186	-0.107	-0.058	-0.076	-0.043	0.057	0.044	0.279	-0.051	0.06	-0.004	0.918	1				
15	Familiarity (out-of-sample article level)	.041	.06	0	.54	0.002	0.204	-0.12	-0.057	-0.099	-0.031	0.095	0.054	0.31	-0.059	0.08	-0.005	0.935	0.917	1			
16	Creativity (in-sample paragraph level)	.012	.04	0	.6	-0.033	0.486	0.046	0.116	.	.	0.042	-0.106	-0.001	0.025	0.213	0.065	0.121	0.113	0.142	1		
17	Creativity (in-sample article level)	.013	.04	0	.6	-0.024	0.29	0.045	0.096	.	.	0.021	-0.067	-0.11	-0.089	0.095	0.01	0.05	0.066	0.082	0.767	1	
18	Creativity (out-of-sample article level)	.058	.80	-3.984509	26.5625	-0.096	0.104	0.03	0.04	0.027	.	0.002	0.011	0.01	0.014	0.074	0.018	0.006	0.022	0.021	0.243	0.103	1

Table 4. Examples of labels in the familiarity-originality grid*

	Low familiarity	**Medium Familiarity**	**High familiarity**
Low originality	Multimedia Handset	Portable Device	Business Phone
Medium Originality	Messaging Device	Smartphone	Mobile Office
High originality	Wireless Handheld	Communications Device	Social Phone

* Based on scores in the familiarity and originality metrics, as defined in the paper.

Table 5. Regression analysis for Familiarity

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Link	Log Link	Log Link	Log Link	Poisson	NegBin
DV: log count (1,2,3,4) count (5,6)						
Familiarity		-0.060	0.224	0.554**	16.060***	10.314***
		(0.04)	(0.16)	(0.17)	(2.62)	(2.63)
Familiarity^2			-0.354*	-0.539**	-20.58***	-11.049**
			(0.17)	(0.17)	(5.51)	(3.44)
Two-words compound	0.041*			0.043*	1.068***	0.789***
	(0.02)			(0.02)	(0.25)	(0.16)
>two-words compound	-0.038			-0.048*	-1.242***	-1.309***
	(0.02)			(0.02)	(0.34)	(0.40)
Words per label	-0.060***			-0.066***	-1.265***	-0.821***
	(0.01)			(0.01)	(0.12)	(0.10)
Large Label Introducer	0.754***			0.754***	1.279***	2.220***
	(0.04)			(0.04)	(0.18)	(0.19)
Generational Technology	-0.042***			-0.038**	-0.746***	-0.835***
	(0.01)			(0.01)	(0.20)	(0.24)
Technological Features	-0.035*			-0.024	-0.311	-0.058
	(0.02)			(0.01)	(0.19)	(0.20)
OS Reference	0.106***			0.119***	1.244***	1.686***
	(0.03)			(0.03)	(0.25)	(0.37)

-powered, -enabled suffixes	-0.003			-0.007	-0.105	-0.437
	(0.01)			(0.01)	(0.25)	(0.25)
Usage Type Reference	-0.134***			-0.155***	-2.377***	-2.034***
	(0.02)			(0.03)	(0.39)	(0.46)
Derivation label	-0.150***			-0.143***	-1.670***	-2.060***
	(0.02)			(0.02)	(0.33)	(0.29)
Adds new words to the pool	-0.025			-0.021	0.092	0.048
	(0.01)			(0.01)	(0.20)	(0.15)
Label contains a trademark	-0.139***			-0.138***	-1.671***	-2.068***
	(0.02)			(0.02)	(0.21)	(0.29)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	3751	3794	3794	3751	3751	3751
R^2	0.199	0.0573	0.0579	0.201		

Notes: Standard Errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Constant omitted from the table. Data range from 1998 to 2011. The dependent variable is the (log) number of times the label has been cited by a company in products' press releases. Robust standard errors.

