1-1-2016

Relatedness, National Boarders, Perceptions of Firms and the Value of Their innovations

Adam Castor
University of Pennsylvania, acastor@wharton.upenn.edu

Relatedness, National Boarders, Perceptions of Firms and the Value of Their innovations

Abstract
The main goal of this dissertation is to better understand how external corporate stakeholder perceptions of relatedness affect important outcomes for companies. In pursuit of this goal, I apply the lens of category studies. Categories not only help audiences to distinguish between members of different categories, they also convey patterns of relatedness. In turn, this may have implications for understanding how audiences search, what they attend to, and how the members are ultimately valued.

In the first chapter, I apply incites from social psychology to show how the nationality of audience members affects the way that they cognitively group objects into similar categories. I find that the geographic location of stock market analysts affect the degree to which they will revise their earnings estimates for a given company in the wake of an earnings miss by another firm in the same industry. Foreign analysts revise their earnings estimates downward more so than do local analysts, suggesting that foreign analysts ascribe the earnings miss more broadly and tend to lump companies located in the same country into larger groups than do local analysts.

In the second chapter, I demonstrate that the structure of inter-category relationships can have consequential effects for the members of a focal category. Leveraging an experimental-like design, I study the outcomes of nanotechnology patents and the pattern of forward citations across multiple patent jurisdictions. I find that members of technology categories with many close category 'neighbors' are more broadly cited than members of categories with few category 'neighbors.' My findings highlight how category embeddedness and category system structure affect the outcomes of category members as well as the role that classification plays in the valuation of innovation.

In the third chapter, I propose a novel and dynamic measure of corporate similarity that is constructed from the two-mode analyst and company coverage network. The approach creates a fine-grained continuous measure of company similarity that can be used as an alternative or supplement to existing static industry classification systems. I demonstrate the value of this new measure in the context of predicting financial market responses to merger and acquisition deals.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Management

First Advisor
Witold Henisz

Keywords
Categories, Classification, Corporate Strategy, Innovation, Mergers & Acquisitions, Organization Theory

Subject Categories
Business Administration, Management, and Operations | Management Sciences and Quantitative Methods | Social Psychology | Sociology

RELATEDNESS, NATIONAL BOARDERS, PERCEPTIONS
OF FIRMS AND THE VALUE OF THEIR INNOVATIONS

Adam R. Castor

A DISSERTATION

in

Management

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

____

Witold Henisz, Deloitte & Touche Professor of Management

Graduate Group Chairperson

____

Catherine Schrand, John C. Hower Professor; Professor of Accounting

Dissertation Committee

Mauro Guillén, Dr. Felix Zandman Professor, Professor of International Management

Tyler Wry, Assistant Professor of Management

Adam Cobb, Assistant Professor of Management

RELATEDNESS, NATIONAL BOARDERS, PERCEPTIONS OF FIRMS AND THE VALUE OF THEIR INNOVATIONS

DEDICATION

To my grandmother, Carmen Castor, who lost her battle to cancer during my time at Wharton. She didn't have the opportunity to go to college and proclaimed that she was 'terrible with fractions,' which was sometimes apparent when she scaled down recipes. But, she knew the value of higher education and couldn't be more proud of me to be the first in my family to earn a PhD.

ACKNOWLEDGMENT

I would like to especially thank my amazing advisors for their always helpful guidance, feedback and support. I am deeply indebted to Vit, Tyler, Mauro, and Adam who were always incredibly supportive and had an unwavering faith in me. I was incredibly fortunate to work with, and be trained by, such wonderful researchers, academics, and people, who were also a major force in nurturing and shaping my scholarly identity. I consider myself very lucky to have them as both colleagues and friends. I would also like to thank other faculty members including Valery Yakubovich, Ivan Barankay, Lori Rosenkopf, Harbir Singh, Samir Nurmohamed, and the late Johannes Pennings for generously giving their time, help, feedback, and emotional support during this arduous journey.

It was clear almost from the first day of the program that I couldn’t have started at Wharton with a better cohort of fellow students. JR Keller, Justin Berg, Shefali Patel, and Patia McGrath were not only strong catalysts in advancing my learning and development as a scholar, they still continue to inspire me to reach new heights. I also greatly benefitted by being surrounded by many other talented and generous colleagues and friends including Tucker McGrimmon, Daniel Albert, Kate Odziemkowska, Sarath Balachandran, Tracy Anderson, Shinjae Won, Luis Ballesteros, Ram Ranganathan, Nick Lobuglio, Andrew Boysen, Julianna Pillemer, Basima Tewfik, Danielle Tussing, and Phebo Wibbens, among many other future scholars.

Finally, I would like to thank my parents, Robert and Ruth, my sister Ashely, and my grandfather, Robert. Despite being unwitting passengers in this academic journey of mine, they were nonetheless always there to support me and help me move forward.

ABSTRACT

RELATEDNESS, NATIONAL BOARDERS, PERCEPTIONS OF FIRMS AND THE VALUE OF THEIR INNOVATIONS

Adam R. Castor

Witold Henisz

The main goal of this dissertation is to better understand how external corporate stakeholder perceptions of relatedness affect important outcomes for companies. In pursuit of this goal, I apply the lens of category studies. Categories not only help audiences to distinguish between members of different categories, they also convey patterns of relatedness. In turn, this may have implications for understanding how audiences search, what they attend to, and how the members are ultimately valued.

In the first chapter, I apply incites from social psychology to show how the nationality of audience members affects the way that they cognitively group objects into similar categories. I find that the geographic location of stock market analysts affect the degree to which they will revise their earnings estimates for a given company in the wake of an earnings miss by another firm in the same industry. Foreign analysts revise their earnings estimates downward more so than do local analysts, suggesting that foreign analysts ascribe the earnings miss more broadly and tend to lump companies located in the same country into larger groups than do local analysts.

In the second chapter, I demonstrate that the structure of inter-category relationships can have consequential effects for the members of a focal category. Leveraging an experimental-like design, I study the outcomes of nanotechnology patents and the pattern of forward citations across multiple patent jurisdictions. I find that members of technology categories with many close category 'neighbors' are more broadly cited than members of categories with few category 'neighbors.' My findings highlight how category embeddedness and category system structure affect the outcomes of category members as well as the role that classification plays in the valuation of innovation.

In the third chapter, I propose a novel and dynamic measure of corporate similarity that is constructed from the two-mode analyst and company coverage network. The approach creates a fine-grained continuous measure of company similarity that can be used as an alternative or supplement to existing static industry classification systems. I demonstrate the value of this new measure in the context of predicting financial market responses to merger and acquisition deals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1: INTRODUCTION

It has been well established that classification systems affect people's perceptions and behaviors. What is less well understood is how classification systems might differ across audiences and how these differences affect audience perceptions and behaviors. Investigating these differences in perceptions and their effects on important firm outcomes such as corporate earnings forecasts and the valuation of intellectual property is the primary goal of my dissertation. Of particular interest is the important role that national boarders play in these processes.

The primary contributions in this dissertation are made to the category studies literature. Though Porac et al.'s (1989) seminal paper, pulling from cognitive psychology, was the first to address categories in the management literature, more recent work in category studies has largely tackled categories through a sociological lens. The dominant focus on sociological approaches, which has generated some very powerful and important findings, has presented a huge opportunity to augment these contributions by integrating more work taking a micro approach as highlighted by Vergne & Wry (2014). This is a nascent stream of work that I buttress and extend primarily by drawing on and applying important incites and ideas not only from social psychology, but also from other management areas. In general terms, the major concepts that I import are distance / similarity and structure. While these concepts are not entirely new to the category studies or management literatures, the novelty is where they are used and the way that they are used.

A major underlying theme throughout this dissertation is the importance of categorization multiplexity: companies or objects can be classified along different dimensions and as a consequence, individuals, or different groups of individuals, may categorize companies and their outputs in different ways. This can be applied to both stakeholder groups and even individuals within a given stakeholder group. For instance, for the former, the implication is that different stakeholder groups might group companies in different ways and this may affect how they perceive and process new events, such as the potential synergies of a merger or the ascription of new informational shaping patterns of information spillovers.

The importance of distance and the different dimensions of distance, is largely imported from the international business literature's focus on cross-national distance. Hymer (1960) first highlighted the importance of cross-national distance in shaping the internationalization of the firm via the 'liability of newness.' This was then built upon by many others who, with an emphasis on the multidimensionality of distance, indirectly (e.g., Johanson & Vahlne, 1977) and directly (Dunning, 1993; Hofstede, 1980; Schwartz, 1992; 1994; Ghemawat, 2001; House et al., 2004; Berry, Guillén, & Zhou, 2010) identified different relevant dimensions of cross-national distance. The importance of many of these different dimensions of cross-national distance was also well established and include political (Delios & Henisz, 2000), economic (Campa & Guillén, 1999), financial (Capron & Guillén, 2009), and cultural (Kogut & Singh, 1988) distance among others.

Surprisingly, distance / similarity among categories (paralleling distance among countries) has received relatively little attention in the category studies literature, with very few exceptions - e.g., Kovács & Hannan (2015) who use a proxy for distance solely to derive extant measures like contrast and niche-width. The stream that most closely touches upon distance in this literature is the work on partial membership, or grade of membership, GoM, (Hannan et al., 2007; Hannan, 2010), where the degree of membership is based on the degree of similarity between objects and the focal category prototype. Likewise, category contrast, an amalgam of GoM's, is a category-level measure that indirectly indicates that a focal category is related to others (e.g., Kovacs & Hannan, 2010). While GoM and category contrast has led to many important insights, especially with respect to penalties from category spanning, what is less understood is the importance of similarity / distance among category dyads as well as the importance of multidimensionality in the measure of distance. In this light, I borrow from international business and investigate categories and category systems with a particular focus on multidimensionality and distance among categories.

For the category studies literature, my dissertation builds upon the social categorization literature by relaxing the assumptions that social categories are largely fixed (i.e., static, resistant to change, and uniform across audiences) and allowing social categories, and the underlying institutional logics (Thornton, Ocasio, & Lounsbury, 2012) that they are based on, to be fluid and potentially vary across individuals. This approach follows in the less-studied branch of category studies that follows in the tradition of Rosch (1978), Porac and colleagues (1989, 1995), and more recently Lounsbury, Wry, and others (e.g., Wry &

Lounsbury, 2013; Vergne & Wry, 2014). By relaxing the usual assumptions, I am able to investigate variance in cognitively employed schema across individual audience members of the same audience to demonstrate the importance of classification structure and embeddedness within that structure.

This is done by building on the institutional logics and attention (Ocasio, 1999) literatures. Category systems and cognitive schema are largely based on social construction (Berger & Luckman, 1966) are invariably linked (Douglas, 1986; Mohr & Duquenne, 1997) and can be viewed as a way that institutions constrain action (Holm, 1995). However, “institutional logics are not static structures impervious to change” (Thornton, Ocasio, Lounsbury, 2012, p. 77), and, via cognitive schemas, may even enable action (Thornton, 2004). Institutional logics drive the attention of stakeholder groups (Thornton & Ocasio, 1999; Thornton, 2001; Thornton, 2002) and even changes in stakeholder groups over time (Glynn Lounsbury, 2005; Rao et al., 2005). Moreover, “If no aspects of highly accessible institutional logics are viewed as applicable or relevant, individuals may rely on other available institutional logics to activate knowledge and information for further information processing.” (Thornton, Ocasio, Lounsbury, 2012, p. 84). But, I envision that there are situations where institutional logics and their associated cognitive schema may not be fine-grained enough to fully suit the task at hand (e.g., technology classification systems only rudimentarily organizing relatedness across technology classes in examiner search or industry classification systems like SIC that are too broad when assessing what rivals are the most relevant). In such situations, I expect actors to augment these classification systems with their own personal heuristics, experiences, and biases. For instance, individual stock market analysts subject to out-group homogeneity / in-group heterogeneity biases based on their own nationality or location when assessing whether new information by other companies should be ascribed to a focal firm when predicting future performance.

Aside from theoretical contributions, this dissertation also yields an empirical contribution. Namely, I leverage a novel and powerful setting, differences in classification systems across three countries, each with their own institutional logics for classifying technology. This yields a quasi-experiment (each country has a different technology classification system) to investigate how the characteristics of a classification system affect cognition, attention, and corporate outcomes. Additionally, I introduce a dynamic approach to measuring company similarity that reflects the mental models of investors, representing the institutional logics that they employ, that could prove to be a powerful tool for other researchers that pursue topics in the area of social category dynamics as well as other fields of management such as those in corporate strategy who study the scope of the firm.

This dissertation also makes contributions to the international business literature. Most notably, I demonstrate how mental models of relatedness differ across external stakeholders located in different countries and the implications this has for multinational firms. In particular, I demonstrate that mental models of relatedness differ due to 1) institutional differences in classification systems, and 2) the geographic location of audience members and their national identities in combination with social psychological processes. Most notably, I demonstrate how mental models differ systematically among individual members of the same stakeholder group. This means that individual members of the same stakeholder group might react differently to the same newly acquired information or in the same situations. My findings also highlight the importance of binary country membership, as opposed to cross-national distance, in shaping the perceptions of multinationals and the ascription of newly acquired information. Finally, the fine-grained analyst-based company similarity measure and industry groups that I introduce could greatly benefit multinational management researchers and constitute an empirical contribution to researchers in this area in two major ways: 1) by allowing researchers to directly compare companies that operate in different countries without the challenges and drawbacks of needing to map classification systems from one country or region into another and 2) as a potentially better measure of similarity among multinationals and companies that operate in multiple industries.

I demonstrate these contributions throughout the dissertation though three separate chapters (i.e., chapters 2 through 4). In chapter 2, I investigate how a seemingly homogeneous stakeholder group might systematically differ in their mental models of companies. Specifically, I investigate how one audience member characteristic, their geographic location and national identity, affects sell-side stock market analysts' industry classification schemes and how this translates into different reactions to new information a la informational spillovers and company earnings projections in an international sample. I find that the mental models of locals differs from members of the same stakeholder group located abroad. This has important implications for how the actions and events of multinationals are perceived by stakeholders in different countries.

In chapter 3, I investigate technology classification in three countries / patent jurisdictions: the U.S., Germany, and Japan. Each of these countries employ different classification systems in their respective patent offices. Leveraging an experiment-like design, – i.e., by studying how the exact same inventions or technologies are situated in each jurisdiction - I examine the outcomes of nanotechnology patents and the pattern of forward citations within these three patent jurisdictions. In essence, each jurisdiction acts as a sort of separate experimental condition or setting and I look at within-invention variation across systems. Ultimately, I find that the technology class that a given invention is assigned to in each system in combination with the classification structure of that system directly affects the number of forward citations received by the associated patent in that jurisdiction. The findings highlight how category embeddedness and category system structure affect the outcomes of category members as well as the role that classification plays in the valuation of innovation. This has implications for understanding how audiences search, what they attend to, and how category members are ultimately valued.

In chapter 4, I propose a novel and dynamic similarity-based measure of companies that is constructed from the two-mode analyst and company coverage network. I call this approach Analyst-Based Similarity. This approach creates a fine-grained continuous measure of company similarity that can be used as an alternative or supplement to existing static industry classification systems. Theoretically, this paper also makes a small contribution to the social categorization literature by deviating from Prototype Theory and reintroducing aspects of Wittgenstein's Family Resemblance logic that have largely been overlooked in the extant literature. Empirically, the approach has a number of benefits over existing and widely used industry classification systems (e.g., SIC and NAICS). One major benefit is the ability to measure similarity between companies dyadically, both within and across industries. Another benefit to this approach is that it is dynamic and captures even subtle underlying changes in industries or perceptions of corporate similarity over time. Aside from bringing forth examples of how this approach differs from SIC and NAICS industry classification, I demonstrate the value of this approach through its ability to predict market reactions to M&A deal announcements.

CHAPTER 2: INTERNATIONAL DIFFERENCES IN CATEGORIZATION AND INFORMATION SPILLOVERS

INTRODUCTION

On Jan. 14th, 2010, Spanish bank Banco Espanol De Credito reported earnings of €1.135 per share, which was €0.10 lower than analyst consensus expectations. A few days later, analysts Francisco Riquel, located in Spain, and Sandra Neumann, located in Germany, both lowered their earnings projections for rival Spanish bank, Banco Santander by €0.03 and €0.17 (from €1.24 and €1.23), respectively. But why the stark difference in the adjustments between the two analysts? Why did the German analyst lower her earnings forecast for Banco Santander much more dramatically when compared to the Spanish analyst? Did the different geographic locations of the two analysts lead to the large difference in their forecast revisions? This is the main question that I hope to investigate in this paper.

It is no secret that stock market analysts exhibit a number of biases in their recommendations and earnings estimates. For instance, Barber, Lehavy, McNichols, & Trueman (2001) found that only 3 percent of all the recommendations covered in their sample between 1985 and 1996 are "sell" recommendations and Chan, Karceski, & Lakonishok (2003) found that analysts overestimated earnings growth rates by over 60% between 1982 and 1998 (median estimates of 14.5% vs. 9.0% actual). While this over-optimism bias might stem from principle agent problems and skewed incentives for analysts, it may also have cognitive roots (Michaely & Womack, 2005). These findings hint that analysts might suffer from other biases as well.

When evaluating many different types of products, both consumers and critics are often at least somewhat cognizant of the nationalities of the corporations whose products they are considering. In some industries, the perceived quality of the product itself is heavily derived from the location of production or the nationality of the producer – for instance, we generally perceive German and Japanese cars to be of higher quality than those from other countries, as we similarly do for cigars from Cuba, or caviar from Russia. Oftentimes, the corporate brands for these products are closely tied to their corporate national identities (i.e., home country of origin). This can be both a benefit and a curse. On one hand, multinational subsidiaries can benefit greatly from their corporate national identities through the positive cognitive associations that these identities evoke in critics, customers, and other external stakeholders in foreign markets. On the other hand, corporate national identities may also evoke cognitive associations which negatively affect the perceived quality or value of the product in the mind of these same stakeholders. Moreover, national identities are one natural dimension by which critics, consumers, and potentially other stakeholders, may cognitively group corporations.

Finally, the positive and negative associations that corporate nationalities evoke in external stakeholders can change over time. In particular, isolated events in a given country can quickly change the perceptions of products, brands, and of the multinationals themselves who are tied to the given country. It is the negative effects of these events that will be the focus of this paper. While it is obvious that corporate earnings misses by one company can create negative externalities for other companies in an industry, I hypothesize that the degree of these negative externalities will be more pronounced for companies that share the same national identity as the announcing company compared to those with different corporate nationalities.

More importantly, I also hypothesize that it isn't just the nationality of the company being evaluated that affects how the company is perceived and evaluated. The relationship between the countries of the company and that of analysts also matters. In this paper I argue that the relative location of the analyst with respect to the target company being covered significantly affects the group of peer companies that the analyst links to the target company. When new events or new information associated with other event firms emerges, this grouping plays a key role in determining if this new information is ascribed and applied to the target company. If this new information is tied to an event company (i.e., the company that missed earnings) that the analyst groups together with the target company, the analyst will ascribe this new information to the target company. Conversely, if the analyst does not group the event and target companies together, this new information will not be ascribed to the target company. Thus, if analysts differ in the way that they group the same set of companies, they will sometimes respond differently to new information and events.

THEORY & HYPOTHESES

Home Country & Cross-National Distance

In the field of multinational management, a large literature has developed with the purpose of better understanding how a multinational's home country affects various dimensions of the multinational's foreign subsidiaries. Schollhammer & Nigh (1984) lay down a complex, yet encompassing theoretical model that not only includes the local political environments of both the multinational's home and host countries, it also takes into account the political environments of third countries (i.e., countries that are neither the home or host country) along with the effects and importance of the political relationships among all of these countries (i.e., home, host, and third countries). Schollhammer & Nigh argue that both intra-country and inter-country political events can alter the political landscape and ultimately affect foreign multinationals at home and abroad. The literature in the field has generally addressed separate parts of Schollhammer & Nigh's model and can be broken down into three major streams. The first stream is primarily concerned with the political environment of the multinational's host country, paying particular attention to host country political risk. The second stream is primarily concerned with the institutional and political environment of the multinational's home country. The third stream is primarily concerned with the differences between the multinational's home and host countries on both political and non-political dimensions like culture. In this paper, I am primarily concerned with the third stream.

This third stream of literature focuses heavily on the differences between a multinationals home and (potential) host countries. These different measures have been employed to show that cross-national distance plays an important role when it comes to the behavior of multinationals operating abroad. For instance, Kogut & Singh (1988) found that increased cultural distance between a multinational's home and host countries increased the proportion of joint ventures as the entry mode used by multinational entrants. Similarly, Barkema, Bell, & Pennings (1996) found that cultural distance decreased the longevity of joint ventures when used as an entry mode.

Following in the footsteps of these scholars, this paper aims to go one step further and investigate if cross-national distance plays a role in how non-political events may be interpreted and perceived differently by actors in different countries, subsequently leading to different responses or behaviors among these different actors. Additionally, this paper aims to add to the multinational literature by demonstrating how new information concerning one multinational might negatively impact other multinationals that are based in the same country. This paper takes the prescribed empirical approach of evaluating the impact of events on multinational firms (Kobrin, 1979); while Kobrin advocates looking at the impact of political events, I instead focus on non-political events. In the next section I lay out the hypothesized underlying mechanisms which are based on the cognitive processes of categorization.

Categorization and Spillovers via Home Country

Zerubavel (1996) describes the innate cognitive processes of lumping and splitting which people naturally do to interpret and make sense of the unmanageable continuous world, filled with infinitesimal shades of grey, by translating it into a more manageable discrete world built on socially constructed categories. It is this translated discrete world, filled with "islands of meaning" (Zerubavel, 1991), that people experience. Lumping describes the process by which people mentally cluster similar things together into groups. Splitting describes the process by which people mentally separate and highlight differences among objects that are members of different groups. The end result of these two processes is a categorization schema where the categories described by this schema represent collections of members with shared attributes. Through the process of lumping and splitting we tend to cognitively overemphasize the similarity of items in the same mental group while simultaneously overemphasizing the differences between items in separate mental groups (Zerubavel, 1991).

In the context of multinationals, this means that we may cognitively perceive multinationals from the same home country to be more similar than they actually are, and likewise perceive multinationals with different home countries to be less similar than they actually are. Mullainathan, Schwartzstein, & Schleifer (2008) refer to this phenomenon more generally as "coarse thinking" and proposed a model that describes one form of such thinking, i.e. transference, which operates on "co-categorized" situations – situations where two events or objects that have been mentally lumped into the same cognitive category. Transference occurs because individuals "fail to differentiate between co-categorized situations and use one model of inference for all situations in the same category" (Mullainathan, et al., 2008). Consequently, new information that is received pertaining to the focal element in a category is similarly associated with all other elements in the same category. Scholars in other disciplines refer to this process as similarity-based reasoning, which is a direct application of the Principle of Similarity – if two objects share similarities on some dimensions then they are likely to share similarities on other dimensions (Peski, 2011). For instance, Yu, Sengul, & Lester (2008) theorized that an organizational crisis is more likely to spillover to other organizations with a similar organizational form, especially when the shared organizational form is simple and there are sharp distinctions among organizational forms (i.e., organizational forms do not blur together). Interestingly, transference may also occur even when the information being transferred is un-informative or even potentially wrong (Zhao, 2009; Fryer & Jackson, 2008).

There is some direct and indirect empirical evidence of transference. Jonnson, Greve, & Fujiwara-Greve (2009) present direct evidence that the stigma of a scandal will spread to organizations with similar organizational forms and similar organizational characteristics as the socially deviant organization. When Swedish insurance firm Skandia AB was embroiled in scandal pertaining to the misuse of company real estate, the subsidiary mutual fund providers of other insurance firms of similar size that also held real estate were penalized more through investor redemptions than those subsidiaries of dissimilar insurance firms. Similarly, Pontikes, Negro, and Rao (2011) find indirect evidence consistent with the transference hypothesis in their research on moral panics and the black-listing of communist sympathizing artists. If one film artist is black-listed, other artists who worked on the same films as the black-listed artist have a harder time finding work and are therefore subject to the negative externalities stemming from the black-listing of their former cast-mate. These negative externalities occur even if the artists never worked together again after or event at the time of the black-listing. Essentially, artists are perceived to be similar to other artists they work with, including black-listed artists, in which case they are perceived to be more likely to also have communist sympathies and thus suffer more than artists that were not in the same films. Pontikes, et al. highlight affiliation as the mechanism leading to the transmission of this stigma. The implication is that affiliation is one means of categorization thus leading to transference; artists who work together on a given film are similar and thus if one artist has a given political ideology, other artists working on the film are likely to as well.

Similarly, stock market analysts may cognitively lump multinationals together along a myriad of dimensions. In this paper, I contend that analysts, and stakeholders more broadly, may also categorize multinationals along the dimension of corporate nationality. In other words, on some level, stakeholders will naturally group and split multinationals according to their home or headquarters' countries, even if the firms are dissimilar in other respects. For instance, consumers and the media tend to lump multinationals within industries according to their corporate nationality: e.g., GM & Ford, Toyota & Honda, and BMW & Mercedes as American, Japanese, and German car manufactures, respectively). Moreover, when consumers and other external stakeholders think of a given company, they might be more inclined to cognitively link it to other companies with the same home country than to companies that are similar along other dimensions yet with different home countries. In a competitive landscape, Porac, Thomas, Wilson, Paton, & Kanfer (1995) found some evidence that competitors may be partially grouped by geographic location.

If multinationals are categorized by their home countries, coarse thinking will dictate that new information pertaining to a given company, usually via some event pertaining to that company, will tend to be transferred and ascribed to other multinationals that share the same home country. Jonnson et al. (2009) found that the negative spillovers surrounding Swedish insurance firm Skandia AB affected other Swedish insurance firms while mostly sparing their non-Swedish equivalents. I expect to see the same reactions by stock market analysts who mediate these investment markets.

> **Hypothesis 1 (H1): After a disappointing earnings announcement by an event firm, a similar target company has its earnings estimates reduced more when the target company shares the same headquarters country as the event company than if it doesn't.**

Categorization and Analyst Location

Most of the previous research on social categorization either implicitly or explicitly assumes that there is only a single dominant categorization scheme used by an audience. There are a couple of explanations why this might be the case. Rosa, Porac, Runser-Spanjol, & Saxon (1999) argue that in product markets, categorization schema evolve through the interactions of producers and consumers and eventually reflect a collective agreement among transaction partners. Agreement may also be due to the presence of market mediators that do the majority of the interpretation and evaluation labor. This can be done indirectly by mediators such as the media (Kennedy, 2005), or directly by mediators such as critics (Rao, Monin, Durand, 2005) and sell-side stock analysts (Zuckerman, 1999).

However, there is a much smaller branch of work that has addresses situations where audiences may have different categorization schema. In general, this work has partitioned larger audiences that apply different schema into separate audiences, each with a single commonly used categorization schema. For instance, Zuckerman & Kim (2003) separate the mass-market film audience and its critics from the independent film audience and its critics while Hannan, Polos, & Carroll (2007) separate firm “insiders” from firm “outsiders.” By partitioning audiences into separate sub-audiences within which all audience members are unified by a singular categorization schema, scholars can focus on the effects of individual schemas in isolation. This has allowed scholars to better identify some of the more rudimentary mechanisms and effects of social categorization. Conversely, this has taken the spotlight away from understanding why subsets of an audience may use different categorization schema and what the effects might be of their doing so. This is a gap in the literature that I hope to draw attention to and begin to bridge by looking at one way in which different audiences or sub-audiences might have categorization schema that differ in a systematic way. I do this by looking at the relative location of the perceiver or evaluator with respect to an event or object that is to be interpreted, evaluated, or cognitively processed. By location, I simply mean that the perceiver themselves fall somewhere along a dimension that is part of the categorization schema. In this context, I focus on the nationality as the dimension shared by both the evaluators and the objects they are evaluating; both stock market analysts and the companies they cover can be mapped in this space. This allows a link to be created to well-documented findings in other fields that may shed some light on social categorization processes and outcomes.

