

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Sociology

2012

WHERE DO SOCIAL CATEGORIES COME FROM? A COMPARATIVE ANALYSIS OF ONLINE INTERACTION AND CATEGORICAL EMERGENCE IN MUSIC AND FINANCE

Goldberg, Amir

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp01wp988j838>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

**WHERE DO SOCIAL CATEGORIES COME FROM? A COMPARATIVE ANALYSIS OF
ONLINE INTERACTION AND CATEGORICAL EMERGENCE IN MUSIC AND FINANCE**

Amir Goldberg

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF SOCIOLOGY

Adviser: Paul DiMaggio

April 2012

© Copyright by Amir Goldberg, 2012. All rights reserved.

Abstract

Social systems of classification have always captivated the sociological imagination. Collectively agreed-upon ways of sorting the world into different categorical boxes – such as those distinguishing people by ethnicity or occupation – play important roles in sustaining social order. Yet while sociologists have thoroughly investigated how divisions reified by social categories have concrete consequences for the unequal distribution of resources, a central question remains largely unanswered: how do these categories emerge?

This dissertation tries to address this question by comparatively investigating the classificatory dynamics in two social network websites: an online music community in which individuals share their musical compilations, and a financial investment website where users exchange information about their investment portfolios. Existing sociological scholarship suggests a co-evolutionary mechanism whereby category systems are inter-subjectively coordinated through interaction, shaping the patterns of interpersonal exchange which sustain them. This dissertation dissects this proposition analytically, translates it operationally, and tests it empirically.

I argue that, whether deciding which music to listen to or what stocks to invest in, the categorical distinctions people use to make sense of the world are inherently related to the people with whom they interact. My results demonstrate a persistent overlap between structures of interpersonal interaction and categorical meaning: when people interact in clearly bounded groups, they also tend to draw stronger categorical boundaries. Change emerges on the seams connecting these groups. As I demonstrate, when an artist is listened to or a stock invested in by people outside the group they are associated with, these objects' meanings begin to transform. Status plays an important role in providing incentives and rewards for people to defy categorical boundaries. Though there are obvious differences

between musical consumption and financial investment, this dissertation is concerned with highlighting the fundamentally similar ways in which meaning is socially constituted in both.

Acknowledgements

Writing a dissertation is a pretty lonely affair. But I could not have done it without the persistent support of a variety of people, to whom I am immensely grateful. My biggest debt is owed to my adviser, and mentor, Paul DiMaggio. I met Paul on a visit to Princeton six years ago, and was immediately captivated by his kindness and friendliness. Little did I know what a towering intellectual figure he is; his modesty provided no empirical evidence for his academic rock-star status. No other person has had such a profound influence on my development as a sociologist. His constructive but demanding critique of my work – always peppered with anecdotes about forgotten bluegrass pioneers or avant-garde seventies comics – has pushed me to think clearly about my research. Whenever I felt I was hitting a wall, he patiently guided me out of the quagmire. I was also fortunate to discover, time and again, Paul's humanity, especially as he provided thoughtful advice when I was at critical and difficult junctures in my life.

Other committee members have each provided their own indispensable forms of support. Delia Baldassarri showed me, with her Mediterranean no-nonsense energy and sharp intellect how to synergistically collaborate on an academic project. Though I benefitted from this relationship far more than she did, she always generously gave me credit even where it wasn't due. Martin Ruef taught me almost all that I know about organizations. His unparalleled analytical precision was invaluable in helping me jump through the hoops of journal submissions. This dissertation would not have come to life without Matt Salganik, who suggested studying an online music community and introduced me to the owner of the website I ended up researching. Matt's own online-based work served as an inspiration for this project. Finally, though not officially on my committee, Viviana Zelizer had an enormous impact on my intellectual trajectory. Her creative and groundbreaking work in economic sociology inspired me to

think of financial investment in cultural terms. Her door was always open, and her advice – whether sociological or personal – always wise and helpful.

I could not have survived through the Sisyphean task of writing a dissertation – mostly in a windowless room very low on oxygen – without the support of a community of friends, colleagues and future sociological field-changers. Hana Shepherd, Danny Schneider, Yael Berda, Jeff Lane, Bart Bonikowski, Emily Marshall, Pierre-Antoine Kremp, Cristina Mora, Shirley Gaw (an honorary sociologist), Cristobal Young, Stephanie Schacht, Carol Ann MacGregor, Sergio Galaz-García and María Abascal, each in his or her own way, provided constant intellectual stimulation, occasional geek-chic comic relief, and unconditional friendship.

My decision to pursue a PhD in sociology would never have materialized were it not for my family, forced by my academic aspirations to travel frequently across the Atlantic or otherwise receive unsatisfying bits of information about my wellbeing over a spotty internet connection. I owe my curiosity, and my ability to imagine the world through a sociological lens, to my parents. From my mother, Neora, I learned the value of creativity, and that there are rewards for thinking outside the box. And my father, Zvika, instilled in me a drive for dissecting reality with a sharp analytical knife. Both have always had absolute faith in my ability to succeed; without it, I would have never dared try. I am also grateful to my sister Ye’ela, who has enthusiastically encouraged me to follow my passions, and was always happy for me when doing so paid off. And to my grandmother, Regina, a wise and distinguished graduate of the University of Life, who has enjoyed watching me get the education she never had the opportunity to receive. I was also lucky to receive unqualified support from my parents-in-law, Dina and Itzhak Vardi, and their trust that whatever risks I had taken would one day prove to be worthwhile.

During the course of my PhD studies I became a father. Through Alma’s naïve and knowledge-voracious eyes, I was able to rediscover the world’s simple beauty. I hope that one day, when she skims

through these pages, she will find some interest – and dare I write: pride – in her dad’s musings. Finally, there is my wife, Gil-li. Had it not been for her example success, relentless encouragement, and selfless willingness to subject herself to my endless pseudo-intellectual meanderings, I would never have become a sociologist, let alone produced a dissertation. Through her modest intelligence, kindness, beauty, and elegance – especially in face of adversity – she reminds me every day that while sociology is a powerful model of the world, there are certain aspects of the human condition that can only be described with poetry. This dissertation is dedicated to her.

Table of Contents

Abstract.....	iii
Acknowledgements.....	v
Chapter 1 Introduction	1
1.1 How Categories Work in Everyday Life	1
1.2 What is Social about Classifications?	6
1.3 Chapter Structure.....	14
1.4 On the Classification of Things and People.....	17
Chapter 2 Measuring Practical Classifications: Theory, Data and Methods.....	19
2.1 Classification and Social Cognition.....	19
2.2 Practical Classifications	22
2.3 Data	31
2.4 Measuring Classifications in Practice	41
2.5 Final Notes about Data and Methodology.....	52
Chapter 3 The Social Construction of Categories through Interaction.....	57
3.1 Introduction	57
3.2 What is a Social Construction?.....	60
3.3 Networks.....	62
3.4 Empirical Implications	76
3.5 Discussion.....	96
Chapter 4 How Object Meanings Change.....	100
4.1 Introduction: How Apple and Jazz Straddle Different Meanings.....	100
4.2 Boundaries and How They Are Traversed.....	105
4.3 Measurements	118
4.4 Analysis	125
4.5 Discussion.....	137
Chapter 5 Performance, Status, and Non-conformity.....	141
5.1 Introduction: Status and Performance-Measurability in Markets.....	141
5.2 Theory	143
5.3 Measurements	155
5.4 Analysis	167
5.5 Conclusion: Innovation and Performance-Measurability	180
Chapter 6 Conclusion: Change and Inter-subjective Synchronicity.....	183
Appendix A: Data Management.....	194
References Cited	197

List of Tables

Table 2.1: Data Types by Data Source.....	33
Table 2.2: Summary Statistics for Units of Analysis.....	35
Table 2.3: Common Symbols	56
Table 3.1: Poisson and logit models estimating interaction distance/dyad between pair i - j in S	89
Table 3.2: GLM estimating actor interactional clustering in S	91
Table 4.1: OLS regression estimating semantic stability η between times t and $t+1$ for MOC artist i	132
Table 4.2: OLS regression estimating semantic stability η between times t and $t+1$ for FOC asset i	133
Table 4.3: OLS regression estimating semantic stability η between times t and $t+1$ for object i	135
Table 5.1: Examples of high and low coherence object collections in the MOC and FOC.....	162
Table 5.2: Fixed-effect estimates for MOC user success (standardized popularity delta)	171
Table 5.3: Fixed-effect estimates for FOC user success (standardized popularity delta).....	172

List of Figures

Figure 2.1: Visualization of fuzzy categories as a Venn diagram..	24
Figure 2.2: Hypothetical network of associations between firms traded on Wall Street	26
Figure 2.3: Illustration of the network structures included in the data	36
Figure 2.4: Visualization of adjacency matrices for the interaction network (S), user-to-object network (A) and object co-occurrence network (W)	37
Figure 2.5: Illustration of the sliding window technique	39
Figure 2.6: Snapshots of network W for MOC and FOC data	42
Figure 2.7: Assortative mixing in MOC co-occurrence network	44
Figure 2.8: Heavy-tailed distributions of node degrees in networks A (objects only) and S.	45
Figure 2.9: Distributions of standardized co-occurrences in simulated and observed data	47
Figure 2.10: Categorical similarity during time t_1 vs. Tag similarity during time t_2	52
Figure 3.1: Two hypothetical consumption spaces	74
Figure 3.2: Stylized examples of two hypothetical high-school networks.	75
Figure 3.3: Two hypothetical interaction networks of size $N=500$	85
Figure 3.4: Categorical proximity vs. cosine similarity and interactional distance	87
Figure 3.5: Poisson model of interactional distance as a function of categorical proximity	90
Figure 3.6: General linear model of interactional clustering as a function of categorical cohesion	92
Figure 3.7: Network segmentation as function of time for the MOC and FOC data	93
Figure 3.8: Segmentation in interaction and object co-occurrence networks	95
Figure 4.1: Categorical similarity time between Apple, Dell and Google	113
Figure 4.2: Semantic stability over time for four exemplary MOC artists	120
Figure 4.3: Illustration of boundedness	123
Figure 4.4: Distribution of semantic stability in both communities	125
Figure 4.5: The effects of artist success and share volatility on semantic stability	133
Figure 4.6: Estimated semantic stability as a function of boundedness	137
Figure 5.1: Snapshot of a typical FOC user homepage	154
Figure 5.2: Distribution of coherence in both communities.	158
Figure 5.3: Network of categorical similarity between the 400 most traded securities	159
Figure 5.4: Success as a function of status and performance.	173
Figure 5.5: MOC success, as a function of an interaction between coherence and status	175
Figure 5.6: FOC success, as a function of an interaction between coherence and performance	176
Figure 5.7: FOC success, as a function of an interaction between performance and coherence	179
Figure 5.8: Who innovates?	182
Figure 6.1: Inter-subjective synchronicity and the cycle of change.	189

Chapter 1

Introduction

“The classification of things reproduces [the] classification of men”

Durkheim and Mauss, 1903 (1963: 11)

1.1 How Categories Work in Everyday Life

When Standard & Poors – one of the world’s three leading credit rating agencies – downgraded the US government’s credit rating from a perfect AAA to AA+ on August 5th 2011, many commentators raised an eyebrow with skepticism.¹ Nevertheless, investors interpreted the unprecedented downgrade – the US government’s first in history – as an ominous sign about the prospects of economic recovery from the financial crisis which had pushed the global economy into a deep recession three years earlier.² Stock markets tumbled the following Monday, with the Dow Jones Industrial Average losing 5.54%, its sixth largest daily point loss in history. Citing the weakened “effectiveness, stability, and predictability of American policymaking and political institutions”, Standard & Poors decision followed a last minute

¹ Economist, Nobel Prize laureate and New York Times financial columnist Paul Krugman, for example, wrote on that same day that “everything I’ve heard about S&P’s demands suggests that it’s talking nonsense about the US fiscal situation”. Retrieved from <http://krugman.blogs.nytimes.com/2011/08/05/sp-and-the-usa/>

² Interestingly, investors did not interpret the downgrade at face value. US Treasury bond interest rates continued to decline throughout the weeks that followed the announcement, suggesting increased trust in the US government’s ability to repay its debt, with 10-year bond rates dropping from a high of 3.03% on July 21st to 1.98% by September 6th 2011 (data from the US Department of The Treasury’s website, <http://www.treasury.gov>). It appears that while investors did not accept Standard & Poors’ assessment of the US government’s solvency, they nevertheless understood the downgrade to have real implications on its performance.

agreement reached in congress to raise the US government's debt ceiling.³ It reflected growing fears about the economic implications of political brinkmanship and polarization in the US.

The debt ceiling debate and ensuing financial plunge are instructive of why classifications matter. By 2011, the subprime mortgage crisis of 2008 had already potentially demonstrated rating agencies' poor ability of assessing debtors' solvency; all major agencies, most notoriously Standard & Poors, provided extremely high credit ratings for investment banks that would days later collapse spectacularly under the weight of overexposure to subprime debt. Yet three years later investors still heeded, en-masse, Standard & Poors' reclassification. This response seems particularly irrational given that investors appear to have already been painfully aware of the US government's acute political deadlock and its hazardous economic implications: during the two weeks preceding the rating agency's announcement the Dow Jones Industrial Average had already dropped more than 10%. As if providing an official stamp of economic malaise, the downgrade appears to have cemented investors' hitherto impalpable suspicions about the US government's inability to act decisively in order to prevent the emergence of a second recession. Like a magic wand, it generated coherent meaning out of otherwise difficult to decipher signals about the US economy's alleged poor health.

Sociologists have always been fascinated with the social implications of category systems. Starting with Durkheim and Mauss (1963) more than a century ago, sociological studies have demonstrated how categories relating to a variety of social domains – from ethnicity (Brubaker et al 2004; Schwartzman 2007) and gender (West and Zimmerman 1987) to art (DiMaggio 1987; Lena and Peterson 2008) and science (Gieryn 1983) – create and sustain social order. Contemporary work on categories and their implications is informed by two tangential traditions. One strand of literature is concerned with how categories reinforce symbolic boundaries between different social groups (Lamont and Molnár 2002). By

³ Full announcement text available at:
<http://www.standardandpoors.com/ratings/articles/en/us/?assetID=1245316529563>

distinguishing between different types of people, categories also make the different privileges and obligations these people enjoy seem natural; how people and objects are classified is therefore part and parcel of broader mechanisms that produce and sustain social inequalities (Douglas 1986; Bourdieu 1984; Tilly 1999; Ridgeway 1991). A second strand focuses on the role of categories in structuring market behaviors. This body of research explores how categorizations are employed by market actors when they evaluate one another, and the consequences of behaviors that challenge established category systems (Hannan 2010).

Both literatures are concerned almost exclusively with how institutionalized categories - such as those produced by state bureaucracies (Montgomery 2011), market analysts (Zuckerman 1999) or industrial producers (Peterson and Kern 1996) – enforce social boundaries; relatively little sociological attention has been paid to folk taxonomies (Bowker and Star 1999). But recent scholarship suggests that the practical ways in which people use categorical distinctions in everyday life differ markedly from the neat and mutually exclusive distinctions produced by formal classifications. In fact, most of the time, we make snap judgments drawing on implicit and vaguely defined categorical distinctions that are automatically and subconsciously activated.

This does not mean that formal classifications are irrelevant: to the contrary, when Standard and Poors downgraded US bonds, investors reacted with hysteria because this decision reified and legitimized their incipient fears. But in everyday situations, such as when deciding what music to listen to or what stocks to invest in, there are usually no clear-cut definitions at our disposal to draw on. Nevertheless, our decisions are still invariably informed by heuristics that distinguish between, for example, cool and uncool pop singers or between high and low risk investments. These decisions are based on things we “know tacitly about how to ‘go on’ in the contexts of social life without being able to give them direct discursive expression” (Giddens 1984: xxiii).

This study seeks to explore the social processes that give rise to classifications in practice. Rather than examining how boundaries are imposed from above by institutions, it asks how they arise organically through social interaction; under what circumstances they change; and when actors are likely to successfully transgress them. Drawing on social constructionist theories and insights from cognitive sociology, this project is informed by three basic propositions. First, it assumes that categories are implicit in the semantic relationships between the concepts they describe. An entity is not simply meaningful in and of itself; rather meaning is “constituted through the systematic distinctions that differentiate words (or sounds, or signs) from one another” (Mohr 1998: 351). When trying to understand how people associate things with meaning we therefore need to shift our analytical attention from the things themselves to their relationships with one another. Because I am concerned with socially diffuse categories, as opposed to idiosyncratic ways of making sense of the world, I look for manifestations of such relationships in the aggregate.

I argue, second, that these relationships are better excavated from routine individual behaviors, than discerned from these individuals’ conscious descriptions. When providing retrospective explanations for our behaviors we tend to impose an abstract order on them. In practice, however, we rely on fuzzy templates to make sense of the world. Our practical cognition lacks the crispness of analytical logic. Rather than exclusively exploring the post-hoc narratives with which people give justificatory meaning to their behaviors (Boltanski and Thévenot 2006), we need also to examine these behaviors in everyday settings.

Finally, I contend that individual actors infer meaning from the recurring patterns of relationships they observe in others’ behaviors. Meaning-making is a continuous and predominantly unconscious process; in order for meaning to be sustained, the semantic relationships underpinning it need to be constantly rehearsed. We seamlessly identify these conceptual associations in the aggregate actions of

those we are in contact with on a daily basis. We learn through observation, for example, to identify certain musical and fashion attributes with the category 'hip-hop,' and to associate this label with particular social roles. Who we interact with and the categorical prisms through which we interpret the world around us are therefore constitutively related to one another. The ultimate goal of this dissertation is to demonstrate how patterns of interaction between individuals have a bearing on the underlying cognitive schemas that structure their behaviors.

The empirical setting is the online social networking world. In the chapters that follow I comparatively investigate two online communities: one dedicated to music and the other to finance. Both communities are social network sites in which users can follow one another, and share information with others. In the first community this information consists of their musical compilations; in the second, their investment portfolios. While online data are clearly not without problems, they are exceptionally useful for the purposes of this study as they provide unprecedented levels of detail on the full set of interactions and behaviors of a complete population, and how those change over a defined period of time. Studies relying on such high-quality data have rarely been used in sociological studies, especially in order to answer the questions with which this research is concerned.

The comparison between music and financial investment websites is useful for several reasons. These two communities are centered on different types of activities which are, in the popular imagination at least, normally conceived of as inherently different: one is squarely within the realm of emotive 'culture,' the other of 'rational' behavior. Moreover, music consumption and its relationship with social distinctions have received a considerable amount of attention from sociologists in recent years (Roy and Dowd 2010), whereas consumer (as opposed to corporate) finance has largely been neglected.

Despite these differences, existing research in economic and cultural sociology points to similarities in the social mechanisms operating in both realms. First, both the fields of artistic (Bourdieu 1984; DiMaggio 1987) and economic (Veblen 1991; Podolny 2005; McKensie and Millo 2003; Callon 2008) production are pervaded by processes of social construction: in the same way that creditworthiness – as the US debt downgrade exemplifies – can often seem an arbitrary designation, so is the distinction between high and low quality in art an historically contingent process (Baumann 2007). Second, both finance and music are anchored in institutions that rely on categorical distinctions to differentiate between various types of artistic objects (Lena and Peterson 2008; Hsu et al 2009) or financial assets (Zuckerman 2004; Ruef and Patterson 2009). In music, these distinctions slice different genres; in finance, they differentiate securities on the basis of their industrial sectors.