Table 6. Regression for familiarity using different levels of analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	Log-link	Log-link	Log-link	NegBin	NegBin	NegBin
	Paragraph	Article in-sample	Article population	Paragraph	Article in-sample	Article population
DV: log count (1,2,3) count (4,5,6)						
Familiarity	0.554^{**}	0.049	1.351^{***}	10.314^{***}	0.522	19.129^{***}
	(0.17)	(0.15)	(0.32)	(2.63)	(1.91)	(3.80)
Familiarity^2	-0.539^{**}	-0.043	-2.865^{***}	-11.049^{**}	-1.285	-46.347^{***}
	(0.17)	(0.16)	(0.73)	(3.44)	(2.10)	(12.18)
Two-words compound	0.043^{*}	0.041^{*}	0.051^{**}	0.789^{***}	0.799^{***}	0.871^{***}
	(0.02)	(0.02)	(0.02)	(0.16)	(0.16)	(0.17)
>two-words compound	-0.048^{*}	-0.039^{*}	-0.055^{**}	-1.309^{***}	-1.171^{**}	-1.358^{***}
	(0.02)	(0.02)	(0.02)	(0.40)	(0.38)	(0.37)
Words per label	-0.066^{***}	-0.060^{***}	-0.072^{***}	-0.821^{***}	-0.725^{***}	-0.822^{***}
	(0.01)	(0.01)	(0.01)	(0.10)	(0.10)	(0.10)
Large label introducer	0.754^{***}	0.754^{***}	0.756^{***}	2.220^{***}	2.196^{***}	2.258^{***}
	(0.04)	(0.04)	(0.04)	(0.19)	(0.18)	(0.21)
Generational Technology	-0.038^{**}	-0.041^{***}	0.003	-0.835^{***}	-0.872^{***}	-0.102
	(0.01)	(0.01)	(0.02)	(0.24)	(0.25)	(0.36)
Technological Features	-0.024	-0.034^{*}	-0.020	-0.058	-0.311	-0.101
	(0.01)	(0.02)	(0.02)	(0.20)	(0.20)	(0.19)

OS Reference	0.119***	0.106***	0.109***	1.686***	1.468***	1.394***
	(0.03)	(0.03)	(0.03)	(0.37)	(0.40)	(0.36)
-powered, -enabled suffixes	-0.007	-0.002	0.005	-0.437	-0.311	-0.274
	(0.01)	(0.01)	(0.01)	(0.25)	(0.25)	(0.25)
Usage Type Reference	-0.155***	-0.136***	-0.164***	-2.034***	-1.801***	-2.030***
	(0.03)	(0.02)	(0.03)	(0.46)	(0.40)	(0.48)
Derivation label	-0.143***	-0.150***	-0.141***	-2.060***	-2.211***	-1.967***
	(0.02)	(0.02)	(0.02)	(0.29)	(0.29)	(0.29)
Adds new words to the pool	-0.021	-0.025	-0.029*	0.048	-0.038	-0.064
	(0.01)	(0.01)	(0.01)	(0.15)	(0.15)	(0.15)
Label contains a trademark	-0.138***	-0.139***	-0.139***	-2.068***	-2.145***	-2.042***
	(0.02)	(0.02)	(0.02)	(0.29)	(0.32)	(0.28)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	3751	3751	3552	3751	3751	3552
R^2	0.201	0.199	0.194			

Notes: Standard Errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Constant omitted from the table. Data range from 1998 to 2011. The dependent variable is the (log) number of times the label has been cited by a company in products' press releases. Robust standard errors.

Table 7. Results for the originality regressions

	(1) Log-link	(2) Log-link	(3) Log-link	(4) Poisson	(5) NegBin
DV: log count (1,2,3) count (4,5)					
Originality	2.698	8.100**	7.525***	66.361***	50.312***
	(1.43)	(2.35)	(1.99)	(10.79)	(8.87)
Originality^2		-14.653**	-15.429***	-186.836***	-153.164***
		(4.37)	(4.07)	(33.39)	(27.43)
Familiarity			4.147***	110.605**	50.655***
			(0.99)	(38.20)	(11.76)
Familiarity^2			-16.332***	-624.064*	-214.219***
			(3.88)	(271.69)	(53.20)
Controls	No	No	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes
N	784	784	783	783	783
R^2	0.0957	0.126	0.243		

Notes: Standard Errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Constant omitted from the table. Data range from 1998 to 2011.The dependent variable is the (log) number of times the label has been cited by a company in products' press releases. Robust standard errors. Controls are: Two-words compound, >two-words compound, Words per label, Label Creator Size, Generational Technology, Technological Features, OS Reference, -powered, -enabled suffixes, Usage Type Reference, Derivation label, Adds new words to the pool, Label contains a trademark

Table 8. Results of the originality regressions using different levels of analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	Log-link	Log-link	Log-link	NegBin	NegBin	NegBin
	Paragraph	Article in-sample	Article population	Paragraph	Article in-sample	Article population
DV: log count (1,2,3) count (4,5,6)						
Originality	7.525***	8.936***	2.229***	50.312***	37.525***	28.778***
	(1.99)	(2.04)	(0.66)	(8.87)	(4.95)	(4.43)
Originality^2	-15.429***	-14.848***	-3.079**	-153.164***	-64.097***	-46.692***
	(4.07)	(3.65)	(0.94)	(27.43)	(10.94)	(8.03)
Familiarity	4.147***	3.264***	1.483	50.655***	45.450***	45.060***
	(0.99)	(0.95)	(0.92)	(11.76)	(8.42)	(11.21)
Familiarity^2	-16.332***	-13.500***	-12.805*	-214.219***	-197.880***	-327.206***
	(3.88)	(3.49)	(6.38)	(53.20)	(34.33)	(83.04)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	783	826	1167	783	826	1167
R^2	0.243	0.262	0.311			