Social psychologists working in the area of social categorization have unearthed a number of findings with regard to the evaluation of members within categories depending on the relative location of the evaluator and the category member being evaluated. One major finding is the out-group homogeneity principle (Park & Rothbart, 1982) or out-group homogeneity bias. According to this principle, individuals will tend to perceive the members of the social group to which they belong, or in-group members, as being relatively heterogeneous and correspondingly perceive members of groups to which they do not belong, or out-group members, as being relatively homogeneous (Park & Rothbart, 1982; Judd & Park, 1988). In other words, out-group members are perceived to be more similar to one another while in-group members are perceived to be more diverse or less similar than one another. This out-group homogeneity bias can strongly affect how perceivers interpret new information and attribute this information to in-group or out-group members.

To illustrate this point, consider an example using two different groups of occupations, investment bankers and offshore oil rig workers. It is likely that investment bankers will feel that there are many different types of jobs within investment banking and that each entail very different things (e.g., front office, middle office, and back office jobs, etc…). Moreover, investment bankers might even separate their jobs even further (e.g., middle office jobs can be separated into risk management, treasury management, internal controls, and corporate strategy). Thus, investment bankers will view their investment banking colleagues jobs as being relatively dissimilar as they perform a variety of very different types of jobs within an investment bank. In contrast, offshore oil rig workers might focus less on the more subtle differences between individual investment banking jobs and instead focus on the larger differences between offshore oil rig workers and investment bankers more broadly. Thus, they would subsequently perceive investment bankers as being quite similar to one another. If the roles are reversed, we would find the same thing. Offshore oil rig workers would likely focus on all the different types of jobs on the oil rig (e.g., rotary drill operators, derrickhands, motorhands, roughnecks, and roustabouts) highlighting the differences among themselves, whereas investment bankers would likely focus on the differences between the two groups and overlook how different these different offshore oil rig jobs are. Thus, from the perspective of the in-group offshore oil rig worker, offshore oil rig workers would appear to be more heterogeneous than from the perspective of the out-group investment bankers and vice versa.

These differences in perception may translate into different cognitive groupings by the different groups of actors. If in-group members perceive their group to be more heterogeneous, it would follow that they would be more likely to create finer-grain cognitive categories, or sub-categories, for these in-group members. When thinking about classification schema in terms of hierarchies, one might expect that in-group members are more likely to rely on sub-classifications for their own in-group category than they would for out-group categories. Consequently, in-group members are likely to ascribe new information associated with another in-group member to a smaller subset of in-group members than would out-group members that are less likely to cognitively invoke out-group sub-categories.

> **Hypothesis 2 (H2): When an event company announces disappointing earnings, foreign analysts will reduce their earnings forecasts more so than will local analysts for a target company with the same headquarters country as the event company.**

> **Hypothesis 3 (H3): When an event company announces disappointing earnings, foreign analysts will reduce their earnings forecasts to the same degree as local analysts for a target company with a different headquarters country than the event company.**

While the out-group homogeneity principle focuses on systematic differences between in-group and out-group audiences, there might also be systematic differences among different out-group audiences (who differ in their relative location to the event or evaluation target). If social categories have multiple levels of hierarchy, the same out-group homogeneity bias might also create systematic differences among perceivers embedded in different places throughout the social category hierarchy.

Suppose that the measure of closeness of two actors in a hierarchy is measured by the lowest level of the hierarchy where both actors are members of the same sub-category. Relative to some focal sub-category member, perceivers that are out-group members at the sub-category level, yet in-group members at higher levels of the categorization scheme, are closer to the objects being evaluated and will more likely cognitively highlight the differences among these objects than would evaluators that are further from these objects (i.e., out-group members at similar or higher levels of the hierarchy). To put this differently, perceivers might view their in-group as being higher in the hierarchy and this means the out-group homogeneity bias will operate in the same way, just at a broader level.

To make this more concrete, consider the Bureau of Labor Statistics' Standard Occupational Classification system (SOC). The SOC has four levels of hierarchy or aggregation, which they refer to from broadest to narrowest as major group (first two code digits), minor group (third code digit), broad occupation (fourth and fifth code digits), and detailed occupation (sixth digit). According to this classification system, the occupation "Rotary Drill Operators" (47-5-01-2) is in the same major group as "Electricians" (47-2-11-1), but not in the same minor group or broad occupation group. The occupation "Derrick Operators" (47-5-01-1), on the other hand, is in the same broad occupation as Rotary Drill Operators. Thus, according to this classification scheme, "Rotary Drill Operators" is closer to "Derrick Operators" than it is to "Electricians." Thus, Electricians are more likely to cognitively lump together Rotary Drill Operators and Derrick Operators than would Rotary Drill Operators or Derrick Operators themselves.

In essence, people will tend to have cognitive categorization schema that have finer or smaller categories that include objects or actors which are proximate to their own location on given dimensions and broader and more encompassing categories that involve objects or actors which are more distant. Additionally, this distance-based lumping effect suggested by the out-group homogeneity bias might be strengthened through differences in exposure among actors. Fryer & Jackson (2008) found that subjects tend to more coarsely categorize objects that they come across less frequently. Thus, if agents have less exposure to actors that are socially more distant, they will tend to lump them into broader categories.

Finally, if the distance between the perceiver and the event (or target agent of the event) affects how narrowly or broadly the target actor is lumped together with other actors, then this will translate to differences in how narrowly or broadly this information will be ascribed to other actors. The more distant the perceiver is from the focal actors, the broader the information pertaining to that event will be ascribed to these actors.

> **Hypothesis 4 (H4): The distance between a foreign analyst and a target company will be positively related to the degree that the foreign analyst reduces the earnings forecasts for the target company. The greater the distance, the larger the reduction in the earnings forecast.**

DATA & METHODS

I test my hypotheses in the empirical setting of international stock market investing of publicly traded firms on global stock exchanges. While the main audience in this setting is stock market investors, I instead focus on the sell-side stock market analysts, experts who mediate the stock market and directly shape the opinions and actions of these investors.

Data Sources

The primary source of data used in this study come from Thomson Reuters' Institutional Brokers Estimate System data files (IBES). IBES covers over 70,000 publicly traded companies from more than 2700 global stock exchanges spanning 90 countries over 5 continents North America, Europe, the Middle East, Latin America, and Asia. This data includes detailed historic individual analyst forecasts of company earnings and other important financial metrics including buy-sell-hold recommendations, future stock target prices, basic contact information for analysts, and the date and content of publicly disclosed earnings announcements.

This study period covers all earnings reports between Jan 1st, 2001 and Aug. 31st, 2011. The beginning of 2001 was chosen for one main reason. On Aug. 10th, 2000, the US Securities and Exchange Commission (SEC) instituted the Selective Disclosure and Insider Trading Rule which included Regulation Fair Disclosure (Reg FD) that took effect on Oct. 23, 2000[1]. Reg FD "requires that when an issuer makes an intentional disclosure of material nonpublic information to a person covered by the regulation [e.g. securities market professionals & institutional traders], it must do so in a manner that provides general public disclosure" (SEC, 2000). In other words, Reg FD prohibits publicly traded firms from selectively disclosing information to anyone that would trade on the basis of that info. This regulation reduced the informational asymmetries among analysts by ensuring that all analysts have equal access to the same corporate information at the exact same time. Prior to this regulation, corporate executives might disclose private material information selectively to certain analysts such as those with strong

---

[1] It should be noted that Reg FD only pertains to companies that trade on US stock market exchanges, making this law not applicable to many of the companies covered in the analysts. However, many international companies, particularly those that are cross-listed on multiple exchanges, voluntarily adopted the same rules (Crawley, Ke, &Yu, 2009). While this doesn't rectify the potential issue that only part of the sample is covered by Red FD, only using data after Red FD nonetheless minimizes private information for at least part of the sample.

industry reputations, or those who's longstanding coverage of the company's stock afforded them special corporate access (Fleischer & Baum, 2010). In this study, I am only using data after the enactment of Reg FD. This better ensures that all analysts in the sample will be acting on the same a priori information and will all received the exact same corporate information at the exact same time. While Reg FD went into effect in late Aug. of 2000, starting the data window in Jan of 2001 yields a 4-month period between the enactment of the regulation and the beginning of the collection of analyst reports. This allows the brokerage industry time to adapt to this new regulation before earnings reports are included in the study (Mohamram and Sunder, 2006). The end of the study period, Aug. 31st, 2011 is the last day that analyst report data was available.

Analyst-Based Industry Groups

To identify similar companies in the study, I constructed analyst-based industry groups which are based on observable patterns of overlapping analyst coverage by grouping together companies that share coverage from the same analysts. I opted to use analyst-based industries because existing industry classifications such as the SIC system or NAICS may be problematic in information spillover studies due to issues regarding such things like the level of aggregation of some industries, classification systems heavily based on the supply side and not demand side, and the diversification strategies adopted by many firms, issues highlighted by Foster (1981) and Guenther & Rosman (1994). Similarly, these classification systems may be too broad or miss the mark when identifying companies which are similar (Clarke, 1989). In contrast, analyst-based approaches are more likely to yield groupings of similar companies; investment banks staff their analyst coverage in a manner that minimizes information acquisition costs to analysts by assigning analysts to cover sets of similar firms (Ramnath, 2002). Moreover, analyst-based approaches to industry groupings yield additional advantages in informational spillover studies since information transfers should be more likely when analysts cover all of the firms within an industry (Ramanth, 2002).

To construct the analyst-based industry groups used in this study, I followed a methodology that mirrors Ramnath (2002). I applied this methodology to analyst coverage patterns in 2006, which corresponds to the middle of the study period. To construct these industry groups, I followed two simple rules: 1) all companies in an industry group must have at least three analysts in common and 2) each industry group must have at least 3 companies. Finally, it should be noted that while inclusion into an industry group requires that the company has at least three analysts in common with all of the other companies in that industry group, all analysts are included in the analyzed data, not just the analysts that are all common to all of the companies. For instance, Banco Espanol De Cridito and Banco Santander are in the same industry group with Sabadell, Bankinter, Banco Popular, Banco Espirito, Banco Pastor, and Banco Bilbao Vizcaya Argentaria. It should be noted that these two rules in combination did not make it possible to assign all the companies in the sample into industry groups[2].

[2] Of the 7,656 companies in the original sample, 3,421 companies (~45%) could not be assigned to industry groups and were excluded from the analysis.

Disappointing Earnings Announcement Events

The purpose of this study is to investigate how earnings announcements from companies that are the first to report earnings in a given industry (i.e., the event company) affect analyst earnings projections for companies that later report earnings for the same fiscal period (i.e., target companies). For each fiscal quarter during the study period, the first announcer and subsequent announcers are determined by the sequence of earnings announcement dates for the companies in each industry group. The event company is simply the company that is the first to report earnings for that fiscal quarter and the date that they report earnings is the event date[3]. All of the other companies in the industry group, which by definition report earnings after the first announcement date, are all considered target companies. In the analysis that follows, I only look at earnings projection changes for the current reporting quarter for target companies. Only companies that have earnings announcement information (e.g., information on the date of their own earnings announcement) in a given quarter will be included in the analysis.

Only industry group fiscal quarters with first announcement dates that occur between Jan 1st, 2002 and July 31st, 2011 are included in the analysis. Jan 1st, 2002 correspond to one year after the beginning of the data range, allowing enough time to include all pre-event analyst earnings projections. July 31st, 2011 corresponds to one month before the end of

[3] If multiple companies report earnings on the same first announcement date, then the observations for the industry group for the given quarter are excluded from the analysis.

the data range, which is sufficiently large enough to gather all relevant post-event earnings projection changes.

Dependent variables

The main dependent variable used in the analysis, *Scaled Change in EPS Estimate*, is a measure of the relative change in individual analysts' projected earnings per share (EPS) estimate for the target companies for the current fiscal year (i.e., the next fiscal year to be reported by the company for a given report) divided by the share price of the underlying company's stock just before the event date. The EPS estimate change is calculated by subtracting the analyst's most recent projection before the event date from the first projection after the event date. This EPS estimate change is divided by the stock price in order to scale EPS numbers and make them comparable across companies. This approach was used by Ramnath (2002) and was chosen in this study over EPS percentage changes to avoid scaling issues when EPS estimates are close to zero. In some cases (12.6% of the sample), analysts do not issue new earnings forecasts after the event date. In these cases, it is assumed that the analyst did not wish to change their standing EPS estimate, thus the EPS estimate change is set to 0.

Independent variables

There are two main set of independent variables of interest used in this study. The first set of independent variables of interest is based on the relationship between the analyst's geographic location country (which I will refer to as the analyst country) and the headquarters country of the target company (or target country). The *Local Analyst* variable is a binary variable that has the value 1 when the analyst is located in the same country as the headquarters of the event company. The *Analyst Distance to Target* variable is a continuous measure of the geographic distance between the analyst country and the headquarters country of the target company. For robustness, I incorporate the other 8 different measures of cross-national distance as summarized by Berry, Guillen, & Zhou (2010). These variables were calculated by taking the cross-national distance of each of these measures between the event country and the target country. The 9 measures of cross-national distance are: administrative distance, cultural distance, demographic distance, economic distance, financial distance, geographic distance, global connectedness, knowledge distance, and political distance.

Analyst countries were determined using analyst contact information provided by Thompson Reuters separate from the I/B/E/S data. This contact information data contained one observation for each analyst-broker pair and included analysts' contact telephone numbers with international country telephone dialing codes which I mapped into ISO country codes. There were two issues when doing this mapping. First, only 50.8% of the analysts in this database had telephone contact numbers that were populated in the data. Second, some of the telephone numbers in the data could not be mapped to a specific country because the country dialing code could not be clearly determined from the phone number in the data.[4] In both of these cases, the analyst's geographic location

[4] This was usually due to formatting issues for some entries whereby the international country code was either omitted or not clearly separated from the rest of the supplied telephone number This was especially problematic because there is no international standard for telephone numbers. Even within some countries telephone numbers might be of different lengths.

was considered to be unknown (52.5% of the analyst-broker pairs). Analysts whose geographic location could not be determined where excluded from the analysis.

The second set of independent variables used in the analysis is based on the relationship between the headquarters countries of the event and target companies (i.e., event country and target country, respectively). The IBES data includes headquarters company information for nearly all of the companies in the database. This information is coded using Thompson Reuters own IBES country coding system. These codes were directly translated into ISO 3166-1 alpha-2 country codes (a standard published by the International Organization for Standardization) using mappings provided by Thompson Reuters. Using these codes I create a binary variable, *Local Event Company*, that has a value of 1 if the event country is the same as the target country (i.e., the event and target companies share the same headquarters country), and zero otherwise.

Control variables

Most of the control variables created in this study attempt to control for observable differences among stock market analysts. I control for whether the analyst also covered the event company (i.e. a binary variable indicating if the analyst actively covered the event company in the given fiscal period, *Analyst Covers Event Company*). I control for the number of companies that an analyst covers in that fiscal period, *Analyst Coverage Load* (Herst et al., 2004). I also control for previous analyst experience along two dimensions: the number of prior years that the analysts has covered the target company, *Analyst Target Company Experience*, (Mikhail et al., 1997) and the number of years that the analyst has worked as an analyst, *Analyst Career Experience*, based on the first analyst report issued by that analyst that appears in the data (Clement, 1999; Jacob, Lys, & Neale, 1999). Additionally, I control for the analyst's report frequency (i.e., the average number of analyst reports that the analyst issues per year for the target company) (Jacob et al. 1999) along with the size of the brokerage firm the analyst works for, *Analyst Broker Size*, as measured by the number of analysts issuing reports for that broker in the data (Jacob et al., 1999). I also include a variable measuring the size of the event firm a la market capitalization, *Event Firm Size*, along with a dummy variable indicating if the target company is headquartered in the US, *Target Firm US Indicator*. Finally, a set of dummy variables for all of the analyst countries are included in the analysis.

Unit of analysis

Since the focus of the paper is to better understand the pattern of informational spillovers from one company to another, the analysis centers on how analysts change their EPS projections for target companies after the announcement of an earnings miss by the event company. Thus, the unit of analysis is the analyst-target company pairing for each fiscal quarter. The data from each fiscal period and industry group is pooled and the analysis is performed where there is one observation per analyst / target company / fiscal quarter combination.

Empirical Strategy

I model analysts' EPS estimate changes using a random effects panel model. I do this to control for variation in the degree to which analysts react to a given event by specifying a random effect for analyst-event pairs. In order to ensure that the coefficient estimates provided by the random effects model are not biased, I use a Hausman specification test. The p-value of the Hausman specification test comparing a fixed effects model to the random effects model is 0.857. This high p-value suggests that a random effects model will not produce biased estimates.

RESULTS

Tables 2-1 & 2-2 display descriptive statistics and correlations for the variables in the model. Table 2-3 gives a breakdown of the number of observations by the 4 different combinations of relative locations of the analyst and event company. Roughly 72.5% of the observations in the study are for analysts that are located in the same country as the headquarters of the target company. Additionally, we see that about 67.5% of the observations are for target companies that share the same headquarters country as the event company. Table 2-4 and Figure 2-1 show the raw mean EPS forecast change broken down by the relative location of the analyst to the target company and the relative location of the event company to the target company.

Table 2-5 displays the results for the random effects regressions on the change in analyst EPS forecasts following an event company earnings misses. Model 1 is the baseline model that includes all of the control variables. Models 2 through 5 include each of the main independent variables separately. Model 6 is the full model that includes all independent variables and control variables and is used to test all of the hypotheses in this study.

Hypothesis 1 proposes that a target company will have its earnings projections reduced more when the event company (i.e., the company with the earnings miss) is located in the same country. Empirically, Hypothesis 1 would predict a negative coefficient on the *Local Event Company* variable - a negative coefficient indicates that earnings misses by local companies will cause analysts to reduce EPS forecasts more than earnings missed by foreign companies. The negative and highly statistically significant coefficient of -1.785 ($p < 0.001$) on the *Local Event Company* variable lends strong support to Hypothesis 1. To put this number into context, let's continue with the Banco Santander example. The event company in this example is Banco Espanol De Credito, which is a local (i.e., Spanish) bank. Since the event company is local, the model predicts that analysts reduced their EPS projections by an extra €0.08 (i.e., by an extra 7.14% of the original consensus estimate) than they would have had the event company, Banco Espanol De Credito, been foreign.

Hypotheses 2 & 3 propose that foreign analysts will reduce their EPS forecasts much more than local analysts, but will do so only when the earnings miss is by an event company that is local (i.e., located in the same country as the target company). To put this another way, these hypotheses together propose a positive relationship between local analyst and earnings forecast changes that is completely moderated by the relative location of the event company to the target company (e.g., the effect exists only when the event is local). These hypotheses are tested using the *Local Analyst* and *Local Event Company & Analyst* variables.

Hypothesis 2 predicts a positive coefficient on *Local Event Company & Analyst*. Given that I model reactions to earnings misses by similar companies where the average earnings forecast change is negative, a positive coefficient on *Local Event Company & Analyst* would suggest that, when the event company is local, local analysts have relatively more positive (i.e., smaller negative) earnings forecast changes than foreign analysts. The significant coefficient of 1.489 (p-value = 0.006) for the *Local Event Company & Analyst* variable lends strong support to Hypotheses 2. Again, using the Banco Santander example, when the event is local (as it is here), the model predicts that a foreign analyst will reduce their EPS projection for Banco Santander €0.0738 (i.e., 5.96% of the original consensus estimate) more than would a local analyst. Hypothesis 3 predicts a coefficient of zero for *Local Analyst*. A positive (negative) coefficient on *Local Analyst* would indicate that local analysts have less negative (more negative) changes in their earnings forecasts than foreign analysts when the event company is foreign. The model yields a statistically insignificant coefficient of 0.046 (p = 0.891) for the Local Analyst variable. However, the lack of statistical significance cannot itself lend support for Hypothesis 3. To determine if this hypothesis has support, I conducted a power test to determine if the sample sized used in the study has enough power to be conclusive. Conservatively assuming that the added explained variance of including local analyst is 0.1%, and using the baseline model 1 as a baseline, the sample sizes required to test with 0.7, 0.8, and 0.9 power are 6136, 7800, and 10400, respectively. Given that the sample used in the study is 6,147, the non-statistically significant finding lends very weak support; for power 0.7, we can have confidence in their not being a difference, but not for power 0.8 and above. Conservatively I would conclude that the test in not powerful enough to detect a small effect from local analyst with enough confidence (i.e., equal to or greater than 80% power), and thus the test is largely inconclusive, or at best very weakly supported.

To better illustrate these findings that were predicted by Hypotheses 1 and 2, I constructed Table 2-6 and Figure 2-2 using scaled earnings forecast change predictions from the full regression model (i.e., model 6).[5] Essentially, these are predictions for four different scenarios (i.e., local or foreign analyst X local or foreign event company). Following the hypotheses, there are three things worth pointing out. The negative slopes of both the foreign and local analyst lines graphically represent that EPS estimate changes are more negative when an event is local (as predicted by Hypothesis 1). Also, the slope of the foreign analyst line is much more negative than the slope of the local analyst line (Hypothesis 2). This graphically shows that foreign analysts revise their earnings estimates much more strongly downward than do local analysts when the event company is local. Again, it should be noted that although there visually doesn't appear to be much of a difference between local and foreign analysts for foreign events, statistically the test is not powerful enough lend strong support that this is indeed the case, and thus only yields very weak support.

[5] The predicted scaled earnings forecast changes from Table 6 and Figure 2 correspond to the unadjusted average scaled earnings forecast changes from the data detailed in Table 4 and Figure 1.

These results are highly consistent with the social categorization theory outlined in this paper. In general I hypothesize that analysts will tend to group companies located in the same country were they reside more finely and into smaller groups due to an in-group heterogeneity bias and will also group companies located in different countries than where they reside into larger and broader groups due to an out-group homogeneity bias. If an event company is foreign and isn't located in the same country as the target company, both the local analyst and the foreign analyst will perceive the event company to be different than the target company. Thus, both local and foreign analysts will react similarly and will only slightly reduce their EPS estimates when the event company is foreign. Conversely, if the event company is local, the local analysts will more likely view the event company as being different than the target company, while the foreign analysts will likely perceive the event and target companies to be relatively similar. Thus, the foreign analysts will revise their EPS estimates downward much more than local analysts.

The pattern of EPS forecast changes as predicted by the model also casts doubt on one major alternative explanation for why foreign analysts would react more strongly than would local analysts. If one makes the case that foreign analysts are less informed or have inferior and asymmetric information compared to local analysts, then we should see foreign analysts reacting more strongly to an earnings miss by the event firm and subsequently revise their EPS forecasts more strongly downward than local analysts regardless of the location of the event company. Moreover, this alternative explanation is even more doubtful given that the study is based on only post Reg FD data (see

Mohanram & Sunder, 2006). Additionally, the models used in the study controlled for Analyst Target Company Experience and Broker Size which other scholars have argued are the main sources of private information and informational advantages among stock market analysts (Dugar and Nathan, 1995; Michaely & Womack, 1999; Fleischer & Baum, 2010).

However, it is possible the analyst experience might moderate the degree that foreign and local analysts might respond differently to negative earnings surprises. As a robustness check, and to investigate this issue, I ran model 6 and included interaction terms of the three local analyst and event variables. The only interaction that was statistically significant was the triple interaction of local event, local analyst, and analyst career experience. Model 7 in Table 2-7 include the results of this additional regression. As we can see from the results, the coefficients of the local event and analyst variables are materially unchanged. However, the statistically significant (at the 0.01 level) coefficient of the triple interaction variable (i.e., local event * local analyst * analyst experience) is statistically significant (at the 0.01 level) and negative. To demonstrate this effect, I used model 7 to predict the EPS forecast changes depending on event location, analyst location, and analyst experience. Figure 2-3 and Figure 2-4 graphs the predicted EPS forecast changes for inexperienced and experienced analysts, respectively. The pattern for inexperienced analysts is similar to what was predicted by the original model. However, the pattern for experienced analysts tells a slightly different story insofar as local experienced analysts will revise their forecasts more downward than inexperienced analysts, but not as much as inexperienced or experienced foreign analysts. While this doesn't change the original finding, it does suggest that analyst experience shapes the degree by which local analysts might suffer from the in-group heterogeneity bias. Moreover, these results suggest that experience might affect how analysts, and potentially other actors leverage category systems or are subject to bias. This is a potentially interesting avenue for future research.

Finally, does analyst distance matter? Hypothesis 4 proposes that analysts should reduce their EPS forecasts more as the distance between them and the target company increases and predicts a negative coefficient for the *Analyst Geographic Distance to Target*. However, the model prediction of a statistically insignificant positive coefficient of 1.371 ($p = 0.056$) lends no support to Hypothesis 4. Though not significant, a positive coefficient implies that as the distance between the analyst and the target company increases, analysts will reduce their earnings forecasts less (i.e., smaller downward revisions). To investigate this further, I ran the same regression (i.e., model 6) using each of other 8 different measures of cross national distance. The coefficients for the cross-national distance variables took on both positive and negative values and were statistically insignificant for all eight measures. These robustness checks cast even more doubt on Hypothesis 4. It should be noted that it is indeed rather surprising not to find a significant effect for any dimension of analyst-target country distance given the importance of cross-national distance in the international business literature (see Berry, Guillén & Zhou, 2010 for a recent breakdown of multidimensional measures). In particular, distance is expected to affect the flow of information between countries, where less information flows to more distant countries leading to increased uncertainty for more distant agents (Johanson & Vahlne, 1977; Barkema et al., 1996; Kogut & Singh, 1988). But, surprisingly, in this context, this doesn't seem to be the case. Perhaps this might be due to the fact that the agents are this context are professionals, and in light of the laws in many countries (i.e., RegFD), are privy to the same relevant information.

DISCUSSION

Much of the work in social categorization has rightly narrowed its theoretical and empirical attention to contexts where audiences all share a common categorization scheme. Conversely, a smaller body of work has also addressed contexts where different interacting audiences have different categorization schema (e.g., mass-market and independent critics in Zuckerman & Kim, 2003) to understand how these different schema affect various outcomes. This project makes contributions to the latter body of work by beginning to address the following two questions. First, how might categorization schema systematically differ among seemingly homogenous audience members, different audiences, and especially market mediators bridging the same market? Second, how do these systematic differences affect audience or mediator perceptions and behaviors?

This project also contributions to the multinational management literature by demonstrating how an audience's or market mediator's nationality affects how multinational firms, and their products, are perceived and understood. Moreover it demonstrates that different stakeholders, or even the same stakeholders spread out geographical, may interpret and react to the same information differently. Finally, this project hopes to uncover additional cognitive biases associated with social categorization when audience members and market mediators are themselves partially embedded in overarching schema.

In summary, while multinational scholars have already benefited greatly by applying Neoinstitutional theory to the field of multinational management, there is still more than can be done. The most recent literature in categorization may be particularly relevant to the context of a multinational's national identity and the interpretation and ascription of events. In this paper, I argued that informational spillovers, or cognitive transfer can occur via multinationals' shared home countries, even when the information is uninformative. More importantly, I show that the relative location of external stakeholders, stock market analysts in this case, plays an important role in determining how these stakeholders perceive and ascribe information from new events. This has very important implications for multinational management research and suggests that multinational scholars be cognizant of the relative locations of external stakeholders when determining how the views of these different stakeholders may change in light of new events. A worthwhile avenue of future research would be to investigate how the same set of stakeholders located in different countries might evaluate the countries themselves and how these evaluations might diverge in light of political and non-political events.

Finally, by exposing how categorization schema systematically differ among members of the same stakeholder group, we might expect to see even more variation if we look at differences across stakeholder groups. Generally speaking, the focus on the reactions of different stakeholder groups has been on the different needs and wants of the groups.

However, systematic differences across groups might also help explain why stakeholder groups act in the ways that they do. For instance, Durand & Paolella (2013) showed that goals affect how audiences categorize objects. Since different stakeholder groups have different goals, they may focus on different sets of firm characteristics when categorizing companies and multinationals. Looking at this in more detail, which could extend the international business literature by adding an additional mechanism that helps explain stakeholder group behavior, is an avenue of future research.