Though substantively different, the two online communities under investigation are functionally practically identical: in both users exchange collections of objects, and in both they are restricted to similar types of interaction. These differences and similarities provide fertile ground for a comparison aimed at delineating the underlying mechanisms that are operating in both, and that are distinctive in each. That both communities constitute enclosed social worlds provides the opportunity for observing how classifications grow endogenously. This dissertation seeks to trace this process.

1.2 What is Social about Classifications?

Bowker and Star (1999: 10) define a classification as a “spatial, temporal, or spatio-temporal segmentation of the world” which ideally exhibits three properties: (1) unique and consistent classificatory principles; (2) mutual exclusiveness of categories; and (3) complete coverage of the world it describes. In reality, most classifications never meet this ideal. But because human beings are innately predisposed to categorical thinking (Fiske and Taylor 2008; Macrae and Bodenhausen 2001; Fiske and

Linville 1980), classifications are ubiquitous. We constantly distinguish between things in order to give them meaning: we separate family from friends and colleagues, distinguish between furniture used for sitting on or as ornamentation, and differentiate between behavior that is appropriate and inappropriate in public. Such distinctions are so commonplace that they are mostly used seamlessly. Situations which make them visible – such as when a cleaner in a Dortmund museum scrubbed clean a work of art valued at more than a million dollars which she mistook for dirt⁴ – are often infuriating, or comical, because they defy boundaries which are normally consistently maintained.

Whereas some divisions of the worlds are universal – presumably because they originate from natural or objective criteria – others are inherently social. Consider the standardization of time as a quintessential example (Zerubavel 2003). While the daily cycle of light is a function of the earth's orbiting around the sun, nothing in this planetary configuration necessarily entails dividing time into discrete weekly sequences. Yet we all unselfconsciously accept that the 48-hour period arbitrarily referred to as the 'weekend' should – at least ideally – be devoted to rest, and agree to calibrate it with respect to the sun reaching its highest point in the sky over the royal observatory in Greenwich.⁵ Classifications are social insofar as they are shared by a group of people and implement a division which is, to some extent at least, a subjective (and potentially arbitrary) interpretation of the worlds they describe.

Social classification is most fundamentally apparent in language. The words we use, and the ways in which we arrange them into statements and sentences, divide the realities they describe into bounded entities, charging them with distinctive meanings. These meanings embody fundamental systems of thought (Foucault 1972). By classifying things with labels discourse also “produces the effects

⁴ See the incident reported by the Guardian at:

http://www.guardian.co.uk/artanddesign/2011/nov/03/overzealous-cleaner-ruins-artwork?CMP=tw_t_gu

⁵ The fact that the division between the workweek and weekend differs across national contexts – the majority of countries in the Arab world, for example, follow a Friday-Saturday weekend – is a testament to this arbitrariness.

that it names” (Butler 1993: 2), that is, creates a perceptual sieve through which the world is experienced in terms of the meanings it presupposes. Consider, for example, the differentiation between money given as a gift and money received as wages. Though both are, ultimately, just money, people often set aside gifts for indulgent expenses, while using their paycheck to buy groceries. Thus earmarking infuses otherwise colorless currency with social meaning (Zelizer 1997).

Classification is “a process of actively ‘sculpting’ islands of meaning rather than simply identifying already-existing natural ones” (Zerubavel 1997: 67). If not natural, then where do these islands come from? Why do certain criteria for demarcating categorical boundaries become consensual, and how do the categories they delineate become associated with collectively agreed-upon meanings? Sociologists have argued that this sculpting always occurs in a social context. In *The Elementary Forms of Religious Life* Durkheim influentially asserted that the fundamental distinction between the sacred and the profane, which, he argued, characterizes every religion, serves as a means for drawing symbolic boundaries between different social groups. More broadly, Douglas (1966; 1986) posits that social classifications emerge and are maintained by institutions. Institutionalized social arrangements, she contends, and the different roles and privileges they entail, are most effectively maintained when they are believed to be grounded in the natural or supernatural world. Indeed a variety of studies demonstrate how classificatory debates are rooted in structural conflicts (Gieryn 1983; Strand 2011), and how categories are harnessed when existing institutions are threatened by challenges (Ben Yehuda 1980; Johnson 2008).

1.2.1 Categories and Networks

Institutions do not emerge from nowhere; nor do they uninterruptedly impose a classificatory order from above. Rather, meanings emerge and are constantly negotiated in concrete social settings

(Eliasoph and Lichterman 2003). What, then, are the micro-dynamical foundations upon which social classifications, and the institutions they support, take form? Most empirical studies eschew this question because they draw upon settings in which these institutions are already in place. When addressing processes of institutional emergence, they often do so from broad historical perspectives that cannot account for the fine-grained dynamics that give rise to new social categories.⁶

Berger and Luckmann's (1966) theory of social construction provides what is perhaps the most promising theoretical framework in this regard. Cognitive categories, they argue, emerge and are intersubjectively coordinated through interpersonal interaction. Through repetition and habituation, these categories are eventually reified as institutions. Understanding under what circumstances categories emerge and become salient therefore necessitates taking into account how people interact. Surprisingly, this task has rarely been taken on empirically.

Social network theory points to two rudimentary mechanisms whose universality has been demonstrated in different social contexts, and that are presumably relevant for the emergence of categories. Social influence, or the tendency of people to mimic the behaviors of those they interact with, suggests that how social networks are structured has an influence on whether or not certain behaviors or preferences diffuse (Granovetter 1978; Christakis and Fowler 2007; Denrell and Le Mens 2007; Friedkin 1993; Watts 2003). Beliefs about the social meanings of categorical attributes – eg. that a certain social attribute signifies high status – spread and are reinforced through interaction (Gould 2002; Ridgeway 1991). Homophily, on the other hand, or the tendency to form ties with similar others, suggests that network formation is structured by social categories (McPherson et al 2001). Homophily can be the result of preexisting network structures limiting interaction opportunities to similar others,

⁶ There are, naturally, some exceptions. See Mohr and White (2008) for a review.

but is also affected by individuals' socially conditioned predispositions to interact with those like them (McPherson and Smith-Lovin 1987).

Both social influence and homophily studies normally rest on an implicit assumption that categories, and their social meanings, are fixed and predate social interaction. Yet, if Berger and Luckmann's constructivist theory is correct, then how we classify things and whom we interact with are inextricably and co-evolutionarily interlinked (Mohr and White 2008). This link is elegantly captured in White's (2008) under-explored concept of catnets, or the idea that mutual group membership short-circuits social interaction. Like Breiger's (1974) notion of the duality of persons and groups, catnets suggest a duality between networks and classifications such that categories depend on the people who make them up to the same extent that these people's categorical identities are formed through the intersections of the groups to which they belong. The classification of people into different categories is therefore an emergent order of social interaction, the patterns of which are shaped by the very order it gives rise to.

I extend this co-evolutionary framework by integrating insights from network and cognitive theories, and apply them to music consumption and financial investment. I argue that people observe regularities in others' behaviors, and learn to associate objects – whether music or financial assets – that co-occur frequently with one another, as making up different categories. They are also sensitive to the patterns of interaction between one another. Thus social classification entails a structural homology between patterns of associations between objects, and patterns of interaction between people.

1.2.2 Categories in Music and Finance

An obvious reason for choosing music and finance as empirical sites is that they are so inherently different. Music is personal and expressive, a “medium with a capacity for imparting shape and texture

to being, feeling and doing” (DeNora 2000: 152). It projects meaning onto our most trivial and routine activities, and serves as a cultural intermediary that can facilitate, or hinder interpersonal relationships. As illustration, try to imagine your favorite movie without its accompanying soundtrack. We are so heavily socialized to associate music with meaningful narration that an action sequence with mellow music in the background, or a poignant ending scene supported by upbeat music, would simply seem out of place.

Money, on the other hand, has no smell; it is, so we tend to believe, an impersonal and interchangeable medium of exchange. Indeed, according to mainstream neoclassical economic theory, *homo economicus* is a rational actor. Rather than expressing his personality or cultural identity, the rational investor presumably pursues a singular objective: maximizing profits. Hence, unlike music, financial assets should not convey social meanings; they either pay off, or they do not. Of course, there is more to financial behavior than mere calculative profit-seeking. But there is some truth in this characterization, at least insofar as financial investment is popularly understood. Thus if we find similarities between the processes operating in our two empirical settings, these can be attributed not to the substantive nature of the information being exchanged, but to the social dynamics sustaining this exchange.

As much as they are different, however, both musical and financial markets share an important similarity: the two are highly unpredictable. Whether an artist will become commercially successful, or whether a financial investment will bear fruit, are almost impossible to predict. This is because quality is very elastically defined in both contexts. And faced with unpredictability, consumers often resort to socially available information as a means to mitigate uncertainty (DiMaggio and Louch 1998). In fact, in both markets, people interpret others’ decisions as quality signals. In the same way that investors use stock price as information (Malkiel 2003), so do music consumers understand song popularity as an

indication about its quality (Salganik et al 2006). But collective behaviors are often too crude a measure of underlying quality: Greek government bonds – as the recent Eurozone crisis has agonizingly demonstrated – can suddenly become worthless, and popular rock stars can quickly fade out of fashion. People therefore rely on mediated information provided by experts – whether music critics or market analysts – and the formal taxonomies they rely on, in order to make sense of the complex realities they need to untangle. Consequently, category systems play a central role in structuring meaning both for music consumers and financial investors (Hsu et al 2009).

The institutionalization of music into a system of tonality that can abstract and reproduce pieces composed centuries earlier is, in and of itself, a quite extraordinary social accomplishment (Roy and Dowd 2010). Over the years it has evolved into the organization of musical production and consumption around the concept of genre (DiMaggio 1987; Lena and Peterson 2008). Genre categories do not merely classify music but also define different subcultures which are often characterized by distinctive rituals and fashions. In other words, a genre simultaneously delineates a set of musical practices and an art-world of musicians, producers, distributors and avid consumers (Becker 1982).

Since the publication of Bourdieu's *Distinction* (1984), music has served as a primary sociological example for how the classification of 'things' also reproduces the classification of 'men' (Durkheim and Mauss 1963). Bourdieu (1984, 1993) argued that genre classifications create taste hierarchies that serve to legitimate and reinforce social boundaries between their consumers, as well as to imbue works of art with social significance that makes them transformable into cultural capital. In the years since, however, mounting empirical evidence has persistently documented a shift in the orientation of high-status individuals toward an inclusive range of cultural preferences that traverses the traditional boundaries between highbrow and lowbrow genres (Peterson 1992, 1997, 2005). It suggests that today musical

preferences function as social markers not through the distinction between elitist and popular musical forms but by differentiating between those with high and low levels of cultural tolerance.

The rise of the so-called cultural omnivore seems inconsistent with Bourdieu's theory of cultural distinction, which expects to find exclusive correspondence between cultural taste and social position (Bryson 1996; Erickson 1996). But as Peterson and Kern (1996) argue, cultural omnivorousness does not negate Bourdieu's general theory; rather, it introduces one's range of preferences as a new criterion for drawing symbolic boundaries. More importantly, it points to the brittleness of categorical boundaries, highlighting musical genres as quintessential examples of socially constructed category systems.

Financial assets, on the other hand, should, in theory, be objectively classifiable. Modern portfolio theory, which remains the pervasive paradigm in financial theory, posits that investors seek to maximize future gains as a function of their tolerance for risk (Markowitz 1991). Returns (or the ratio of money gained or lost on an investment) and volatility (normally calculated as the standard deviation of a return) are therefore the natural criteria for classifying financial assets; in principle, investors should not be concerned with the identity of a stock.

But a central component of portfolio theory is the concept of diversification: the reduction of risk through investment in a varied set of uncorrelated securities (Bodie et al 2004). In theory, diversification relates to an asset's expected return and volatility. But because future performance is very difficult to extrapolate from the present, investors often assume that firms belonging to the same industry are likely to exhibit similar short term performance. Diversification therefore practically implies mixing different asset types. How assets are classified thus becomes highly consequential for investors' behaviors.

Indeed, as Zuckerman (1999, 2000, 2004) demonstrates, how a firm is classified affects how it is interpreted by investors: firms that are ambiguously categorized tend to be undervalued and overtraded

by investors who struggle to make sense of these companies' business models, and how they should be evaluated. Formal classification schemes provide investors with cognitive energy-reducing tools to assess financial products. This often leads them to evaluate securities on the basis of their industrial sector identity, rather than 'real' financial attributes. In fact, modal investors do not behave as rational action theory expects them to on a variety of measures (Kahneman and Tversky 1979; Goetzmann and Kumar 2008; Coval and Moskowitz 1999). While these divergences are often explained in economic models as cognitive 'biases' or 'preferences,' sociological work (Ruef 2000; Hsu et al 2009; Smith 2010) suggests that how assets are inter-subjectively categorized plays an influential role in producing these supposed 'anomalies.'

1.3 Chapter Structure

The studies cited above provide ample evidence for the role that institutions, and the formal classification they produce, play in shaping the behaviors of music consumers and financial investors. The objective of this dissertation, however, is – having established that categories matter in music and finance – to explore how these categories emerge and change through interpersonal interaction, long before becoming institutionalized in formal classification schemes. Of course our two online network sites are not ideal lab-like settings: no one joins either community without having been socialized to think of the world in conventional categorical terms; it is unlikely that the average internet user is entirely unfamiliar with the difference between classical music and rock 'n' roll, or hasn't been exposed to media descriptions of the auto industry's tribulations or the technology sector's robustness. This prior knowledge serves as these actors' initial points of reference. What is of interest here, however, are the ways in which this knowledge is collectively explored and transformed.

Chapter 2 lays out the theoretical and methodological foundations of this dissertation. This chapter's objectives are twofold. It begins with an overview of relevant literature on social cognition, and ties it to recent sociological debates on practice theory and its methodological implications. My main argument is that because people employ a fuzzy categorical logic when making sense of the world, our methods should take this fuzziness into account. The chapter's second objective is therefore outlining a method for inferring categorical structures from actors' collective behaviors. This method conceives of the data as a bi-partite network whereby individuals are connected to the objects they consume. Its main operational building block is the concept of categorical similarity: the proximity between two objects in this bi-partite network. Categories, I argue, inhere in clusters of objects that are unusually proximal to one another.

This definition may seem confusing. Where are the categories I am referring to, readers might ask. Most sociological (and for that matter, social scientific) studies rely on clear distinctions between categorical constructs, treating them as separate social 'things.' Yet these categories are analytical abstractions that often do not exist as such in the real world. Nor are they used by people with the surgical precision that imposes boundaries on what is more likely experienced subjectively as a continuum. It may be analytically convenient to make clear-cut divisions between 'junk' and 'investment grade' bonds, or between 'hip-hop' and 'R&B' musicians, but in reality people employ a multi-layered and flexible logic whereby such boundaries are contingent and implicit. Hence, I refrain from drawing such boundaries. Rather, I examine semantic associations between objects and their collective patterns, arguing that these are structured by implicit fuzzy categories, but stop short of delineating these categories and treating them as objective constructs.

In the third chapter I apply this method in order to trace the process of social construction. I argue that a socially constructed category system can be understood as a structural homology between two

types of networks: a network of interpersonal interaction, and a network of categorical similarity between objects. I demonstrate, in both communities, a time-invariant homology between people's interaction patterns and the schematic structures that orient their behaviors. I examine the empirical evidence for this homology on three structural levels: the individual actor, the dyad connecting two actors, and the network topology as a whole. Together, these three angles of analysis demonstrate that the more people interact in cohesive clusters, the more they tend to think of the world through crisply demarcated categories, and vice versa. This finding is consistent with a co-evolutionary model of interaction and categorization whereby the two mutually structure one another. To my knowledge, such a duality – though often discussed in sociobiological literature – has not been empirically demonstrated in such a systematic way.

If interaction and categorization are consistently homologous, then how does change arise? Chapter 4 addresses this question by shifting the analytical focus from structural stability to dynamism. It explores under what circumstances the categorical meanings of objects change. I argue that an object's meaning stability relates to its social boundedness: the extent to which it is associated with a singular social group. When a new group of people begin adopting an artist so far associated with another group, for example, this artist's meaning begins to change. I conceptualize change in meanings as categorical dissimilarity between an object and itself over time, and demonstrate, in both communities, that when an object's boundedness changes, it is likely to be subsequently reinterpreted.

The existence of change poses another fundamental question: if people's behaviors are funneled through categorical structures external to them, then what incentives do they have to defy these boundaries? In chapter 5 I argue that this is related to the availability of a consensual measure for performance. In the financial investment community, one's portfolio rate of return serves as such a measure. In the music community, on the other hand, there is no agreed-upon method for

differentiating people by performance. I demonstrate that performance-measurability affects which actors are rewarded and which are penalized for displaying unconventional combinations of objects. When performance-measurability is low, status becomes consequential: high status actors are rewarded for being entrepreneurial, and low-status actors are penalized. But in the financial investment community status is insignificantly related to rewards or penalties for nonconformist behavior. I argue that this is related to a complex tradeoff between performance and status, and how both are used for making sense of actors who appear to be unconventional.

1.4 On the Classification of Things and People

In their 1903 essay on ‘primitive classifications’, Durkheim and Mauss made a provocative and bold statement about the very basics of human cognition. The classification of things, they argued, reproduced the division of people into groups. What they were trying to assert was that social structures, and the hierarchies they entail, produce a cosmology through which non-social things are perceived. Members of an Australian tribe divided into two moieties, they maintained, perceive the animal kingdom through dichotomous distinctions that reflect this tribe’s fundamental social divisions. Social structure thus becomes an epistemological model through which the world is experienced.

More than half a century later, Foucault (1970) began his ‘archeology of knowledge’ with an entertainingly bizarre animal taxonomy taken from a “certain Chinese encyclopedia” conjured in a short story by Borges. Resonating with Durkheim and Mauss’s principles of classification, this example illustrates Foucault’s thesis that our knowledge of the world is grounded in our experiences of social order. Classifications are, essentially, social constructions.

In both cases, social order is assumed to be projected on the world like a sense-making template. Yet a different way of thinking of this classificatory reproduction is to reverse the arrow: to see the

classification of non-social things as creating symbolic boundaries that reinforce and sustain what are otherwise difficult to defend social divisions. Much of the recent sociological literature on categorization espouses – whether explicitly or implicitly – this approach. This is also the position I take in this study.

It would be somewhat of a long shot to argue that the social structures emerging over a period of several months in the online communities I study produce different epistemologies through which music and finance are understood. As Needham aptly notes in the introduction to the 1963 edition of their essay, to demonstrate Durkheim and Mauss's cosmological argument one would need to show cross-culturally how similar modes of societal organization are reflected in corresponding non-social classifications. My objective is more modest. I explore the micro-dynamics of online interaction with the expectation of finding that when objects are consumed predominantly by a singular group of people, these objects are perceived as making up a distinctive category; through its association with this group the category becomes its symbolic marker. This, I argue, is how social and categorical boundaries co-evolve to reproduce one another.