Notes: Standard Errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Constant omitted from the table. Data range from 1998 to 2011.The dependent variable is the (log) number of times the label has been cited by a company in products' press releases. Robust standard errors. Controls are: Two-words compound, >two-words compound, Words per label, Label Creator Size, Generational Technology, Technological Features, OS Reference, -powered, -enabled suffixes, Usage Type Reference, Derivation label, Adds new words to the pool, Label contains a trademark

Table 9. Results of regression controlling for interaction between familiarity and originality

	(1) Log-link Paragraph	(2) Log-link Article in-sample	(3) Log-link Article population	(4) Poisson Paragraph	(5) Poisson Article in-sample	(6) Poisson Article population	(7) NegBin Paragraph	(8) NegBin Article in-sample	(9) NegBin Article population
DV: log count (1,2,3); count (4,5,6,7,8,9)									
Familiarity*Originality	137.428*	115.076**	-21.766	950.816***	250.454	96.752	1049.965***	115.076**	4.933
	(55.47)	(41.51)	(11.62)	(196.13)	(151.47)	(123.31)	(140.95)	(41.51)	(85.09)
Originality	2.985	4.479	2.974**	3.130	21.871**	24.411***	-12.750	4.479	28.657***
	(2.07)	(2.30)	(0.93)	(12.07)	(7.66)	(5.19)	(6.98)	(2.30)	(4.76)
Originality^2	-7.463*	-8.091	-3.651***	-30.270	-42.400***	-47.834*	10.805	-8.091	-46.605***
	(3.53)	(4.15)	(1.07)	(29.71)	(11.80)	(19.94)	(12.56)	(4.15)	(7.86)
Familiarity	4.060***	3.425***	1.909*	36.558	42.694	49.423***	41.200***	3.425***	44.876***
	(0.97)	(0.98)	(0.82)	(27.52)	(26.18)	(12.44)	(10.80)	(0.98)	(10.99)
Familiarity^2	-26.081***	-23.149***	-10.304	-260.877	-245.912	-377.164***	-239.364***	-23.149***	-327.507***
	(7.04)	(5.89)	(6.97)	(193.03)	(180.31)	(113.12)	(50.07)	(5.89)	(84.50)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	783	826	1167	783	826	1167	783	826	1167
R^2	0.263	0.283	0.317					0.283	

Notes: Standard Errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01. Constant omitted from the table. Data range from 1998 to 2011.The dependent variable is the (log) number of times the label has been cited by a company in products' press releases. Robust standard errors. Controls are: Two-words compound, >two-words compound, Words per label, Label Creator Size, Generational Technology, Technological Features, OS Reference, -powered, -enabled suffixes, Usage Type Reference, Derivation label, Adds new words to the pool, Label contains a trademark

REFERENCES

Anderson, P., and Tushman, M. L. (1990). Technological discontinuities and dominant designs: A cyclical model of technological change. *Administrative science quarterly*, 604-633.

Berger, J. L., Addis, M. E., Reilly, E. D., Syzdek, M. R., and Green, J. D. (2012). Effects of Gender, diagnostic Labels, and causal theories on Willingness to report symptoms of depression. *Journal of Social and Clinical Psychology*, *31*(5), 439-457.

Bingham, C. B., and Kahl, S. J. (2013). The process of schema emergence: Assimilation, deconstruction, unitization and the plurality of analogies. *Academy of Management Journal*, *56*(1), 14-34.

Bowker, G. C., and Star, S. L. (2000). Invisible mediators of action: Classification and the ubiquity of standards. *Mind, Culture, and Activity*, *7*(1-2), 147-163.

Fleischer, A. (2009). Ambiguity and the equity of rating systems: United States brokerage firms, 1995–2000. *Administrative Science Quarterly*, *54*(4), 555-574.

Downing, P. (1977). On the creation and use of English compound nouns. *Language*, 810-842.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management science*, *47*(1), 117-132.

Fleming, L., Mingo, S., and Chen, D. (2007). Collaborative Brokerage, Generative Creativity, and Creative Success. *Administrative Science Quarterly*, 52(3): 443–475.

Granqvist, N., Grodal, S., and Woolley, J. L. (2013). Hedging Your Bets: Explaining Executives' Market Labeling Strategies in Nanotechnology. *Organization Science*, *24*(2), 395-413.

Grodal, S., and Thoma, G. (2014). Cross-pollination in science and technology: concept mobility in the nanobiotechnology field. *Annales of Economics and Statistics,* Forthcoming.

Grodal, S., Gotsopoulos, A., and Suarez, F. (2014). The co-evolution of categories and designs during industry emergence. *Working paper.*

Hannan, M. T., Polos, L., & Carroll, G. R. (2007). *Logics of organization theory: Audiences, codes and ecologies*. Princeton and Oxford: Princeton University Press.