CHAPTER 3: COGNITIVE NEIGHBORHOODS AND THE VALUATION OF INNOVATION: A CROSS-NATIONAL ANALYSIS

INTRODUCTION

In 2001, LG Electronics developed a new method to produce carbon nanotube transistors and subsequently patented their new invention in many countries including the U.S. and Japan. Oddly, the patent that was granted in the U.S. received almost 3 times as many forward citations than the equivalent patent granted in Japan, even after adjusting for country differences in citation rates (194 vs. 69[6]). Conversely, in 1997, Merck patented a new liquid-crystalline medium for thin-film transistors and the Japanese patent received more than 5 times as many forward citations than the U.S. equivalent patent (28 vs. 5). It is unexpected that there would be such stark differences in the number of forward citations across patents granted in different countries for the exact same invention. This paper aims to understand why such differences may exist. In doing so, I argue that these differences are explained in part by differences in the technology classification systems used in each country. In particular, I highlight the important role that category relatedness and horizontal category structure play in shaping citation patterns across technologies and the realized value of innovation. Empirically, I leverage a novel

---

[6] The adjustment multiple for the Japanese number is the average number of forward citations for all U.S. patents divided by the average number of forward citations of all Japanese patents. This multiplier accounts for baseline citation differences between the U.S. and Japan.

experiment-like design that predict changes in citation rates and citation patterns for the patents of the exact same invention that are granted in different countries. This novel empirical approach has a number of strong advantages to the usual approaches used in most of the extant literature.

Over the past few decades, an interesting and ever growing body of literature has emerged in the management literature investigating the role that social categories play in valuation by external audiences. The most dominant thrust of this literature has focused on the need for objects to fit within categories or else be subject to a discount, i.e. the categorical imperative (Zuckerman, 1999). This is the result of a theorized two-step process (Phillips & Zuckerman, 2000). In the first step, the audience assigns the focal object to a given category, e.g. a firm to an industry. In the second step, the audience then evaluates the focal object based on the expectations or criteria associated with the assigned category. While many important insights have been gained from this general approach, not all valuation settings follow something that is akin to this two-step process where conformity is needed.

When novelty is the goal, as is the case with a variety of different settings including innovation, music, and even academic research itself, a very different search and valuation process unfolds. In these settings, broad or widespread appeal is paramount, and this requires gaining the attention of diverse audiences that transcend a single category. In fact, in the innovation literature, the impact and importance of a patent is measured by how broadly applicable it is, e.g. Trajtenberg, Henderson & Jaffe’s (1997) generality measure, which is based on the number of different categories who’s members cite the given patent. These types of measures, and the underlying logic for their use, are strong reminders that the ultimate value, impact, or influence of novel objects, like patents, are not determined solely by their influence on members of the same category, but also by their influence on members of other categories. To put this another way, valuation doesn't always take place at the individual category level (Vergne & Wry, 2014, p. 76). What is needed is broad appeal from and influence over members of multiple categories. The more categories the broader the appeal and the stronger the influence.

But from a category perspective, what is needed for broad appeal and influence? The answer stems from category relatedness. Category systems are used as a heuristic that shapes cognition by narrowing focus and directing attention (Zerubavel, 1997). Categories often exhibit varying degrees of relatedness with other categories. Often, categories are arranged into hierarchies that denote varying degrees of relatedness among them at a given level of the hierarchy (Rosch et al, 1976); categories that are members of the same superordinate category are more similar than those that aren't. Similarly, categories can be seen as being more or less related, but still distinct, as has been shown in the case of movie genres (Hsu, 2006) and patent categories (Wry & Lounsbury, 2013). Moreover, category relatedness plays an important role in regulating attention across categories which may lead to influence across categories. The more related categories appear, the greater amount of attention that will be directed across category boundaries. If a focal category is more similar to many other categories, its members will receive more attention and enjoy a broader and larger potential audience than members of an isolated category. Thus, for a given agent or object, the imperative is not to fit neatly into a single category, but rather to be associated with a category that is relationally similar to many other categories. Accordingly, the focus of this paper is to demonstrate the critical role that category structure, and the relatedness among categories at the same hierarchical level, plays in generating attention and appeal, which ultimately shapes valuation.

So how might we best go about demonstrating the relationship between category structure and valuation? In the context of innovation, the gold standard would be an experiment where we take an invention, with an assigned focal category, and systematically vary the relationship structure of the surrounding category system - specifically varying the number of related categories to the focal category and the degree of relatedness – and see how this affects the realized value of the invention. While this isn't financially practical nor feasible, the natural variation in technology classification systems across patent jurisdictions, and the stark differences in their underlying structure, does allow for a similar type of setting that is not too far off from the ideal experiment. Moreover, this novel approach constitutes a major empirical contribution to both the category studies and innovation literatures. Typically, the findings from prior studies are based on explaining variation across objects or within objects over time all taking place within a single category system where great effort is needed to control for quality and other object-level and time-variant factors. Instead, my approach models variation within the same objects across category systems at the same point in time, allowing me to perfectly control for quality and other object-level unobservables. This yields a much stronger identification approach than serves as an improvement that was has traditionally been done in the past. In the case of patents, I model how a given invention's realized value changes when it's embedded in different technology classification systems.

Leveraging this very novel and improved empirical approach, I find that the structure of the classification system, and the position of a focal category embedded in that structure, greatly impacts the value of the focal category's members. Specifically, an invention whose patent is assigned to a technology class that is highly related to other technology classes will tend to receive more forward citations overall than if it was assigned to a class that is related to only few other technology classes. This increase is entirely due to increased citations from patents in other technology classes. Getting back to the LG carbon nanotube transistor example, LG's U.S. patent was assigned to a more highly related and more central technology class than was the equivalent Japanese patent and this resulted in relatively more forward citations despite being the exact same invention. Similarly, Merck's Japanese patent for a new liquid-crystalline medium was assigned to a more central and more related technology class and ultimately received more forward when compared to the U.S. equivalent.

THEORY & HYPOTHESES

Over the past 25 years, starting with Porac et al.'s (1989) seminal paper on competitive groups, research on the implications of categories in the study of management has steadily grown. In this time, two separate approaches have emerged that borrow from different root disciplines: the psychology approach following in the traditions of Porac et al. (1989) and the economic sociology approach following in the tradition of Zuckerman (1999). While important findings have been uncovered by both approaches, the latter approach has been the most dominant in the past 15 years and the research on categories has become much more synonymous with Sociology. More recently, however, some scholars have endeavored to marry both approaches into a single cohesive research domain, called category studies (Vergne & Wry, 2014). One of the main benefits of doing this is to help the field move forward by shifting part of the attention back on potentially fruitful insights from psychology, which have recently been mostly absent in the literature to augment the more developed insights from sociology. This paper leverages both literatures: category similarity at the category dyad level, and category structure from the psychology literature.

Sociological Foundation: The general approach taken by the sociology perspective is illustrated in the two-stage audience evaluation model outlined by Phillips & Zuckerman (2001) and is based on evoked sets. In the first stage, the audience filters potential offerings by screening out those that do not fit within the category of interest in order to facilitate comparison. Membership in the focal category is necessary as cross-category comparisons are inherently difficult or impossible and non-members are seen as impure or illegitimate. In the second stage, the audience selects the best offering among the remaining legitimate offerings which are all now deemed full members of the focal category. This two-stage model epitomizes the sociological approach via the importance of membership to a focal category, the categorical imperative, and the penalties levied on category-spanners (Zuckerman, 1999). Much of the work following in this tradition has identified the moderators and boundary conditions of this categorical imperative as well as the importance of the degree of membership.

One interesting common starting point of research based on this sociological perspective is that the focal category used for evaluation is generally given. This makes sense in the more narrow empirical contexts chosen by researchers where the evaluators have expertise, preferences, or an orientation toward a single category such as stock market analysts (Zuckerman, 1999), specialized film critics (Zuckerman & Kim, 2003), food critics (Rao et al. 2005; Durand et al., 2007; Kovacs & Hannan, 2010), and e-bay buyers (Hsu et al., 2009), etc. However, this has limited scholars to the investigation of individual categories in isolation, or category dyads, resulting in a greatly limited scope leaving this stream of research largely silent when it comes to the influence of larger category structure.

Cognitive Psychological Foundations: In contrast, research in the cognitive psychological literature, with a stronger emphasis on understanding how objects are classified, has investigated category structure in much more detail. Of particular interest is Eleanor Rosch's work on category hierarchy. Rosch (1978) proposed that category systems are formed on the bases of two general principles: 1) cognitive economy - a category system should provide the most information with least amount of effort and 2) perceived structure - objects in the perceived material world have a high correlational structure.

Rosch argued that the combination of these two rules necessitated category systems that were hierarchical and contain both vertical and horizontal dimensions. The vertical dimension captures different levels of inclusiveness while the horizontal dimension captures the partitioning of categories at the same level of inclusiveness. Her main takeaways were that 1) categorization can take place at different levels, and these levels are not equally effective or helpful, and 2) that categories tend to be defined in such a way that category members possess many items in common, yet possess few items with members in other categories - i.e., horizontal categories should be maximally distinct. Both takeaways underscore the necessarily hierarchical, or taxonomic, nature of category systems and the important role that this structure plays in cognition.

While this idea hasn't been completely absent in categories research in management, it certainly hasn't received much attention until very recently. This is rather surprising given that the very first article to incorporate categorization in the management literature (i.e., Porac, Thomas, & Baden-Fuller, 1989) picked up on the importance of hierarchical structure to explain managerial cognition, perceptions of relatedness, and the identification of rivals. Most recently, the importance of category hierarchy has been revived by work in organizational identity that focuses on relatedness. For example, Wry & Lounsbury (2013) demonstrated the importance of multiple levels of relatedness whereby vertical and horizontal linkages in technology influence organizational outcomes.

Other recent research in management grounded in cognition has also helped to understand the importance of relatedness and clarify the different limitations of the categorical imperative. Kennedy (2005; 2008) revealed how media coverage shapes manager perceptions of relatedness, the identification of rivals, and the emergence of new categories. Ruef & Patterson (2009) showed that boundary spanning is not penalized when the category system itself is emergent or in flux. Kennedy et al. (2010) found that when category meaning changes it can reverse the fortunes of initially conforming and non-conforming product offerings. Granqvist, Grodal, & Woolley (2013) demonstrated that managers strategically use market labels and hedge when labels are ambiguous or create credibility gaps. Finally, Fleischer (2009) found that conflict-of-interest relationships between producers and the firms that rate their products manifest ambiguity in the rating system as a whole. While this research leads to a number of valuable insights, it nonetheless overlooks the importance of category structure.

Work in psychology has firmly established that category systems used in cognition are hierarchical (Rosch, 1978) and structure cognition in an efficient and consistent way. Thus, it is no wonder that effectively all cultures' classifications of the natural world have a hierarchical structure (Berlin, 1992). Most importantly, these category structures, via the positioning of categories within the hierarchy, reflect patterns of relatedness among categories (Rosch, 1976). For instance, categories that share a common superordinate, or higher level, category are seen by both audiences (Rosch & Lloyd, 1978; Wry et al., 2011) and managers (Porac & Thomas, 1990) as being more directly related.

The important point to take away from this stream of research is that category structure is critical when it comes to cognition. While much of the work in this tradition has focused on category hierarchy, their findings elude to the implicit relational structure of categories at the level of the hierarchy, particularly at the basic level. If category structure plays such an important role in cognition overall, we would also expect structure to play a similarly important role across horizontal (i.e., categories at the same hierarchical level) categories. Thus, the focus of this paper is to understand the relationship between horizontal category structure and influence and valuation.

Innovation, Nanotechnology, and Examiner Search: The empirical setting chosen in this paper, namely prior art search for patents by examiners, differs somewhat from the usual empirical settings of much of the research on categories. The main difference lies in the use of the category system and the type of evaluation being conducted. Instead of categories being used as a way to filter, compare, and evaluate patents, they are instead used to assist patent examiners in identifying related technologies. While patent examiners must ultimately evaluate whether a patent application is novel enough to merit being granted, I take novelty as given by looking only at granted patents (i.e., innovations that have already been judged as novel) and investigate the earlier step of the search process for prior art. It must be noted that the unique function of categories in this setting is different enough from previous literature to warrant caution in the form of potential boundary conditions when extending these empirical findings to more generalized setting. Nonetheless, the main idea that larger category systems play an important role in categorization and evaluation should still generally hold.

When searching for relevant prior art, a patent examiner firsts starts by generating a field of search, which outlines exactly where the examiner will search. Usually this field of search is described by identifying potentially relevant USPTO classes and subclasses (e.g., "424/270, 272, 273"). These fields of search are important because they help isolate the potentially most relevant innovations to make prior art search more feasible when the examiner faces limits to both time and effort. They are also critical because they ultimately help determine what prior art is cited by the examiner, and as a consequence, the set of future citations. But how are these fields of search determined?

In this paper I model patent examiner cognition and potentially relevant prior art via the use of patent classes to predict forward citation rates. While there are likely a number of factors that determine the field of search that is chosen by the examiner, I argue that the main characteristics of the innovation, as identified by the patent's primary class, is one critical source for identifying related domains of technology. While fields of search naturally include prior art in the ascribed primary class, I am more interested in understanding what prior art in other patent classes are seen as potentially relevant. Thus I investigate classes at the same hierarchical level. Psychological research has demonstrated numerous empirical advantages of "basic level" categories, which are situated in the middle level of the categorization hierarchy, when making similarity judgments (Rosch et al., 1976; Markman & Wisniewski, 1997), thus I use 3-digit USPTO classes for U.S. patents and 4-digit IPC classes for German and Japanese patents, which represent a middle level of the taxonomy and also have roughly the same number of classes across the systems.

I argue that patent examiners have some conception of the larger classification structure in terms of the degree of relatedness among categories. I operationalize this idea by constructing a cognitive map of category relatedness and apply it in a broad way to patent classification. I argue that patent examiners use these maps, to identify potentially related domains of technology. Using the assigned primary class, patent examiners identify other potentially relevant domains through technology neighborhoods. The neighborhood is constructed by selecting other technology classes that are most psychologically similar to the focal class in the cognitive map. If we picture the cognitive map in physical space, the examiner draws a circle with the primary category at its center. This is the cognitive neighborhood and all patent classes within the circle are considered to be part of the technological neighborhood and share some similarity to the primary class. It then follows that patents which are members of classes within the technological neighborhood are included in the field of search and have the chance to be cited as relevant prior art.

If a focal patent class has a dense technological neighborhood with many similar other classes, then patents belonging to the focal class will be seen as related to a larger number of domains, reviewed more often in future prior art searches, and will, ceteris paribus, collect more forward citations. Conversely, patents that are members of classes with sparse or empty technology neighborhoods should ultimately receive fewer forward citations.

> **Hypothesis 1 (H1): The members of categories with many neighboring technology classes will accrue more forward patent citations than members of categories with fewer neighboring technology classes.**

Moreover, technology neighborhoods should affect not only the overall citing of a patent, but also the pattern of citations. In particular, there are potentially two competing forces with respect to technology neighborhoods that should affect forward citation rates for a given patent. To highlight these different forces, I make a distinction between forward citations from subsequent patents with the same primary technology class and forward citations from patents that are assigned a different primary technology class, which I call in-class and out-class citations, respectively. When an examiner searches for relevant prior art, they not only search in the primary class, they also search a number of other classes. For a focal patent application, the pool of potentially relevant in-class patents that reside in the focal class is by definition fixed. However, the size of potentially relevant out-class patents that are considered is dependent on the number of out-classes (i.e., other classes) that are identified in the field of search. Fields of search that include many out-classes are likely to include a much larger consideration set of prior art than fields of search that include fewer out-classes. However, examiner time and attention is a limited resource when they conduct their prior art search. This finite attention is spread, evenly or unevenly, over all of the prior art in the field of search. Since the amount of in-class prior art is fixed, I expect that they will receive less attention for wider fields of search that include many out-classes in comparison to fields of search that are more narrow and include fewer out-classes. For instance, take the extreme example of a field of search that only includes prior art in the focal application's primary technology class, or in-class. In this case, in-class prior art will receive all of the examiner's finite attention. However, if the examiner additionally included a number of out-classes in their field of search, then the in-class prior art would receive less attention. Essentially, out-classes steal examiner attention away from in-class prior art by directing it toward out-class prior art. Thus, more out-classes in the examiner's search should, on average, lead them to cite fewer in-class patents while simultaneously citing more out-class patents. The measure of technology neighbors that I construct in this paper is a reflection of the underlying examiner cognition concerning the relatedness and relevance among patent classes and approximates the number of potentially relevant out-classes. This leads to two separate hypotheses related to the frequency of in-class and out-class citations as predicted by cognitive neighborhoods.

> **Hypothesis 2 (H2): The members of categories with many neighboring technology classes will accrue fewer forward patent citations from patents with the same primary classification (i.e., in-class forward citations) than will members of categories with fewer neighboring technology classes.**

> **Hypothesis 3 (H3): The members of categories with many neighboring technology classes will accrue more forward patent citations from patents with different primary classifications (i.e., out-class forward citations) than will members of categories with fewer neighboring technology classes.**

DATA & METHODS

Data:

Sample: The empirical setting of this paper is nanotechnology (Nano) patents between 1995 and 2010. The field of nanotechnology is an ideal setting for this study for two reasons. First, nanotechnology, broadly defined, is applicable to a staggering number of domains, from energy storage to sporting goods, biomedical imaging to water filters, and electronics to pharmaceuticals. Second, nanotechnology is still a relatively new and emergent field. In fact, it wasn't until 2004 that the USPTO established a new crossreference class (class "977") to help identify and represent nanotechnology-related innovations.

To create the population of Nanotech patents, I used the European Patent Office’s 2014 Autumn edition of the Worldwide Patent Statistical Database (i.e., PATSTAT). PATSTAT was created by the EPO for the OECD Taskforce on Patent Statistics to help support and aid empirical and statistical research on patents. PATSTAT is a comprehensive database of patent applications and grants from over 100 different countries across the globe and includes information on more than 35 million granted patents.

PATSTAT organizes individual patent applications and granted patents across countries into patent families. Each patent family represents the set of patent documents (i.e., applications and grants) that have been issued in different countries (i.e., different patent jurisdictions) to protect a single invention. Identifying patent families is done by linking patents to a common first patent application in a given patent jurisdiction, called a priority, which established a priority date for the claim of the invention. The Paris Convention for the Protection of Industrial Property allows applicants up to twelve months to file patent applications in other member countries while still claiming the same initial priority date (setting the date when intellectual property protection begins) by linking their application to the same initial priority. In some situations, when applicants file a patent application they link the application to multiple prior priorities. This might be due to the timing of the applications (e.g., if they simultaneously filed initial applications in two different countries then later apply to a third), or when they make different claims. This can lead to different patent family definitions. In this study I identify patent families using the 'Espacenet patent family' which requires that all documents (e.g., applications and grants) are linked to exactly the same priority, or set of priorities. This is the most restrictive definition of a patent family. For the purposes of this study, this most restrictive definition ensures that the patent grants issued in different countries pertain to the exact same invention and set of claims.

To identify nanotechnology patents I used the StarTechZD nanotech database (STARTECH). The STARTECH database is supported by the United States National Science Foundation and identifies and catalogs nanotechnology patents granted by the USPTO from 1976 through 2010 using a number of different methods. As the STARTECH database only identifies patents in the US, I then linked these U.S. patent grants to their patent families via PATSTAT, which identified grants in other countries based on the exact same invention. I then select the set of nanotech patent families which were granted simultaneously in three different countries: the U.S., Germany, and Japan. The PATSTAT database has comprehensive data on all three countries through 2013. I then selected those patent families with applications that were filed or published in or before 2010 in all three countries. I chose 2010 as the cutoff to allow each patent in the sample enough time to collect at least three years' worth of forward citations.

Variables:

Dependent variables*:* The dependent variable in the analysis is a simple three-year count measure of forward patent citations based on the publication date of the initial application. Because I only have citation data through 2013, the year that a patent appears in the sample greatly affects the number of forward citations it receives. This is purely a function of patents in earlier years having more time to collect citations by patents in later years. For instance, I would only have one year's worth of forward citations data for a patent that is granted in 2012, whereas I would have 3 years' worth for one that is granted in 2010. To standardize this dependent variable across patent families and across countries, I simply count the number of times the focal patent (or its application) is cited by patents issued in the same country as the focal patent and that were granted in the three-year window directly following the publication of the focal application or grant. I do this three different ways, leading to three different dependent variables. First, I count all citations (i.e., 'all' forward citations). Second, I count only citations made by subsequent patents that share the same primary technology class (i.e., 'in-class' forward citations). Third, I count only citation made by subsequent patents that have different primary technology classes (i.e., 'out-class'). I define the primary technology class as the first three digits of the primary USPTO classification for patents issued in the U.S., and the first 4-digits of the primary IPC code for patents issued in Germany or Japan.

Independent variable: The main independent variable, *neighbors*, is a measure of the number of other related and technologically proximate patent classes to a focal patent's primary class. The purpose of this measure is to quantify the number of other technology classes that are likely to be included when an examiner conducts search for relevant prior art when examining a patent application. This measure is constructed at the level of the technology class and is assumed to change over time and vary across patent systems.

This is a measure that is calculated for each patent technology class (i.e., 3-digit USPTO class for U.S. and 4-digit IPC class for Germany & Japan) for each year, and separately for each of the three countries in the sample. The construction of technology neighbors involves two steps. Both steps are done separately for each of the three countries and for each year in the sample.

In the first step, I represent the technology landscape and the relationships among patent classes as a weighted network. In network terminology, the nodes of this network are patent classes and the edges connecting these nodes signify the degree to which the connected classes are technologically related. I calculate the degree of relatedness, or edge weight, of two focal patent classes as a proportion of the frequency by which the two classes are listed simultaneously on granted patents. The degree of relatedness, P_{ij}, for patent classes *i* and *j* is calculated as:

$$P_{ij} = \frac{1}{2}\left(\frac{C_{ij}}{C_i} + \frac{C_{ij}}{C_j}\right)$$

Where C_{ij} is a count of the number of times patent classes *i* and *j* are co-listed in a patent in the prior year. C_i and C_j are the counts of the number of patents that classes *i* and *j* are listed on in the prior year, respectively. By construction, the degree of relatedness falls somewhere on the unit interval. This proximity measure is the same measure used by Lo & Kennedy (2015).

In the second step, I calculate the number of neighbors using degree centrality. In a binary (read un-weighted) network, the degree centrality for a given node is simply the number of ties to the node. For a node in a weighted network, this measure has been extended by Barrat (2004) and Newman (2004) and is the sum of all of the weights for edges connected to the node. Thus, I calculate the number neighbors for patent class *i* using:

$$N_i = \sum_{for\ all\ j} P_{ij}$$

Figure 3-1 depicts the distribution of the neighbors measure for the sample. The majority of patents reside in patent classes that have between 1 and 4 neighbors.

Control variables: To control for observed and unobserved patent quality differences within a country, I include patent family fixed effects. Because I am including these family fixed effects, I am unable to also include any control variables that do not vary across patents in a given patent family as these effects are already being captured by the patent family fixed effects. However, I do include a number of control variables that do potentially vary across patents in the same family which have been identified in prior patent research. These include the number of technology classes listed on the patent, *classes*, and a count of the number of unique backward citations to prior patents and literature, *backward citations*. I also include a count of the inventors listed on the granted patent, *number of inventors*, as well as the number of listed applicants, *number of applicants*. Both the number of inventors and applicants can vary across patents in the same family, though they only do so in roughly 3% and 5% of the patent families in the sample. Since I measure the number of forward citations starting from the date that the application for a granted patent is published, I control the number of years (days / 365) between the application and the grant. Given that my DVs are forward patent citations, I want to control for the size of a patent's primary class. Thus, I included a count of the number of patent applications in the focal patent's country with the same primary patent class in the year the focal patent application was published, *primary class size*. Finally, prior research has shown the average forward citation rates can differ substantially across countries (Yasukawa and Kano, 2015), so I include country fixed effects by including dummy variables for Germany and Japan, making U.S. the baseline.

Empirical Model

All three dependent variables used in the analysis are counts of forward citations. As is usually the case with forward citations, all three count variables exhibit a high degree of overdispersion due to the high frequency of zeros. Typically, researchers model forward citations and attempt to account for this overdispersion by using negative binomial regressions. I also model forward citations using a negative binomial regression. However, because I am also including fixed-effects in the model, I take a slightly different approach. Allison & Waterman (2002) demonstrated that the fixed-effects negative binomial model introduced by Hausman, Hall, and Griliches (1984) does not control for fixed effects in the same way as they do in OLS regressions, in most statistical software implementations like in Stata[7]. This is due to the non-linear nature of the model in combination with the mathematical parameter estimation approach. Thus, instead of using a fixed-effects negative binomial panel model regression, I use what Allison &

[7] The negative binomial panel model is implemented in Stata using the command 'xtnbreg'.

Waterman (2002) propose as a hybrid approach that negates this problem. Essentially what this approach does is account for the fixed-effects prior to running the model. In my case, to control for the patent family fixed-effects, I transform all of the variables that vary across patent family members (i.e., *neighbors, classes, backward citations, inventors, applicants, time to approval, primary class size, and grant year*). For each of these independent variables, this transformation is done by subtracting the mean of the focal variable across the patent family from the individual values associated with the patent family. In other words, you are simply accounting for the average for each patient family, as a fixed effect would normally do. To complete this hybrid approach, the parameters of the model are estimated using a negative binomial regression of forward citations on the transformed independent variables (if they vary across patent families) and untransformed independent variables (if they do not vary) along with patent family random effects.

Table 3-1 lists the descriptive statistics of the independent and untransformed dependent variables. Table 3-2 lists the correlation coefficients for all of the variables in the model and includes the transformed version of the variable when used. There is only one set of variables with particularly high (i.e., >0.60) correlation, which are *time to approval* and *grant year*. Removing either variable from the regressions didn't materially change the results, nor did it significantly change the point estimates or significance of either variable in any of the models. Since both variables are not relevant to any of the hypotheses, and since they both remain highly statistically significant in the models, I thus decided to keep both variables as they are. Aside from these two variables there are no multicolinearity issues with the main independent variables or the controls.

RESULTS

Table 3-3 shows the coefficients and standard errors for the regressions of the hybrid fixed-effects negative binomial models (see Allison & Waterman, 2002) on the three different forward citation count variables. Models 1, 3, and 5 are baseline models for each of the three dependent count variables and only include control variables. The predictions from these models yield the generally expected effects for most of these controls that are statistically significant. Backward citations and primary class size are positively related to the different types of forward citations where longer approval times are negative related. Grant year also has a negative coefficient in all three baseline models. This is expected because after adjusting for patent family fixed effects, later grant years generally indicate longer approval times as well. The number of inventors and the number of applicants also have negative coefficients. Normally, we might expect both to be positively related as they proxy patent quality. However, the interpretation of both coefficients is different than is normally the case because the patent family fixed effects already controls for the unobserved quality of the patent. It is unclear exactly what is driving the negative effect and there isn’t any prior research that might suggest why this is the case. Finally, the dummy variables signaling the German and Japanese equivalent of the same invention have large negative coefficients suggesting that German and Japanese patents have significantly fewer citations than their U.S. peers. Again this is to be expected as the overall number of citations, and hence the average forward citations rates, are much lower in Japan and Germany when compared to the US.

The coefficient for *classes* is positive when predicting all forward citations, as is usually expected in the literature. However, this obscures much of the actual picture. While the coefficient for *classes* is also positive when predicting out-class citations, the coefficient is large and negative when predicting in-class citations. There are two possible explanation for this, one perceptual and the other real. The first is that categorical imperative-like effects (Zuckerman, 1999) appear when examiners are searching within the primary class of a given application. Prior art considered by the examiner is seen as less related and relevant if it lists many secondary classes leading the examiner to either skip over it or discount its relevance in some way. If the first is actually the case, then the findings across all three models signals an important boundary condition for categorical imperative effects. Namely, that the effects only hold for searches within the same primary class. Moreover, the categorical imperative effect is actually reversed when examiners are searching outside of the primary class of the focal application. This makes sense. If a given patent lists few secondary classes, then it is less likely to be included in prior art searches initiated in other primary classes.