Chapter 2

Measuring Practical Classifications: Theory, Data and Methods

2.1 Classification and Social Cognition

Society and individual cognition interpenetrate one another through the symbolic division of the world into different categories (Durkheim 2001; Brubaker and Cooper 2000; Zerubavel 1997).

Categorical perception is a basic and universal human cognitive strategy for imposing discrete order on the inherently messy realities each and every one of us encounters on a daily basis. Our knowledge about the world is cognitively represented using complex mental structures commonly referred to as schemas. Schemas organize knowledge that is abstracted from our daily experiences and function as filtering mechanisms through which new experiences are processed; they are, in essence, cognitively-implemented systems of classification (Fiske and Linville 1980; D'Andrade 1995). When we listen to a new band, or consider investing in a firm we encounter for the first time, we interpret this new information in terms of the schemas it activates. Salient features, such as musicians' hairstyle and instruments, or a company's logo, serve as cues for this activation. Schemas therefore function as cognitive filing systems for storing and retrieving information about the world.

In order for it to be useful, schematic categorization needs to conform to two principles: cognitive economy, and empirical reassertion (Rosch 1978). First, categories need to provide individuals with the ability to process sensory input with as little cognitive energy as possible. They need to be, on the one hand, parsimonious representations of the world, but at the same time sufficiently detailed so as to

effectively capture meaningful differences between relevant phenomena. Such an optimization is only useful insofar as it corresponds to how the world is ontologically experienced. The second principle of categorization is therefore that the correlational structure assumed by the classification system be borne out empirically. The category 'heavy metal', for example, is useful only if its activating cues – distorted guitars and vocals, tattooing, leather clothing and dark makeup – co-occur empirically. Category systems are therefore always defined by a tension between reliable representation, and economic reduction.

Sociologists have long been interested in the categorical inner-workings of social cognition. When shared by more than one person, categories become intermediaries between individual cognition and larger social structures. This is what sociologists often term 'inter-subjectivity' (Scheff 2006). Inter-subjectively shared categories make social action workable. They coordinate people's expectations from one another, and their behaviors toward each other (Berger and Luckmann 1966; Zerubavel 1997). But because categorical perception is by definition reductive, it necessarily distorts the reality it perceives. Thus when categorical schemas are replicated across actors, social structure becomes subjectively embodied.

Existing sociological work on classification and its social ramifications is hampered by two key limitations. First, most work, despite recognizing that categories are socially constructed, treats them empirically as objective constructs. Consequently, it also essentializes them as fixed. But a variety of studies demonstrate that categorical boundaries are permeable, and therefore demarcate fuzzy domains (Hannan et al 2007; Rao et al 2005; Peterson and Kern 1996). Recent work in organizational ecology has made great contributions to moving beyond such reductive assumptions by examining how categories emerge and erode (Ruef 2000; Hsu et al 2009; Zuckerman 2004). Yet it still conceives of reality through institutionalized classificatory schemes, such as music industry's assignment of artists to

genres or financial analysts' assignment of firms to sectors. This second limitation, namely the focus on institutions rather than individual actors, does not account for the different ways in which people use categories in practice.

Practical categorization systems are different from formal classifications in three important ways. First, unlike discrete taxonomies which create neat and mutually exclusive divisions, in practice, people rely on porous and often ambiguous classification schemes. Rather than using clearly defined binary oppositions they employ a fuzzy logic (Bowker and Star 1999; Hannan et al 2007). Second, the meanings that are attributed to a categorical label do not inhere in the category in and of itself but rather in its association with other categories. That is, categories are only meaningful insofar as they relate to one another (Mohr 1998; Lévi Strauss 1963; Emirbayer 1997). Finally, when used in everyday life, categorical identities are not discerned using a sequential flow-chart-like process but are rather seamlessly activated using reflexive heuristic logic.⁷ People are not always aware or capable of articulating the categorical underpinnings of their behaviors.

These principles imply that understanding how people use categories inter-subjectively in real life requires conceptualizing classification systems as networks of associations between concepts. Rather than examining how analysts, specialists and practitioners formally classify the world, we need to investigate these patterns of association in people's everyday practical behaviors. This chapter provides an overview of the methodology used in this study for achieving this goal. It begins with a summary of the theoretical motivation for the methods used, elaborating on the three characteristics of practical classifications and how they relate to sociological practice theory. The next section offers an overview of the music and finance data used in this study and describes the various network-based structures employed to analyze it. The chapter continues with an outline of the methodology used for

⁷ Reflexive here is meant in the sense of reflex-based automatic behavior, not self-aware reflexivity in its methodological and ideological sense.

operationalizing practical classifications, and provides empirical evidence for its usability and construct validity. It concludes by discussing the advantages, and inevitable limitations, of the data and methods used in this study.

2.2 Practical Classifications

Formal classification systems are usually defined by precise principles by which different entities – be they objects, people, or situations – are assigned a singular exclusive label. Bureaucracies rely on such classificatory schemes for systematized procedures. Different employees in an insurance company, for example, follow the same procedural codes in categorizing claim eligibility. Social scientists similarly commonly reduce complex realities to simple, mutually exclusive categorizations (Hannan 2010). But folk taxonomies are never that clear-cut. Consider the seemingly simple task of deciding whether or not to add an acquaintance to one's pool of Facebook friends. The social networking online application requires enforcing a sweeping dichotomy – between those who are friends and those who are not – on a far more nuanced reality whereby the boundary distinguishing between friends and mere acquaintances is almost impossible to define. Folk taxonomies differ from formal taxonomies because human cognition operates in a manner incongruent with the precision of formal logic. "Practice", as Bourdieu puts it, "has a logic which is not that of the logician" (1990: 86). In the following paragraphs I review three dimensions that characterize categorical perception – (1) fuzziness, (2) relationality, and (3) reflexivity – and their implications on how people use categories in practice.

2.2.1 Fuzziness

In an influential experiment, Rosch (1975) asked subjects to rate different objects in terms of their typicality of the category 'furniture'. She found that rather than being a binary distinction, category

membership is graded: whereas some objects are central to a category and therefore serve as its prototype, others are more marginal. Rosch and other colleagues' findings resulted in a fundamental departure from hitherto prevailing models of categorical perception. According to the traditional Aristotelian logic, an object is classified based on a set of binary characteristics such that it either exhibits the necessary conditions for membership in a category, or does not (Bowker and Star 1999). Prototype theory, in contrast, conceptualizes of categorical membership as a graded property, whereby objects can be assigned to multiple categories with varying degrees of strength. It moves from thinking about categorical assignment in terms of simple binary designations to a fuzzy-set logic allowing for gradation (Hannan 2010).⁸

Prototype theory assumes that when classifying information we rely on an abstract mental image of what a prototypical member of a category is, and use logical extensions and metaphors to consider whether an object under investigation can reasonably be assigned with this categorical label (Bowker and Star 1999). Prototypical extension satisfies the need for cognitive economy, as categories are represented only by their most typical members, and at the same time provides enough elasticity for the classificatory logic to be empirically reasserted (Rosch 1978).

The implications of prototypical classification are twofold. First, prototypical perception suggests that boundaries demarcating categories are porous and fuzzy, and therefore that categories can overlap (Rosch 1978; Hannan 2010). If we imagine a classification system as a two-dimensional space, we can think of the categories it is made up of as Venn circles with faint boundaries. Some areas of the space are exclusively colored by one circle, whereas others are liminal zones corresponding to overlapping categorical identities (Figure 2.1). Note that unlike traditional Venn diagrams, the circles have no clear boundaries. Thus each object is understood in terms of its distance from the circles' centers, or from the

⁸ This theoretical move was largely inspired by Ludwig Wittgenstein (Rosch 1978; Hannan 2010). In *Philosophical Investigations* (1953), Wittgenstein very persuasively argues that people communicate intelligibly without having clear definitions for words.

various prototypes around which this categorical landscape is structured. The second implication is therefore that objects can vary in the level of their categorical distinctiveness. Whereas some objects are close to a singular prototype (eg. object A), others may be located in marginal areas that are not exclusively dominated by a single categorical identity (eg. object B).

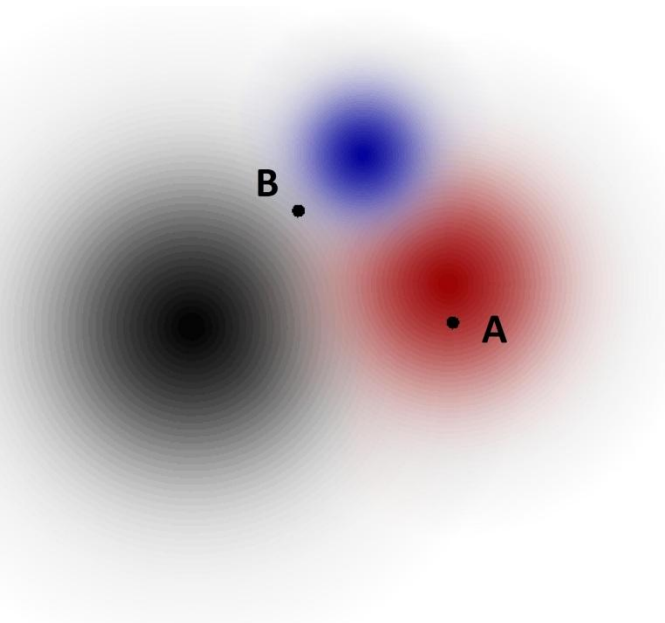


Figure 2.1: Visualization of fuzzy categories as a Venn diagram. Each colored circle represents a category, with its prototype located in the center. Color intensity corresponds to the distance from the center. Whereas object A has a distinctive categorical identity, object B's identity is categorically ambiguous.

Classifications are therefore characterized by two levels of fuzziness: object ambiguity, and category ambiguity. Hannan et al (2007) term the latter 'contrast', and suggest operationalizing it as the average member's grade of membership in a given category; the higher the average, the higher that category's contrast (and hence the lower its ambiguity). Such an operationalization necessarily implies a clear demarcation of a boundary; otherwise, it is impossible to calculate the average member's grade of membership. Yet prototypical thinking suggests an even more radical conclusion: if boundaries are vague, then categories do not exist in and of themselves. They are analytically useful perhaps, and can serve as ex-post explanations for actors' classificatory behaviors. In practice, however, they function as anchors – but not impermeable molds – for action.

Such an image is consistent with Swidler's (1986) conceptualization of culture as 'toolkit' (DiMaggio 1997). In her view, culture is experienced subjectively as a loose and incoherent hodgepodge of meanings that are translatable into various 'strategies of action', depending on whether or not these strategies are facilitated by the environment. Toolkit theory explains how context can prompt an actor to interpret the same reality in different ways. It conceives of the process of cognitive meaning-making not as a highly structured pattern of individual reasoning but as an externally stimulated selection from a multidimensional amalgam of possibilities. Such variability accounts for chronic inconsistencies in individual behaviors (Swidler 2001). If categories are fuzzy, then reality is amenable to multiple understandings. Culture provides the toolkit, but not the prescription, for such interpretation.

2.2.2 Relationality

The problem of categorical fuzziness is compounded by the relational nature of meaning: not only are categories ambiguously delineated constructs, they are also embedded in complex semantic networks. The meanings that people associate with a concept do not derive from its intrinsic properties but rather inhere in its multiple relationships with other concepts. To understand how an object is interpreted by an observer it is necessary to examine that object's perceived associations with other objects (Lévi-Strauss 1963; Emirbayer 1997; Mohr 1998; Mohr and White 2008).

Consider, for example, the online retailer Amazon. Some investors might think of Amazon primarily as a retail business; others might think of it as an internet-based service. Imagine that we were to ask a group of investors about the similarities between a number of firms, including Amazon, that are traded on Wall Street. Figure 2.2 illustrates the results of such a hypothetical survey. Retail companies, such as Wal-Mart, Target and Costco are strongly associated with one another, as are online-based companies such as Netflix, eBay and Yahoo. Amazon falls in between the two clusters that make up this

semantic network. Neither exclusively a retailer nor exclusively an online service, Amazon is associated with both types of business activities. The two unnamed categories, 'retail' and 'internet', are latent in the web of relationships between these various firms.

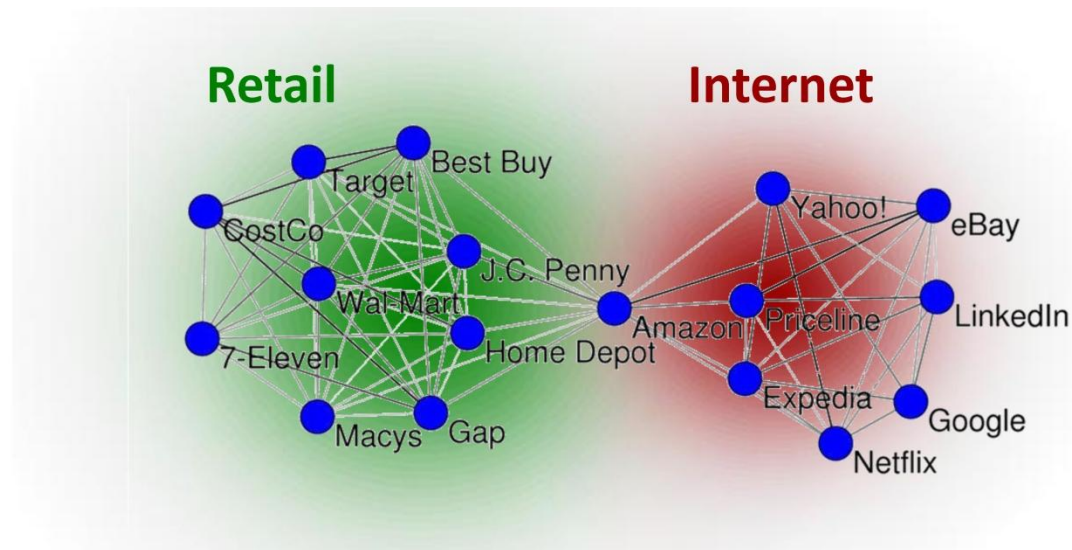


Figure 2.2: Hypothetical network of associations between firms traded on Wall Street. Latent categories illustrated with Venn circles. Edges vary in width and shade as a function of the strength of association between the two firms they connect.

The idea that meaning is constituted relationally was introduced by European structurlists more than half a century ago (Mohr 1998). Most influentially, Lévi-Strauss (1963) developed formal models of interpretative analysis aimed at identifying underlying structural principles in the patterns of relationships between concepts and cultural objects. Recent experimental work reaffirms his and others' assumptions about the relational nature of meaning-making. Naccache and Dehaene (2001), for example, demonstrate that unconscious semantic priming occurs when subjects are asked to rank numbers, suggesting the existence of an underlying latent category structure. Such reaffirmation is also derived from examining collective patterns of behavior. Cattuto et al (2009), for example, analyze the dynamics of online social tagging where users annotate websites with categorical labels. They find that various aspects of these so-called folksonomies can be explained in terms of a collective, albeit uncoordinated, exploration of a network-like semantic space.

Because classification is, in essence, the act of assigning meaning, it calls for a network approach (Mohr 1998). If meaning inheres in relationships between concepts then networks, such as the one depicted in Figure 2.2, are useful for representing and analyzing structures of meaning. Indeed sociologists have used network techniques as a means for deducing and interpreting semantic organization. Bearmen and Stovel (2000), for example, conceptualize narratives as networks connecting words to one another, and use them in order to analyze the self-reported process of identifying as a Nazi. Ghaziani and Baldassarri (2011) similarly analyze public debates on national LGBT marches and map the associations between major themes over the years, demonstrating how certain themes remained identity anchors for the LGBT movement throughout the two-decade period under investigation.

Ultimately, networks are mathematical tools for representing relational data. Yet merely capturing semantic data in the form of a network says nothing about their underlying meaning structures. Network analysis techniques are therefore necessary for uncovering the foundational structures that push certain nodes closer to one another. Such a structure is readily apparent when visually examining the network illustrated in Figure 2.2.⁹ In this stylized example, the naked eye easily discerns a division into two clusters of nodes, each corresponding to a different latent category. In most cases, however, the underlying categorical organization is more nuanced and difficult to detect. The task of the analyst is therefore to look for categories that cohere in patterns of clustered associations between nodes.

⁹ This visualization is not neutral: it was produced using the Fruchterman-Reingold force-based algorithm for determining nodes' location in the two-dimensional space. As such, it already includes a layer of analysis.

2.2.3 Reflexivity

What is perhaps most astonishing about human cognition is how seamless and effortless it appears to be in practice. For the analyst, interpreting network structures of meaning can be a very challenging and taxing experience. In real life situations, however, people cannot afford devoting much time or cognitive energy for performing such fundamental tasks. While the act of understanding intuitively implies a reflective process of deliberation, such meditation is, in actuality, quite rare. Rather, the prevailing consensus among cognitive scientists is that information processing is handled by two concurrent cognitive mechanisms. Only a small fraction of the information that we process is analyzed consciously and deliberately. The majority of our ongoing experiences of reality are automatically processed by subconscious, or reflexive, cognitive systems (Chaiken and Trope 1999; Lieberman et al. 2002; Galdi, Arcuri, and Gawronski 2008).

Schematic perception is central to this automated processing capacity. It allows us to interpret new information in familiar terms, rather than analyzing it anew. When such automatic recognition patterns cohere and diffuse inter-subjectively, society becomes internalized as unconscious individual dispositions. This is precisely what Giddens (1984) calls 'practical consciousness', and Bourdieu (1990) terms 'habitus' (Strauss and Quinn 1997): the mechanisms through which socialized behaviors become automated second nature.

One important implication of automated cognition is the use of heuristics. Because people do not have the time to thoroughly analyze new information, they need to pick up on salient cues and use them as interpretative shortcuts. When scanning the radio tuner on a rental car just picked up at the airport, for example, we look for very specific tonal characteristics to help us figure out quickly, and with minimal interference to our driving, whether the music now playing is of the kind we like. Consequently, employing our categorical perception makes it easier for us to interpret information that is congruent

with our schematic representation of the world, and filter out information that is not (DiMaggio 1997). As long as the world corroborates our schematic blueprints, our categorical worldviews become reinforced. But such reinforcement is constantly necessary. Schemas that are rarely primed by the environment gradually diminish in importance (Lieberman et al 2002). That is, our categorical perception is dependent on that which it attempts to perceive. When incongruences arise, our schematic perception is updated accordingly.

Our reliance on heuristics therefore entails that the environment external to our subjectivity plays a constantly causal role in shaping our behavior. It is not merely that we imbibe, through our conscious cognition, cultural schemas that are then integrated into our reflexive behavioral routines. Our automatic consciousness is continuously sensitive and reactive to information unconsciously received from the environment. The patterns and regularities in the social structures outside us therefore funnel and catalyze our individual action. The semantic network structures of the kind depicted in Figure 2.2 are not merely analytical abstractions of aggregate data. They also correspond to the regularities that individual actors observe in others, and that consequently shape their own categorical understanding of the world. In other words, it makes empirical sense to look for network structures of categorical meaning in collective behavioral patterns.