Hargadon, A. B., and Douglas, Y. (2001). When innovations meet institutions: Edison and the design of the electric light. *Administrative science quarterly*, *46*(3), 476-501.

Hsu, G. (2006). Evaluative schemas and the attention of critics in the US film industry. *Industrial and Corporate Change*, *15*(3), 467-496.

Hsu, G. and Grodal, S. (2013): “Category Taken-for-grantedness as a Strategic Opportunity: The Case of Light Cigarettes, 1964-1993", *Working Paper.*

Hsu, G., Hannan, M. T., and Koçak, Ö. (2009). Multiple category memberships in markets: An integrative theory and two empirical tests. *American Sociological Review*, *74*(1), 150-169.

Kaplan, S., and Tripsas, M. (2008). Thinking about technology: Applying a cognitive lens to technical change. *Research Policy*, *37*(5), 790-805.

Kennedy, M. T. (2008). Getting counted: Markets, media, and reality. *American Sociological Review*, *73*(2), 270-295.

Kennedy, M. T., and Fiss, P. C. (2013). An ontological turn in categories research: from standards of legitimacy to evidence of actuality. *Journal of Management Studies*, *50*(6), 1138-1154.

Lieber, R. (1983). Argument linking and compounds in English. *Linguistic inquiry*, 251-285.

Lounsbury, M. (2001). Institutional sources of practice variation: Staffing college and university recycling programs. *Administrative Science Quarterly*, *46*(1), 29-56.

Lounsbury, M., and Ventresca, M. (2003). The new structuralism in organizational theory. *Organization Science*, 10(3), 457-480.

Lounsbury, M., Ventresca, M., and Hirsch, P. M. (2003). Social movements, field frames and industry emergence: a cultural–political perspective on US recycling. *Socio-Economic Review*, *1*(1), 71-104.

Navis, C., and Glynn, M. A. (2010). How new market categories emerge: Temporal dynamics of legitimacy, identity, and entrepreneurship in satellite radio, 1990–2005. *Administrative Science Quarterly*, *55*(3), 439-471.

Negro, G., Hannan, M. T., and Rao, H. (2011). Category reinterpretation and defection: Modernism and tradition in Italian winemaking. *Organization Science*, *22*(6), 1449-1463.

Orlikowski, W. J., and Gash, D. C. (1994). Technological frames: making sense of information technology in organizations. *ACM Transactions on Information Systems (TOIS)*, *12*(2), 174-207.

Peirce, C. S. 1931. *Collected papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press

Pieters, R., Warlop, L., Wedel, M., (2002) Breaking Through the Clutter: Benefits of Advertisement Originality and Familiarity for Brand Attention and Memory. *Management Science* 48(6):765-781

Pontikes, E. G. (2012). Two Sides of the Same Coin How Ambiguous Classification Affects Multiple Audiences’ Evaluations. *Administrative Science Quarterly*, *57*(1), 81-118.

Rosa, J. A., Porac, J. F., Runser-Spanjol, J., and Saxon, M. S. (1999). Sociocognitive Dynamics in a Product Market. *Journal of Marketing*, *63*(4).

Ruef, M., and Patterson, K. (2009). Credit and classification: The impact of industry boundaries in nineteenth-century America. *Administrative Science Quarterly*, *54*(3), 486-520.

Santos, F. M., and Eisenhardt, K. M. (2009). Constructing markets and shaping boundaries: Entrepreneurial power in nascent fields. *Academy of Management Journal*, *52*(4), 643-671.

Schumpeter, J. A. (1939). *Business cycles* (Vol. 1, pp. 161-74). New York: McGraw-Hill.

Seidel, V. P., and O'Mahony, S. (2014). Managing the Repertoire: Stories, Metaphors, Prototypes, and Concept Coherence in Product Innovation. *Organization Science*.

Suarez, F. F., Grodal, S., and Gotsopoulos, A. (2013). Perfect Timing? Dominant Category, Dominant Design, and the Window of Opportunity for Firm Entry. *Strategic Management Journal*.

Utterback, J. M. (1996). *Mastering the dynamics of innovation*. Harvard Business Press.

Wry, T., Lounsbury, M., and Jennings, P. D. (2013). Hybrid Vigor: Securing Venture Capital by Spanning Categories in Nanotechnology. *Academy of Management Journal*, amj-2011.

Yamauchi, T., and Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and language*, *39*(1), 124-148.

Zuckerman, E. W. (1999). The categorical imperative: Securities analysts and the illegitimacy discount. *American journal of sociology*, *104*(5), 1398-1438.

Zuckerman, E. W. (2000). Focusing the corporate product: Securities analysts and de-diversification. *Administrative Science Quarterly*, 45: 591-619.