Hypothesis 1 predicted that members of categories with many neighboring technology classes will accrue more forward patent citations than members of categories with fewer neighboring technology classes. The coefficient for *neighbors* in model 2 is the test of this hypothesis. Indeed, the positive and statistically significant ($p < 0.001$) coefficient lends strong support for H1. It should be noted that the *neighbors* variable has been de-meaned within each patent family, which is the same as including patent family fixed effects. Thus, after controlling for patent family (or invention) quality along with all other non-varying (across family) unobservable characteristics, a patent will accrue more forward citations if it is assigned a primary technology class with many proximate neighbors. To put this coefficient into perspective, a one standard deviation increase in the number of neighbors will, holding other variables constant, increase the number of 3-year forward citations by approximately 7.3%.

Hypothesis 2 predicted that members of categories with many neighboring technology classes will accrue fewer forward patent citations from patents with the same primary classification (i.e., in-class forward citations) than will members of categories with fewer neighboring technology classes. The coefficient for *neighbors* in model 4 is a test of this hypothesis. While the coefficient is negative, as predicted, it is small in absolute value and not statistically different from zero at even the 5% level. Thus, I do not find support for hypothesis 2.

Hypothesis 3 predicted that members of categories with many neighboring technology classes will accrue more forward patent citations from patents with different primary classifications (i.e., out-class forward citations) than will members of categories with fewer neighboring technology classes. The coefficient for *neighbors* in model 6 is a test of this hypothesis. The large positive and highly significant ($p < 0.001$) coefficient lends strong support for H3. After controlling for the quality of the invention as well as other non-varying (across family) unobservable characteristics, a patent will accrue more out-class forward citations if it is assigned a primary technology class with many proximate technological neighbors. Again, in practical terms, a one standard deviation increase in the number of technological neighbors will increase the expected number of forward citations by out-class patents by 19.3%.

Taken together these results suggest that technological neighbors play an important role in shaping not only the frequency of forward citations, but also the pattern of citations. In particular, category neighbors increase forward citations from patents with different primary classes, though seem to have no effect on the frequency of forward citations by patents with the same primary class. Thus, the presence of category neighbors will increase attention from other technology fields.

Robustness Checks

I include two extra analyses to lend additional confidence to the findings. The first check examines alternative neighborhood measures. The second check examines two-country subsamples to ensure the findings aren't due to nuances from a single country's patent system.

For the first check, I constructed six alternative neighborhood measures, each of which is based on a different method of calculating the relatedness among category dyads (See Table 3-4 for results). The original measure calculates the neighbors measure of a given category as the degree centrality of a weighted proximity network of categories where the weight of an edge between two categories is the given category pair's proximity score. For all of the alternative neighborhood measures, I simply use different proximity scores, and thus different weights. For the first 5 of these alternative measures, I use relatedness scores that are based on the frequency of technology class co-listings within granted patents as I did with the original measure. For the first alternative measure, 'neighbors^2,' the relatedness scores are simply the squared proximity scores of the undirected category-to-category proximity network. By using the squared proximity scores as the edge weights, the resulting neighborhood measure is much more heavily influenced by very proximate categories. As Table 3-4 shows, the coefficients for this first alternative measure is extremely similar to the original measure when used in the full models; that is to say that they are positive and highly significant when predicating either all citations or out-class citations, but close to zero when predicting in-class citations.

The next four alternative measures are based on the proximity scores of the directed proximity network, including a simple and squared version for each direction. That is to say that the underlying proximity measure is based on the direction of category co-listings on patents where the proximity from a category A to category B is calculated based on the frequency that category B is co-listed when category A is the primary class (note, the original measure calculated proximity based on frequency of co-listings irrespective of which category was the primary class). The alternative measures 'Neighbors to' and 'neighbors to^2' are based on the frequency, and squared frequency of co-listings, when the focal class is listed as a secondary class. Again, for both of these alternatives we see a positive and statistically significant relationship when predicting all citations and out-class citations and statistically insignificant when predicting in-class citations. Conversely, the alternative measures 'Neighbors from' and 'neighbors from^2'

are based on the frequency, and squared frequency of co-listings, when the focal class is listed as the primary class.

Finally, for the last alternative measure, I calculated neighbors using a relatedness measure that was based on the past frequency of citations between category pairs using bibliographic coupling. Instead of basing category similarity on patterns of category co-listings, this final measure is based on prior citation patterns. Again, this yields the same pattern as the original measure.

For the second check, I ran the same models predicting all citations, in-class citations, and out-class citations using subsets of the sample that included two countries at a time. The logic for doing this is that one country might be driving the overall results. However, this does not seem to be the case. In Table 3-5, the analysis of each subset with the full model yields largely similar results to those of the full sample. The coefficients for neighbors is positive and highly significant in all three two-country subsamples when predicting all citations and out-class citations.

DISCUSSION

In this paper I showed that the position of a category within a larger structure, via the number of proximate neighbors, affects the outcomes of its members. This demonstrates that larger category structure can play an important role in cognition and categorization processes at the individual actor level. These findings give credence not only to the importance of larger category structure in shaping cognition and perceptions of relatedness, but also to the idea that novel insights can be gained by importing findings from cognitive psychology in the pursuit of a better understanding of categorization in management research as espoused by Vergne & Wry (2014).

The importance of category structure is abundantly clear in the psychological tradition of category research. For instance, one well established and widely tested model is the Generalized Context Model, or GCM (Nosofsky, 1992; Nosofsky & Palmeri, 1997). The GCM has its roots in what psychologists call the exemplar approach. According to this approach, categories are represented by individual exemplars. Objects are evaluated and classified on the bases of comparisons to these known exemplars. The more similarity between the object and a given exemplar, the more likely the object is categorized as belonging to the category that exemplar represents. What is most striking about the GCM model, particularly in comparison to sociological approaches to categories, is that the larger category structure, via comparison with all exemplars, affects classification. The probability that an object will be seen as belonging to a focal category is affected by the presence or absence of other nearby category exemplars. This model also allows for the possibility of individual categories being represented by multiple exemplars, which has yet to be considered in the category studies literature.

By taking a larger macro view and looking at categories as the unit of analysis, I open the door to the possibility that individual categories may be aligned with different institutional logics. This would allow for a more nuanced way to map the effects of institutional logics on individual cognition and action. One potentially fruitful example of this would be to consider how different levels of meaning may lead to different predictions of value. For example, Wry & Lounsbury (2013) and Wry, Lounsbury, & Jennings (2014) distinguish between two different logics: "science" and "technology." Mapping these logics onto technology classes may help us uncover how the presence or absence of classes with similar or different logics in technology neighborhoods affect the standing of the focal category and the frequency that its members are cited. For instance, members of patent classes associated with a science logic might be cited much more frequently when neighboring classes are associated primarily with a technology logic.

Finally, in this context, not only are patent examiners embedded in different categorization systems across countries, each with different logics, examiners in the same system are also field experts that are embedded within the national system. Porac and colleagues (1989) showed that actor cognition is shaped by their location within the system, resulting in socially constructed strategic groups. This would suggest that the cognitive schema of technology relatedness used by examiners might also by shaped by their area of expertise and the institutional logics associated with that expertise. Moreover, it is quite likely that the institutional logics used differ systematically across stakeholder groups, which would also significantly shape the cognitive schema that each group employs. Investigating how these cognitive schema vary across stakeholder groups and why (e.g., different institutional logics leading to different categorization schema) is a potentially fruitful avenue of future research.

One important theoretical boundary condition that is important to note is that the context for this setting is fundamentally different from the settings traditionally used by researchers studying categories in management. Namely, I investigate the role of category structure and category embeddedness in the context of search. Again, this differs fundamentally from most extant literature in the area. In traditionally studied settings, it is often implicitly assumed that the use of categories follows a two-step process as described by Phillips & Zuckerman (2000). In the first step, actors narrow the consideration set by keeping only those objects that fit within the focal category. Conversely, in the context of search and innovation, this two-step process doesn't seemingly apply. While understanding what processes unfold and the way that categories and category systems are used in search processes is an exciting direction for future research, we must be careful in assuming that the findings in this chapter hold in traditionally studied settings like valuation. Determining whether this is indeed the case and improving our understanding of the role that larger category structure and category embeddedness plays in the traditional settings could prove to be an important and rewarding challenge for future work.

CONCLUSION

This study took the novel approach of investigating the role that broader category structure has in shaping the relatedness of individual categories and the subsequent evaluation of their members. In pursuing this work, I make three important contributions. First, I demonstrate how the structure of technology classification, and embeddedness within the classification system not only affects the valuation of intellectual property, it also shapes patterns of forward citations.

Second, I add to the small, but growing literature that views categories through the lens of psychology. By doing so, I highlight the importance of larger category structures in the field of category studies. I also introduce network methods to aid in the understanding of these larger structures and leverage patterns of category relationships that these structures represent to understand how category structure affect the outcomes of category members. These larger structures have received very little attention, particularly in the last 15 to 20 years.

Third, I use an empirical design, namely a cross-country analysis of patent equivalents (i.e., patent family members) to investigate questions relating to both the valuation of innovation as well as the effect of categories. Using this approach, I am able to ameliorate many of the issues of unobserved quality differences in inventions. Essentially, I am able to see how the exact same inventions fare in different patent systems. This is also important for the category studies literature. Traditionally, research in this area is forced to investigate a single category system in isolation. However, the approach that I take more resembles an experiment whereby I can investigate differences in different category systems and make more direct comparisons. Essentially I use an identification strategy that is useful for understanding differences across category systems, as opposed to differences with a given system.

CHAPTER 4: THE ANALYST-BASED SIMILARITY APPROACH

INTRODUCTION

When pressed to think of the companies that are similar to, or directly compete with, Blockbuster and Netflix, most consumers would likely cite Amazon (i.e., Amazon Prime) and Redbox, among others. This same pattern would emerge when looking at analyst reports or news stories, whereby analysts and financial reporters would directly cite Blockbuster, Netflix, Amazon, and Redbox as major direct competitors. We might also conclude that these four firms are similar and compete if we looked at the movement of executives and employees among these four companies For instance, Mitch Lowe, a co-founder of Netflix, was a COO and eventually President of Coinstar.[8] Again, a similar pattern in asset acquisitions; Redbox purchased Blockbuster Express in 2012. Finally, the same pattern would again emerge if we looked at overlap in stock market analyst coverage whereby these four competitors are jointly covered by many of the same stock market analysts. However, if one were to determine Blockbuster's direct competitors by looking at the list of companies that share its primary SIC or NAICS Codes, even at the broadest level, then we might be surprised to find out that Redbox would not be on that list.[9]

---

[8] As of Feb. 2010, Redbox is a fully owned subsidiary of Coinstar. In fiscal year 2012, Redbox accounted for more than 86% of Coinstar's consolidated revenue.

[9] According to COMPUTSAT, Blockbuster's primary SIC code is 7841, "Video Tape Rental", and NAICS code is 53223, "Video Tape and Disc Rental." Coinstar's primary SIC code is 3578, "Calculating and

This Redbox example highlights the broadly known reality that industry classification systems, such as the SIC or NAICS, are imperfect representations of the actual underlying industries and competitive landscapes. Despite this frustration, researchers, especially those conducting large-scale cross-industry research, have largely had their hands tied as few, if any, better feasible alternatives exist. Channeling Winston Churchill, Management researchers might complain that SIC and NAICS codes are the worst classifications systems, except for all of the others. In the pages that follow, I hope to convince readers that there is still hope by introducing a new empirical approach that not only creates a fine-grain continuous measure of similarity among pairs of companies, but also produces an industry classification of companies that scholars may use in conjunction with, or as an alternative to, current industry classification systems such as SIC or NAICS.

Empirically, the approach that I introduce, which I call Analyst-Based Similarity (ABS), leverages the coverage patterns of sell-side stock market analysts. Imbedded in these coverage patterns are analyst assignment decisions made by investment banks that consciously aim to reduce information acquisitions costs by assigning analysts to similar sets of firms (see Zuckerman, 1999 and Zuckerman, 2000 for a detailed explanation). These decisions result in analysts coverage patterns that reflect underlying similarities among firms in the eyes of investment banks and the broader investing community. The

---

Accounting Equipment", and NAICS code is 333318, "Other Commercial and Service Industry Machinery Manufacturing." Additionally, according to the COMPUSTAT Industry Segments File, Coinstar's "Redbox" business segment is assigned SIC code "3578" and NAICS code 333313, "Office Machinery Manufacturing."

underlying logic for the ABS is simple: If two firms are covered by the same analyst(s), then this would signal at least some degree of similarity between these two firms. Thus, patterns of overlapping coverage can be used to construct a similarity measure among all possible combinations of firms. The resulting matrix of similarity measures can then be used in conjunction with empirical clustering techniques to create groupings of similar companies.

Theoretically, I build ABS from the ground up based primarily on the theory and insights of both newer and older research in the social categorization literature. While this measure in many ways complements the research on Prototype Theory (Rosche, 1973; Rosch & Mervis, 1975) and subsequent work on fuzzy categories and grades of membership (GoM) (see Hannan, 2010 for an overview), it also fundamentally deviates from this work by going back to Wittgenstein and implementing his original conception of Family Resemblance (Wittgenstein, 1853) that does not rely on category prototypes or category ideals. There are two main ideas that I introduce in this paper. First, classification and categorization may benefit when based on Family Resemblances and local comparison. Second, overlap in stock market analyst coverage implies local similarity between companies and these coverage patterns can be used to construct a similarity measure that has a number of empirical benefits over existing classification systems.

I structure the rest of the paper in the following manner. In the following section, I discuss the main limitations of traditionally used industry classification systems, i.e. SIC and NAICS, and explain why an Analyst-Based Similarity approach might prove to be advantageous as an alternative or complement in many empirical or theoretical contexts. I then review some of the literature in social categorization and discuss the polythetic approaches that have been introduced. I also situate Analyst-Based Similarity (ABS) in the most recent empirical polythetic approaches and highlight how ABS integrates and extends these approaches along with the benefits that this integration entails. I then describe in detail the empirical methodology of the ABS approach in constructing both dyadic measures of similarity among companies, as well as how it can be used to generate industry classifications and present examples that demonstrate the value of this measure in comparison to SIC and NAICS. I then detail the empirical setting and approach that I take to validate this measure where I use ABS, as well as SIC and NAICS-based measures, to predict market reaction to M&A deal announcements. In the last two sections, I describe the results from the empirical test and discuss the implications and potential uses for this new measure.

EMPIRICAL MOTIVATION AND THEORHETICAL FOUNDATION

Limitation of SIC / NAICS and Similar Classification Systems

The Standard Industrial Classification (SIC) and its newer replacement, the North America Industrial Classification System (NAICS), have worked overtime for scholars conducting empirical research in the area of corporate strategy. In particular, researchers studying corporate diversification have relied almost exclusively on SIC/NAICS as a taxonomy of industrial output to determine relevant sets of peer companies (i.e., industry groups) and to measure and evaluate corporate diversification. Indirectly, SIC/NAICS codes have also played a key role in measuring the average value of firms in a given industry, which is then used to calculate "excess value" for multi-industry firms, which is often the central dependent variable in much diversification research. While using SIC codes have greatly benefited scholars, I propose an alternative and potentially complementary approach for classifying and grouping companies based on overlapping stock market analyst coverage that offers a number of theoretical and empirical advantages over the SIC, the NAICS, and potentially other similar systems in use today.

Theoretically, much of the diversification literature is concerned with how firms are perceived by financial markets. A taxonomy of companies based directly on the coverage patterns of stock market analysts, important and highly influential mediators of these financial markets, should more closely resemble the actual underlying mental models of financial market participants and more aptly represent an investment-based view of firms. In particular, investment banks assign analysts to cover sets of firms they view as similar to reduce information acquisition costs (Ramnath, 2002).

This approach is in contrast to the SIC system (and its newer replacement, the NAICS) which was created by the U.S. Office of Management and Budget as a way of classifying economic activity for government reporting and arguably represents a more production-based view of firms.[10] Indeed, the SIC and NAICS are based on input-output tables. This is not to say that the SIC system and other production-based taxonomies are inferior. Rather, I argue that these different taxonomies and approaches best represent the mental

[10] The NAICS was created in collaboration with Statistics Canada and Mexico's Instituto Nacional de Estadistica y Geografia.

models of different stakeholder groups. Specifically, the Analyst-Based Similarity taxonomy that I propose better represents investors and other external stakeholder groups while SIC/NAICS potentially better represent internal stakeholder groups. Depending on the research questions and empirical context of the researcher, either the SIC system or analyst-based system may be more appropriate. However, when researchers are most concerned with financial market perceptions, I argue that the Analyst-Based Similarity approach introduced here is theoretically more attractive than the SIC and other similar system approaches as the Redbox, Netflix and Amazon example illustrates.

Empirically, the Analyst-Based Similarity approach also offers some advantages over SIC/NAICS, or at least does not suffer from some of the SIC/NAICS' known faults. First, the SIC system (and its newer replacement, the NAICS) only covers companies that report in North America. This means that SIC codes aren't always readily available for foreign companies. Moreover, even when SIC codes are available, they might not be consistent among different sources which may have a material impact on the results of studies depending on what source researchers use (Guenther & Rosman, 1994).[11] Aside from being beholden to data providers to provide industry classifications, researchers might also encounter problems with correspondences between the SIC system and other classification systems. This can be particularly problematic for scholars working in the area of Multinational Management, or those conducting analyses on firms that span many

[11] In fact, Guenther & Rosman (1994) find a surprisingly large degree of disagreement in the SIC codes assigned to companies between the COMPUSTAT and CRSP databases. This disagreement stems from the different approaches taken by COMPUSTAT and CRSP's data provider in determining the primary SIC code of a company (see Guenther & Rosman, 1994, p. 118).

countries and regions. The Analyst-Based Similarity approach avoids this issue as analyst coverage is international in scope.

Second, SIC codes, even at the 4-digit level, are still very broad and only allow for very rough measures of similarity among firms (often these measures are discrete). For instance, the similarity of two companies is often based on whether they share the same 2-, 3-, or 4-digit SIC code (or 2-, 4-, or 5-digit NAICS code). Furthermore, scholars need to be especially creative when measuring similarity among firms that share the same 4-digit SIC code or 5-digit NAICS code, which are both still very large groupings. In other words, the SIC/NAICS offer only broad groupings of firms. Conversely, the proposed Analyst-Based Similarity approach yields smaller, more fine-grained groupings of firms allowing researchers to focus on potentially more relevant peer groups. Moreover, the approach also yields fine-grained, continuous dyadic measures of similarity among firms and even among industry categories.

Third, the Analyst-Based Similarity approach can better accommodate firms that are related vertically or those that participate in many different industries (i.e., multi-business firms). This is due to the fact that if the same set of analysts cover companies across vertically related industries, then the Analyst-Based Similarity measure will reflect this (see a more in-depth discussion of this later in the text). SIC and NAICS, on the other hand, treat vertically related companies as belonging to separate industries and deems them to be unrelated. Furthermore, scholars have also identified substantial issues with classifying multi-business firms into a single SIC code (e.g., Clarke, 1989). The SIC system determines a company’s primary SIC code by the industry segment that has the highest sales. This can be very misleading for large multi-business firms such as Virgin group which operate in many different businesses. The Analyst-Based Similarity approach avoids this problem as it is based entirely on analyst coverage. If large, multi-business firms are covered by the same analysts, then the proposed approach will group them together. In other words, if investment banks who employ analysts view companies as being similar enough to warrant coverage by the same analysts, then the companies are likely to be highly similar and are grouped together accordingly.

Finally, the Analyst-Based Similarity approach is dynamic in nature allowing for groupings and the overarching hierarchy to adapt and change naturally over time following changes in the underlying industries. Likewise, the approach can accommodate even yearly changes in the classification taxonomy. An example of this is how analyst coverage changed with respect to Coinstar and Netflix between 2008 and 2012 (see appendix 1 figure A). Redbox is an automated retail kiosk company whose largest business is DVD and Video Game rental kiosks that compete directly with Netflix and Blockbuster. In 2013, Redbox had more than 50% of the US disk rental market. In 2005, Coinstar bought a large 47% stake in Redbox. In 2009, Coinstar acquired the remaining shares of Redbox, making it a fully owned subsidiary. Prior to 2009, there has never been any overlapping coverage between Coinstar and Netflix. However, in 2009, 4 analysts start covering both companies simultaneously, with 11 and 10 analysts having mutual coverage in the following years representing approximately 20% of the analysts covering either firm. With the full acquisition of Redbox by Coinstar in 2009, Coinstar became a more direct competitor with Netflix, and this was captured by the increase in analyst coverage in the same and following year. As the Analyst-Based Similarity measure is directly derived from the rate of overlapping coverage, it is able to capture this change over time and does so with little delay. On the contrary, the SIC was last updated in 1987 and its replacement, the NAICS, created in 1997 is only updated every 10 years. Given that these systems are updated infrequently (and SIC not being updated at all), it is unlikely that they will reflect more recent changes in the industry landscape, especially more recent changes. Even today, SIC and NAICS systems indicate that Coinstar and Netflix operate in completely unrelated industries.

Social Categorization Theory and Grades of Membership

Earlier research in the area of the social categorization of markets focused on the importance of categorical boundaries and the penalties faces by agents that spanned these boundaries. This area of work was largely kick-started by Zuckerman's seminal paper on the "Categorical Imperative" (Zuckerman, 1999), which argued that actors that crossed categorical boundaries, i.e., boundary spanners, were less cognitively understood, garnered less attention, and ultimately fell victim to an illegitimacy discount. Subsequent work sought to gain a better understanding of how the properties of the boundaries themselves were due to institutional factors (Zhao, 2008), characteristics of the boundary spanners such as identity with an organizational form (Carroll & Swaminathan, 2000), or past boundary spanning behavior (Rao, Monin, Durand; 2005) affected the relative size of the illegitimacy discount.

An important implicit assumption of much of this work was that objects or actors were either category members or they were not. The approach taken by these scholars was fundamentally monothetic in nature. Categories are defined on the basis of a set of required characteristics. If an object had all of these required characteristics, then the object is considered to be a member of the category. On the other hand, if an object has either none of the characteristics, or only a subset of the characteristics, then the object is not considered to be a member of the category. To put it a different way, a monothetic approach does not allow for varying degrees of membership in a single category as category members much posses all the defining characteristics that define the category.

More recent research, on the other hand, has turned toward a polythetic approach; categories are defined on the basis of objects possessing a subset of commonly held characteristics, without the requirement of objects collectively possessing any specific individual characteristic. In this recent move toward a polythetic approach, scholars have shifted their attention away from studying the boundaries themselves and have instead focused on issues surrounding membership within category boundaries. For instance, this very recent area of work has largely fallen under the umbrella of what might be called the prototype view of categories (Rosch, 1973; Rosch & Mervis, 1975), which takes a decidedly more lenient view of category membership than in prior work and stems from the work of Wittgenstein (1953).

The main assertion of the prototype approach is that category membership may potentially be only partial, and categories themselves fuzzy. Membership is graded and based on the degree that an object shares common characteristics with a real or fictional category prototype, or category ideal; the category prototypes posses all of the fundamental characteristics that describe the category. So, instead of category members all possessing all of the fundamental characteristics (and thus being either a full category members or not), category membership is based on the degree of similarity between objects and the category prototype. To put it another way, category membership is based on the number and importance of the fundamental category characteristics, as embodied by the prototype, that the object posses. This means that an object might not be a category member, might have varying degrees of partial membership in the category, or might be considered as a full member. In math parlance, set membership isn’t a binary zero or one, but rather measured on a continuous unit scale between zero (not a member) and one (full member). As the degree of similarity between an object and the category prototype increases, so does the partiality, or grade of membership, of the object with respect to the category. Objects with a high grade of membership or typicality are seen as more typical of the given category and are more easily identified as category members.

As Rosche's original research in prototype theory was based on Wittgenstein's (1953) Family Resemblances ("Familienaehnlichkeit"), both her work and the research that followed in her tradition share many similarities with that of Wittgenstein. For example, membership in a category isn’t entirely determined by possessing all the fundamental characteristics of the category. However, there is one fundamental deviation. Wittgenstein's family resemblance does not assert that categories have a category prototype or category ideal by which all category members are compared. In fact, the concept of family resemblance emphasizes that there is no specific set of characteristics that define a category, and thus no category ideal or category prototype must exist.

Essentially, the difference between Prototype Theory and Family Resemblance is where the comparisons are made. In the former, comparisons are made between an object and a category prototype. In the latter, comparisons are made between the considered objects themselves. When the compared objects are highly similar (i.e., share a number of characteristics), then local connections are made. These connections lead to chains of objects which all belong to the same category via local resemblances.

This is echoed in Wittgenstein's (1853) example of the category "games." In the example, Wittgenstein highlights that it is impossible to come up with a common set of attributes that encapsulate all different types of games (e.g., board games, card games, Olympic games, ball games, etc…). Implicitly, this means that no category prototype, or set of fundamental characteristics can be articulated. Rather, we see the category games as a collection of dyadic-level similarities. As you move through the dyadic similarities – e.g., board games, which share many similarities to card games, card games which share many similarities to ball games, etc… - you end up with a chain of dyadic relationships where the elements at one end of the chain may not share the same characteristic features with the objects at the other end of the chain. Similarly, as one moves from one end of the chain to the other, the characteristic features or shared characteristics may change. Thus, categories based on family resemblance may include objects that don't necessarily share the same set of characteristics. While Prototype Theory and GoM research allow for partial category membership (i.e., when they posses some, but not all of the prototype's characteristics), they require that such a category ideal exists and that there is only one ideal for each category. This is not the case with Family Resemblances. Rather, the power of the Family Resemblance logic is that there need not be any category prototype or ideal. While the Analyst-Based Similarity measure is constructed using a Family Resemblance logic that differs in some ways from Prototype Theory, empirically they share a number of the same challenges.

Empirical Approaches to Prototype Theory, Fuzzy Sets, and Grades of Membership

The emergence of the prototype approach to categories has brought with it a number of empirical challenges that scholars are now addressing. In particular, scholars have had to move from using crisp sets to fuzzy sets, which has required a number of necessary adjustments and changes to the empirical approaches taken.[12] Hannan (2010) describes fuzzy sets in detail and how to port some of the past incites and theories from the work based on crisp sets into a fuzzy set framework. Hannan (2010) also outlines some of the empirical approaches to measuring grades of membership that scholars have used in the last few years and suggests a typology of these approaches by grouping them into 5 different themes that I will discuss in turn: inference, similarity, self-claim/label, ties, and audience/critic assignment.

[12] A fuzzy set is a set in which elements may have varying degrees of membership in the set, usually measured on the continuous interval [0,1]. A crisp set is a set where membership is binary, representing either no membership or full members. A crisp set is a special case of fuzzy sets.

The first theme, inference, relies on researchers to infer category membership based on the observable characteristics of companies or products. This approach requires researchers to not only make strong assumptions about not only which characteristics are the most salient to audience members and producers, but also about which categories even exist or are relevant. Scholars then use these salient characteristics of the companies or products to asses and measure the degree of partial membership in the social categories they identified. For instance, Negro et al. (2008), identified two important categories of wines, Barolo and Barbaresco, along with the most salient characteristic differentiating between these two categories of wine as the type of barrel used to age the wine. The grades of membership for wineries for these two types of wine were then calculated as function of the relative use of the two different types of barrels that were representative of the two different styles among the portfolio of wines offered by the winery.

The second theme, similarity, calculates grades of membership based on the degree of objective similarities in the characteristics of the producers or product offerings. For instance, Pontikes (2008; 2012) includes a similarity-based measure of grades of membership on category labels as determined by the similarity of firms in a knowledge space constructed from patent citations. Similarly, Boone and colleagues' (2009) use formally defined characteristics of music (i.e., conception of music, technical ingredients, and style) in their investigation of the failure of modernistic music to emerge as a legitimate category in inter-war Brussels. A related approach couched in the theme of similarity would be to construct a grade of measure based on the agreement from different sources on the assignment of products or companies into different categories. For instance, Hsu (2006) who measured a general film-level grade of membership to film genres, audience consensus, based on genre assignments from three different external and independent sources. Audience consensus for each film is calculated based on an aggregated measure of pairwise agreement among the three external sources. The more similar were the assignments to genres by the different sources, the higher the measure of audience consensus.