2.2.4 Uncovering Classifications in Practice

As many observers have noted, the contingent and fragmented nature of human cognition is highly consistent with sociological theories that focus on practice as a central arena of sociological explanation (Strauss and Quinn 1997; DiMaggio 1997). What is often termed ‘practice theory’ is not a coherent and formalized theoretical set of assumptions about society but rather refers to a more loosely defined theoretical approach that sees action as a nexus of social structures (Schatzki 2001). Much has

been written about practice theory – especially that which is associated with Pierre Bourdieu’s work (see for example Martin 2003) – and its influence on American sociology. It would be redundant to review this literature here. Suffice it to point out how this approach, particularly as it is manifested in Bourdieu’s (1990) work and Swidler’s (1986, 2001) cultural toolkit theory, corresponds to the three characteristics of categorical perception outlined above: unlike the highly socialized individual in Parsonian theory, the actor in Swidler’s and Bourdieu’s worlds acts upon fuzzy and inconsistent schemas; constructs meaning from relational regularities in fields of human practice (Bourdieu 1984; Martin 2003); and embodies external social structures through automatic cognition that makes action predominantly reactive.

The latter stipulation, namely, that inter-subjectivity operates primarily through providing ex-post justificatory logic for action, has recently been contested by sociologists of culture. Vaisey (2009) makes a highly persuasive argument that dual-processing theories of cognition imply that culture also provides explicit motivations for action, at least for some forms of action. But research in cognition continues to provide ample evidence that motivation is often consciously inaccessible, and that action is catalyzed by automatic reactions to the environment (Galdi et al 2008; for a review, see Custers and Aarts 2010). External cues unconsciously prime categorical thinking that results in reactive behaviors. The presence of a briefcase, for example, makes group members more competitive (Kay et al 2004), and a light scent of detergent prompts people to maintain the cleanliness of their immediate surroundings (Holland et al 2005). Swidler (2001) finds incoherencies in people’s conscious accounts for their motivations for romantic relationships, similarly arguing that catalysts for action can be found in social institutions, not conscious motivation. These findings suggest that if sociologists are interested in understanding the categorical logics that inform people’s behaviors they are better served by examining these behaviors than asking their practitioners to explain them.

This focus on action locates the interface between individual subjectivity and larger social structures in collectively routinized practices. If individual actors are incoherent and incapable of representing highly systematized cultural logics, then structural regularities should be looked for outside them. This is what Lizardo and Strand (2010) mean when they argue that both Bourdieu's and Swidler's are scaffolding theories of cognition. They conceive of objectified cultural structures (embodied primarily in institutions) as supports for action that funnel systematized behaviors where such systematicity is unattainable by individual human cognition and its inherent limitations. The majority of sociological work on categorization looks for such systematic patterns in institutionalized taxonomies produced, for example, by credit agencies (Ruef and Patterson 2009), art critics (Hsu et al 2009), and industrial bureaucracies (??). They mostly assume that individual actors reproduce these taxonomies. Yet such highly formalized reifications of schematic categorization cannot account for the fuzziness, relationality and reflexivity of human cognition. Understanding how categories mediate between social structures and human action therefore requires searching for patterns in actors' collective practical behaviors. In the remainder of this chapter I describe the data and methods used in this study to accomplish this task.

2.3 Data

The data used in this study include the full sets of activities in two online social network sites over a period of 10 and 17 months respectively: a music community (henceforth MOC) and a financial community (henceforth FOC). These data have been stored by the respective website owners and have been voluntarily shared by them for academic purposes. All the data used in this study are publicly accessible to site members online, and have been authorized for use by Princeton University's Human Subjects Review Board.

Both the MOC and FOC are typical examples of what boyd and Ellison (2007) call ‘social network sites’ (SNS): web-based services that allow individual users to construct a public profile, articulate a list of other users with whom they are connected, and view their and others’ lists of connections. Both websites can be defined as Dedicated SNSs: unlike generic SNSs such as Facebook, each of the two websites is dedicated to a particular form of activity, and is designed accordingly. The MOC is designed to allow users to share music, and the FOC is designed to allow them to share information about financial investments.

2.3.1 Overview of Data Types

The MOC allows users to register as DJs, and publish mixes of their own making on their individual homepages. Its stated purpose is to provide people with a platform for “sharing and discovering” music. Each mix contains eight tracks or more; usually, these tracks were produced by established (though not necessarily mainstream) musicians, not the users themselves. Users can also add to their homepage a photo, a short biographical description, and their location; not all users choose to do so. They can listen to other DJ’s mixes, and may comment on them. They also designate other users as people they follow. The data acquired include all user profiles, their mix compositions, what users they follow, what mixes they listened to, what mixes they commented on and the contents of these comments.

The FOC allows users to virtually trade stocks with the stated purpose of “aggregating investment knowledge to the benefit of our community members”. Each selling or buying activity a user makes is immediately posted on that user’s homepage. Users can only invest in stocks or exchange-traded funds (ETF) that are traded in US exchanges. Derivatives and other financial instruments cannot be included in a user’s portfolio. At each moment in time, a user’s homepage reflects her investment portfolio. When viewing another user’s profile, one is presented with the full asset breakdown, trading history, as well as

performance metrics (returns, volatility, beta coefficient and S&P correlation) and portfolio breakdown by sector and market capitalization. Users can designate up to three other users they follow, as well as an unlimited number of other users in their feed, meaning that they get notified for any activity that user has. On their homepage, users can add a photo and a philosophy statement. They can also add opinion notes, which other users can comment on.

Despite their substantive differences, and their different look-and-feel, the MOC and FOC are practically identical in terms of structure. The types of data included in each SNS can be categorized as follows:

1. Affiliation data: users can nominate other users as people whom they follow. Such nominations do not require reciprocation. In effect, affiliation data make up a directed network structure whereby users are connected to one another.
2. Collection data: composition of users mixes (MOC) or financial portfolio (FOC).
3. Conversational data: comments users leave on each other's mixes (MOC) or opinion notes (FOC).
4. Personal data: such as photos, biographical statements and location.

Table 2.1 summarizes these data types. The analyses that are included in this study rely primarily on the first two data types. See Appendix A for more information about the data and procedures implemented to clean and structure them.

Table 2.1: Data Types by Data Source

Data type	MOC	FOC
Affiliation*	Followers	Followers
Collection*	Artist composition	Stock/ETF composition
Conversational	Comments on mixes	Comments on opinion notes
Personal	Photo, biographical statement, location	Photo, personal statement

* data used in empirical analyses contained in this study

An important way by which the two communities differ is that the FOC ranks users by performance, whereas in the MOC no such ranking exists. In other words, consensual evaluation criteria for financial assets are explicitly built in to the FOC platform. Nevertheless, while the MOC does not explicitly rank users, it still lists users' mixes by default according to their popularity. Hence in both websites users are directed by the website to access other users who enjoy high popularity.

All data types except for personal data include timestamps indicating the exact time in which the user activity happened, for example, when a user designated another user as someone she follows, or added a new mix to her music collection. The data included in this study correspond to activity performed on both websites between 2008 and 2009, ranging from both SNS's launch dates and until the data were provided by their respective owners.

These data lend themselves to two types of units of analysis: (1) the individual user and (2) the object. Summary statistics for these two types of units of analysis are reported in Table 2.2. Users correspond to real-life individuals who have registered on the website and created an account¹⁰. Objects correspond to these users' collection compositions. In the FOC, collections are made up of financial assets, and hence each object corresponds to a unique financial asset. In the MOC, collections are made up of different tracks. Overall, users included 205,832 unique tracks in their collections, 80.13 percent of which appear only once, while 99.17 percent appear no more than five times. Only ten tracks appear in more than twenty mixes. Given these very low frequencies, it makes little sense focusing on tracks as units of analysis. Consequently, the analyses included in this study treat MOC collections as comprised of different artists. Such a focus makes theoretical, not only practical sense. Classical music has traditionally been identified with its composers, and popular music with its performing artist. Artists anchor meaning, providing a flesh-and-blood embodiment linking music with social identity (DeNora

¹⁰ It is possible for one individual to create more than one user account. Such multiple account membership is not controlled by the website.

2000). This focus on artists as objects allows bridging what Roy and Dowd (2010) call the textualist and contextualist approaches to music analysis.

Table 2.2: Summary Statistics for Units of Analysis

	MOC	FOC
Dates of operation	August 2008 – June 2009	May 2008 – October 2009
Users	27,584	5,383
Objects	Artists	Financial Assets
Count	16,541	5,249
Objects per user (mean)	20.88	9.77
Objects per user (standard deviation)	35.22	30.84
Statistics are aggregated over the full time period		

2.3.2 Network Structures

As their name suggests, social network sites are, in essence, network-based data structures. The Dedicated SNSs used in this study can be represented as two interrelated networks: (1) The *Interaction Network* which connects users to one another, and (2) the *User-to-Object Network* which connects between users and objects. These network structures are illustrated in Figure 2.3, which depicts a small subset of the FOC data. The complex graph illustrated in this figure connects two types of nodes, corresponding either to users (green circles), u_i , or objects (blue squares), o_j . Blue edges connect users to one another and make up the interaction network, and black edges connect users to objects and make up the user-to-object network.

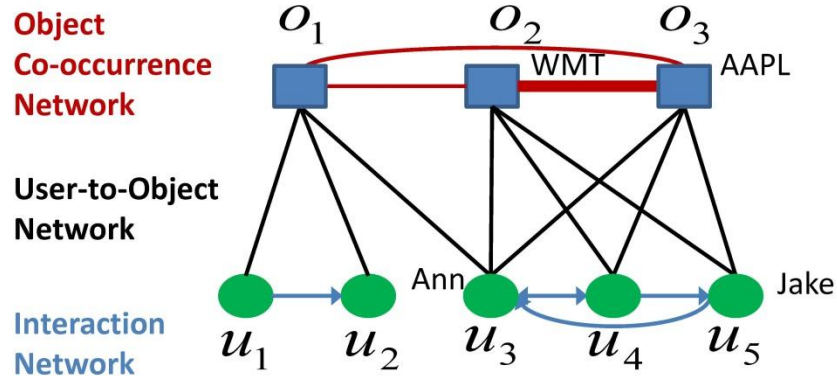


Figure 2.3: Illustration of the network structures included in the data

Let N be a set of users, and K a set of objects. The interaction network is a binary directed network connecting users in N . Its edges correspond to users' nominations of other users whom they follow. When Jake decides to follow Ann, for example, an edge connecting the two nodes corresponding to these users is added to the network. The interaction network can be represented as an adjacency matrix S of size $N \times N$ such that $S_{ij} = 1$ if user u_i follows user u_j .

The user-to-object network is an undirected binary bi-partite network connecting users in N with objects in K . It corresponds to users' collection compositions. A bi-partite network is a network comprising of two subsets of nodes such that each edge connects nodes in both subsets (Wasserman and Faust 1994). When Jake adds Apple stock to his investment portfolio, for example, an edge is added to the network connecting Jake's user node with the object node corresponding to Apple's stock. The user-to-object network can be represented as an $N \times K$ sized adjacency matrix A such that $A_{ij} = 1$ if user u_i adds object o_j to his portfolio, that is, if i invests in stock j .

The relationship between two objects in A , for example between Apple and Wal-Mart stock, can be conceptualized as the number of users who have both objects in their collections, or the number of collections in which both objects co-occur. In the example illustrated in Figure 2.3, the number of such co-occurring collections for Apple and Wal-Mart is three. The bi-partite network A can therefore be

transformed into a one-dimensional co-occurrence network W which connects objects to one another.

This *Object Co-occurrence Network* can be represented as a $K \times K$ sized weighted matrix such that

$W_{ij} = \sum_{u \in U} A_{ui} A_{uj}$. This network is illustrated in red in Figure 2.3. Figure 2.4 visualizes the matrices

used to represent the three different networks.

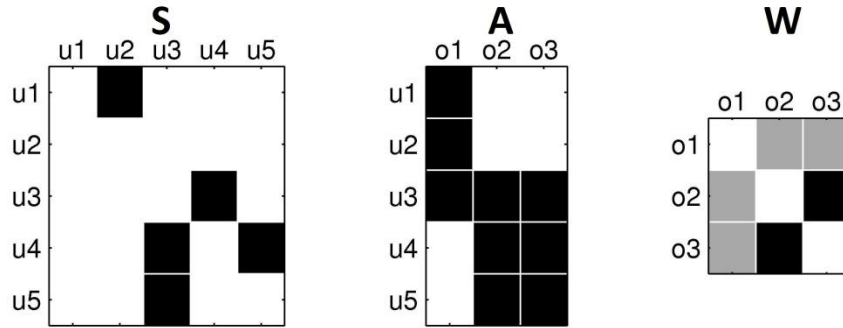


Figure 2.4: Visualization of adjacency matrices for the interaction network (S), user-to-object network (A) and object co-occurrence network (W). Shading, where exists, corresponds to edge weights. Matrices correspond to data in Figure 2.3.

2.3.3 Accounting for Network Evolution

The structures outlined in the previous section describe static networks. The edges drawn in the user-to-object network in Figure 2.3, for example, describe the existence of a connection between the user and object, but are indifferent to the time this connection was formed. Yet time is important. The timing of a stock investment is obviously consequential, as it depends on the asset's market price in that moment. Not dissimilarly, MOC users are influenced by events in the music market.¹¹

More importantly for the purpose of this study, however, if we assume that the actions people take are influenced by their observations of other people then time becomes analytically crucial. Whether two users invest in the same stock within a few minutes or within a few months from one another implies different interpretations about the extent of similarity in their behavioral motivations.

¹¹ A release of an artist's new album, or that artist receiving media attention, may have an effect on users' overall awareness of that artist.

Similarly, two users may be adding a specific artist to their respective collections. If they do so within months from one another, the meaning collectively associated with this artist – and therefore the symbolic significance of liking this artist – may have substantially changed. As illustration, consider the different meanings associated with Led Zeppelin in 1969 and in 2010. Whereas in the late sixties the band still enjoyed an aura of youth counter culture, forty years on it has become a symbol of mainstream rock music. Though the MOC covers a period of roughly one year – not four decades – timing is nevertheless similarly important for deducing meaning.

A common strategy for addressing network dynamics over time is to divide the full time-range into T equally-sized time-slots of length α , resulting in T static networks. Each network represents all edges that were formed within the time interval $[t, t + \alpha)$ as if they were created simultaneously (Moody et al 2005). A simple strategy of this kind is to divide time into non-overlapping slots, such that each event falls exclusively within a singular time-slot. Such an approach necessarily imposes arbitrary divisions over time, such that events that occurred within moments of one another might fall on different sides of the division fault line. To prevent such an arbitrary imposition, I adopt the sliding window approach which is often used in studies of network dynamics (Kossinets and Watts 2009). Imagine that time is visualized as a straight line progressing from left to right. The sliding window is a filter of a fixed size, α , that is slid over this timeline in fixed intervals of size δ , such that $\delta < \alpha$ (see Figure 2.5). Like a moving average, the sliding window filter generates a smooth transition from one network snapshot to the other, such that each pair of sequential networks overlap by $\alpha - \delta$. The sliding window method results in the creation of T adjacency matrices for each matrix type, such that $S_{ijt} = 1$ if user i is following user j at time t , $A_{ijt} = 1$ if user i includes object j in his collection during time t , and $W_{ijt} = 1$ if objects i and j co-occur during time t .

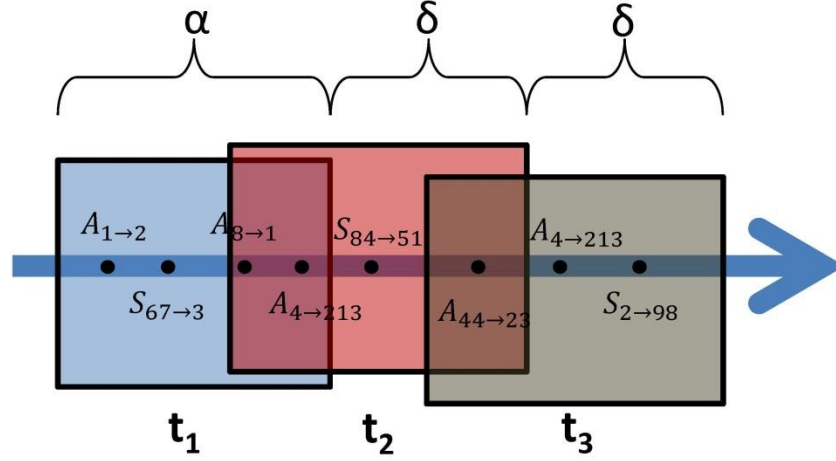


Figure 2.5: Illustration of the sliding window technique for transforming continuous data into discrete network snapshots. Each rectangle represents a time-slot. Dots represent a user action. For example, user 84 follows user 51 during time 2.

Before deciding on the sizes of the two parameters defining the sliding window specifications, we need first to define more clearly what the T network snapshots constitute. In the interaction network, S , an edge corresponds to a user following another user. The network S at time t should therefore correspond to users following one another during this time. Intuitively, such a requirement may imply that S_t include only edges that were formed during time t , i.e., decisions by users to follow other users during that time-slot. Such an operationalization would disregard all edges created in previous time-slots. But there is no reason to believe that follower relationships necessarily decay over time.¹² Consequently, network S_t is constructed such that it includes all edges formed during the time period $[1, t)$, i.e. from the moment activity began in the community and until time-slot t ends (unless an edge was actively removed sometime during that period by a user stopping to follow another user).

The user-to-object and object co-occurrence networks relate to users' collection compositions. The networks reflect these users' underlying practical classification schemes. Unlike with the interaction relationships, an artists that was recently added to a collection more strongly reflects a user's orientation than one added several time-slots ago. Hence with network A and its one-dimensional

¹² For example, in the MOC, users continue to comment on mixes even after a long time has elapsed since they began following their owners.

projection W , a time-slot constitutes only relationships that were created (or re-created) between users and objects (and by extension, between objects in network W) during that time. In other words, edges can be removed from the network between time-slots without users having to take any action. In Figure 2.5 user u_1 includes object o_2 in her collection at time t_1 , but not in later time-slots. The corresponding edge is added to network A at time t_1 and removed later. Conversely, the edge connecting user u_4 to object o_{213} remains in the network throughout the three time-slots.

Because edges in network S cumulate over time, α and δ are only consequential for the two other networks. There is no readymade solution for determining these parameters' sizes (Moody et al 2005). The window size, α , needs to correspond in a meaningful way to the type of activity being represented by the network structure in question. Because we are interested in exploring how people's categorical evaluations of objects change over time, α should ideally be sized such that it corresponds to the maximal amount of time during which such change does not occur. Moreover, to be analytically useful, α needs to be sufficiently large to include enough data per time-slot, but sufficiently small to create a large enough number of time-slots.