The third theme, ties, bases measures of grades of membership on the relationships among focal actors and external actors or organizations. The assumption in using this approach is that similar actors and organizations will have similar associations with both external and other internal actors and organizations (DiMaggio, 1986). In investigating the legitimation of the accountant label among early Dutch accounting firms, Bogaert et al. (2010) calculate the grade of membership of each accounting firm to each of 16 different professional accounting organizations based on the proportion of ties that the firm has with each of these professional organizations; each tie is represented by an individual accountant working at the firm being a member of one of the 17 professional associations. Firms where a large number of its employees were members of a given professional association had a high grade of membership with regard to that association. Kovacs (2009) takes a similar approach and introduces a similarity measure of universities based on internet links from common third-party webpages and demonstrates that this measure of similarity is positively related to the degree of competition over the same students. As the relative number of common ties from third-party webpages increases, so does the measure of similarity between two universities.

The fourth theme, self-claim/label, calculates grades of membership based on the claims and associations that producers or sellers themselves make to different category labels. In investigating the success of online peer-to-peer lending applicants, Leung & Sharkey (2013), measured grade of membership via the number of categorical affiliations that applicants could claim on their loan applications. The more categorical affiliations that a loan applicant claims, the lower the applicant's grade of membership in any individual category. Pontikes (2008) also included a grade of membership measure that is based on self-claims by analyzing the press releases of software producers. Grade of membership for a software producer in a category label was measured by the relative frequency by which the producer included the category label in their press releases – i.e., the number of times referencing a given category label divided by the number of references to any category label. Also, Kuilman & Wezel (2008) measured grades of membership in the passenger airline industry based on a weighted measure of the inclusion of seven different category labels (e.g., airlines, airways, sky) in the company's name in the early nascent period of the aviation industry.

The fifth and final theme, audience/critic assignment, bases measures of grades of membership on either external audience or market critic's assignments to genres. In the food service industry, Kovacs & Hannan (2010) measure the grade of membership as a function of the category labels assigned by a third-party review website and the similarity of category labels based on the frequency that labels appear together. Hsu et al. (2009), using similar data, but a somewhat different approach than in previous work (i.e. Hsu, 2006), measure a film's grades of membership in a given genre by the proportion of three archival external sources that associate the film with that genre.

The Analyst-Based Similarity Approach

The Analyst-Based Similarity measure that I introduce here integrates a number of the prototype theory approaches taken in the previous literature: namely ties, audience/critic assignment, and similarity, yet stays true to Wittgenstein's original conception of family resemblances. ABS measures similarity among companies based on the direct and indirect ties stemming from financial market critics. First, ABS leverages sell-side stock market analysts, important market mediators in financial markets (Zuckerman, 1999; 2000; 2003). This approach follows the suggestion made by Hannan (2010) that the more promising avenues of future research in determining categories and measuring category membership lie in the audience/critic assignment and self-claim/label approaches. However, instead of following past researchers who have used audience/critic assignments of organizations to categories directly, I instead look at the company coverage patterns of these critics as an implicit measure of the degree of similarities among companies. This allows for a dynamic and less restrictive measure of categories. If perceived category membership changes over time, or if the meanings of category labels changes over time, analyst coverage patterns will change, and the ABS measure will incorporate these changes in a seamless way that can't be duplicated to the same degree by externally supplied classifications, whether from critics, audience members, or third parties, which are typically static in nature.

Second, in practice, the method is quite similar to the tie approach taken by Kovacs (2009), who measures the dyadic similarity of universities based on common links by third-party websites. I, on the other hand, measure the similarity of publically traded firms based on common coverage by third-party stock market analysts. In the financial market setting, stock market analysts are specialized by industry to reduce information acquisition costs (Zuckerman, 1999; 2000; 2003). Therefore, if two firms are covered by the same stock market analyst, those two firms are likely to be similar and compete in some of the same markets.

One of the main benefits of a tie based approach is that it can easily be dynamic in nature and allow for a greater degree of freedom in not only the measures of similarity among producers, but also in terms of changes in the meanings of categories themselves. The tie approach only looks at similarity and is not based directly on category characteristics, producer characteristics or category meanings. This means that as the symbolic meaning of categories change, it will lead to changes in the associations among producers, and in turn, be picked up and reflected in the tie approach. As Pontikes (2008) points out, similarity is the foundation by which audience members can converge on shared label meanings; classification emerges from audience perceptions of similarity (DiMaggio, 1986) especially when groups are small or new and emergent (Blau, 1982). The dynamic aspect of this approach is an especially important characteristic when the aim of researchers is to understand the emergence and dissolution of categories and labels, which Kennedy and Fiss (2013) highlight as one important avenue for future of category research which is also of high importance to strategy researchers.

Third, ABS ultimately measures dyadic similarity among firms. The end result is a continuous measure of the perceived similarity of firms from the perspective of financial market critics and, by extension, customers in financial markets. Again, the ABS approach parallels the work using the similarity approach, but differs in one important way. Instead of measuring similarity from objective characteristics of the firms in the same way that Boone et al. (2009) do for music and Hsu (2006) does for films, I derive similarity from the critics themselves. The major benefit in avoiding objective measures is that ABS does not rely on potentially risky assumptions about which characteristics are the most salient for audience to categorize firms. The ABS is also theoretically more appealing as it fits closer with Wittgenstein's conception of family resemblances whereby no single characteristic, or set of characteristics, can fully specify category membership. Instead, similarity is constructed from a number of local dyadic comparisons and companies are seen as related even if they are linked indirectly through chains of family resemblances. For instance, if one analyst covers companies A and B, another analyst covers companies B and C, and another covers companies, C and D, this approach will pick up the fact that companies A and D are somewhat related indirectly via family resemblances, even if they don't share in common the attributes that make them separately similar to companies B and C, respectively. The result is a more fine-grained measure of similarity that will also pick up similarities among indirectly related, yet still somewhat-similar producers or products that standard similarity approaches will overlook.

In summary, the Analyst-Based Similarity approach follows in the traditions of the work in fuzzy categories, particularly that of Wittgenstein Family Resemblances and makes some empirical headway in this area of scholarly inquiry. Integrating some of these different approaches taken by previous scholars – specifically, combining the tie and similarity approaches and using data from critics - yields many of the benefits from the different approaches while avoiding many of the pitfalls when the same approaches are used in isolation. The Analyst-Based Similarity approach results in a dynamic, fine-grained measure of the similarity of companies that reflects the perceptions of important critics in financial markets both within and across companies in classically defined industries.

Construction of the Analyst-Based Similarity Measure

The Analyst-Based Similarity Measure was constructed using the Thomson Reuters' Institutional Brokers Estimate System data files (IBES). IBES covers over 70,000 publicly traded companies from more than 2700 global stock exchanges spanning 90 countries and 5 continents. This data includes sell-side stock market analyst coverage dates of individual stocks – i.e., the dates when an individual analyst initiates coverage on an individual stock along with coverage termination dates, when coverage by the same analyst ceases. The data used to construct the similarity measure will include all analyst coverage between 1990 and 2012.

Broadly speaking, local similarity between firms will be measured by the overlap in sell-side stock market analyst coverage. I start off by constructing a simple symmetric measure of similarity between publically traded firms based on the Jaccard (1901) similarity index that includes only direct overlapping coverage by analysts. The Jaccard measure of similarity, $s_{j,k}$, between company j and company k is calculated as:

$$s_{j,k} = \frac{\left\| \mathrm{A}_j \cap \mathrm{A}_k \right\|}{\left\| \mathrm{A}_j \cup \mathrm{A}_k \right\|}$$

Where A_j is the set of analysts covering company j and A_k is the set of analysts covering company j. The matrix S is a company-analyst similarity matrix whose j,k'th element is equal to $s_{j,k}$. The numerator of the similarity measure represents the number of analysts that both company i and j have in common. The denominator represents the total number of analysts covering either company j or company k or both. The similarity measure can take values anywhere on the unit interval. High values indicate high similarity between companies with a value of 1 indicating perfect overlap in analyst coverage. As the proportion of analysts that companies j and k have in common increases, so does the measure of similarity between the two companies. Holding fixed the number of analysts that companies j and k have in common, an increase in the number of analysts covering either company j or k that don't also cover the other company will cause the similarity measure to decrease. In essence, the larger the proportion of common analysts to total analysts, the larger the measured similarity between the two companies. For example, if two companies were covered individually by 15 analysts, of which 10 analysts covered both companies, then the similarity measure would equal $\left(\frac{10}{20}\right)$ or 0.50. If one of the analysts covering company j started covering company k, then the similarity measure would equal $\left(\frac{11}{20}\right)$ or 0.55. Alternatively, if 5 extra analysts started covering company j but didn't also start coverage for company k, then the similarity measure would drop to $\left(\frac{10}{25}\right)$ or 0.40. Appendix 1 figure A shows how this measure changes for the similarity between Coinstar and Netflix over a 6 year period where the '% of total' column is the measure of similarity, s, between these two companies.

I then incorporate Wittgenstein's Family of Resemblance logic. If we think about the Jaccard measure as measuring the degree of local similarity between firm dyads, we can then compute similarity among firms that don't directly share analysts in common via chains of local dyad similarities. Consider three example firms, A, B, and C who each are covered by 4 analysts. Suppose that firms A and B have 2 analysts in common, analyst similarity = $\left(\frac{2}{4}\right) = 0.5$. Also suppose that firms B and C have 1 analyst in common, analyst similarity = $\left(\frac{1}{4}\right) = 0.25$. I operationalize the Wittgensteinian logic by multiplying the similarities along the path from A to B and B to C to get the analyst similarity between A and C = $\left(\frac{2}{4}\right) * \left(\frac{1}{4}\right) = \left(\frac{1}{8}\right)$. For any given pair of firms, I then use this multiplicative similarity logic to calculate the indirect similarity along all possible paths between the two firms, keeping the highest value as the measure of similarity. In rare cases this value is higher than the direct measure, which I then replace. After doing this calculation for all possible company pairs, I'm left with an similarity matrix, AS, whose i,jth element is equal to the maximum relatedness score between firms i and j. This matrix is symmetric and contains values that reside on the unit interval, i.e. from 0 (highly dissimilar) to 1 (highly similar). For companies that don't have analysts in common, the resulting value can be interpreted as the inferred number of analysts they should have in common. A similarity measure of 1 represents maximum similarity (i.e., perfect overlap in analyst coverage) between two firms, where a value of 0 indicates maximum dissimilarity where the companies not only have no analysts in common, but also are not indirectly related by having analysts in common with any possible string of third-party companies.

Analyst-Based Industry Groups

For many empirical settings, I envision the use of the dyadic-level similarity measure, in conjunction with some cutoff value, as a potentially fruitful way to generate sets of related firms that are not necessarily mutually exclusive. However, some settings might require a mutually exclusive taxonomy of companies similar to SIC and NAICS. Again, while the primary benefits of this approach is its dyadic-level, company-to-company, measure of relatedness, for demonstration purposes, I used the analyst based measure to construct a mutually-exclusive analyst-based industry groups for comparison purposes. There are a number of different ways that this may be done, I demonstrate one such approach that can be easily duplicated by other scholars, but easily tailored to their specific needs.

With the analyst similarity matrix in hand, I construct analyst-based industry groups using hierarchical clustering techniques. I constructed analyst-based industries using hierarchical average-linkage agglomerative clustering analysis.[13] This method requires the number of clusters as an input when producing the groupings. Since this approach does not automatically determine the number of groups to be generated, I can create a classification based on any number of groupings. For example, you can generate a classification with 300 groups, which would be comparable to 3-digit SIC codes and 4-digit NAICS codes that have a similar number of classifications. Again, depending on the research question or setting, research can easily use the underlying dyadic similarity measures to create industry groupings with whatever scope is desired (e.g., broader groups or very fine-grained groupings).

In appendix 2, I use the Calinsky & Harabasz clustering criterion to create mutually exclusive groupings and show how these groupings differ from 3 example NAICS industries. Table A1 presents clustering results for the NAICS industry 'Semiconductor and Related Device Manufacturing' (334413). The table displays the Analyst-Based Similarity measure for each combination of 7 companies in the NAICS industry. A value of 1 represents perfect overlap in analyst coverage. The values are symmetric and are, by definition, equal to 1 when comparing a company with itself (i.e., the values on the diagonal). The Analyst-Based Similarity Measure partitions companies in this industry into 3 separate groups (partitioned by lines); note that these groups contain companies with other NAICS classifications which are not shown. The first group includes Intel,

---

[13] This approach can be implanted in Stata using the cluster linkage command where "averagelinkage" is indicated along with a dissimilarity matrix

Texas instruments, and AMD, which are three of the largest semiconductor manufactures in the world. The second group consists of only 8x8 Inc. (at least from NAICS 334413), which is a fabless vendor of semiconductor products that specializes in VoIP and cloud-based communication solutions. The third group includes First Solar, Sunpower, and Evergreen Solar, which manufacture solar panels, which are a specialized semiconductor product. While there is a fair amount of overlapping analyst coverage among all of these companies (evidenced by high analyst-based similarity values), the three groups that the analyst based measure produces appear to partition the industry into even more sensible sets of competitors. Table A2 displays similar partitions for the more generic NAICS industry ‘R&D in the Physical, Engineering, and Life Sciences’ (541710). As we can see from the included descriptions of the companies, this industry contains many seemingly unrelated groupings of companies, which is reflected by the low analyst similarity values between companies in the indicated groupings. For instance, this NAICS industry contains pharma and biotech companies (i.e., group 1: Arena, Exelixis, and Keryx), companies that assist pharma and biotech companies (group 2: Covance, Charles River Labs, and Pharma Products Development). However, this NAICS industry also includes companies that seem unrelated which are grouped by themselves: group 3 which includes Metabolix (develops bio alternatives for the plastics, chemical, and energy industries), group 4 which includes Microvision (a high-resolution laser display company), and group 5 which includes Syntroleum (develops synthetic liquid hydrocarbons). On first blush, while the groupings seem reasonable, there doesn’t appear to be much competition across these groups, as reflected by low values of analyst similarity among companies in different groups, despite all belonging to the same 6-digit NAICS industry.

Table A3 displays a similar partitioning of companies in the SIC industry ‘Miscellaneous Amusement & Recreational Services (7990). The three groupings include Casino Hotels (MGM Resorts, Wynn Resorts, and Las Vegas Sands), Casinos (Pinnacle entertainment, Boyd Gaming, Ameristar Casinos, Penn National Gaming and Isle of Capris Casinos), and skiing and outdoor operators (Vail Resorts). Again, we see similar patterns with the other two examples, where the analyst-based industry groups are very sensible.

Finally, appendix 3 table A4 lists all of the companies that are most similar to MGM resorts in descending order according to their analyst-based similarity. The second column list the similarity measure of the given company to MGM Resorts (i.e., ‘D(MGM)’). The third and forth columns include the NAICS industry of the given company along with the NAICS code description. As we see from the table, the two closest companies are Las Vegas Sands and Wynn resorts, direct competitors which also operate casino hotels. The next two companies, Isle of Capris Casinos and Pinnacle Entertainment, which only operate casinos. Next is Starwood Hotels & Resorts, which operates hotels and motels. Next are a collection of companies that are vertically related to the hotel or casino businesses. All of these companies seem to be related to MGM Resorts in intuitive ways. However, if we look at the associated NAICS codes to these companies, only 2 share the same 5-digit NAICS, one shares the same 4-digit NAICS, and the rest share at most 2-digits. So according to NAICS, most of these companies appear to be seemingly unrelated. This example shows not only how the analyst based measure can capture vertically related companies, but can also be used with different value thresholds to capture larger and larger groupings of companies by relatedness. If one wanted to create very narrow groupings, and a highly-related comparison set to MGM, a threshold of 0.9610 would capture other casino and hotel operators. If one wanted to include vertically-related companies, a lower value would result in a grouping that includes support companies.

Overall, these examples demonstrate some of the benefits of the Analyst-Based Similarity approach. Appendix 2 show how the analyst approach results in more sensible and smaller, more fine-grained industry groupings than is possible using 6-digit NAICS or 4-digit SIC codes. Appendix 3 shows how the Analyst-Based Similarity approach captures highly-related companies that NAICS misses as well as the flexibility to use different similarity threshold values to create groupings of different degrees of relatedness, or to additionally capture vertically related companies. Also, as discussed previously, appendix 1 demonstrated the dynamic nature of the Analyst-Based Similarity approach and how it can capture subtle changes in industry membership on a yearly basis. Next, I validate the underlying dyadic-level measure used to create these industry groupings.

DATA & METHODS

Empirical Approach

To demonstrate the value of this new measure, I apply the dyadic-level Analyst-Based Similarity measure to the setting of financial market reactions to M&A announcements.

In particular, I follow the methodology used by Schijven & Hitt (2012), but use a more recent data set. Due to data limitations or potential coding issues, I wasn't able to include all of the variables used in the original study. However, I was able to include most. Moreover, the base model that I use has nearly the same adjusted r-squared as the comparable model used by Schijven & Hitt.

Data Sample

First, I collected the set of completed merger and acquisition deals using Thompson Financial's SDC Platinum database. I included all deals from 2000 through 2012 undertaken by publically-traded U.S. acquirers across all industries that resulted in complete, 100%, ownership of other publically-traded U.S. company targets. Second, I collected information on the acquirer and target companies for these deals using Standard & Poor's Compustat North America database. Third, I calculated abnormal stock market returns using daily stock market price data from the Center for Research in Security Prices (CRSP). This resulted in a set of 1,611 deals with complete information on all included control variables. However, not all acquirer and target firms had sell-side analyst stock market coverage, and analyst similarity measures, in the year prior to the acquisition. This led to a final set of 1,027 deals made by 662 unique acquirers with full information for all variables, including the Analyst-Based Similarity measure.

Variables

Dependent variable: Acquirer's 3-day abnormal stock market return. I follow Schijven & Hitt (2012), who used acquirer abnormal stock market returns to proxy for investor stock market reactions to deal announcements. In particular, I used the cumulative abnormal return (CAR) for the acquirers' stock price starting from the day before the announcement through the day after the announcement, or 3-day CAR. CARs represent the unexpected returns from a stock over a given period of time, i.e. the returns that are not explained by general market movements or systematic risk. CARs are calculated by subtracting the expected return of a stock (based on the capital asset price model) from the actual return. The 3-day CAR for each deal was calculated using daily stock market price data from CRSP with the asset price model calibrated using the acquirer's stock price data for the 365 calendar days ending one week prior to the deal announcement.

Explanatory variable: Similarity: I included the Analyst-Based Similarity measure of corporate relatedness as well as two alternative measures of industry similarity for comparison based on the SIC and NAICS industry classification systems. The latter two capture the degree to which the target and acquiring firms share the same primary SIC or NAICS code. *SIC similarity* takes the values of: 1 when both primary SIC codes are the same, 0.67 when they have the same first 3-digits, 0.33 when they share the same first 2-digits, and 0 otherwise. *NAICS similarity* takes the values of: 1 when both primary NAICS codes have the same first 5 digits, 0.75 when they have the same first 4 digits, 0.50 when they share the same first 3 digits, 0.25 when they share the same first 2 digits, and 0 otherwise. Higher values of both variables indicate that the acquiring and target firms belong to more similar industries. I also included *Analyst similarity*. As described above, this variable is based on patterns of stock market analyst coverage and takes values on the unit interval. Larger values indicate higher degrees of similarity between the acquiring and target firms.

Control Variables: I included a number of variables that have been used in the extant literature to model market reaction to M&A deal announcements. In particular, I use Schijven & Hitt's recent (2012) analysis as a starting point for the base model of my analysis. Thus, all of the control variables used in this analysis are taken from their model, with only a few modifications. I used the same variable names to more easily enable comparison and lend to continuity. I should note that I was not able to include a handful of their variables due to data limitations – i.e., 'relative size' and 'acquirer debt-to-equity' were sparsely populated for many deals.

I included a total of 15 control variables, which were directly taken or derived from data from SDC platinum, Compustat, and CRSP. *Merger of equals* is a dummy variable that indicates that the acquirer and target companies have similar market capitalization. *Tender offer* is a dummy variable that indicates that the acquiring firm is buying the target's shares directly from the shareholders. *Divestiture* is a dummy variable that indicates that the target is part of a third party firm's divestiture. *Geographic proximity* is a dummy variable that indicates that the acquiring and target companies are headquartered in the same U.S. state. *Number of competing bidders* is a count of the number of other third-party firms that have made bids for the target company. *Related to recent deals* is a dummy variable that indicates if the focal deal is related to prior deals initiated by the same acquirer. *Target product scope* is a count of the number of SIC codes identified by the target firm. *Number of target advisors* is a count of the number of advising firms that are assisting the target company in the focal deal. *Payment method* is an ordinal variable that takes three possible values (Hayward & Hambrick, 1997): 1 for cash-only deals, 2 for cash & stock deals, and 3 for stock-only deals. *Involvement of acquirer advisors* is a count variable representing the number of stages of the acquisition process that acquirer advisors were active participants (Schijven & Hitt, 2012). *Pre-existing partial ownership position* is a dummy variable that indicates if the acquiring company had at least a 10% prior stake in the target company before the deal. *Defense tactics* is a dummy variable that indicates that the target company adopted 'antitakeover defense mechanisms' in response to the acquirers bid (Schijven & Hitt, 2012); these include golden parachutes, poison pills, etc… as indicated by SDC Platinum. *Acquirer acquisition experience* is a count of the number of completed acquisitions made by the acquiring firm starting from 1990 until the date of the deal announcement. Finally, the a*cquisition premium* is calculated as the percentage difference between the offer price and the target stock price 1 week prior to the deal announcement.

Empirical Model

I used Ordinary Least Squares (OLS) estimation to predict 3-day Cumulative Abnormal Returns (CARs) for the acquirer related to the announcement of M&A deals. Because some companies engaged in multiple acquisitions during the analysis period, I use robust Huber-White standard errors and cluster the errors on the unique acquirer. Doing so allows for a more conservative test of the industry similarity variables. I also include year fixed-effects and acquirer industry fixed-effects. For the latter, I used the 3-digit primary SIC code of the acquiring firm – alternatively using the acquiring firm's primary NAICS code yielded materially similar results for all models. Moreover, the results don't materially change for any of the industry similarity variables even when the acquirer industry fixed-effects are excluded.

RESULTS

Tables 4-1 and 4-2 display variable summary statistics and correlations, respectively. The highest correlation in magnitude is 0.34 among all variables. This suggests that multicolinearity should not be an issue. However, as a precaution for including many interaction variables, I de-mean all of the variables before including them in regression models. However, I should note that I display the original means of all of the variables in Table 4-1.

Table 4-3 presents the OLS regression results. Model 1 is the base model and does not include any of the continuous industry relatedness variables. The only control variable that is statistically significant across all models is ‘payment method.’ The coefficient for this variable is negative, as is expected. Higher values of this variable indicate that the deal is financed to a larger extent via stock. In general, deals that are heavily financed by the acquirer’s stock are worrisome to investors as it is interpreted as a signal that management thinks their stock is more overvalued.

Regression models 2 through 4 separately include the three different industry similarity variables and their interactions with acquisition premium. Models 2 and 3 include industry similarity based on primary SIC and NAICS, respectively. The coefficients for all industry similarity variables in both models are not statistically significant at the 0.05

percent level. Model 4 includes the Analyst-Based Similarity measures and its interaction with acquisition premium, which are statistically significant at the 0.001 and 0.01 levels, respectively. The positive coefficient of 0.087 for analyst similarity indicates that the acquirers stock performs better when the acquirer and target firms are more similar. The magnitude of this effect is quite striking. The acquirer's 3-day CAR following the deal announcement will be 8.7% higher if the companies are very similar (i.e., analyst similarity = 1) than if the companies are highly dissimilar (i.e., analyst similarity = 0). To put this number in perspective, the average 3-day CAR in the sample is -1.6%. Given that the negative coefficient for the interaction term of analyst similarity and acquisition premium was also significant, Figure 4-1 graphs the predicted CARs for three different values of acquisition premium (i.e., small, medium, and large representing values that are 2 standard deviations below the mean, at the mean, and 2 standard deviations above the mean) over the range of possible analyst similarity values with all other variables at their means. As we can see from the graph, irrespective of acquisition premium, the stock of acquirers performs much better when they acquire target companies that are highly similar according to the analyst measure. It is also worth noting that due to this large effect, the acquirer's stock tends to greatly outperform the market in the wake of deals with a highly similar target companies. Conversely, acquirer's stock tends to underperform the market after deal announcements with dissimilar targets.

As a robustness check, I split deals into 4 separate groups corresponding to deals where the acquirer and target are dissimilar, slightly similar, similar, and highly similar, and used dummy variables that flag each group and included them in model 5 of Table 4-4. As we can see from the coefficients, the effect of analyst similarity appears to be largely linear, which suggests that the model is correct in only included a linear term. Additionally, Figure 4-2 depicts the predicted acquirer 3-day CAR for each of these groups when all other variables are at their means. Again, as we can see from the figure, the effect is linear. We also see that highly similar mergers lead to positive acquirer's stock performance but all others lead to negative acquirer stock performance.

Altogether, the results lend strong support to the validity of the Analyst-Based similarity measure. Analyst-Based Similarity is a better predictor of stock market reactions to M&A deal announcements then those based on SIC or NAICS industry classification systems. This suggests that researchers might also benefit from using Analyst-Based Similarity as an alternative or supplement to SIC or NAICS industry classification.

DISCUSSION

I employ simple network methods and leverage the two-mode network of stock market analysts and covered companies to generate a dyadic-level measure of company relatedness that I call Analyst-Based Similarity. As stock market analysts specialize by industry, coverage of two companies by the same analyst indicates that the companies are related, at least in the eyes of the investment firms. As the overlap in analyst coverage by two companies increases, so does the implied degree of relatedness between the two companies. Network methods are then used to estimate the degree of relatedness among companies that don't share any analysts in common using Wittgenstein's logic of family resemblance. The result of such an approach is a dyadic company-to-company level measure of corporate relatedness that possesses many advantageous properties when compared to traditionally used measures of relatedness used by management researchers such as SIC and NAICS.

Moreover, as the Coinstar and Netflix example demonstrates, the measure is calculated on a yearly basis and changes dynamically over time capturing even subtle changes in industries or industry membership. The classification is also based on coverage decisions by third-party financial market mediators. This means that the measure is not based on company self-reporting, but rather on the perceptions of external stakeholders. As such, the measure is a potentially very valuable tool in settings regarding financial market outcomes, such as shareholder reactions to corporate actions, both by the companies themselves and their perceived peers.

I demonstrated that value of the Analyst-Based Similarity approach in one such setting, prediction financial market returns to M&A deal announcements. The measure proved to be highly statistically significant and explains large differences in 3-day abnormal returns on the order of 2 percent, which is materially substantial. Moreover, the proposed measure performed far better than traditionally used measures of industry relatedness including SIC and NAICS, both of which had little to no explanatory power in this setting, suggesting that the analyst-based measure could be very valuable to management scholars, especially those studying financial market responses, or who conduct event studies based on stock price movements.

One potential downside to the Analyst-based measure is that it is not available for all companies. In particular, the measure can't be calculated among firm dyads where at least one of the companies is not covered by a stock market analyst, or is not covered over the entire window of study. The measure is also less accurate when the company has few analysts, i.e., only one or two. This means that the measure will not be available for many small companies, nor for companies that are not publically traded. In the setting that I used to analyze financial market returns to M&A deal announcements, almost 28% of the deals had to be excluded from the sample because either the acquirer, or much more likely the target, was not covered by at least one analyst in the year prior to the deal announcement. For some situations, the loss of these observations and the potential bias that this creates can be problematic.

Regardless, while the measure is likely to prove valuable in a number of empirical settings, the external shareholder perspective captured by the Analyst-Based Similarity approach may also integrate nicely with other measures of relatedness that are based on the perceptions of other corporate stakeholder groups. For instance, the analyst approach may integrate nicely with measures that are based on internal stakeholder perceptions, such as technological relatedness. Used together, they may be able to offer another explanation as to why managers engage in corporate actions, such as M&A deals, that are viewed unfavorably by shareholders. Accordingly, one potentially fruitful avenue of future research is to study situations where relatedness measures based on different stakeholder groups yield different results.

The findings and arguments made in this and previous chapters also hint at an important caveat for management researchers. Different stakeholder groups may employ different institutional logics when making sense of and evaluating corporate actions and outcomes. Aside from accounting for this theoretically, scholars need to be diligent in ensuring that the industry similarity / company similarity measures that they employ in their empirical research also reflects the appropriate logic by actors in the focal setting. For instance, when investigating target choice and governance form in corporate M&A and alliance deals, the more appropriate measures of relatedness is likely to be grounded in organizing production (e.g., technology relatedness), which best reflects the institutional logics employed by managers. Conversely, when explaining external stakeholder reactions, such as financial market reactions to M&A deals, measures which best reflect the institutional logics used by the relevant actors (e.g. analyst similarity or financial market relatedness) are most appropriate to explain stock price movements or shareholder activism.

CHAPTER 5: CONCLUSION

The main goal of this dissertation was to shed light on both the category studies and multinational management literatures. In addressing the category studies literature, I used two separate studies and empirical settings to demonstrate 1) how mental models of relatedness, or categorization, can differ systematically across members of the same audience, in this case, sell-side stock market analyst, and 2) that the larger structure of category or classification systems and embeddedness within that structure has a heavy hand in directing the attention of audience members, shaping their perceptions of relatedness, and ultimately affecting the realized value of innovations and intellectual property.

In addressing the multinational management literature, I demonstrated how national boarders and the geographic location of external stakeholders can shape external stakeholder perceptions of corporate relatedness and subsequently affect patterns of informational spillovers among multinationals. In the stock market analyst setting, I show that analysts that are collocated with the firms they cover tend to group companies more narrowly and thus ascribe and integrate new information they deem relevant for their earnings forecasts of a focal firm from a smaller set of peers, as compared to analysts who are not co-located. In the setting of innovation, I demonstrated how one seemingly unimportant country-level institutional factor, namely a country's patent office's technology classification system, can affect both the prior art that is deemed relevant as well as the local perceived value of an innovation.

However, there are a number of limitations and boundary conditions to all of this work, particularly with respect to the study involving the valuation of innovation. First, the setting of innovation is somewhat unusual insofar as category and classification systems serve a different function than they have in the traditionally studies settings most commonly seen in the literature. In the setting of innovation, classification systems are used by experts to conduct exhaustive searches through massive amounts of existing prior art. It is critical in this setting that examiners scour all potentially relevant prior art in order to determine the novelty of a given invention. In other words, the classification system, and the underlying mental models of relatedness held by these examiners direct their attention not only to prior art in the same technology classes, but more importantly to other potentially relevant technology classes as well. Additionally, the examiners themselves do not directly asses the value of the inventions that they evaluate. Instead value is determined indirectly through the citations of other examiners. The focus on intentional search outside the domain of a focal category in combination with indirect valuation differs substantially from settings that have traditionally been used. In much of the extant literature, settings generally dictate that audiences are trying to narrow their attention and evaluate objects directly according to the criteria of a single category or to select a single object from a set of options. Thus, one potential boundary condition is that these findings, at least in terms of the importance of category structure and embeddedness, might only hold in settings based on search.

This presents both challenges and opportunities. While the scope of my findings may be limited, they do hint at the need to further study categories and classification systems in the context of search. Moreover, my findings may contrast to the generally accepted and nearly taken-for-granted tenets associated with Zuckerman's (1999) categorical imperative – i.e. that actors benefit most when they fit nicely into a single category; instead my findings show the benefits of category relatedness and multiple-category membership. Additionally, more recent work has also uncovered a number of settings where the categorical imperative doesn't necessarily hold (e.g., Ruef & Patterson, 2009; Granqvist, Grodal, & Woolley, 2013). Taken together, these findings in combination with my own suggest the need to study the different ways that individuals leverage categories and classification systems, which should shed light on the conditions under which these different, and sometimes opposing, incites best apply. My intuition, based on my findings here, is that the context of search differs fundamentally from contexts involving only evaluation. In which case, the former could prove to be a fruitful area of future inquiry as it has received very little attention to date.

Altogether, my findings also complement the work of Vergne & Wry (2014) and other research in underscoring the need to integrate sociological approaches to categories with the less well-explored psychological approaches to categories into a cohesive 'category studies.' In that regard, my chapter on how the geographic location, and potentially the national identity, of stock market analysts affects informational spillovers points to another promising direction for future research. Namely, that individual characteristics and the past experiences of audience members may augment employed category systems and further shape attention and perceptions of relatedness. For instance, I would expect the career trajectories and past experiences of patent examiners to shape what technologies and technology classes they see as relevant to a particular domain. Examiners with experience or exposure to nanotechnology, a very broad area with many applications across myriad fields, might have different mental models compared to those that have experience in narrower, but still related domains.

Accordingly, a very promising direction for future work is to investigate how individual characteristics, experiences, etc… may lead to augmented category schema, particularly with respect to the associations made among the categories themselves, which vary across individuals. Indeed, the additional robustness analysts conducted in Ch. 2 regarding analyst experience, would point me in that direction. While category studies scholars have started to unpack differences in schemas over time (e.g., Glynn Lounsbury, 2005; Rao et al., 2005) and across audience subgroups (e.g., Zhao, 2005), much can be gleaned by also unpacking individual differences in the schemas themselves or their use. This is also a promising avenue for future research.

Switching gears, there is also another more subtle, but potentially critical implications stemming from the analyst-based similarity measure that I introduced that hasn't yet been mentioned. Initially, one major driver to develop this new measure, aside from general dissatisfaction and frustrations with SIC and NIACS classification, was the idea that mental models of industry and corporate relatedness may differ substantially across stakeholder groups. Discrepancies across stakeholder groups might offer another, and more benign, explanation for why managers and corporate executives make decisions that are perceived negatively by shareholders and other external stakeholders. In particular, I intended to argue that my analyst-based measure best captures shareholder perceptions of corporate relatedness, whereas other measures, such as SIC and NAICS systems, which are based on input output tables and production, might better capture internal stakeholder perceptions of relatedness. In other words, irrespective of different stakeholder goals and wants, different stakeholder groups have different perspectives – in this case different perceptions of the corporate landscape and the relatedness of firms – which may affect their behaviors and actions. Thus, scholars should use the most appropriate measure based on the stakeholder group that the dependent variable or outcome under study best represents. A good example of the discrepancy between stakeholder perspectives is the proposed 2004 merger of VNU and IMS Health.

To the corporate executives of both VNU (now The Nielsen Company) and IMS Health, the proposed $6.7 billion merger between the two companies in 2004 made perfect sense. As both companies were large players in the market research space, merging the two would certainly manifest numerous synergies given the large overlap in their businesses.[14] IMS Health and VNU share the same 5-digit NAICS code.[15] IMS Health and Nielsen Media (a subsidiary of VNU) used to have the same corporate parent, Dun & Bradstreet, and were spun off together to form Cognizant.[16] AC Nielsen (also a subsidiary of VNU), IMS Health, and Dun & Bradstreet were even jointly sued by a rival firm for anticompetitive behavior and bundling. Surely the corporate executives thought

---

[14] VNU (Verenigde Nederlandse Uitgeverijen) is an international market research firm that owns Nielsen Media and ACNielsen, which track television audiences and supermarket sales. IMS Health is a market research firm that specializes in the pharmaceutical industry.
[15] NAICS: 541910 (Marketing Research and Public Opinion Polling)
[16] On Nov. 1st, 1996 Dun & Bradstreet spun off IMS Health, Nielsen Media Research, and the Gartner Group together to create Cognizant Corporation.

that given this corporate overlap the sum of the two companies would be larger than the two separate parts. To the executives' surprise, shareholders did not agree, leading to a shareholder revolt ending in a cancellation of the deal and the ousting of VNU's CEO.[17] How could corporate executives have so badly misjudged shareholder reactions? I argue that the answer is simple. Internal stakeholders saw VNU and IMS Health as being highly similar and related given synergies that would develop on the production side of the equation where IMS Health could leverage VNU's expertise and data and apply it to their own industry as indicated in public filings. Conversely, shareholders indicated that the industries that these two companies operated in were too different and dissimilar and that a merger would be foolish.

This example highlights one important issue that has received relatively little attention in the multinational and strategy literatures: that perceived corporate relatedness is subjective and that different stakeholders have different mental models of industries and the corporate landscape. In other words, there may be disagreement across stakeholder groups as to the relatedness among firms, the scope of industries, what companies are members of what industries, and to what degree they are members. It seems true that such differences and disagreements of corporate relatedness across stakeholders exist, as evidenced by all of the different measures of industry relatedness, and this can help explain why corporations engage in activities that are penalized by shareholders.

[17] Major shareholders Templeton Global Advisors, Fidelity Investments International, & Knight Vinke Asset Management (representing 48% of VNU shares) voiced public opposition to the deal.

The broader empirical implications of this finding for scholars in strategy, management, and related disciplines are substantive. If different stakeholders see the corporate landscape and corporate relatedness differently, then researchers will be better served by using empirical measures that best approximate the stakeholder groups most relevant to the research question and empirical setting at hand. In fact, pioneering work by Stimpert & Duhaime (1997) showed that business relatedness depends largely on three different attributes: resource attributes, product market attributes, and value chain attributes. They also found evidence that while one objective measure of relatedness, the Herfindahl Entropy measure, was correlated with manager perceptions of relatedness, managers additionally saw other dimensions of relatedness. Moreover, Nayyar (1992) uncovered empirical evidence that the usual objective measures of relatedness did not capture managerial assessments of relatedness.

These findings, that managers perceive relatedness as multidimensional, highlight the empirical challenge faced by strategy scholars to measure industry and company relatedness when trying to explain diversification decisions and the direction of firm growth. The multidimensionality of relatedness means that the relevant dimension of relatedness that applies to a given diversification decision changes as context and resources change. So in one situation, technological relatedness may be the primary dimension used by managers, whereas product market relatedness may be the primary dimension in another. Naturally, this creates a huge challenge for researchers who may not be able to determine what dimensions most apply.

One solution to this challenge was put forth by Bryce and Winter (2009) who constructed a General Interindustry Relatedness Index. To overcome this challenge, Bryce and Winter use the survivor principle (Stigler, 1968) as a theoretical foundation to measure the relatedness of 4-digit SIC industries. The logic of their measure is that the activity patterns of existing firms best reflect the underlying relatedness of resource baskets among industries through experience and knowledge accumulation. Thus, if many companies span the same two 4-digit SIC industries, then the relevant resource baskets of the two industries must be similar and the two industries must be highly related. Similarly, if few companies span the same two industries, then the relevant resource baskets have little overlap and the two industries must be unrelated. Essentially what Bryce and Winter do is aggregate the diversification decisions of managers, which incorporate the contextual idiosyncrasies of each individual growth decision, and distill them into a general index of industry relatedness. Bryce and Winter argue that this approach captures the structure of resource and knowledge relatedness across industries.

However, since their measure is built on individual managerial decisions of corporate diversification, it can be conceived that their approach and measure actually capture general managerial, or internal stakeholder, perceptions of industry relatedness. Or, given that mergers require shareholder approval, as the VNU example highlights, the measure might be some amalgam of internal and external stakeholder perspectives, potentially, and problematically, merging the two. Moreover, their approach relies on SIC categories, which may not match up with the industries and constituent members conceived by managers and other stakeholders. In this light, and consistent with my expectations that substantial benefits will arise when stakeholder perspectives are measured separately and directly, I conducted my own exploratory analysis using the analyst-based approach that I introduced.

Accordingly, I expected to find that my analyst-based measure, representing the external stakeholder perspective, best predicts financial market reactions to M&A announcements but SIC & NAICS, representing the internal stakeholder perspective, best predicts M&A target choice and governance form choice (e.g., acquisition vs. alliance, or the degree of integration). As I show in chapter 4, the analyst-based measure does indeed strongly predict market response, whereas SIC & NAICS measures are statistically insignificant, as I expected. In the preliminary analyses that I conducted investigating M&A target choice and governance choice, the analyst-based measure proved to be a moderately strong predictor in both settings (though much weaker than in the M&A response setting). However, again SIC & NAICS demonstrated almost no predictive power in the other two settings, which was not expected. I should also note that the Bryce and Winter measure also fared rather poorly in all three settings. While this might suggest a flaw in the underlying logic that there should be a number of different industry and corporate relatedness measures and that scholars would be best served by using the most appropriate and relevant one, my intuition is that this failure to find what I expected was due to the fact that SIC & NAICS were either poor measures in general or, despite the production IO logic for their construction, poorly represent the perspectives of internal stakeholders (the General Interindustry Relatedness Index would also suffer as it is based on SIC). Perhaps a better measure to capture internal stakeholder perspectives, especially in the context of M&A target and governance choice, might be based on technological relatedness such as Silverman's operationalization (1996, 1999), which is something that I am currently investigating, or at least using a similar relatedness measure that is based on patents. Regardless, there is strong theoretical grounding for expecting such a finding and is a promising avenue for future research, not to mention the huge ramifications that this would have for scholars in many areas of management research, especially corporate strategy and the scope of the firm, and multinational management.

A final potential avenue of future research worth highlighting is directed at the international business literatures. In Ch. 2 I showed that a stock market analyst's nationality / geographic location affects their cognitive schema of company relatedness. It would be interesting to explore other individual characteristics, particularly those that are steeped in national differences (e.g., information asymmetries, linguistic, cultural, etc…) that may also affect how stakeholders located in different countries might have different cognitive schema pertaining to perceptions of similarity among companies or even the institutional environments and attractiveness of different countries. In this regard, Porac et al. (1989) demonstration that Scottish producers had an inaccurate representation of the competitive landscape due in part to their geographic location and Jonsson et al.'s (2009) demonstration that local shareholders responded differently in light of a corporate scandal lend evidence that this line of enquiry might prove to be very important.

To conclude, this dissertation has made theoretical and empirical contributions to both the category studies and multinational management literatures. The proposed analyst-based measure in combination with findings throughout the chapters also hint at additional contributions on the near horizon to the corporate strategy literature, especially with regards to the scope of the firm, and other areas of management research. Finally, this dissertation is yet another reminder of the explanatory power in understanding macro-level outcomes by investigating individuals and leveraging individual-level data.

Table 2-1: Variable Summary Statistics

Variable	Description	N	Mean	StdDev	Min	Max
Scaled Change in EPS Estimate	Measure of change in analyst's quarterly EPS estimate. (New estimate – Old estimate) x 100 / (Share price)	6,147	-0.632	7.488	-453.13	62.42
Local Event Company	Indicator if headquarters countries of event and target companies are the same	6,147	0.675	0.468	0.00	1.00
Local Analyst	Indicator if analyst if located in the same country as the target company's headquarters country	6,147	0.725	0.447	0.00	1.00
Local Event Company & Analyst	Indicator if headquarters countries of event company and target company are the same as the analyst's country	6,147	0.577	0.494	0.00	1.00
Analyst Distance to Target	Geographical distance between analyst's country & target headquarters' country – rescaled to [0,1]	6,147	0.041	0.115	0.00	1.00
Event Firm Size	Market capitalization of event firm (bil. USD)	6,147	0.494	3.124	0.00	63.15
Analyst Covers Event Company	Indicator if analyst covers event company	6,147	0.043	0.204	0.00	1.00
Analyst Coverage Load	Number of companies covered by analyst	6,147	12.849	7.459	1.00	84.00
Analyst Target Company Experience	Analyst experience covering target company (years)	6,147	2.432	2.774	0.00	24.86
Analyst Career Experience	Analyst career experience (years)	6,147	6.019	4.519	0.00	29.56
Analyst Broker Size	Number of other analysts working at the same broker.	6,147	11.459	11.452	1.00	52.00
Target Firm US Indicator	Indicator if target company is based in the US	6,147	0.206	0.404	0.00	1.00

Table 2-2: Variable Correlations

		1	2	3	4	5	6	7	8	9	10	11
1	**Scaled Change in EPS Estimate**	1.00										
2	**Local Event Company**	0.00	1.00									
3	**Local Analyst**	0.02	0.41	1.00								
4	**Local Event Company & Analyst**	0.01	0.82	0.71	1.00							
5	**Analyst Distance to Target**	0.00	-0.20	-0.59	-0.42	1.00						
6	**Event Firm Size**	0.00	0.05	0.03	0.05	-0.02	1.00					
7	**Analyst Covers Event Company**	0.01	0.05	0.05	0.06	-0.04	0.01	1.00				
8	**Analyst Coverage Load**	0.00	0.20	0.13	0.21	0.01	-0.04	0.00	1.00			
9	**Analyst Target Company Experience**	0.00	0.03	0.07	0.05	-0.07	0.02	0.00	0.06	1.00		
10	**Analyst Career Experience**	0.00	0.00	0.03	0.02	-0.04	0.01	-0.02	0.20	0.45	1.00	
11	**Analyst Broker Size**	-0.01	-0.19	-0.24	-0.24	0.04	-0.02	0.00	-0.12	0.04	0.06	1.00
12	**Target Firm US Indicator**	0.00	0.30	0.20	0.33	0.00	-0.05	-0.11	0.37	0.07	0.13	-0.19

Table 2-3: Sample Breakdown by Event & Analyst Location

		Event Country & Target Country are the Same		
Target Country & Analyst Country Same		**No**	**Yes**	**Both**
	No	1,090 (17.7%)	603 (9.8%)	1,693 (27.5%)
	Yes	907 (14.8%)	3,547 (57.7%)	4,454 (72.5%)
	Both	1,997 (32.5%)	4,150 (67.5%)	6,147 (100.0%)

Table 2-4: Mean Change in EPS Forecast by Event & Analyst Location

Target Company & Analyst Same Country		Event Company & Target Company Same Country: No	Yes	Both
	No	-0.383 (4.353)	-1.709 (8.978)	-0.855 (6.425)
	Yes	-0.408 (3.905)	-0.582 (8.577)	-0.546 (7.854)
	Both	-0.395 (4.154)	-0.746 (8.644)	-0.632 (7.488)

Numbers in parentheses are standard deviations

Table 2-5: Fixed Effects Regressions on Change in EPS Forecasts

	(1)		(2)		(3)		(4)		(5)		(6)	
Analyst Covers Event Company	0.280	(0.184)	0.395*	(0.191)	0.275	(0.185)	0.366	(0.193)	0.278	(0.184)	0.347	(0.192)
Analyst Coverage Load	-0.011	(0.016)	-0.007	(0.016)	-0.012	(0.016)	-0.009	(0.016)	-0.011	(0.016)	-0.010	(0.016)
Analyst Company Experience	0.000	(0.025)	0.000	(0.025)	-0.001	(0.025)	0.000	(0.025)	0.000	(0.025)	0.000	(0.025)
Analyst Career Experience	0.024	(0.016)	0.020	(0.015)	0.024	(0.016)	0.017	(0.016)	0.025	(0.016)	0.018	(0.016)
Analyst Broker Size	-0.023	(0.016)	-0.024	(0.016)	-0.022	(0.017)	-0.022	(0.017)	-0.023	(0.016)	-0.021	(0.017)
Event Firm Size	0.010	(0.014)	0.015	(0.014)	0.010	(0.014)	0.019	(0.014)	0.010	(0.014)	0.019	(0.014)
Target Firm US Indicator	-0.301	(0.499)	0.077	(0.507)	-0.371	(0.470)	-0.160	(0.466)	-0.295	(0.496)	-0.245	(0.467)
Local Event Company			- 0.836***	(0.239)			-1.776***	(0.516)			-1.785***	(0.516)
Local Analyst					0.181	(0.233)	-0.138	(0.29)			0.046	(0.333)
Local Event & Local Analyst							1.418**	(0.53)			1.489**	(0.538)
Analyst Distance									0.538	(0.484)	1.371	(0.719)

Dependent variable = ([post EPS estimate] - [pre EPS estimate]) * 100 / [share price].
Standard Errors are clustered by analyst and are in parentheses.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 2-6: Change in EPS Forecasts Regression Predictions

Target Company & Analyst Same Country		Event Company & Target Company Same Country: No	Yes	Both
	No	-0.355 (0.052)	-2.140 (0.249)	-1.560 (0.126)
	Yes	-0.310 (0.051)	-0.606 (0.029)	-0.655 (0.023)
	Both	-0.322 (0.029)	-1.248 (0.047)	-0.668 (0.015)

Numbers in parentheses are point estimate standard errors

Table 2-7: Fixed Effects Regressions on Change in EPS Forecasts

	(7)	
Analyst Covers Event Company	0.351	(0.193)
Analyst Coverage Load	-0.009	(0.016)
Analyst Target Company Experience	0.000	(0.025)
Analyst Career Experience	0.064 *	(0.026)
Analyst Broker Size	-0.021	(0.017)
Event Firm Size	0.021	(0.014)
Target Firm US Indicator	-0.246	(0.467)
Local Event Company	-1.758 ***	(0.516)
Local Analyst	0.060	(0.332)
Local Event & Local Analyst (LELA)	1.907 ***	(0.573)
LELA * Analyst Career Experience	-0.072 **	(0.026)
Analyst Distance to Target	1.466 *	(0.719)

Dependent variable = ([post EPS estimate] - [pre EPS estimate]) * 100 / [share price].
Standard Errors are clustered by analyst and are in parentheses.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 3.1: Variable Summary Statistics

	Mean	SD	Min	Max	Source
1. 3-year forward citations (all)	0.85	2.01	0.00	55.00	PATSTAT
2. 3-year forward citations (in-class)	0.36	1.11	0.00	35.00	PATSTAT
3. 3-year forward citations (out-class)	0.50	1.38	0.00	35.00	PATSTAT
4. Neighbors	2.44	1.73	0.00	11.81	CONSTRUCTED
5. Classes	1.78	1.00	1.00	9.00	PATSTAT
6. Backward citations	4.30	7.69	0.00	390.00	PATSTAT
7. Number of inventors	2.60	1.73	0.00	21.00	PATSTAT
8. Number of applicants	1.09	0.40	0.00	12.00	PATSTAT
9. Time to approval	4.72	2.95	0.51	15.87	PATSTAT
10. Primary class size	11.19	16.31	0.00	136.25	PATSTAT
11. Grant year	2003.00	3.89	1995.00	2010.00	PATSTAT
12. German Patent System	0.33	0.47	0.00	1.00	PATSTAT
13. Japanese Patent System	0.33	0.47	0.00	1.00	PATSTAT

Table 3-2: Variable Correlations

	1.	**2.**	**3.**	**4.**	**5.**	**7.**	**8.**	**9.**	**10.**	**11.**	**12.**	**13.**
1. 3-year forward citations (all)												
2. 3-year forward citations (in-class)	0.75											
3. 3-year forward citations (out-class)	0.85	0.30										
4. Neighbors	0.04	0.05	0.02									
5. Classes	0.00	0.05	-0.04	0.02								
6. Backward citations	0.36	0.28	0.30	0.03	-0.03							
7. Number of inventors	-0.03	0.00	-0.04	0.02	0.01	-0.02						
8. Number of applicants	0.00	0.01	-0.01	0.00	0.01	0.08	0.02					
9. Time to approval	-0.34	-0.25	-0.30	-0.05	0.09	-0.41	0.01	0.02				
10. Primary class size	-0.04	-0.05	-0.02	0.43	-0.02	-0.10	0.01	0.02	0.14			
11. Grant year	-0.26	-0.19	-0.22	-0.03	0.05	-0.26	0.01	0.01	0.64	0.08		
12. German Patent System	-0.21	-0.15	-0.19	-0.05	-0.04	-0.22	-0.03	-0.05	0.16	-0.35	0.12	
13. Japanese Patent System	-0.22	-0.17	-0.19	-0.01	0.09	-0.42	0.01	-0.01	0.52	0.42	0.30	-0.50

Table 3-3a: Hybrid Negative Binomial Regressions on Forward Citations:

	All Citations (1)		All Citations (2)	
Neighbors			0.046	***
			(0.011)	
Classes	0.117	**	0.106	*
	(0.043)		(0.043)	
Classes^2	-0.013		-0.012	
	(0.008)		(0.008)	
Backward citations	0.016	***	0.016	***
	(0.002)		(0.002)	
Number of inventors	-0.153	***	-0.158	***
	(0.043)		(0.043)	
Number of applicants	-0.180	***	-0.180	***
	(0.049)		(0.049)	
Time to approval	-0.077	***	-0.076	***
	(0.007)		(0.007)	
Primary class size	0.000		-0.003	*
	(0.001)		(0.001)	
Grant year	-0.043	***	-0.043	***
	(0.004)		(0.004)	
German Patent	-1.461	***	-1.473	***
	(0.036)		(0.036)	
Japanese Patent	-1.550	***	-1.526	***
	(0.044)		(0.044)	
Constant	87.389	***	86.359	***
	(7.800)		(7.802)	
N	25,215		25,215	
Groups	8405		8405	
N per Group	3.00		3.00	
ChiSq-statistic	8017		8055	
ChiSq p-value	<0.0001		<0.0001	
Log Likelihood	-25954		-25946	

Note: Two-tailed t-tests: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$
Following Allison and Waterman's (1995) hybrid approach to fixed-effects negative binomial regression, all variables that vary within patent families are de-meaned by subtracting their family averages. Additionally, the model includes patent family random effects and robust standard errors.

Table 3-3b: Hybrid Negative Binomial Regressions on Forward Citations:

	In-Class Citations				Out-Class Citations			
	(3)		(4)		(5)		(6)	
Neighbors			-0.002				0.116	***
			(0.014)				(0.016)	
Classes	-0.256	***	-0.256	***	0.651	***	0.622	***
	(0.054)		(0.054)		(0.062)		(0.062)	
Classes^2	0.030	**	0.030	**	-0.078	***	-0.074	***
	(0.010)		(0.010)		(0.011)		(0.011)	
Backward citations	0.014	***	0.014	***	0.021	***	0.021	***
	(0.002)		(0.002)		(0.002)		(0.002)	
Number of inventors	-0.135	*	-0.135	*	-0.168	**	-0.180	**
	(0.055)		(0.055)		(0.064)		(0.064)	
Number of applicants	-0.281	***	-0.281	***	-0.075		-0.076	
	(0.065)		(0.065)		(0.069)		(0.069)	
Time to approval	-0.073	***	-0.073	***	-0.078	***	-0.076	***
	(0.009)		(0.009)		(0.010)		(0.010)	
Primary class size	0.004	**	0.004	*	-0.005	***	-0.013	***
	(0.001)		(0.002)		(0.002)		(0.002)	
Grant year	-0.044	***	-0.044	***	-0.046	***	-0.045	***
	(0.005)		(0.005)		(0.006)		(0.006)	
German Patent	-1.609	***	-1.609	***	-1.264	***	-1.295	***
	(0.047)		(0.047)		(0.051)		(0.051)	
Japanese Patent	-1.543	***	-1.544	***	-1.502	***	-1.440	***
	(0.055)		(0.055)		(0.064)		(0.064)	
Constant	88.166	***	88.204	***	92.679	***	89.471	***
	(9.709)		(9.714)		(11.068)		(11.039)	
N	25,215		25,215		25,215		25,215	
Groups	8405		8405		8405		8405	
N per Group	3.00		3.00		3.00		3.00	
ChiSq-statistic	4976		4976		3919		3998	
ChiSq p-value	<0.0001		<0.0001		<0.0001		<0.0001	
Log Likelihood	-19160		-19160		-15855		-15827	

Note: Two-tailed t-tests: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$
Following Allison and Waterman's (1995) hybrid approach to fixed-effects negative binomial regression, all variables that vary within patent families are de-meaned by subtracting their family averages. Additionally, the model includes patent family random effects and robust standard errors.