Users in the MOC who created at least two mixes waited 18.1 days, on average, between creating one mix and the other. Consequently, I define $\alpha=31$ for the MOC, roughly twice the average interval between creating two mixes.¹³ People's evaluations of music do not change very quickly, hence a month appears to be a reasonable specification. One would imagine that dynamics in the FOC are substantially faster. Given that people are trading stocks, the value of which can change significantly in minute-sized time frames, the equivalent reference size should be much smaller in the FOC. But FOC users are predominantly lay investors. Unlike Wall Street traders, their reaction times and investment horizons

¹³ In order to optimize T and data size per time-slot, I use a number slightly smaller than double the average.

are much longer than minutes. In fact, it took FOC users on average 9.24 days between any two actions in their investment portfolios. Due to the FOC's relative small data volume, I define $\alpha=25$.¹⁴

Theoretically, one can choose δ to equal the finest time resolution available. Such an implementation would, however, be computationally unfeasible ($\delta=1$ sec would result in $T \approx 4 * 10^7$), and highly inefficient. A better choice for δ is a large enough number to allow for some sort of meaningful change between two time-slots. To estimate this size, I calculated the median time that transpires between any two co-occurrences for all pairs of objects in the MOC. This time interval was discovered to be 23. To optimize for analysis, I defined $\delta=21$ in the MOC. In order to maintain the same δ/α ratio in the FOC for comparability purposes, it is defined as $\delta=18$.

2.4 Measuring Classifications in Practice

As discussed in section 2.2, the characteristics of categorical perception make networks ideal tools for analyzing how people use categories in practice. Reflexivity suggests that categories inhere in users' aggregate behaviors; relationality implies looking for meaning in networks of associations between concepts; and fuzziness suggests that categories are latent in these network structures. Practical social classification systems can therefore be deduced from networks representing patterns of association between objects, as they cohere in people's aggregate daily behaviors. The object co-occurrence network satisfies these requirements.

Figure 2.6 illustrates two representative snapshots from network W for the MOC and FOC (at times 4 and 9 respectively). Recall that each node in this network corresponds to an object, and edge weights correspond to pairs of objects' co-occurrences. As is clearly visible, the two networks are highly

¹⁴ All the analyses reported in this study were also conducted with different α and δ specifications (both larger and smaller in size than those reported here), and results remain robust to different specifications.

centralized, such that the majority of nodes appear in the center. Though it is difficult to notice due to node overcrowding, high-weight edges are concentrated in the network center. In both networks there is no clear pattern of association between objects dividing the network into different clusters of association. Unlike the network illustrated in Figure 2.2, there seems to be no underlying latent categorical logic structuring users' decisions about what objects to include in their networks.

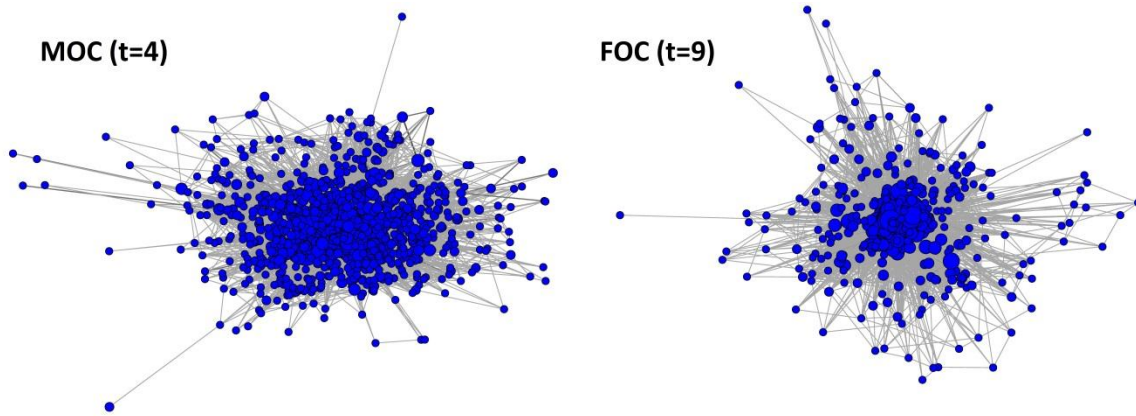


Figure 2.6: Snapshots of network W for MOC and FOC data. Node size is proportional to its object's popularity. Edge shade corresponds to its weight.

2.4.1 Standardizing Co-occurrences

This centralized structure is consistent throughout the T timeslots in both communities. Does this consistency suggest that users are not following some sort of semantic order but rather are more or less arbitrarily adding objects to their collections? As we shall see in the next paragraphs, this seeming arbitrariness is attributable to the unequal distribution of object popularity, or the number of collections in which an object appears. Formally, an object's popularity is defined as its node degree, $d_i^A = \sum_{j \in N} A_{ji}$ in network A , or the number of user nodes adjacent to it. If objects enjoy different levels of popularity, then even if users randomly choose objects into their collections, we should still expect to find that

highly popular objects co-occur more frequently with other objects. Co-occurrences are therefore very coarse measures of semantic association.

Indeed, Figure 2.6 demonstrates that popular nodes are exclusively concentrated in both networks' cores. In fact, as Figure 2.7 illustrates, central nodes in network W also tend to enjoy high popularity. The scatter plot in Panel A plots object popularity against its corresponding node's strength in network W . Like node degree, strength measures the sum of the weights of the edges connecting a node to other nodes.¹⁵ Node strength, $s_i^W = \sum_{j \neq i} W_{ij}$, is a simple measure of network centrality. The correlation between the two measures is extremely high (0.8538), suggesting that popular objects also tend to have heavy-weighted edges with other objects.

It appears that popularity is a strong driver in the formation of this network: nodes with high popularity tend to co-occur with similarly popular nodes, generating a centralized network structure. Panel B of Figure 2.7 provides evidence that indeed high-degree nodes tend to co-occur with one another. It plots nodes' weighted average nearest neighbor degree as a function of node degree. In network W , node degree is defined as the number of edges in that node's immediate neighborhood, $d_i^W = \sum_{j \neq i} (W_{ij} > 0)$. Node i 's weighted average nearest neighbor degree is defined as the average node degree in i 's neighborhood (i.e. nodes connected to i), weighted by edge weights (Cattuto et al 2009), or formally, $d_{nn,i}^W = \frac{1}{s_i^W} \sum_{j \neq i} W_{ij} d_j^W$. By extension, $d_{nn,i}^W(d)$ is defined as the average $d_{nn,i}^W$ for all i for which $d_i^W = d$. As plot B illustrates, nodes tend to be connected with nodes of similar degree.¹⁶

¹⁵ For binary networks, where edge weights are strictly 1, node degree and strength are identical.

¹⁶ Though the figure refers to one network snapshot, similar trends were observed in all network snapshots in both communities.

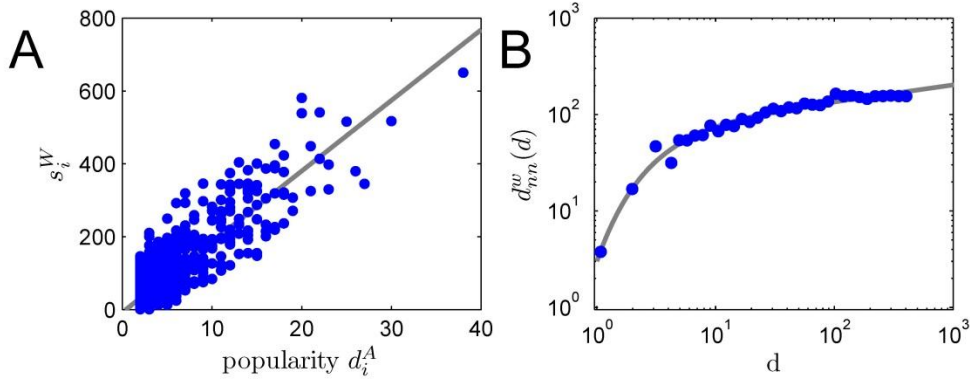


Figure 2.7: Assortative mixing in MOC co-occurrence network for $t=4$ (corresponding to Figure 2.6). Panel A plots object strength as a function of popularity. Panel B plots average $d_{nn,i}^W$ for nodes of a given degree. Grey lines plot fitted curves.

Assortative mixing by degree, as this phenomenon is often referred to (Newman 2002), is consequential when degrees are unevenly distributed. As Figure 2.8 demonstrates (upper panels), object node degrees appear to be following a heavy-tailed power-law distribution in both communities. Power-law distributions are hallmarks of complex evolving system. As Barabási and Albert (1999) influentially demonstrate, their emergence in networks can be associated with a generative process whereby new nodes preferentially attach to already connected nodes. But a thorough analysis of these distributions suggests that while degree distribution in the interaction network follows a power-law (bottom panels), in the user-to-object networks it does not.¹⁷ In other words, while the patterns of interaction in both communities are consistent with a preferential attachment model, they are not so in terms of how users add objects to their collections.¹⁸

¹⁷ I use methods described in Clauset et al (2009) to estimate a minimal cutoff and exponent for fitting a power-law distribution, as well as the simulation-based Kolmogorov-Smirnov test they describe to estimate the likelihood that the observed distributions are significantly different from power-law distributions. Only the interaction networks reasonably fit power-law distributions.

¹⁸ This does not suggest that nodes do not preferentially attach to one another. Indeed it is highly plausible that users are more inclined to include in their collections artists that are popular with other users. Nevertheless, these results suggest that preferential attachment does not fully explain the network's evolution.

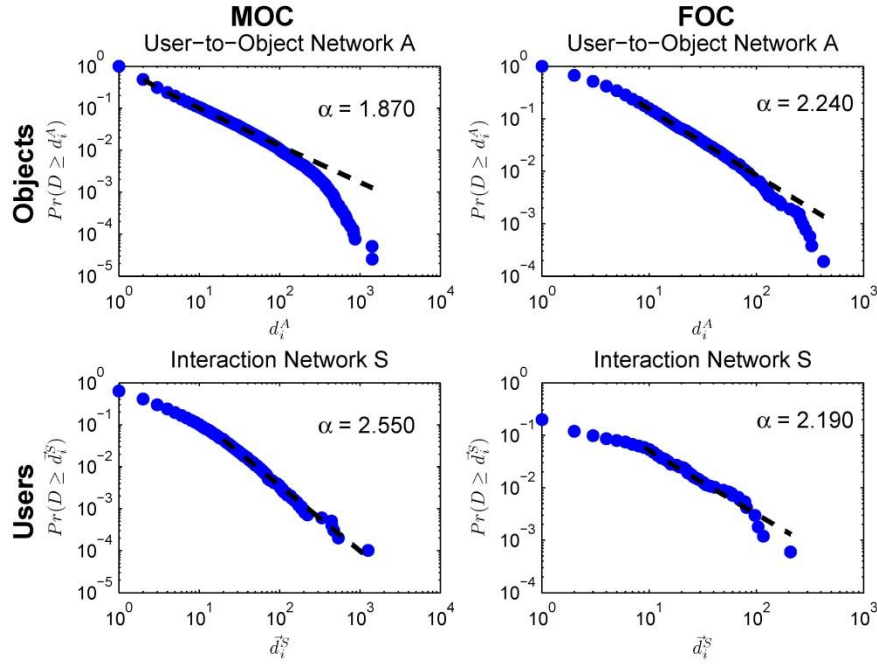


Figure 2.8: Heavy-tailed distributions of node degrees in networks A (objects only) and S. Both degrees are calculated for the whole time period, such that $T=1$. Dashed lines plot the fitted power-law. The α parameter is estimated to fit the power-law model $\Pr(x) \propto x^{-\alpha}$.

Object popularity is clearly shaping the patterns of object co-occurrences. It is not surprising that the Beatles, often considered the most influential act in popular music history, is the second most popular artist in the MOC; similarly Apple, which has received continued media attention throughout the late 2000s, is the second most traded stock on the FOC. But something more than mere exogenous popularity is driving network dynamics; to account for it, it is necessary to standardize weights in W by their respective objects' popularities.

Let us assume that popularity is defined exogenously.¹⁹ If users were to add objects to their collections based solely on these objects' popularities, we would expect the weight of an edge connecting two nodes in network W to correspond to both objects' popularities. Because $s_i^W \propto d_i^A$, we can use node strength in W to estimate the effect of popularity on edge creation. Imagine W as a multi-

¹⁹ i.e., that the Beatles are more likely to appear in a random collection than, say, they krautrock band Cosmic Jokers, net of any endogenous social influence.

graph, such that weighted edges are replaced by multiple binary edges. If we maintain the distribution of node strength constant, but rewire the multi-graph's edges at random, we should expect the number of edges connecting nodes i and j , or the weight of the edge connecting them in network W , to be as follows (Newman 2004):

$$(2.1) \quad E(W_{ij}) = \frac{s_i^W s_j^W}{\frac{1}{2} \sum_{ij} W_{ij}}$$

We can therefore standardize the observed weight by subtracting the expected weight from it:

$$(2.2) \quad \tilde{W}_{ij} = W_{ij} - \frac{s_i^W s_j^W}{\frac{1}{2} \sum_{ij} W_{ij}}$$

$\tilde{W}_{ij}$ is the weight of an edge connecting two object nodes in the co-occurrence network, net of effects of their mutual popularities. Values greater than zero suggest that the two objects co-occur more frequently than expected at random, and values smaller than zero suggest that they occur less than expected. The standardized weight captures information about the association between two objects above and beyond the likelihood of their co-occurrence as a function of their popularity.

To test the usability of standardized co-occurrence, I generated 1000 synthetic user-to-object networks using computer-based simulations. Each simulated network was identical to network A in the MOC in its dimensions and distribution of object node degrees.²⁰ Edges were created at random, while maintaining constant degree distribution. I then created corresponding object co-occurrence matrices for each simulated object-to-user object. Averaging the values of each cell in these synthetic co-occurrence matrices, I generated matrix ψW . Cell values in the synthetic and observed matrices are highly correlated: $\rho(\psi W, W) = 0.8298$. When standardizing both matrices, however, this co-variance disappears, such that $\rho(\psi \tilde{W}, \tilde{W}) \approx 0$. In other words, object popularity accounts for a great deal of

²⁰ A is computed for the full time range such that $T=1$.

object co-occurrence. Figure 2.9 plots the distribution of cell values in $\tilde{W}$ for the synthetic and observed data. While in synthetic data $\tilde{W}$ is distributed log-normally, in the observed data the distribution is double-peaked, with a second peak centering on 1. These cell values diverge from the underlying normal distribution, corresponding to co-occurrences that are either greater or smaller than expected at random. Net of popularity effects, these object associations, I argue, are driven by semantic relationships.

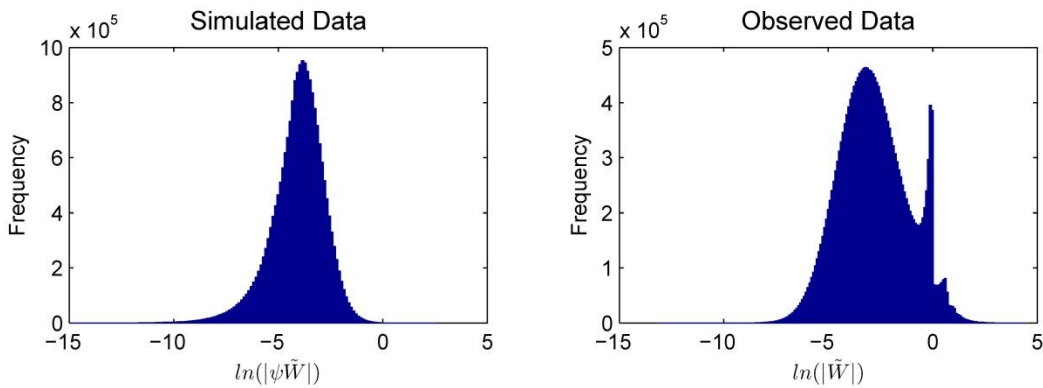


Figure 2.9: Distributions of standardized co-occurrences in simulated and observed data. Standardized co-occurrence is transformed by its absolute value, and then by its natural logarithm.

2.4.2 Measuring Categorical Similarity

The standardized weight is a high-definition tool for separating semantic association from its background noise. Nevertheless, it depends on objects co-occurring with one another. Yet categorically similar objects are not necessarily frequently included in the same collections. DJs, for example, are often motivated by a desire to adorn their mixes with musical breadth and unexpectedness; to create compilations of identical artists is, in essence, to empty DJing from its art. Not dissimilarly, financial investors are guided, at least in theory, by the axiom of diversification; a portfolio made up of highly similar assets carries a high risk of substantial loss. That two objects do not co-occur therefore does not necessarily imply that they are categorically distant. Consider a visualization of a hypothetical object co-

occurrence network. Imagine further that there exists a cluster of objects in this network, corresponding to a latent category, which includes two nodes that are not connected to one another. Though the two nodes do not co-occur with one another, they nevertheless co-occur with more or less the same other objects. In network terms, they are structurally equivalent.

Structural equivalence is a prominent concept in social network analysis (Wasserman and Faust 1994). When applied to networks representing social relationships, structural equivalence is commonly conceptualized as the relational manifestation of social role similarity. Two actors occupying equivalent network positions in, for example, a professional field, are likely to have attained similar status within that field (Burt 1987). Their similarity inheres not in their relationships with one another but in their relations with others. In a network connecting conceptual entities, structural equivalence similarly implies that two entities carry comparable meanings. Structural equivalence is therefore a means of measuring the extent to which two objects in the co-occurrence network are categorically identical.²¹ Categorical similarity is therefore defined as two objects structural equivalence in network $\tilde{W}$.

Let $\vec{\tilde{W}}_i$ denote a vector of length K , corresponding to the i 'th row in matrix $\tilde{W}$. This vector is, effectively, object i 's set of standardized co-occurrence weights with other objects. Two objects are said to be categorically identical if their respective vectors are identical. Recall that cell values in $\tilde{W}$ are created by subtracting expected co-occurrence from observed co-occurrence, and can theoretically range from very large negative numbers to very large positive numbers. Zero values are extremely rare, and therefore the two vectors are likely to have no or very few empty cells.²² An appropriate measure for comparing such vectors is the Pearson correlation coefficient. Correlation is standardized by the two

²¹ A similar rationale informs Latent Semantic Analysis, a method for extracting underlying meanings from co-occurrence matrices of words in documents (Landauer et al 1998).

²² The denominator in equation 2.1 makes it unlikely that the expected value will be a natural number. Hence the subtraction in equation 2.2 almost always results in a rational number different from zero. In the observed and simulated data described in the previous section, none of the millions of matrix cells equal zero.

vectors' variances, and is therefore insensitive to differences in variance and magnitude between vectors. Formally, categorical similarity between two objects, i and j is defined:

$$(2.3) \quad \tau_{ij} = \rho(\vec{w}_i, \vec{w}_j)$$

where ρ is the correlation coefficient operator. τ_{ij} is a dyadic measure of categorical similarity. It ranges from minus one to one, such that values close to either extreme suggest either strong categorical difference or strong categorical similarity.

2.4.3 Measurement Validation

The skeptical reader might be concerned that the various methods described in this chapter for measuring classifications rely on compounded mathematical manipulations that ultimately depend on a set of theoretical assumptions about the relationships between practice and subjectivity. Luckily, the MOC data include information about users' subjective categorizations of music that can be used for the purpose of validating these assumptions, and the measurements tools that are derived from them.

Five months after the launch of the website, DJs were given the possibility to annotate mixes with tags. Users added tags to 92.9 percent of mixes subsequently created. Tagging is a common online practice, providing users with the possibility of describing entities with discrete and freely-defined categorical labels. In the MOC, only the mix creator is allowed to tag her creation. Tags therefore reflect how a mix is categorically perceived by the DJ who created it.²³ In essence, the addition of the tagging functionality mid-way into the time period constitutes a natural experiment that allows comparing collective patterns before and after tagging was introduced.²⁴

²³ Unlike categorical survey data, where categories are predefined by the researcher and therefore circumscribe the respondent potential answer, with tagging users can freely decide what labels to use.

²⁴ The availability of data before and after tagging was introduced allows controlling for potential effects that tagging may have on practice. When a user is asked to explicitly label a set of objects with a tag, that might

Overall, DJs used 7067 different tags to describe their mixes.²⁵ The most popular tags, such as ‘alternative rock’, ‘hip-hop’ and ‘reggaeton’, correspond to established musical genres. But tags include a variety of dimensions of musical classification, including era (eg. ‘90’s’, ‘sixties’), mood (‘fun’, ‘melancholy’), affect (‘sexy’, ‘cute’), theme (‘religious’, ‘romantic’) and origin (‘Detroit’, ‘Brazilian’). The average mix is annotated by 2.44 tags, suggesting that users conceive of their mixes through fairly narrow categorical prisms.

In order to assess the construct validity of categorical similarity, τ (as defined in the previous section), I divide the MOC data into two time-slots: t_1 covering the time period up until tagging was introduced, and t_2 covering the subsequent time period. The design behind this test is rather simple: if categorical similarity between artists deduced from users’ behaviors during time t_1 corresponds to how they tag their mixes during time t_2 , then we can safely assume that practical classifications reflect subjective categorical perception. Because categorical similarity is inferred from user choices before tagging was made available, tagging could not have played a causal role in shaping it. Such an association would therefore support the underlying assumption that collective patterns of behavior reflect an underlying inter-subjective categorical structure.

Let G denote a $K \times H$ sized artist-to-tag matrix, such that each row represents one of K artists, and each column one of H tags. The value in each cell, G_{ij} , corresponds to the number of mixes that include artist i and that were tagged with tag j . $\vec{G}_i$, the row corresponding to artist i in matrix G , is artist i ’s tag frequency vector. The tagging similarity between two artists is therefore the similarity between their two tag frequency vectors. Because frequency vectors tend to be very sparse – only a small proportion of cells are greater than zero – an appropriate measure of similarity is the cosine similarity coefficient

unconsciously prime her to think about her mix in crisper categorical terms. Thus inducing categorical structures before tagging is practiced enables us to see how categorical thinking shapes practice without categories being explicitly identified or named.

²⁵ Tag frequency follows a fat-tailed Poisson distribution such that the 12 most popular labels accounts for roughly 50 percent of all tags.

(Tan et al 2007). Cosine similarity measures the cosine of the angle between two vectors, and is calculated as follows:

$$(2.4) \quad \cos(A, B) = \frac{\sum_i A_i B_i}{\sqrt{\sum_i A_i^2} \cdot \sqrt{\sum_i B_i^2}},$$

where A and B are two frequency vectors. Tagging similarity between two artists is therefore defined as the cosine similarity between their respective tag frequency vectors:

$$(2.5) \quad \gamma_{ij} = \cos(\vec{G}_i, \vec{G}_j),$$

Figure 2.10 plots the relationship between categorical similarity, τ , inferred from practical behaviors, and tagging similarity, γ , inferred from subjective labeling. These two values were computed for all pairs of artists included in at least two collections during times t_1 and t_2 . As the figure unmistakably illustrates, the two measures are very strongly correlated (with the Pearson correlation coefficient equaling 0.7816). The cubic polynomial model fitted to the data, and plotted in Figure 2.10 explains 72.1 percent of variance.

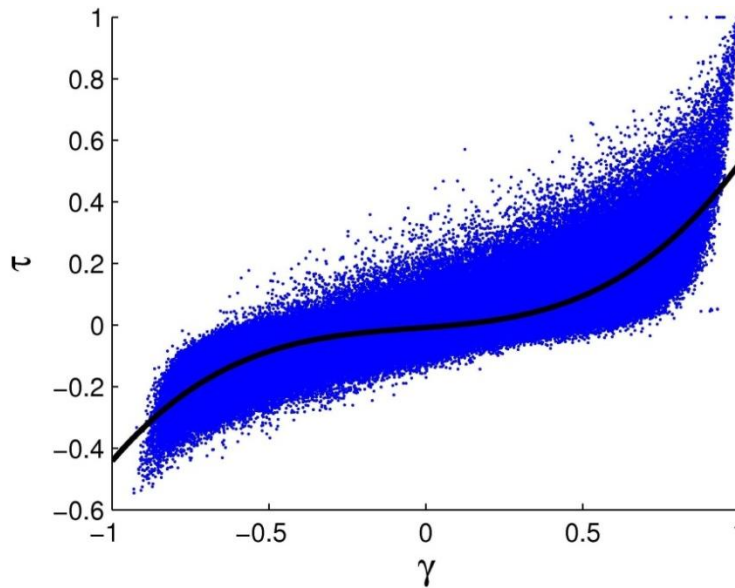


Figure 2.10: Categorical similarity during time t_1 vs. Tag similarity during time t_2 . The black line plots a fitted third-degree polynomial model.

These results strongly validate the assumption that practice mirrors users' categorical logic, and that collective patterns of musical consumption are correlated with inter-subjectively shared systems of classification. The quasi-experimental design precludes the possibility that tagging directly influences users' behaviors.

2.5 Final Notes about Data and Methodology

Suppose we could conduct an experiment in which a random set of strangers would be sent to a desert island. These people's preexisting cultural baggage would no doubt be consequential for their categorical interpretations of each other's behaviors, and ultimately for the status hierarchies and concomitant network structures that would emerge. Presumably, over time these structures of interaction will become habituated and could potentially give rise to new local categorizations, some of which would eventually become institutionalized. New participants arriving on the island would be confronted by these emergent cultural regularities; merely by joining the island they might also unbalance the social structural equilibrium on which they depend.

For obvious reasons, conducting such an experiment is unfeasible. In order to explore the questions this study is concerned with – namely, what are the social interactional processes through which categorical inter-subjectivity coheres – it is necessary to be able to observe a group of people, and their interactions in their entirety, from their inception. The data used in this study – if you will, two online islands of sorts – answer these two requirements. The online medium limits interaction to a very narrow set of inter-personal activities, all of which leave an electronic trace. And because both datasets follow their respective communities from the moment of their launch, there are no preexisting emergent cultural institutions that make it difficult – as is often the case with participant observation – to trace the sources of the social dynamics in question.

Of course, no data are ideal. Critics might rightly point out that the two SNS communities under investigation draw specific audiences and are, under no circumstances, representative of the American population, or any other socially identifiable group for that matter. Yet this study does not purport to be representative in the traditional social scientific sense. Rather, if it manages to demonstrate that similar rudimentary social dynamics are in operation in both communities – despite the obvious fundamental differences between musical and financial consumption – then it will be providing evidence for the universality of social processes of inter-subjective classification.

Other critics may be concerned that data harvested online are not sufficiently ‘real’ for making meaningful inferences that extend beyond the electronically-mediated reality created on the web. Online interaction, by its very nature, cannot provide the depth afforded by co-present inter-personal interaction. But this electronically-imposed leanness serves as an advantage, not an obstacle, for the purpose of this study. Because the two websites strictly limit the types of activities users can pursue, the

data – despite their exceptional volume in sociological terms²⁶ – become analytically manageable. In other non-electronically-mediated settings, the avenues for inter-personal communication are so varied that no researcher can realistically trace them all. It is obviously possible that people who are active in either community also interact with one another outside the online platform; such interactions would be inaccessible for research purposes. Yet users on both websites come from a variety of geographic locations spanning the globe.²⁷ It is unlikely that any user is familiar and meaningfully interacting with more than a handful of other user outside the website; comment threads and biographical statements thoroughly investigated provide no evidence that users are familiar with one another offline.

Online data may also not be ‘real’ in the sense that we cannot know with certitude to what extent users reliably report their behaviors. First, individual flesh-and-blood actors may use different user aliases. While this possibility cannot be ruled out, personal data contained in user profiles on both websites suggests that users do in fact correspond to real people.²⁸ Second, users may provide partial, or deliberately disingenuous portrayals of their behaviors. This is perhaps more pertinent with the FOC: interested parties may use the website for influencing people’s investment behaviors. Because of the website’s relative miniscule size and marginal online impact, it is unlikely that actors interested in affecting stock performance would invest efforts through this outlet. Whether users reliably report their behaviors – i.e. whether they truly invest in the stocks they virtually trade on the FOC, and only those stocks – is of less importance as these data are not being used to make particular arguments about users’ real investment behaviors. What is consequential, however, is that their reported behaviors are

²⁶ Overall, these data encompass millions of data-points, the analysis of which requires computationally intensive methods and hardware. Traditional social scientific software is not designed, and therefore incapable of analyzing such large quantities of data.

²⁷ Users can report their location, and indeed in both communities users self-report location in the US, Europe, Asia and South America. Moreover, technical attributes of users’ internet accounts suggest that they are indeed communicating from multiple locations worldwide.

²⁸ Many user profiles include links to other web-based personal information, such as blogs, personal homepages and other SNS profiles. Due to privacy concerns, these data were not systematically analyzed. Unlike virtual-world SNSs such as Second Life, there is little reason to assume that users intentionally explore alternative identities.

observed by other users, and hence, irrespective of whether or not they correspond to users' 'real' stock market investments or musical preferences, fuel the collective dynamics under investigation.

Ultimately, the data and methods used in this study allow thinking of classification as an uncoordinated collective endeavor, not a social process imposed from above by institutional actors. Such an approach has rarely been used in sociological studies (Hannan 2010). As discussed in section 2.2, this practice-focused approach is consistent with how categorical perception is employed by actors in everyday life; results reported in section 2.4.3 provide evidence that categories induced from practical behaviors correspond to actors' subjective orientations. A variety of recent studies, especially those originating from organizational ecologists, explore the fuzzy application of established categories. Taking this agenda one step forward implies thinking of the labels people use for making sense of the world not in the same way analysts or bureaucrats carve it up into discrete ontological entities but rather through what Wittgenstein (1953) calls 'form of life': the socio-cultural context that makes labels intelligible through the application of heuristic logic. Networks are useful conceptual tools for operationalizing this inter-subjectivity.

Yet, those seeking to harness conventional network analysis techniques for the exploration of categorical meaning face four fundamental challenges in achieving this goal. First, given the high cost of collecting social network data, research often relies on a static snapshot of the network which reflects its structure at a singular moment in time. Such studies are hampered by endogeneity bias that ultimately undermines the possibility of demonstrating causality (Stuart 2007). Second, even when high quality longitudinal data is available, many studies elect to overlook co-evolution by focusing on one direction of causality (Abbott 1988). These models are normally implemented using statistical techniques which necessarily presuppose the dependence of a singular variable on a set of independent variables. Third, embedded into such models is the assumption that entities are fixed, which in turn

allows for the objectification of socially constructed categories. However, the boundaries demarcating different categories, as well as the meanings people associate with them, are in constant flux. Finally, a long tradition of social network exploration using surveys relies on an ego-centered conceptualization of networks, in which unrelated respondents are asked about the people with whom they interact. But ego-centered operationalizations cannot account for the emergence of local patterns. A variety of studies have demonstrated how local patterns aggregate into systemic properties that can give rise to macro phenomena that are greater than the sum of their constituent parts (Watts and Strogatz 1996, Uzzi and Spiro 2005). The data analyzed in this study afford using network techniques in a manner that overcomes these limitations.

The basic data structures and methods used in this study have been outlined in this chapter. Operationalizations specific to the empirical investigations included in the remainder of this study are detailed in their respective chapters. Table 2.3 provides a summary of the symbols defined here and used throughout the analyses contained in the following chapters to represent the various data-structures. Readers may find it useful referring to this table when reading the remainder of this study.

Table 2.3: Common Symbols

Symbol		Meaning		
N	Set of users			
K	Set of objects			
u_i	User i			
o_i	Object i			
T	Number of time-slots			
Networks				
	Name	Size	Edges	Weights
S	Interaction	$N \times N$	$u_i \rightarrow u_j$	Binary
A	User-to-object	$K \times N$	$u_i \leftrightarrow o_j$	Binary
W	Object co-occurrence	$K \times K$	$o_i \leftrightarrow o_j$	Positive Integer
$\tilde{W}$	Standardized co-occurrence	$K \times K$	$o_i \leftrightarrow o_j$	Rational
τ	Categorical similarity	$K \times K$	$o_i \leftrightarrow o_j$	Rational

Chapter 3

The Social Construction of Categories through Interaction

3.1 Introduction

Eskimos, so the urban legend goes, have a large variety of words for describing snow. While this popular piece of folklore turns out to be empirically contested (Pullum 1991)²⁹, it concisely captures something fundamental about social cognition: that different people might experience the same reality through different interpretative lenses. The labels we use to describe these realities reflect an underlying classificatory logic which structures our knowledge about the world in a meaningful way. The Eskimos' alleged multiplicity of terms for snow is intriguing to the rest of us because it implies that what they conceptualize as different types of 'snow', others see merely as a singular and all-encompassing category.

Different groups of people, whether contemporaneously in different parts of the world, or at the same place but during different periods of time, might often employ different cognitive sieves with which they "carve out distinctive islands of meaning out of the same reality" (Zerubavel 1997). This is often the case with contested historically and culturally variable domains such as politics (Ferree 2003) or art (Goldberg 2011). But even such ontologically undisputed domains as the natural world can lend themselves to different forms of interpretation. When, for example, English zoologists encountered

²⁹ Moreover, the term 'Eskimo' refers to various peoples indigenous to North America, Siberia and Greenland. The Sami people of Sweden, on the other hand, apparently do have an unusual variety of words for describing snow. See http://www.arcticlanguages.com/papers/Magga_Reindeer_and_Snow.pdf

animals such as the platypus for the first time in the late 18th century, these animals' unusual anatomical features defied these scientists' basic assumptions about how the animal kingdom was structured.

George Shaw of the British Library was so perplexed by the platypus – possessing both mammalian and avian traits – that he initially assumed it to be a hoax (Ritvo 1997). At the time, British zoologists' understanding of the animal kingdom was informed by Linnaean taxonomy. Over the centuries since, however, European zoologists updated their taxonomies in a manner that is congruent with the ideas of Darwinian evolution theory. Such a way of understanding and classifying animals could accommodate what 18th century scientists considered to be the platypus' irreconcilable peculiarities.

Where do these classification schemes come from? At its core, categorization is a cognitive processing mechanism hard-wired into our brains (Fiske and Taylor 2008; Macrae and Bodenhausen 2001; Fiske and Linville 1980). As discussed in the previous chapter, classifications are constitutive of the process of understanding because we interpret sensory input by reducing it onto prototypical mental representations, which are themselves cumulatively constructed through our daily experiences (Rosch 1978; D'Andrade 1995). These classificatory schemes reify our experiential knowledge of the world; as such, they are economical representations of it. In order to serve this function, however, they cannot be entirely arbitrary, but rather need to correspond to some sort of real, or 'natural', order (Rosch 1973). Animal taxonomies, as the platypus conundrum demonstrates, are only useful insofar as they make distinctions between animal species that can be vindicated empirically.

But being 'natural' does not necessarily imply being objectively rooted in nature (Foucault 1970). As Ritvo (1987) demonstrates, the different categories used by the Victorians to differentiate between animals all too often reflected, and reinforced, a class-based system of social order. Despite the two breeds' apparent external similarity, 19th century English thought of greyhounds and whippets, the former raced by upper-class sportsmen and the latter by working class miners, in very different terms.

This is what sociologists often refer to as a ‘social construction’: a belief about the natural order of things that is inter-subjectively shared by a group of people but that is neither inevitable nor does it correspond to an objective reality (Hacking 1999). From Durkheim’s writing about religion and its foundation in delineations between what is sacred and profane, to more recent accounts of the constructed nature of distinctions such as those between science and pseudo-science (Gieryn 1983), fine and popular art (Bourdieu 1984), or mental health and disorder (Bowker and Star 1999), sociologists have indeed thoroughly investigated how the categories people use in their daily sense-making originate from and serve to reproduce preexisting social configurations.

If not inevitable, these configurations must come from somewhere. Yet, for the most part, sociological studies eschew the question of why specific categorical distinctions emerge in the way they do. Berger and Luckmann’s (1966) theory of social construction provides what is perhaps the most promising theoretical framework in this regard. It argues that cognitive representations are coordinated inter-subjectively through social interaction. These representations are reified – and become codified through institutions – so as to reproduce the very patterns of interaction that give rise to them. This theory has an important, and hitherto empirically undemonstrated implication. If who people interact with and how they perceive the world are related to one another, then we should expect that the patterning of these interactions – the network structures that undergird them – correspond to the structures of their interlocutors’ inter-subjective representations of the world. In other words, interaction and categorization produce coterminous structures.

Yet the study of the social underpinnings of categorization processes, on the one hand, and network analysis on the other, are two sociological subfields that have mostly developed tangentially. In fact, network analysis initially consolidated and, up until recently, evolved as a reaction to traditional sociological models that conceptualize actors in terms of categorically-defined attributes. Rejecting such

conceptualizations as objectified social constructions, network analysis has been informed by what Emirbayer and Goodwin (1994) refer to as an ‘anti-categorical imperative’ that focuses exclusively on inter-individual relationships.

In this chapter I adopt a network-based approach for the operationalization of a social construction, and define it as a structural homology between two types of networks: an interaction network between actors, and a co-occurrence network between the objects they consume. The major assumption informing this chapter is that categories emerge endogenously through social interaction and, as they take form, shape these patterns of interaction; such a coevolution results in a structural homology. I demonstrate that a homology indeed exists in both communities under study – the MOC and FOC – on three structural levels: (1) the dyad, (2) the personal actor, and (3) the network topology.

This chapter begins with a brief overview of sociological literature on social construction and its roots in social interaction. It then discusses social construction through the prism of network theory, and explains how two rudimentary mechanisms – social influence and choice homophily – are consistent with a network-based conceptualization of social construction. In the fourth section, these theoretical expectations are translated into analytically operationalizable constructs and a set of three hypotheses, which are then tested empirically in both the music and finance communities. The final section discusses these findings and their theoretical implications.

3.2 What is a Social Construction?

“The social reality of everyday life” contend Berger and Luckmann “is apprehended in a continuum of typifications” (1966: 33). At the time it was made, this assertion was more of a proposition than an empirically grounded statement. But research in cognitive science that emerged in the decades following the publication of *The Social Construction of Reality* vindicates Berger and Luckmann’s basic

intuitions (DiMaggio 1997; D'Andrade 1995): when people enter a situation they attribute meaning to it by applying categorical labels to the events they experience. Berger and Luckmann contend that we always perceive others in terms of some sort of preexistent categorical identity: a man, a conservative, a musician. The more our interaction with others is mediated, the less opportunity we have to challenge and potentially falsify these categorical designations.