Table 3-4: Regressions of Alternative Distance Measures on Forward Citations

	All Citations (2)		In-Class Citations (4)		Out-Class Citations (6)	
Class-Assignment-Based						
neighbors	0.046	***	-0.002		0.116	***
	(0.011)		(0.014)		(0.016)	
neighbors^2	0.073	***	0.024		0.157	***
	(0.011)		(0.014)		(0.016)	
neighbors to	0.043	***	0.009		0.085	***
	(0.011)		(0.014)		(0.016)	
neighbors to^2	0.030	**	0.017		0.048	***
	(0.010)		(0.013)		(0.015)	
neighbors from	0.004		-0.124	***	0.188	***
	(0.018)		(0.023)		(0.026)	
neighbors from^2	-0.020		-0.171	***	0.178	***
	(0.034)		(0.044)		(0.044)	
Backward Citation-Based						
bibliographic coupling	0.054	***	-0.001		0.101	***
	(0.006)		(0.009)		(0.007)	

Note: Two-tailed t-tests: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 3-5: Two-Country Subsample Regressions on Forward Citations

	All Citations (2)		In-Class Citations (4)		Out-Class Citations (6)	
Germany & Japan Only neighbors	0.374	**	0.307		0.384	*
	(0.118)		(0.159)		(0.170)	
Germany & US Only neighbors	0.169	**	0.024		0.462	***
	(0.057)		(0.072)		(0.085)	
Japan & US Only neighbors	0.388	***	0.192	*	0.755	***
	(0.065)		(0.076)		(0.099)	

Note: Two-tailed t-tests: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 4-1: Variable Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
3-day abnormal return	-0.02	0.07	-0.36	0.30
Year	2006.96	2.69	2003.00	2012.00
Merger of equals	0.02	0.14	0.00	1.00
Tender offer	0.16	0.37	0.00	1.00
Divestiture	0.01	0.11	0.00	1.00
Geographic Proximity	1.26	0.44	1.00	2.00
Number of competing bidders	1.09	0.35	1.00	4.00
Related to recent deals	0.13	0.34	0.00	1.00
Target product scope	3.33	2.04	1.00	17.00
Number of target advisors	1.31	0.62	0.00	6.00
Payment method	1.71	0.75	1.00	3.00
Involvement of acquirer advisors	1.70	1.39	0.00	11.00
Pre-existing partial ownership position	0.02	0.13	0.00	1.00
Defense tactics	0.03	0.16	0.00	1.00
Acquirer acquisition experience	3.33	3.30	1.00	20.00
Acquisition premium	0.03	0.04	-0.08	0.44
SIC similarity	0.57	0.43	0.00	1.00
NAICS similarity	0.62	0.43	0.00	1.00
Analyst similarity	0.20	0.22	0.00	0.97

N = 734

Table 4-2: Variable Correlations

		1	2	3	4	5	6	7	8	9	10
1	3-day abnormal return										
2	Year	0.16									
3	Merger of equals	0.04	-0.05								
4	Tender offer	0.03	0.11	-0.06							
5	Divestiture	0.04	-0.03	-0.02	0.02						
6	Geographic Proximity	-0.04	-0.04	0.05	-0.03	-0.04					
7	Number of competing bidders	0.02	-0.01	-0.01	0.04	-0.03	-0.05				
8	Related to recent deals	0.00	0.03	-0.02	0.05	-0.04	0.00	0.67			
9	Target product scope	0.00	0.16	-0.01	0.02	0.03	-0.10	-0.02	0.01		
10	Number of target advisors	0.01	0.05	0.08	-0.05	0.04	-0.06	0.08	0.16	0.10	
11	Payment method	-0.16	-0.09	0.16	-0.34	0.01	0.14	-0.02	0.01	0.00	0.08
12	Involvement of acquirer advisors	-0.11	0.07	0.08	-0.04	0.01	-0.06	0.07	0.16	0.16	0.32
13	Pre-existing partial ownership position	0.01	0.00	-0.02	-0.03	-0.01	0.10	0.03	0.14	-0.03	0.02
14	Defense tactics	0.03	0.07	-0.02	0.18	0.06	-0.02	0.13	0.09	0.09	0.06
15	Acquirer acquisition experience	0.06	-0.06	-0.07	0.09	-0.04	0.01	-0.04	-0.01	-0.08	-0.04
16	Acquisition premium	-0.06	0.17	-0.06	0.20	-0.04	-0.01	0.06	0.04	0.05	-0.03
17	SIC similarity	0.00	-0.02	0.06	-0.07	0.04	0.03	0.04	0.00	-0.11	0.05
18	NAICS similarity	-0.06	-0.03	0.07	-0.05	0.04	0.03	0.00	-0.06	-0.03	0.05
19	Analyst similarity	0.07	0.08	0.12	-0.15	0.05	0.02	0.10	0.14	0.06	0.28

Table 4-2: Variable Correlations (continued)

		11	12	13	14	15	16	17	18
12	Involvement of acquirer advisors	0.21							
13	Pre-existing partial ownership position	0.06	0.02						
14	Defense tactics	-0.01	0.10	-0.02					
15	Acquirer acquisition experience	-0.22	-0.20	-0.02	0.00				
16	Acquisition premium	-0.12	-0.06	-0.03	0.06	0.03			
17	SIC similarity	0.12	0.10	0.04	-0.01	-0.07	-0.04		
18	NAICS similarity	0.15	0.11	0.00	0.04	-0.02	0.00	0.53	
19	Analyst similarity	0.27	0.44	0.05	0.12	-0.20	-0.11	0.27	0.17

Table 4-3: Regressions on Acquirer's 3-Day Cumulative Abnormal Returns

	(1)	(2)	(3)	(4)
Intercept	0.011	0.003	0.018	0.014
Merger of equals	0.014	0.011	0.018	0.006
Tender offer	0.001	0.001	0.001	0.002
Divestiture	0.012	0.012	0.013	0.013
Geographic Proximity	0.006	0.005	0.005	0.005
Number of competing bidders	-0.010	-0.010	-0.009	-0.007
Related to recent deals	0.015	0.017	0.014	0.014
Target product scope	-0.001	-0.001	-0.002	-0.001
Number of target advisors	-0.003	-0.003	-0.003	-0.005
Payment method	-0.016 **	-0.016 **	-0.016 **	-0.021 ***
Involvement of acquirer advisors	-0.002	-0.002	-0.002	-0.007 *
Pre-existing partial ownership position	0.020	0.021	0.019	0.013
Defense tactics	-0.040	-0.038	-0.041	-0.046
Acquirer acquisition experience	0.003 *	0.003 *	0.003 *	0.002
Acquisition premium	0.279	0.432	0.204	0.155
* Payment method	-0.077	-0.072	-0.076	0.009
* Involvement of acquirer advisors	-0.148 **	-0.161 **	-0.144 *	-0.068
* Pre-existing ownership position	-0.831	-0.869	-0.764	-0.544
* Defense tactics	0.938	0.931	0.966	1.115
* Acquirer acquisition experience	-0.020	-0.025	-0.019	-0.013
Industry similarity variables:				
SIC similarity		0.015		
SIC similarity * premium		-0.246		
NAICS similarity			-0.014	
NAICS similarity * premium			0.111	
Analyst similarity				0.087 ***
Analyst similarity * premium				-1.487 *
Industry fixed effects	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y
r-squared	0.343	0.346	0.345	0.356
Adjusted r-squared	0.161	0.162	0.161	0.175
F statistic	3.018	2.922	2.897	3.272
N	734	734	734	734

Table 4-4: Regressions on Acquirer's 3-Day Cumulative Abnormal Returns

	(1)		(5)	
Intercept	0.011		0.014	
Merger of equals	0.014		0.009	
Tender offer	0.001		0.002	
Divestiture	0.012		0.010	
Geographic Proximity	0.006		0.005	
Number of competing bidders	-0.010		-0.007	
Related to recent deals	0.015		0.013	
Target product scope	-0.001		-0.002	
Number of target advisors	-0.003		-0.005	
Payment method	-0.016	**	-0.018	***
Involvement of acquirer advisors	-0.002		-0.004	
Pre-existing partial ownership position	0.020		0.023	
Defense tactics	-0.040		-0.047	
Acquirer acquisition experience	0.003	*	0.002	
Acquisition premium	0.279		0.242	
*** Payment method**	-0.077		-0.063	
*** Involvement of acquirer advisors**	-0.148	**	-0.142	*
*** Pre-existing partial ownership position**	-0.831		-0.780	
*** Defense tactics**	0.938		1.092	
*** Acquirer acquisition experience**	-0.020		-0.017	
Analyst similarity groups				
0.26 - 0.50			0.013	
0.51 - 0.75			0.028	*
0.76 - 1.00			0.053	*
Industry fixed effects	Y		Y	
Year fixed effects	Y		Y	
r-squared	0.343		0.354	
Adjusted r-squared	0.161		0.171	
F statistic	3.018		3.093	
N	734		734	

ILLUSTRATIONS

Figure 2-1: Change in EPS Forecasts – Unadjusted Data

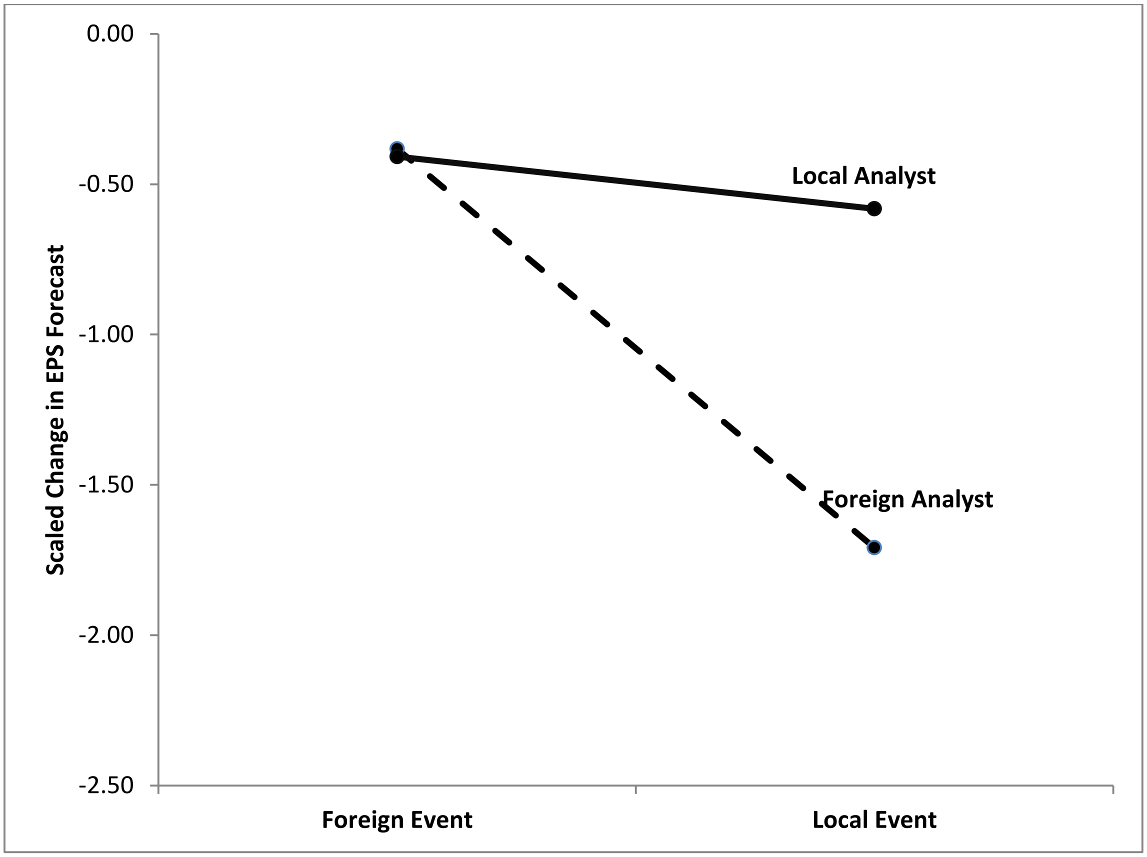


Figure 2-2: Change in EPS Forecasts – Regression Predictions

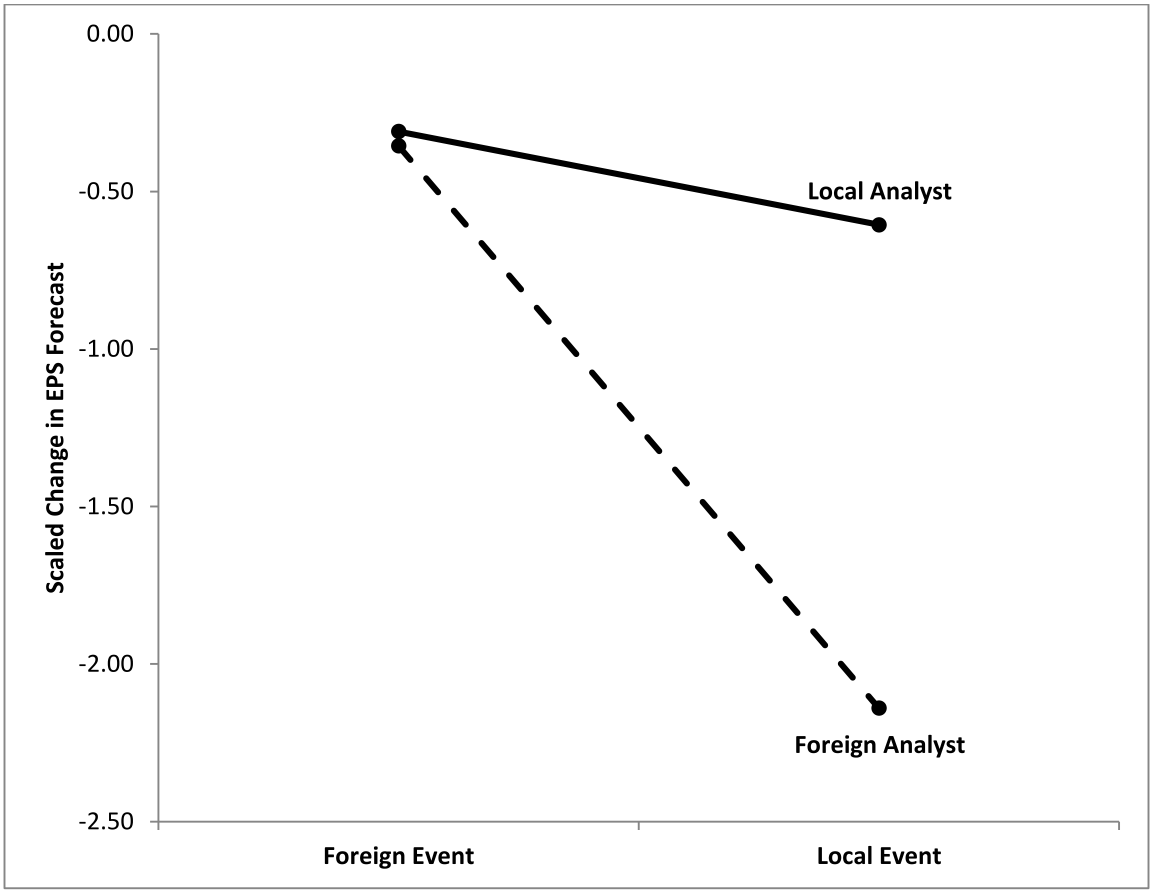


Figure 2-3: Change in EPS Forecasts – Regression Predictions – Inexperienced Analysts

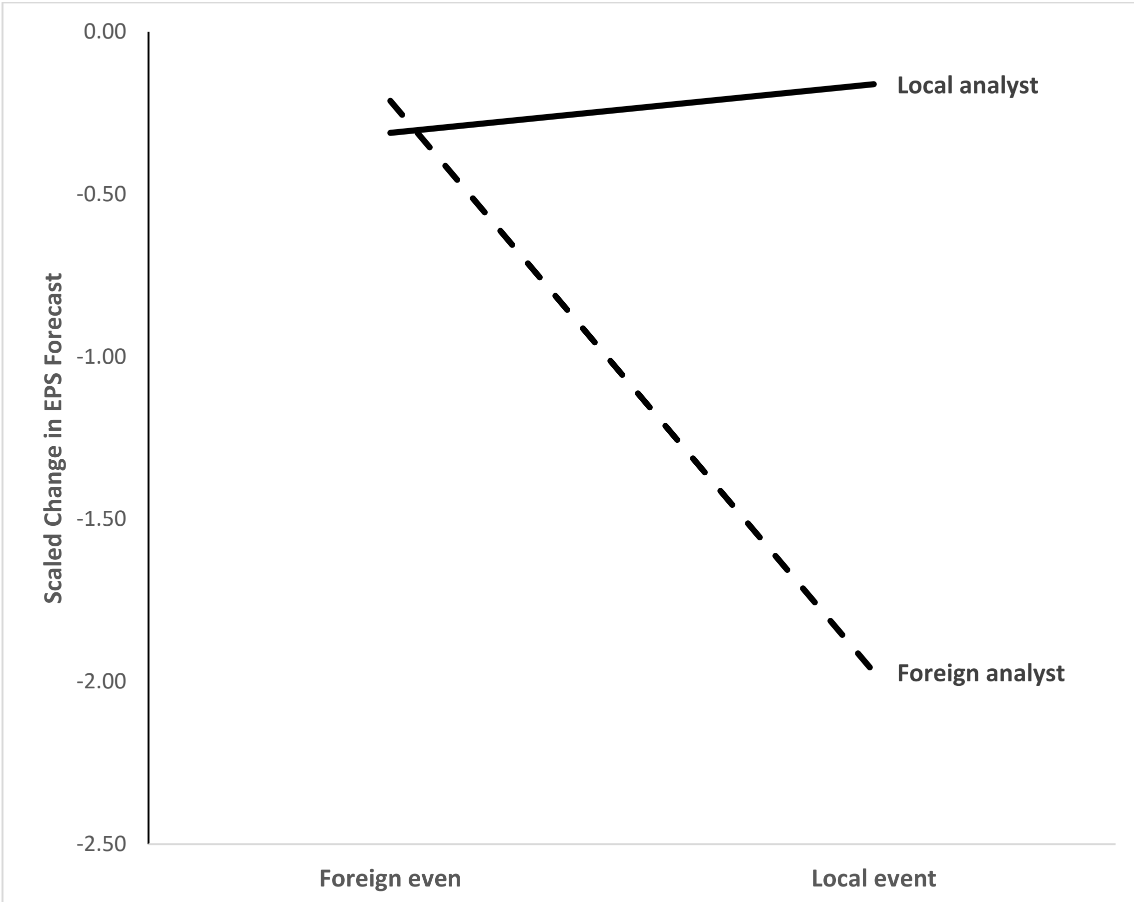


Figure 2-4: Change in EPS Forecasts – Regression Predictions – Experienced Analysts

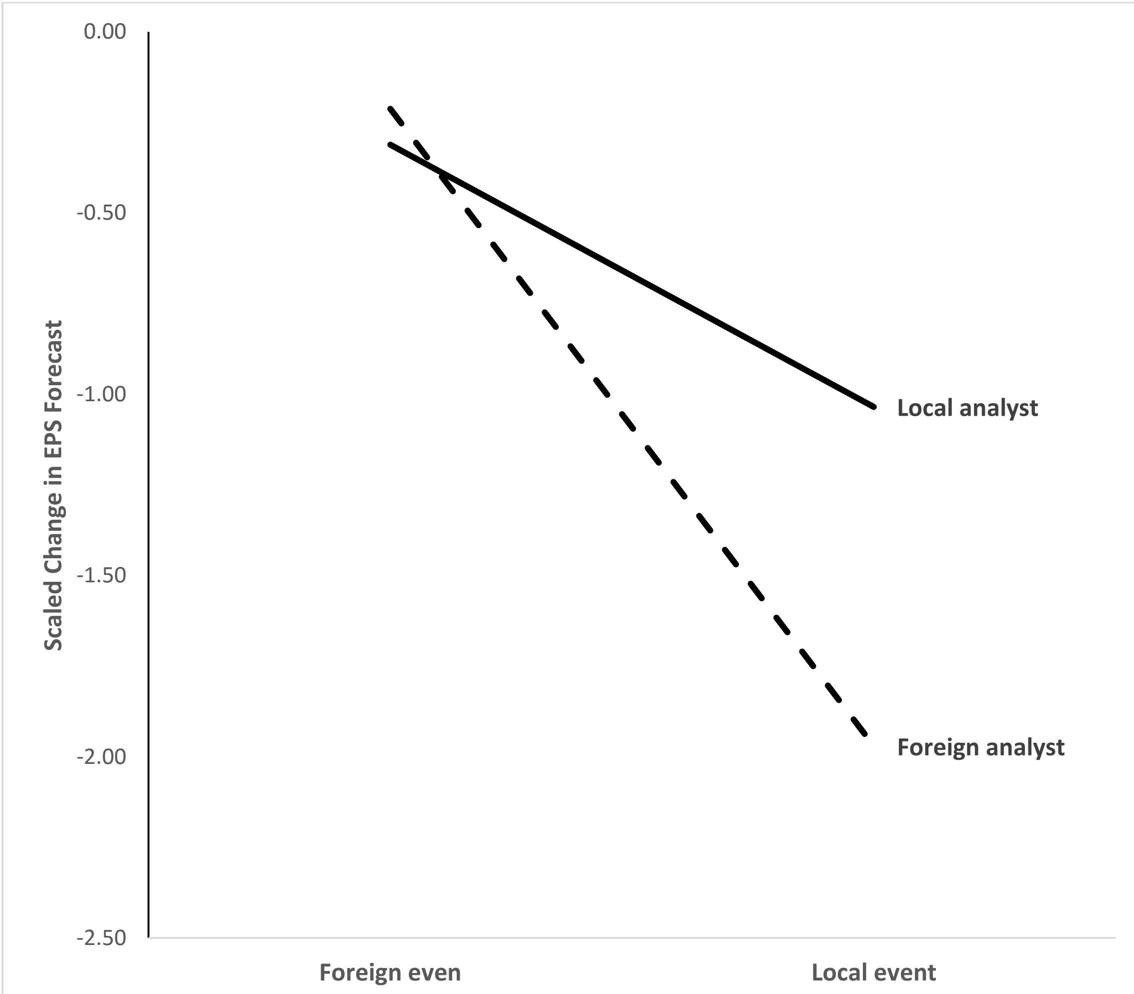


Figure 3-1: Neighbors Measure Distribution (Kernel Density)

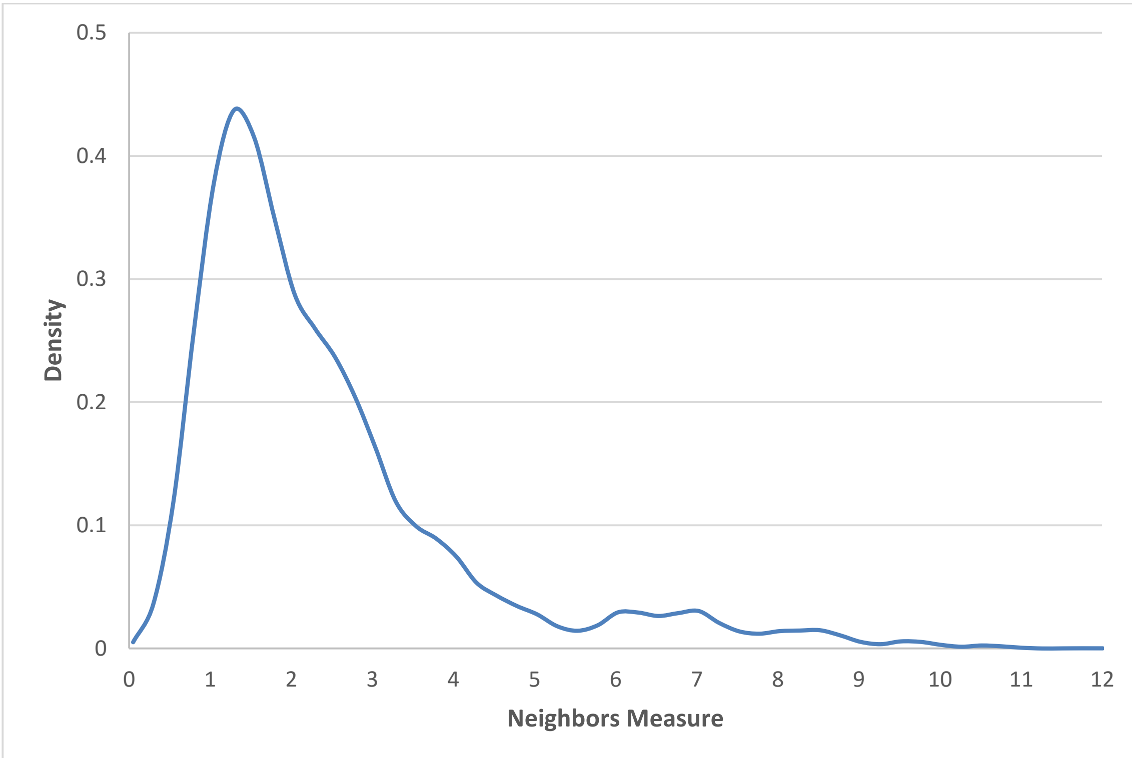


Figure 4-1: Predicted Acquirer CAR by Premium & Analyst Similarity

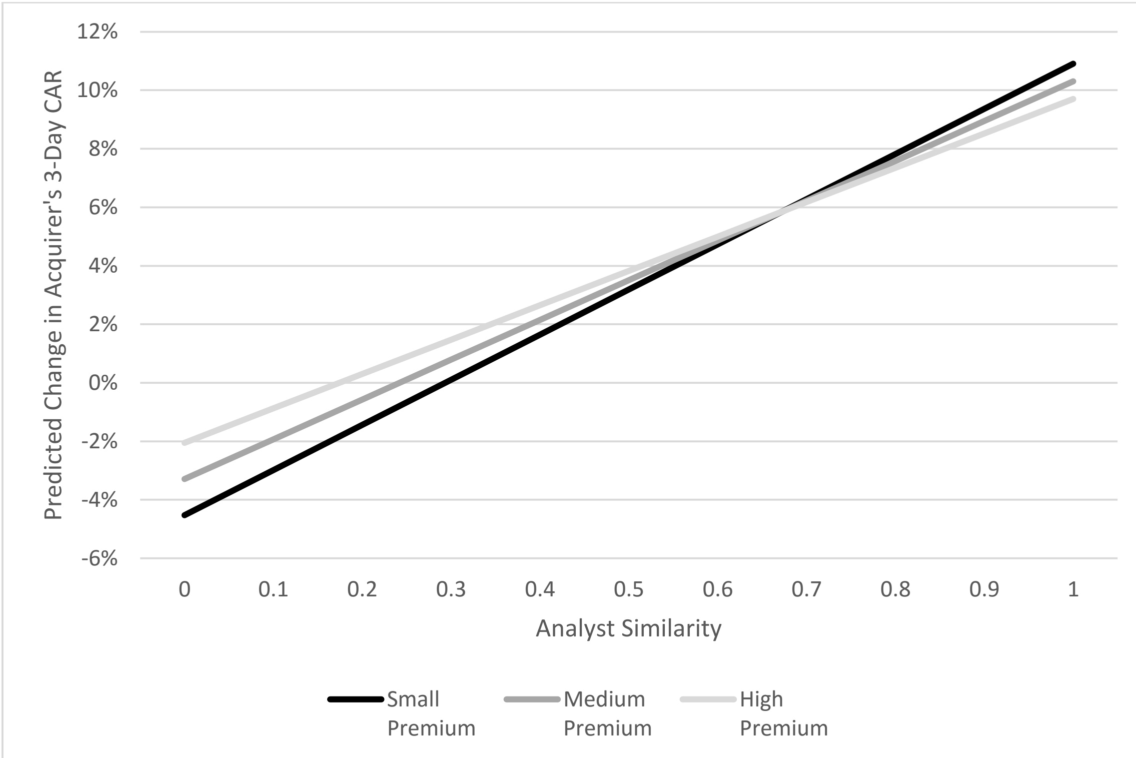
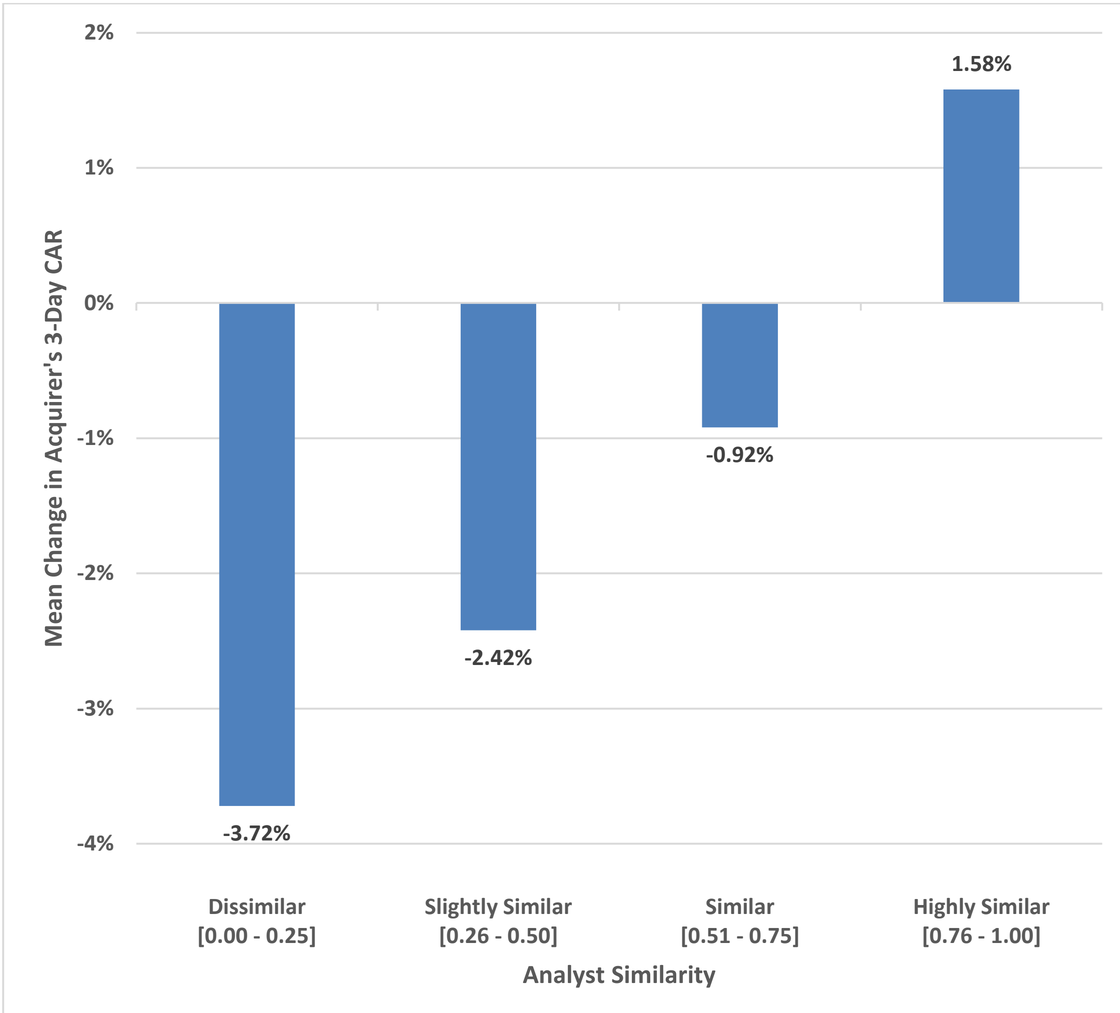