For example, when browsing others' profiles, an MOC user evaluates them based solely on their username, the composition of their musical mixes, and the titles and photos they use to label these mixes. Schema theory suggests that this information is processed in terms of the categorical identities it activates. We can get a glimpse of the cognitive inner workings of this process by examining comments users leave on each other's mixes. "We needed some drum'n'base" writes one user on a fellow DJ's mix; "feelgood", "chillout", "classic" write others to succinctly describe the musical attributes of mixes they are commenting on. All these terms, presumably, relate to the schematic labels users employ when identifying and classifying one another.

Sociologists are often inclined to assume that such typifications are socially constructed (Hacking 1999). But people use schemata as mechanisms for representing reality in its entirety, not only the 'social reality' of everyday life. Our knowledge of the world is, by the nature of our cognitive abilities, constructed. Recent studies seem to suggest that some basic categories of thought are apparent in babies before they are capable of engaging in any sort of socially meaningful interaction (Bergesen 2004). Not all 'typificatory' knowledge of the world is necessarily socially constructed.

What makes a certain belief or disposition *socially* constructed is its dependence on information that we receive from, and inter-subjectively coordinate, with others. Social constructionist theory puts a

particular emphasis on inter-personal interaction³⁰ as the pathway through which such coordination occurs (Ridgeway 1991; Gould 2002). Berger and Luckmann posit that typification - the assignment of interactions into different predefined categories – also prescribes the behavior that is appropriate for a given type of interpersonal exchange. The more this behavior is rehearsed in real life situations, the more embedded these typifications, and their behavioral entailments, become. “Social structure”, they write, “is the sum total of these typifications and of the recurrent patterns of interaction established by means of them” (1966: 33). Surprisingly, especially given *The Social Construction of Reality*’s profound influence on American sociology, little has been done to empirically explore the relationship between patterns of interaction, and the categorical orderings used by agents who do the interacting.

3.3 Networks

This empirical dearth is perhaps best explained by what DiMaggio calls ‘Nadel’s paradox’: that “a satisfactory approach to social structure requires simultaneous attention to both cultural and relational aspects of role-related behavior,” but that there exists a discrepancy between the “taxonomic specificity” of cultural analysis and the abstract structuralism of relational sociology (1993: 119). Social network analysis has traditionally espoused the latter. In what has, over the years, become a fundamental, indeed almost programmatic statement of American social network analysis, White and colleagues write that “social structure is regularities in the patterns of relations among concrete entities; it is not a harmony among abstract norms and values or a classification of concrete entities by their attributes” (1976: 733-4). The scholarship that followed focused almost exclusively on structure, and its interpersonal relational underpinnings (Emirbayer and Goodwin 1994), largely bypassing the “cultural and socio-psychological meanings of actual ties” (White et al 1976: 734).

³⁰ ‘Interaction’ in this context is very broadly defined. It can refer to interpersonal contact mediated only by language and bodily gestures, or to indirect contact mediated by institutionalized symbols.

Nevertheless, social network analysts have not rejected the notion that cultural aspects of social action exist. At the same time, however, they normally conceive of them as causally secondary. When thinking about the relationship between network structure and individual worldviews, network analysts often resort to one of two complementary mechanisms: *social influence*, or the assumption that people mimic one another, and *homophily*, which is people's inclination for propinquity-based association. Both of these theories think of culture in inter-subjective terms. The former focuses on modes of cultural transmission while the latter is concerned with how inter-subjectivity filters patterns of interaction.

3.3.1 Social Influence

Social influence theories conceptualize networks as the social infrastructure upon which ideas, beliefs, and behaviors spread. In their basic formulations, these theories postulate that beliefs and behaviors diffuse through mimesis. They see the dyad as the core conduit for inter-personal transmission. The tendency of people to coordinate their beliefs and behaviors has been empirically demonstrated in a variety of social domains. Voting behavior (Nickerson 2008), lifestyle choices (Christakis and Fowler 2007) and political activism (Zhao 1998), to name but a few, have all been demonstrated to propagate through networks.

A number of studies have looked at the underlying socio-psychological mechanisms that explain how people influence one another through interaction. Some argue that people imitate others because they, often wrongly, perceive others' behaviors or opinions to be more accurate or beneficial. In a classical study of technological diffusion, Coleman et al (1966) argue that doctors, faced with the need to make medical decisions under uncertainty, turned to one another for signals about the usefulness of a particular drug. In a more recent study, Liu et al (2010) argue that information diffusion through social

influence accounts for a significant increase in the prevalence of autism in California during the first half of the previous decade.

Other explanations focus on social desirability and the psychological need to make interaction rewarding. The drive to conform is particularly salient when deviance is sanctioned. A variety of lab experiments demonstrate that people are influenced by others' opinions when conformity is incentivized through sanctioning. A recent study by Edelson et al (2011), for example, demonstrates that people tend to conform to erroneous recollections of other group members in a manner that affects their neuro-physical representation of memory. Moreover, even when they themselves do not hold a particular belief, individuals are nevertheless often inclined, under pressures of conformity, to behave as if they espouse it (Willer et al 2009).

Social influence models normally assume that network structure has casual primacy over meaning-making and subjective representations (Pachucki and Breiger 2010). Taking an epidemiological approach to culture, they often conceive of it in terms of independent memes or ideas that spread through the network like a virus (Watts 1999, 2002). This is particularly the case when the objects being exchanged are bits of information such as knowledge about a new job. Recent scholarship, however, increasingly suggests that diffusion that entails behavioral outcomes, such as adopting new beliefs or partaking in socially unconventional or illegitimate action, propagates through what Centola and Macy (2007) call 'complex contagion.' Unlike disease, such contagions require reinforcement through multiple sources.

In fact, it appears that when such behaviors diffuse they do not necessarily do so through direct mimesis but rather indirectly, and subconsciously, through sampling. When people watch their neighbor cycling to work, or hear a colleague talk about the need to reduce the US external debt, they are exposed to ideas they may have hitherto been unaware of. Mere exposure suddenly makes these ideas

visible. Even without motivation to mimic their neighbor or colleague, multiple exposures to novel ideas and behaviors make adoption increasingly likely (Denrell and LeMens 2007; Denrell 2008).

3.3.2 Homophily

The diffusion of ideas and behaviors through inter-personal interaction can explain one of the most ubiquitous social phenomena: homophily, or the tendency to form ties with similar others (Kossinets and Watts 2009; McPherson et al 2001). Homophily has been observed practically in every field of social activity, from marriage (Kalmijn 1998) and close friendships (Marsden 1987) to the racial makeup of college acquaintanceships (Wimmer and Lewis 2010) and the composition of founding organizational teams (Ruef et al 2003). Behaviors that diffuse through complex contagion are even more likely than those that propagate through simple contagion to produce homophilous outcomes, as the former depend on multiple reinforcement and are therefore likely to spread within tightly knit cliques, but less so across cliques (Centola 2010). I might be influenced by my next-door neighbor to ride my bike to work. Biking will have increased visibility for our other neighbor across the street who might come to realize that riding his bike is a good idea for getting rid of that excess weight he has been battling. My work colleagues, on the other hand, having me as their only example, might think of my cycling as peculiar, preferring the convenience of their air-conditioned commute. Cycling is therefore likely to spread in my neighborhood, but not at work. In theory, complex contagions should make groups become increasingly homogenous, while maintaining an overall inter-group differentiation.

This is what McPherson and Smith-Lovin (1987) refer to as induced homophily: the result of preexisting network structures limiting interaction opportunities to similar others. People's tendency to befriend others with similar political orientations (Lazer et al 2010), for example, might therefore be the result of structural circumstances circumscribing friendship opportunities to people who are more likely

to become exposed to one's sources of political ideological influence. What might seem as a voluntary tendency to select interaction partners on the basis of some sort of individual trait can often be a spurious manifestation of underlying network dynamics.

Yet observed homophily can also be the result of individuals' socially-conditioned predispositions to interact with similar others, net of structural effects. Choice homophily, as McPherson and Smith-Lovin (1987) label it, necessitates that people will collectively privilege certain attributes – eg. skin, but not eye color – and agree what social meanings they signal, as precepts through which to assess other people and pursue social ties. In one of the first systematic studies on the subject, Lazarsfeld and Merton (1954) distinguish between two types of homophily on the basis of the dimensions of similarity along which people tend to associate with one another: status homophily, whereby similarity is based on ascribed or formal status categories such as race and gender, and value homophily, whereby people are similar in their beliefs and moral orientations. Value homophily is often assumed to be particularly potent in influencing tie formation because it reduces ideological friction and facilitates easy communication (McPherson et al 2001).

Imagine a cocktail party where participants come from similar sociodemographic backgrounds, reducing the salience of ascribed characteristics as social differentiators. In such a context, conversational topics like art or politics can play a role in boundary rituals that allow people to make snap judgments about one another without having to resort to probing and cognitively taxing exchange. Similar musical or cinematic preferences, for example, can be interpreted as proxies for more fundamental moral, behavioral or ideological similarities (DiMaggio 1987). This interpretation, as discussed in length in the previous chapter, need not be conscious and purposive. Rather, such expressive features are automatically processed and understood by interlocutors in terms of the cognitive categories that they activate. Others who are perceived to be categorically closer to ego

generate smoother interaction that is likely to persist (Bourdieu 1984, 1990). Choice homophily explanations, especially in regard to value-based similarities, reverse the causal direction assumed by social influence explanations: culture is assumed to predate and channel network formation.

Yet while the effect of value homophily has been demonstrated in experimental studies (Huston and Levinger 1978) it is difficult to prove and quantify in observational data because of the confounding effects of structural constraints (Centola 2010). A few studies have nevertheless attempted to demonstrate that value-based similarities can influence network structure. Using an instrumental variable strategy, Lizardo (2006) demonstrates that artistic preferences are often more resilient than network ties. In a more recent study relying on a longitudinal design, Vaisey and Lizardo (2010) argue that moral orientations are strong predictors of the persistence of ties among adolescents.

Ultimately, conceptualizing network dynamics in either/or terms – as either concepts spreading through interaction or categorical identities predetermining network structure – is to create a false dichotomy (Pachucki and Breiger 2010). People coordinate their beliefs and behaviors through interaction but at the same time also seek interactions with people whom they consider similar on dimensions they perceive as important. Analyzing email exchange between members of a university community, Kossinets and Watts (2009) argue that structural constraints and predispositions to befriend similar others complement one another, such that homophilous preferences are amplified by the structural opportunity to interact with similar others and, in turn, tighten these constraints. Various methodological strategies have been proposed for simultaneously accounting for social influence and choice homophily. Snijders and colleagues (2010) present a stochastic actor-based model for quantifying the relative saliences of influence and self-selection in longitudinal network data.³¹ In one application of this model Mercken et al (2010) find that self-selection has a stronger effect than influence in producing

³¹ Though highly useful and amenable to a variety of theoretical assumptions, this method is highly computational intensive and is ultimately applicable to datasets of no more than several hundred network nodes in size.

observed smoking homophily between adolescents. Examining the adoption of an online application in an instant-messaging network, Aral et al (2010) similarly find that social influence has been overestimated in previous studies. The authors use an ingenious research design whereby they match pairs of users who have similar propensities for having friends who adopt the application at a given time – estimated as a function of their individual attributes and previous behaviors – but who differ in the observed number of such friends. These pairs’ adoption behaviors at a given time are then compared, allowing the authors to differentiate between adoption that is driven by social influence and that which is driven by non-relational inclinations. They find that even in this presumably simple case of contagion, much of the observed homophily in adoption is attributable to choice homophily and not to social influence.

3.3.3 Network Evolution under Conditions of Social Influence and Homophily

The picture that emerges from the interplay between agentic and structural aspects of social network dynamics is one of complexity: choice homophily and social influence cumulatively amplify one another. But what does this imply about the co-evolution over time of interpersonal interaction and inter-subjective processes of categorization? Does the reciprocal intensification between the mechanisms of social influence and homophily necessarily lead to increased fragmentation into ideationally homogenous cliques? Or are there sufficient incongruities between the two to prevent such perfect stability from consolidating?

This question is particularly vexing because modeling such complexity, and tracing it empirically, are very difficult tasks. A variety of models have been suggested as responses to it. The majority bracket questions of choice homophily and examine the circumstances under which diffusion occurs. In a groundbreaking and highly influential study, Granovetter (1978) suggested a simple mechanism for

modeling the diffusion of collective action. The central assumption in his model is that different people have different thresholds of susceptibility for novel behavior; some require little or no influence, whereas others will act only once a large number of others have adopted the behavior. Granovetter demonstrates that diffusion depends on how individual thresholds are distributed. Ridgeway (1991) suggests a similarly parsimonious theory for explaining the social construction of status beliefs. She hypothesizes that attributions of status to various categories of people – e.g. beliefs about the worthiness of people of different races – diffuse through interaction. In a recent study, Mark et al (2009) propose a model based on this theory which rests on two assumptions: that people who have no status beliefs are influenced by interaction partners to adopt them, and that status beliefs can be lost through interactional influence. The authors show that a tipping point is necessarily achieved such that once a status belief becomes more prevalent than its opposite it will eventually become consensual.

Others have looked at the structural circumstances that impede, or enable diffusion. In particular, the idea that small-world networks (Watts and Strogatz 1998) are conducive to the spread of behaviors and beliefs has gained much currency. These decentralized networks, comprising interconnected clusters, have been argued to provide an optimal layout for simple social contagion (Watts 1999, 2002, 2003). Empirical studies that followed suggest that structure is important for system-level outcomes. Kogut and Walker (2001), for example, argue that a small-world structure of corporate ownership in Germany made the German economy resilient to pressures of globalization. Uzzi and Spiro (2005) similarly find that networks of artist collaborations reach their maximal global creativity when they are poised between centralization and balkanization.

Models that take into account both social influence and choice homophily are less common. Simulations based on these models usually find that, under certain conditions, interaction dynamics and cultural phenomena mutually reinforce one another. Such cases are characterized by a co-evolutionary

process that ultimately leads to a close to perfect overlap between patterns of interaction and patterns of ideational or behavioral diffusion (Carley 1991; Mark 1998, 2003). Baldassarri and Bearman (2007) find that in rare cases in which political discussion topics become contentious and polarizing, discussion networks become concomitantly fragmented into ideologically homogenous groups. Similarly, Centola et al (2007) demonstrate that such differentiation is highly resilient to local variation and cultural innovation. Such models have been generally used to explain how social boundaries emerge and why certain categorical distinctions – such as those based on race or gender – remain highly stable even in light of historical shifts.

These computational models more often than not find that networks evolve in a bimodal fashion: they either disintegrate into chaos, or converge toward stable equilibrium (Dodds and Watts 2004). But such complete stability, or disorder, are empirical rarities. In fact, empirical and simulation-based explorations of both theories of social influence and choice homophily implicitly rest on the assumption that social categories are fixed and predate interaction. This is most likely a tactical decision resulting out of necessity – after all modeling interactional complexity is highly challenging without assuming that cultural meanings themselves change – not out of a vehement disavowal of cultural change. It is nevertheless problematic because categories are elastic and constantly negotiated constructs.

This elasticity manifests itself in three ways. First, categories often die and emerge; for example, new organizational forms materialize as novel industrial categories (Ruef 2000; Pontikes 2010), and new musical styles are reified and institutionalized by the music industry as genres (Peterson 1997). Second, even when categorical labels stay fixed, their composition might shift markedly; what was once considered a health disorder might be seen today as normal behavioral variation (Strand 2010), and what was once considered popular entertainment can become sacralized as art (Bauman 2007). Finally, the meanings that people associate with a given category are negotiable and tenuous; children can be

transformed from economic assets to financially worthless but emotionally priceless possessions (Zelizer 1984), and winemaking can become an arena of identity struggles between traditionalist and modernist interpretations of wine as art (Negro et al 2011).

3.3.4 The Duality between Interaction and Categorization

How, then, do such changes in the categorical landscape correspond to changes in network structures? If Berger and Luckmann's social constructionist hypotheses are correct, then these two orders of social life – categorization and interaction – are co-constitutively interlinked. This link resonates with the old Simmelian idea that social identity emerges when patterns of interaction are experienced by actors as objective reality (Simmel 1950; Martin 2009). Breiger (1974) formalized this observation as a duality between persons and groups: that groups are constituted by the individuals who make them up, and at the same time that individuals are perceived by others – as well as themselves – through the different groups to which they belong.

In other words, group membership confers social identity. Such an identity becomes real, in the sense of becoming a 'social fact', when the group that gives rise to it is inter-subjectively objectified by people as making up a distinctive category; when, for example, 1950s Londoners on Vespa scooters become identified as 'mods,' or when the music produced by community of young musicians in late 1970s Manchester is labeled 'punk rock' (Crossley 2009). This is what White (2008) calls a *catnet*: a set of social relationships that is perceived to constitute a category of people. Its cognitive objectification is the moment when, to use Simmel's (1950) imagery, this interactional reality takes on a life of its own.

With the catnet concept, White offers an elegant way of thinking about and locating the duality between the material and symbolic aspects of social life.³² Such a duality often animates sociological theories about the interplay between structure and agency (Bourdieu 1990; Sewell 1992; Giddens 1984; Mohr and Dueqenque 1997; Breiger 2000). Yet it is mostly conceived of in very abstract terms that are difficult to operationalize. White proposes that catnets are, simultaneously, inter-subjectively shared schematic representations and patterns of relationships. This does not mean that categorization and interaction are reflections of the same underlying reality and are therefore reducible to one another; nor does it imply that one causally predates the other. Rather, a duality is a “relationship that inheres within and between two classes of social phenomena such that the structural ordering of one is constituted by and through the structural ordering of the other” (Mohr and White 2008: 490). Though analytically distinctive, categorization and interaction are mutually dependent.

The catnet duality is particularly useful for thinking in concrete terms about the dynamical process that gives rise to this mutual constitution. It is theoretically consistent with social influence and choice homophily; in fact, it implicitly assumes that both mechanisms are in operation. Social influence is the process through which categorical identities, and the behaviors they entail, are transmitted and coordinated inter-subjectively, whereas choice homophily generates the interactive dynamics that push people of an emergent categorical identity to cluster together. The two processes complement and reinforce one another.

This reinforcement occurs because schematic transmission does not operate through simple, virus-like social contagion. Rather, when people are updating their schematic representations of the world they are changing synaptic associations between cognitively embodied concepts in their heads.

³² Ironically, White proposed the catnet concept early in his career, and as he became a leading proponent of structuralism, his early attempts to link material structure with subjective aspects of social action were left forgotten for several decades. Only later in his career did he return to rediscover and develop more nuanced models of social structure (1992).

Such updating occurs, unconsciously and incrementally, as a function of the patterns of association they empirically observe in their daily realities (Rosch 1978; D'Andrade 1995; Strauss and Quinn 1997). Thus, transmission does not happen through willful and active mimesis but rather through passive and repeated exposure to others' behaviors. In other words, people are influenced by their interactional partners to adopt new ways of thinking categorically about the world through sampling others' behaviors (Denrell and Jerker 2008). Because this form of influence requires reinforcement through multiple exposures, it propagates effectively when network structures are clustered. Choice homophily provides the drive for such clustering to emerge.