Figure 4-2: Predicted Acquirer's 3-Day CAR by Analyst Similarity Range



APPENDIX

Figure A1

	NAICS	Description
Netflix	532230	Video Tape and Disc Rental
Coinstar	333318	Other Commercial and Service Industry Machinery Manufacturing

Analyst Coverage (# of Analysts)

Year	Coinstar	Netflix	Both	% of Total
2008	36	21	0	0.0%
2009*	40	28	4	6.3%
2010	37	27	11	20.8%
2011	37	28	10	18.2%
2012	29	22	10	24.4%

Table A1
NAICS 334413 – Semiconductor and Related Device Manufacturing

Company		1	2	3	4	5	6	7
INTEL	1	1.00	1.00	1.00	0.96	0.96	0.96	0.96
TEXAS INSTRUMENTS	2	1.00	1.00	1.00	0.96	0.96	0.96	0.96
AMD	3	1.00	1.00	1.00	0.96	0.96	0.96	0.96
8X8 INC	4	0.96	0.96	0.96	1.00	0.95	0.94	0.95
FIRST SOLAR	5	0.96	0.96	0.96	0.95	1.00	1.00	1.00
SUNPOWER	6	0.96	0.96	0.96	0.94	1.00	1.00	1.00
EVERGREEN SOLAR	7	0.96	0.96	0.96	0.95	1.00	1.00	1.00

* Analyst-Based Similarity measures are between 0 (unrelated) and 1 (related)
** lines represent groups determined by Calinski & Harabasz psuedo F-test

Table A2
NAICS 541710 - R&D in the Physical, Engineering, and Life Sciences

Company		1	2	3	4	5	6	7	8	9
ARENA PHARMA.	1	1.00	1.00	1.00	0.92	0.93	0.92	0.60	0.59	0.39
EXELIXIS	2	1.00	1.00	1.00	0.93	0.94	0.93	0.62	0.60	0.39
KERYX BIOPHARMA.	3	1.00	1.00	1.00	0.94	0.94	0.93	0.62	0.61	0.39
COVANCE	4	0.92	0.93	0.94	1.00	1.00	1.00	0.81	0.80	0.41
CHARLES RIVER LABS	5	0.93	0.94	0.94	1.00	1.00	1.00	0.80	0.79	0.40
PHARMA. PRODUCT DEV.	6	0.92	0.93	0.93	1.00	1.00	1.00	0.81	0.81	0.40
METABOLIX	7	0.60	0.62	0.62	0.81	0.80	0.81	1.00	0.93	0.59
MICROVISION	8	0.59	0.60	0.61	0.80	0.79	0.81	0.93	1.00	0.47
SYNTROLEUM	9	0.39	0.39	0.39	0.41	0.40	0.40	0.59	0.47	1.00

* similarity measures are between 0 (unrelated) and 1 (related)
** lines represent groups determined by Calinski & Harabasz psuedo F-test

Companies 1,2,3 are pharma & biotech companies
Companies 4,5,6 are companies that assist in drug development and discovery
Company 7 develops bioscience alternatives to plastics, chemicals, energy industries
Company 8 develops high-resolution laser displays
Company 9 develops synthetic liquid hydrocarbons

Table A3
SIC 7990 – Miscellaneous Amusement & Recreational Services

Company		6	7	8	1	2	3	4	5	9
MGM RESORTS	6	1.00	1.00	1.00	0.96	0.96	0.94	0.96	0.97	0.76
WYNN RESORTS	7	1.00	1.00	1.00	0.95	0.95	0.92	0.93	0.96	0.72
LAS VEGAS SANDS	8	1.00	1.00	1.00	0.95	0.95	0.93	0.94	0.95	0.72
PINNACLE ENTERTAINMENT	1	0.96	0.95	0.95	1.00	1.00	1.00	0.99	1.00	0.87
BOYD GAMING	2	0.96	0.95	0.95	1.00	1.00	1.00	1.00	0.99	0.87
AMERISTAR CASINOS	3	0.94	0.92	0.93	1.00	1.00	1.00	0.99	0.99	0.88
PENN NATIONAL GAMING	4	0.96	0.93	0.94	0.99	1.00	0.99	1.00	0.99	0.89
ISLE OF CAPRIS CASINOS	5	0.97	0.96	0.95	1.00	0.99	0.99	0.99	1.00	0.87
VAIL RESORTS	9	0.76	0.72	0.72	0.87	0.87	0.88	0.89	0.87	1.00

* Analyst-Based Similarity measures are between 0 (unrelated) and 1 (related)
** lines represent groups determined by Calinski & Harabasz psuedo F-test

Table A4
Most Similar Companies According to MGM Resorts International (NAICS 72112)

Company	D(MGM)	NAICS	NAICS Description
LAS VEGAS SANDS	0.9973	72112	Casino Hotels
WYNN RESORTS	0.9967	72112	Casino Hotels
ISLE OF CAPRIS CASINOS	0.9668	71321	Casinos (except Casino Hotels)
PINNACLE ENTERTAINMENT	0.9612	71321	Casinos (except Casino Hotels)
STARWOOD HOTELS & RESORTS	0.9610	72111	Hotels (except Casino Hotels) and Motels
INT. GAME TECHNOGY[1]	0.9605	33999	All Other Miscellaneous Manufacturing
MULTIMEDIA GAMES[2]	0.9604	71329	Other Gambling Industries
WMS INDUSTRIES INC[3]	0.9604	33999	All Other Miscellaneous Manufacturing
HOST HOTELS & RESORTS	0.9581	52593	Real Estate Investment Trusts
SCIENTIFIC GAMES	0.9578	3341	Computer and Peripheral Equipment Manufacturing
STRATEGIC HOTELS & RESORTS	0.9577	52593	Real Estate Investment Trusts
CHURCHILL DOWNS	0.9561	71121	Racetracks

D(MGM) is the Analyst-Based Similarity to MGM Resorts International.
1. design, development, manufacture, and marketing of electronic gaming equipment and systems products
2. design, development, manufacture, and distribution of gaming machines and systems related to casino operators
3. design, development, manufacture, and distribution of games, video and mechanical reel-spinning gaming machines and video lottery terminals

BIBLIOGRAPHY

Allison, P. D., & Waterman, R. P. (2002). Fixed-Effects Negative Binomial Regression Models. Sociological Methodology, 32(1), 247-265.

Baden-Fuller, C., & Morgan, M. S. (2010). Business Models As Models. Long Range Planning, 43(2), 156-171.

Barber, B., Lehavy, R., Mcnichols, M., & Trueman, B. (2001). Can Investors Profit From the Prophets? Security Analyst Recommendations and Stock Returns. The Journal of Finance, 2: 531-563.

Barkema, H., Bell, J., & Pennings, J. (1996). Foreign Entry, Cultural Barriers, and Learning. Strategic Management Journal, 17: 151-166.

Barrat, A., Et Al. (2004). The Architecture of Complex Weighted Networks. Proceedings of the National Academy of Sciences of the United States of America, 101(11), 3747-3752.

Berger, P. L., & Luckmann, T. (1991). The social construction of reality: A treatise in the sociology of knowledge (No. 10). Penguin UK.

Berlin, B. (1992). Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies. Princeton, Nj: Princeton University Press.

Berry, H., Guillén, M., & Zhou, N. (2010). An Institutional Approach to Cross-National Distance. Journal of International Business Studies, 41: 1460–1480.

Bogaert, S., Boone, C., Negro, G., & Witteloostuijin A. (2010). Organizational Form Emergence: A Meta-Analysis of the Econological Theory of Legitimation. Journal of Management.

Bryce, D. J., & Winter, S. G. (2009). A General Interindustry Relatedness Index. Management Science, 55(9), 1570-1585.

Calinski, T. & Harabasz, J. (1974). A Dendrite Method For Cluster Analysis. Communications in Statistics, 3(1): 1-27.

Campa, J., & Guillén, M. F. (1999). The internalization of exports: Firm- and location-specific factors in a middle-income country. Management Science, 45(11): 1463–1478.

Capron, L., & Guillén, M. F. (2009). National corporate governance institutions and post-acquisition target reorganization. Strategic Management Journal, 30(8): 803–833.

Carroll, G. & Swaminathan, A. (2000). Why the Microbrewery Movement? Organizational Dynamics of Resource Partitioning in the U.S. Brewing Industry. The American Journal of Sociology.

Chan, L., Karceski, J., & Lakonishok, J. (2003). The Level and Persistence of Growth Rates. Journal of Finance, 58: 634-684.

Clarke, R. (1989). Sics As Delineators of Economic Markets. Journal of Business, 62: 17-31.

Clement, M. (1999). Analyst Forecast Accuracy: Do Ability, Resources, and Portfolio Complexity Matter? Journal of Accounting and Economics, 27: 285-303.

Crawley, M., Ke, B., & Yu, Y. (2011). Externalities of Disclosure Regulation: the Case of Regulation Fd. Working Paper No. 1905192, Social Science Research Network.

Delios, A., & Henisz, W. (2000). Japanese firms' investment strategies in emerging economies. Academy of Management Journal, 43(3): 305–323.

Dimaggio, P. (1986). Structural Analysis of Organizational Fields: A Blockmodel Approach. Research in Organizational Behavior.

Douglas, M. (1986). How institutions think. Syracuse University Press.

Duda, R. & Hart, P. (1973). Pattern Classification and Scene Analysis. New York: Wiley.

Dugar, A., & Nathan, S. (1995). The Effects of Investment Banking Relationships On Financial Analysts' Earnings Forecasts and Investment Recommendations. Contemporary Accounting Research, 12(1): 131-160.

Dunning, J. H. (1993). Multinational enterprises and the global economy. Wokingham: Addison-Wesley.

Fleischer, A., & Baum, J. (2010). Leader of the Pack: the Stock Coverage Network As A Source of Security Analysts' Information Leadership. Working Paper. University of Toronto.

Foster, G. (1981). Intra-Industry Information Transfers Associated With Earnings Releases. Journal of Accounting and Economics, 3: 201-232.

Fryer, R., & Jackson, M. (2008). A Categorical Model of Cognition and Biased Decision-Making. Contributions in Theoretical Economics, 8(1): Article 6.

Ghemawat, P. (2001). Distance still matters: The hard reality of global expansion. Harvard Business Review, 79(8): 137–147.

Glynn, M. A., & Lounsbury, M. (2005). From the critics' corner: Logic blending, discursive change and authenticity in a cultural production system. Journal of Management Studies, 42(5), 1031-1055.

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

Guenther, D., & Rosman, A. (1994). Differences Between Computstat and Crsp SIC Codes and Related Effects On Research. Journal of Accounting and Economics, 11: 3-33.

Guillén, M., & Suárez, S. (2005). Explaining the global digital divide: Economic, political and sociological drivers of cross-national internet use. Social Forces, 84(2): 681–708.

Hannan, M. (2010). Partiality of Memberships in Categories and Audiences. Annual Review of Sociology, 36: 159-181.

Hannon, M., Polos, L., & Carroll, G. (2007). Logics of Organization Theory: Audiences, Codes, and Ecologies. Princeton, Nj: Princeton University Press.

Hirst, E., Hopkins, P., & Wahlen, J. (2004). Fair Values, Income Measurement, and Bank Analysts' Risk and Valuation Judgments. The Accounting Review, 79: 454-473.

Hofstede, G. (1980). Culture's consequences: International differences in work-related values. Beverly Hills, CA: Sage.

Hofstede, G. (2001). Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations. Beverly Hills, CA: Sage.

Hofstede, G. (2006). What did GLOBE really measure? Researchers' minds versus respondents' minds. Journal of International Business Studies, 37(6): 882–896.

Holm, P. (1995). The dynamics of institutionalization: Transformation processes in Norwegian fisheries. Administrative science quarterly, 398-422.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P., & Gupta, V. (2004). Culture, leadership, and organizations: The GLOBE study of 62 societies. Thousand Oaks, CA: Sage.

Hymer, S. (1960). The international operations of national firms: A study of direct foreign investment. Cambridge, MA: MIT Press.

Hsu, G. (2006). Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production. Administrative Science Quarterly, 51: 420-450.

Hsu, G., Hannan, M. & Özgecan K. (2009). Multiple Category Memberships in Markets: A Formal Theory and Two Empirical Tests. American Sociological Review, 74: 150-69.

Jaccard, P. (1901), Étude Comparative De La Distribution Florale Dans Une Portion Des Alpes Et Des Jura, Bulletin De La Société Vaudoise Des Sciences Naturelles 37: 547–579.

Jacob, J., Lys, T., & Neale, M. (1999). Expertise in Forecasting Performance of Security Analysts. Journal of Accounting and Economics, 28: 51-82.

Johanson, J., & Vahlne, J. (1977). The internationalization process of the firm: A model of knowledge development and increasing foreign market commitments. Journal of International Business Studies, 8(1): 23–32.

Jonsson, S., Greve, H., & Fujiwara-Greve, T. (2009). Undeserved Loss: the Spread of Legitimacy Loss to Innocent Organizations in Response to Reported Corporate Deviance. Administrative Science Quarterly, 54(2): 195-228.

Judd, C. M., & Park, B. (1988). Out-Group Homogeneity: Judgments of Variability At the Individual and Group Levels. Journal of Personality and Social Psychology, 54(5): 778.

Kennedy, M. & Fiss, P. (2013). An Ontological Turn in Categories Research: From Standards of Legitimacy to Evidence of Actuality. Journal of Management Studies, 50: 1138–1154.

Kennedy, M. (2005). Behind the One-Way Mirror: Refraction in the Construction of Product Market Categories. Poetics, 33(3–4): 201–226.

Kennedy, M., Lo, J. & Lounsbury, M. (2010). ‘Category Currency: the Changing Value of Conformity As A Function of Ongoing Meaning Construction’. in Hsu, G., Negro, G. and Koçak, Ö. (Eds), Categories in Markets: Origins and Evolution (Research in the Sociology of Organizations, Volume 31). Bingley: Emerald Group Publishing, 369–97.

Kennedy, M., & Fiss, P. (2013). An Ontological Turn in Categories Research: From Standards of Legitimacy to Evidence of Actuality. Journal of Management Studies, 50(6), 1138-1154.

Kobrin, S. (1979). Political Risk: A Review and Reconsideration. Journal of International Business Studies. 10(1): 67-80.

Kogut, B., & Singh, H. (1988). The Effect of National Culture On the Choice of Entry Mode. Journal of International Business Studies, 19(3): 411-432.

Kovacs, B. (2009). Essays On the Similarity of Organizations. (Doctoral Dissertation, Stanford University).

Kovács, B., & Hannan, M. T. (2010). The Consequences of Category Spanning Depend On Contrast. Research in the Sociology of Organizations, 31, 175-201.

Kovács, M. & Hannan, M. T. (2015). Conceptual spaces and the consequences of category spanning. Sociological Science, 2: 252-286.

Kuilman, J., & Wezel, F. C. (2008). Reaching Consensus? Naming, Labeling, and Organizational Mortality in the Uk Passenger Airline Industry, 1919–2000. in Organizational Ecology Conference, Antwerp, June.

Leung, M., & Sharkey, A. J. (2013). Out of Sight, Out of Mind? Evidence of Perceptual Factors in the Multiple-Category Discount. Organization Science, 25(1), 171-184.

Lo, J. & Kennedy, M. (2015). Approval in Nanotechnology Patents: Micro and Macro Factors That Affect Reactions to Category Blending. Organization Science, 26(1), 119-139.

Markman, A. B., & Wisniewski, E. J. (1997). Similar and Different: the Differentiation of Basic-Level Categories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23(1), 54.

Meyyappan, M. (Ed.), (2005). Carbon Nanotubes: Science and Applications. Crc Press, New York, Ny.

Michaely, R., & Womack, K. (1995). Conflict of Interest and the Credibility of Underwriter Analyst Recommendations. Review of Financial Studies, 12(4): 653-686.

Michaely, R., & Womack, K. (2005). Market Efficiency and Biases in Brokerage Recommendations. in Thaler, R. H. (Ed.), Advances in Behavioral Finance Ii. Princeton, Nj: Princeton University Press.

Mikhail, M., Walther, B., & Willis, R. (1997). Do Security Analysts Improve Their Performance With Experience? Journal of Accounting Research, 35: 131-157.

Mohanram, P., & Sunder, S. (2006). How Has Regulation Fair Disclosure Affected the Operations of Financial Analysts? Contemporary Accounting Research, 25: 491-525.

Mohr, J. W., & Duquenne, V. (1997). The duality of culture and practice: Poverty relief in New York City, 1888--1917. Theory and society, 26(2), 305-356.

Mullainathan, S., Schwartzstein, J., & Schleifer, A. (2008). Coarse Thinking and Persuasion. Quarterly Journal of Economics, 123(2): 577-619.

Nayyar, P. (1992). On the Measurement of Corporate Diversification Strategy: Evidence from Large US Service Firms. Strategic Management Journal, 13(3), 219-235.

Negro, G., Hannan, M. T., & Rao, H. (2010). Categorical Contrast and Audience Appeal: Niche Width and Critical Success in Winemaking. Industrial and Corporate Change, 19(5), 1397-1425.

Newman, M. (2004). Analysis of Weighted Networks. Physical Review E, 70(5), 056131.

Nosofsky, R. M. (1992). Exemplars, Prototypes, and Similarity Rules. From Learning Theory to Connectionist Theory: Essays in Honor of William K. Estes, 1, 149-167.

Nosofsky, R. M., & Palmeri, T. J. (1997). An Exemplar-Based Random Walk Model of Speeded Classification. Psychological Review, 104(2), 266.

Ocasio, W. (1997). Towards an attention-based view of the firm. Strategic management journal, 187-206.

Park, B., & Rothbart, M. (1982). Perception of Out-Group Homogeneity and Levels of Social Categorization: Memory For the Subordinate Attributes of In-Group and Out-Group Members. Journal of Personality and Social Psychology, 42(6): 1051-1068.

Peski, M. (2011). Prior Symmetry, Similarity-Based Reasoning, and Endogenous Categorization. Journal of Economic Theory, 146(1): 111-140.

Phillips, D. J., & Zuckerman, E. W. (2001). Middle-Status Conformity: Theoretical Restatement and Empirical Demonstration in Two Markets1. American Journal of Sociology, 107(2), 379-429.

Pontikes, E. (2008). Fitting in or Starting New? An Analysis of Invention, Constraint, and the Emergence of New Categories in the Software Industry (Doctoral Dissertation, Stanford University).

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

Pontikes, E., Negro, G., & Rao, H. (2010). Stained Red? A Study of Mere Proximity to Blacklisted Artists and Oscar Winners in Post-War Hollywood. Working Paper, Stanford Graduate School of Business.

Porac, J., & Thomas, H. (1990). Taxonomic Mental Models in Competitor Definition. Academy of Management Review, 15(2), 224-240.

Porac, J., Thomas, H., & Baden-Fuller, C. (1989). Competitive Groups As Cognitive Communities: the Case of Scottish Knitwear Manufacturers*. Journal of Management Studies, 26(4), 397-416.

Porac, J., Thomas, H., & Emme, B. (1987). Understanding Strategists' Mental Models of Competition. Business Strategy and Retailing, 59-79.

Porac, J., Thomas, H., Wilson, F., Paton, D., & Kanfer, A. (1995). Rivalry and the Industry Model of Scottish Knitwear Producers. Administrative Science Quarterly, 203-227.

Ramnath, S. (2002). Investor and Analyst Reactions to Earnings Announcements of Related Firms: An Empirical Analysis. Journal of Accounting Research, 40(5): 1351-1376.

Rao, H., Monin, P., & Durand, R. (2005). Boarder Crossing: Bricolage and the Erosion of Categorical Boundaries in French Gastronomy. American Sociological Review, 70: 968-991.

Rosa, J., Porac, J., Runser-Spanjol, J., & Saxon, M. (1999). Sociocognitive Dynamics in A Product Market. The Journal of Marketing, 63: 64-77.

Rosch, E. (1973). Natural Categories, Cognitive Psychology, 4: 328-350.

Rosch, E. (1978). Principles of Categorization. in E. Rosch and B. B. Lloyd (Eds.), Cognition and Categorization (Pp. 27–48). Hillsdale, Nj: Erlbaum.

Rosche, E. & Mervis, C. B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. Cognitive Psychology, 7(4): 573-605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic Objects in Natural Categories. Cognitive Psychology, 8(3), 382-439.

Ruef, M., & Patterson, K. (2009). Credit and Classification: the Impact of Industry Boundaries in Nineteenth-Century America. Administrative Science Quarterly, 54(3), 486-520.

Schijven, M. & Hitt. M. A. (2012). The Vicarious Wisdom of Crowds: Toward A Behavioral Perspective On Investor Reactions to Acquisition Announcements. Strategic Management Journal, 33: 1247-1268.

Schollhammer, H., & Nigh, D. (1984). The Effect of Political Events On Foreign Direct Investments by Corporations. Management International Review. 24: 18-40.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. Advances in Experimental Social Psychology, 25(1): 1–65.

Schwartz, S. H. (1994). Beyond individualism/collectivism: New cultural dimensions of values. In U. Kim, H. C. Triandis, C. Kagitcibasi, & G. Yoon (Eds), Individualism and collectivism: Theory, method, and applications: 85–119. Thousand Oaks, CA: Sage.

Searle, J. R. (2006). Reality and social construction Reply to Friedman. Anthropological Theory, 6(1), 81-88.

Silverman, B. (1996). Technical Assets and the Logic of Corporate Diversification. Unpublished Ph.D. dissertation, Haas School of Business, University of California, Berkeley, CA.

Silverman, B. (1999). Technological Resources and the Direction of Corporate Diversification: Toward an Integration of the Resource-Based View and Transaction Cost Economics. Management Science, 45(8): 1110-1124.

Stimpert, J. L., & Duhaime, I. M. (1997). In the Eyes of the Beholder: Conceptualizations of Relatedness Held by the Managers of Large Diversified Firms. Strategic Management Journal, 18(2), 111-125

Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University Versus Corporate Patents: A Window On the Basicness of Invention. Economics of Innovation and New Technology, 5(1), 19-50.

Thornton, P. H. (2001). Personal versus market logics of control: A historically contingent theory of the risk of acquisition. Organization Science, 12(3), 294-311.

Thornton, P. H. (2002). The rise of the corporation in a craft industry: Conflict and conformity in institutional logics. Academy of management journal, 45(1), 81-101.

Thornton, P. H. (2004). Markets from culture: Institutional logics and organizational decisions in higher education publishing. Stanford University Press.

Thornton, P. H., & Ocasio, W. (1999). Institutional logics and the historical contingency of power in organizations: Executive succession in the higher education publishing industry, 1958–1990 1. American journal of Sociology, 105(3), 801-843.

Vergne, J. P., & Wry, T. (2014). Categorizing Categorization Research: Review, Integration, and Future Directions. Journal of Management Studies, 51(1), 56-94.

Wittgenstein, L. (1853). Philosophical Investigations.

Wry, T., & Lounsbury, M. (2013). Contextualizing the Categorical Imperative: Category Linkages, Technology Focus, and Resource Acquisition in Nanotechnology Entrepreneurship. Journal of Business Venturing, 28(1), 117-133.

Wry, T., Lounsbury, M., & Jennings, P. D. (2014). Hybrid Vigor: Securing Venture Capital by Spanning Categories in Nanotechnology. Academy of Management Journal, 57(5), 1309-1333.

Yasukawa, S., & Kano, S. (2015). Comparison of Examiners’ Forward Citations in the United States and Japan With Pairs of Equivalent Patent Applications. Scientometrics, 102(2), 1189-1205.

Yu, T., Sengul, M., & Lester, R. (2008). Misery Loves Company: the Spread of Negative Impacts Resulting From An Organizational Crisis. Academy of Management Review, 33(2): 452-472.

Zerubavel, E. (1991). The Fine Line: Making Distinctions in Everyday Life. New York: Free Press.

Zerubavel, E. (1996). Lumping and Splitting: Notes On Social Classification. Sociological Forum. 11: 421-433.

Zhao, W. (2005). Understanding classifications: Empirical evidence from the American and French wine industries. Poetics, 33(3), 179-200.

Zhao, W. (2008). Social Categories, Classification Systems, and Determinants of Wine Price in the California and French Wine Industries. Sociological Perspectives, 51:163-199.

Zhao, W. (2009). Market Institutions, Product Identities, and Caluation of California Premium Wines. Sociological Quarterly, 50(3): 525-555.

Zuckerman, E. (1999). The Categorical Imperative: Securities Analysts and the Illegitimacy Discount. American Journal of Sociology, 104(5), 1398-1438.

Zuckerman, E. (2000). Focusing the Corporate Product: Securities Analysts and De-Diversification. Administrative Science Quarterly, 45: 591-619.

Zuckerman, E. (2004). Structural Incoherence and Stock Market Activity. American Sociological Review, 69: 405-432.

Zuckerman, E., & Kim, T.-Y. (2003). The Critical Trade-Off: Identity Assignment and Box-Office Success in the Feature Film Industry. Industrial and Corporate Change, 12(1): 27-67.