As illustration, consider how music becomes imbued with social meanings in, for example, a small high school community. Imagine a multidimensional space in which each dimension corresponds to a degree of consumption for a particular musical object, e.g. a pop singer. Borrowing from Martin (2000) I refer to this as a 'consumption space'. Each teenager in the school occupies a position in this space as a function of her levels of consumption of different artists. The social significance of this position is not predetermined, but rather defined by the social identities of those who occupy it. Culture, as it inheres inter-subjectively, can be understood as the categorical rules that tie preferences together by restricting movement in this space along certain axes, which demarcate different social worlds (Martin 2002). Figure 3.1 illustrates such a hypothetical space. In panel I, these categorical rules are virtually non-existent; people exhibit a range of idiosyncratic consumption patterns. Panel II, on the other hand, exhibits a highly structured pattern of consumption such that those who consume Artist A also consume Artists B but rarely consume Artist C, and vice versa. The categorical logic informing students' consumption choices in this stylized example seems to be structured on a distinction between Artists A and B forming one category, and Artist C forming another.

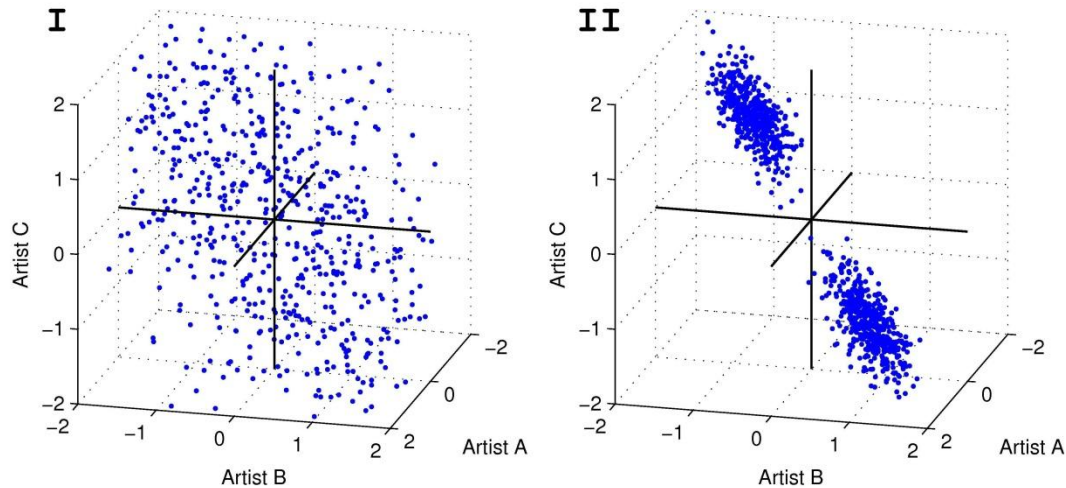


Figure 3.1: Two hypothetical consumption spaces. Each dot represents a hypothetical individual, and is located in the space as a function of that individual's degree of consumption of three hypothetical artists.

Imagine further a new student joining mid-year the school depicted in Panel II. By observing others, she learns to associate musical attributes, such as those characterizing Artists A and B, as delineating groups, or categories, of related artists. Because of her peers' homophilous tendencies, she also notices that those who tend to listen to the same type of music also tend to be each other's friend. She therefore attributes preference for or consumption of such music with emergent social identities. Gradually, as she becomes integrated into the social life of the school, she adopts one of these identities and, accordingly, listens to the music her adopted identity entails. Thus, by reenacting the patterns she observes she also empirically reasserts for other observers the association between musical consumption and social identities that guides her own behavior. In other words, she reproduces the socially constructed division of music into different genres.

These emergent categories are therefore based on a duality in which the classification of music and persons produce one another through a homology between groups of people and musical pieces. They fulfill the two rudimentary principles of categorization systems (Rosch 1978): they reduce cognitive energy by making a musical preference a signal for an otherwise difficult to discern social identity, and as they grow in salience they reproduce the correlational structure between preferences and identities

that they presuppose. These categorizations are social constructions in the sense that, once objectified by actors, they become cognitive filters through which the reality they describe is interactionally reproduced.

The effects of schematic transmission and choice homophily should, over time, produce a transition from an entirely randomly organized structure to a more patterned one. This is illustrated in Figure 3.2. Like Figure 2.3 in the previous chapter, these diagrams illustrate hypothetical interaction (S) user-to-object (A) and object co-occurrence (W) networks. In panel I, students are interacting with one another more or less arbitrarily, and their musical consumption patterns are such that various artists are only weakly associated with one another. In panel II patterns of interactions are clustered and, concomitantly, so are the associations between various artists; two categories of music emerge, symbolically marking two groups of teenagers.

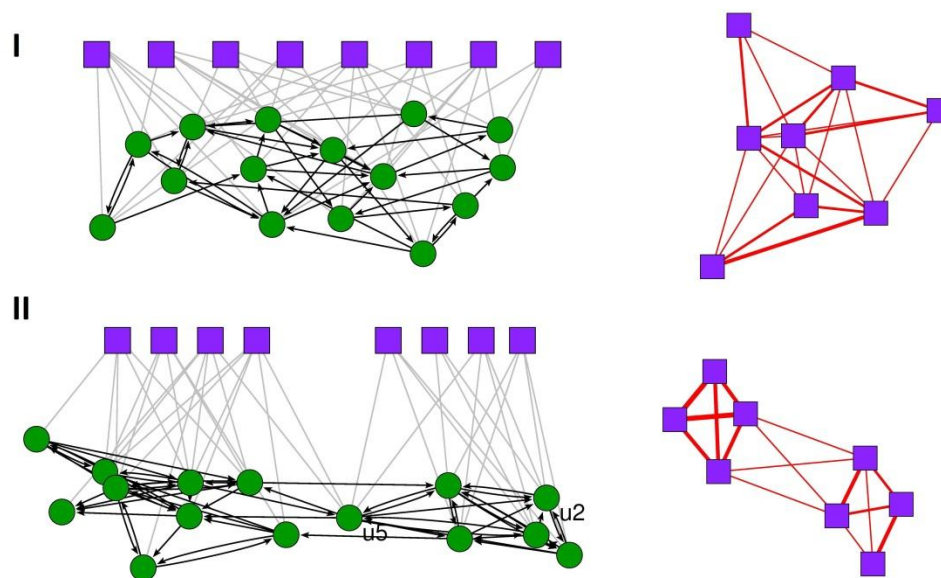


Figure 3.2: Stylized examples of two hypothetical high-school networks. Green circles represent students, and purple squares represent the artists whose music they consume. The interaction (S) and user-to-object (A) are illustrated on the left. Object co-occurrence projection (W) of network A is illustrated on the right.

At the same time, students do not operate in a social vacuum. They join the school with certain preconceptions they had acquired from previous observations about how music is classified and what

social meanings different categories carry. Panel I is rarely a realistic starting point. Because categories themselves are malleable, and constituted through interaction, initial conceptions about what they include and how they are interpreted can change over time. Their composition and meanings are not necessarily reproduced. This is especially the case because the process of schematic transmission is complex and messy, and is constantly prone to error and re-adaptation (Strauss and Quinn 1997). Small variations in the associational patterns different individuals observe may lead to overall different interpretations and hence to new patterns of interaction. A duality between categorization and interaction therefore does not necessarily assume an evolution toward a steady equilibrium.

3.4 Empirical Implications

The idea that the material and cultural aspects of social life are mutually constituted has become increasingly prominent amongst network analysts in recent years (Mohr and White 2008). Nevertheless, because this conceptual complexity is very difficult to formalize and operationalize, empirical studies concerned with demonstrating it have been relatively rare. In fact with the exception of a few ethnographic and historical studies (see for example Eliasoph and Lictherman 2003), the interaction-categorization duality remains a hypothetical proposition rather than an empirically grounded statement.

A common strategy for network analysts to unpack mutually constitutive dependencies has been to use Galois lattice analysis. Galois lattices are descriptive tools that map relationships between sets of objects and attributes. This technique has been used, for example, to demonstrate a duality between institutional categorizations of the poor in New York City and the practices employed to assist them (Mohr and Duquenne 1997), the homology between inter-organizational ties and conceptions of change in Brazil (Mische and Pattison 2000), and the interdependence between the affective meanings

commune members associate with their colleagues and the relationships they have with them, as a means to uncover how 'love' is socially constructed (Yeung 2005).

Lattice-based techniques are very useful for exploring and visualizing the intricate interdependencies between two classes of entities. Yet they are not designed to take into account the relationships between entities within each class. Of course, exploring these intra-class relationships is the substance of classic network methods. However, such techniques are very rarely used to trace the emergence of a co-constitutive relationship between two different modes of social action. One prominent exception is a set of studies by Padgett and colleagues exploring the co-evolutionary dynamics of interpersonal networks and cultural practices in Renaissance Florence. In one study, the authors demonstrate how political power was formed around social identities emerging through intricate inter-familial relationships (Padgett and Ansell 1993). In another, they explore how nascent capitalist conceptualizations of business associations emerged in the interplay between economic, kinship and political networks (Padgett and McLean 2006). Unlike most network models, these studies do not conceive of culture in static terms but rather explore its emergence by penetrating "beneath the veneer of formal institutions ... down to the relational substratum of people's actual lives" (Padgett and Ansell 1993: 1310).

This relational substratum exists both culturally and materially. As Figure 3.2 illustrates, a duality between categories and people can be conceptualized analytically as a relationship between two sets of networks: one connecting people with one another and the other connecting objects (which, in the high school example refer to different artists) with one another. If indeed a mutually constitutive relationship exists between these two networks then we should expect to find that their overall structural attributes change in tandem. If, on the other hand, the categorical logics employed by people are not shaped by their interaction patterns, and vice versa, then we should expect there to be no such relationship.

Unearthing this mutual dependence requires focusing, simultaneously, on the relationships between nodes comprising each of the two networks separately, and those connecting nodes across networks.

But this bilateral comparison cannot be performed simply on a cross-network dyadic level, as there does not necessarily exist a one-to-one relationship between nodes in the two networks. Consider node u_5 as illustrated in Figure 3.2-II. This actor is connected to several object nodes through the bipartite network A . The edges connecting u_5 to other actors in network S do not correspond uniquely to edges connecting object nodes in W . The two networks cannot be compared on a node-to-node or edge-to-edge basis.

A central focus is therefore how, and to what extent, the networks comprise clusters that relate to one another. The Simmelian duality that Breiger and White formalize is not merely one between individual concepts and people but between groups of concepts and groups of people that make up, respectively, distinctive categories and distinctive social communities. The mutual constitution inheres precisely in the overlap between clusters of nodes in these two networks. Indeed, network analysis methods are commonly used to detect groups both in interactional (Rivera et al 2010) and cultural data (Mohr 1998); yet they are seldom used to explore the homology between clusters in two different types of networks.

Clustering, or network partitioning, is a major problem for network analysts; one which draws great methodological attention but which rarely reaches consensus (see Danon et al 2005 and Porter et al 2009 for useful reviews). This challenge is exacerbated when, as is the case with a structural duality, the objective is to find clusters of overlapping nodes in two interconnected networks. Consequently, in order to detect this dependence, rather than adopting one clustering approach I examine the relationship between the two networks from three different vantage points: the dyad, the individual actor, and overall network topology. For that purpose I introduce three measures that relate,

respectively, to each of these three prisms: (1) categorical proximity, (2) categorical cohesion, and (3) topological segmentation. Each corresponds to a different aspect of this structural duality, and leads to different hypotheses.

3.4.1 Measures and Hypotheses

3.4.1.1 The Dyad

Categorical Proximity is a dyadic measure, which is defined as the extent to which two actors overlap in their categorical identities. Network models traditionally measure such proximity using a distance metric, such as Euclidean distance, between two vectors of categorical attributes (e.g. Baldassarri and Bearman 2007; Mark 2003). In our case, however, these categorical attributes are latent in the structure of co-occurrences between objects and cannot be reduced to simple attribute vectors. Consider two hypothetical actors, whose object collections are comprised of different objects. Imagine further that one is a big fan of the rapper Eminem, and the other of hip-hop trio De La Soul, but that neither listens to both artists. A simple measure of similarity would consider the two as dissimilar as Eminem and Céline Dion fans. But Eminem and De La Soul occupy structurally equivalent positions in the object co-occurrence network W , whereas Eminem and Dion do not. As discussed in Chapter 2, such equivalence is indicative of both artists being members of the same latent category. The collections of our two hypothetical actors might have low compositional overlap – they might include different artists – but still be comprised of objects that are categorically similar. Categorical proximity between two actors is therefore calculated as the average categorical similarity between all pairs of objects in two collections:

$$(3.1) \quad P_{ij} = \frac{\sum_{\forall k, l \in K} A_{ik} A_{jl} \tau_{kl}}{\sum_{\forall k, l \in K} A_{ik} A_{jl}}$$

where τ is the categorical similarity between objects k and l .

Intuitively, categorical proximity measures the extent to which two actors tend to consume the same categories of objects. If emergent categories play a causal role in the network dynamics that generate observed homophily, then we should expect to find that actors who interact with one another tend to be categorically closer to one another. Schematic transmission should lead actors who interact with one another to become more categorically proximal, and choice homophily should induce categorically proximal actors to interact with one another. Overall, such a dynamic should manifest itself in a positive relationship between interactional distance and categorical proximity. This leads to:

Hypothesis 3A: *Categorical proximity between a pair of actors decreases as the distance between them in the interaction network increases.*

3.4.1.2 The Actor

Examining relationships at the dyadic level allows comparing the two networks while bracketing questions of structural position: the two nodes are compared exclusively in terms of their distance. Yet structural position is not inconsequential. Consider actors u_2 and u_5 in Figure 3.2. The former is embedded in a cluster of highly connected actors, and concomitantly consumes objects that are categorically proximal to one another. The latter however, functions as a structural bridge that connects two otherwise disconnected clusters. Such actors are commonly referred to as structural hole-spanners. Their position in the networks makes them function as brokers that connect different sources of information and social capital (Burt 1992). They can also function as cultural interfaces between different meanings and practices (Pachucki and Breiger 2010). Indeed, as Figure 3.2 illustrates, actor u_5 consumes a categorically heterogeneous set of objects. The categorical distance between this actor and

his interaction partners is therefore larger than the distance between actor u_2 and her interaction partners, as u_2 connects a categorically homogenous group of interconnected actors. In other words, the relationship between categorical and interactional distance depends on actors' location in the network.

Actors u_2 and u_5 differ in the levels of the *Categorical Cohesion* of their object collections.

Categorical cohesion is defined as the extent to which a set of objects is categorically similar. Actors with high categorical cohesion are those whose object collections are comprised of objects that are categorically similar to one another; those with low cohesion have collections of dissimilar objects. A common method for measuring cohesion in a set is calculating the average distance between all objects comprising the set and the set's geometric center. As discussed in Chapter 2, an appropriate measure for a distance between two vectors is the Pearson correlation coefficient. Categorical cohesion for a collection of L objects is therefore calculated as the average correlation between each object's vector of edge weights with all other objects comprising the co-occurrence network $\tilde{W}$, and the vector of mean edge weights in network $\tilde{W}$ for this collection (which is the set's geometric mean). The average is weighted by the frequency of each object in the collection. Formally:

$$(3.2) \quad \varphi_i = \frac{\sum_{\forall k \in K} A_{ik} \rho(\vec{\tilde{W}}_i, \vec{\tilde{W}}^L)}{\sum_{\forall k \in K} A_{ik}}$$

where ρ is the Pearson correlation coefficient, $\vec{\tilde{W}}_i$ is the vector of edge weights between i and all other objects in network $\tilde{W}$, and $\vec{\tilde{W}}^L$ is the mean vector averaged over all $i \in L$.

Actors u_2 and u_5 also differ in the extent to which they are embedded in densely connected clusters. The clustering coefficient (Watts and Strogatz 1998) is a common tool for measuring such density. Conventionally used to explore triadic closure (Rivera et al 2010), the clustering coefficient measures the fraction of ego's neighbors that are connected to one another. Actors, like u_2 , whose neighbors are densely connected, have high clustering; those, like u_5 , whose neighbors are members of

different cliques, or who are unrelated to one another, will have low clustering. The clustering coefficient in a directed network corresponds to the extent of triadic closure for a given node. In an undirected network, such as the interaction network S , triadic closure is less trivial a concept. In fact, for a set of three nodes in such a directed network, there exist eight different combinations of edges that can form a triangle between these three edges. The directed clustering coefficient for node i is therefore defined as the proportion of such triangles actually realized out of the overall potential of triangles (for more details, see Fagiolo 2007). Formally the directed clustering coefficient is defined as follows:

$$(3.3) \quad C_i^D = \frac{1/2 \sum_j \sum_h (S_{ij} + S_{ji})(S_{ih} + S_{hi})(S_{jh} + S_{hj})}{d_i^T(d_i^T - 1) - 2\vec{d}_i}$$

where $d_i^T = \sum_j A_{ij} + A_{ji}$ is node i 's total in- and out-degree, and $\vec{d}_i = \sum_j A_{ij}A_{ji}$ is node i 's number of reciprocal (or bi-directional) ties.

The interaction-categorization duality suggests that clustering and cohesion should be related to one another. Actors with high interactional clustering should be anchored in a strong categorical identity and therefore exhibit high levels of categorical cohesion. On the other hand, actors who occupy bridging positions in the interaction network should also interface different categorical identities and therefore have lower levels of categorical cohesion.

Hypothesis 3B: *Actors' interactional clustering increases as a function of their categorical cohesion.*

3.4.1.3 The Network Topology

A third level of comparison between both networks is that of overall network topology. *Topological Segmentation* refers to the level of divisibility of a network into groups or communities of nodes that are clustered together. In the empirical context of this study these communities constitute

either different groups of actors who interact with one another more densely than they do with other non-group members, or groups of objects that co-occur in actors' collections more often than expected at random and therefore make up implicit categories.

A common strategy for measuring segmentation relies on a network property known as modularity (Girvan and Newman 2004; Shwed and Bearman 2010; Zhang et al 2008). Modularity measures the extent to which a given partition of a network into groups increases the number of within-group edges compared to what would have been expected had the network been rewired at random and the distribution of node degrees (which is the number of edges per node) maintained constant. Formally, given a partition p of an undirected binary network X into g mutually-exclusive groups, modularity is defined as follows:

$$(3.4) \quad Q(A) = \frac{1}{2m} \sum_{i,j} \left[X_{ij} - \frac{d_i d_j}{2m} \right] \delta(p_i, p_j)$$

where $d_i = \sum_j X_{ij}$ is node i 's degree, $m = \sum_{i,j} X_{ij}$ is the overall number of edges in the network, $p_i \in [1, g]$ is an assignment of node i to one of g groups, and $\delta(x, y)$ is the Kronecker Delta symbol which equals 1 if $x = y$ and otherwise equals 0. Modularity therefore measures the fraction of the edges that fall within the given groups minus the expected such fraction if edges were distributed at random.

In other words, modularity quantifies the strength of a given network partition in relation to the overall structure of the network above and beyond the properties of its individual nodes. Originally formulated for simple undirected binary networks, it can be extended for directed (Leicht and Newman 2008) and weighted networks (Newman 2004). For the interaction network S , which is a binary directed network, modularity is defined as follows:

$$(3.5) \quad \vec{Q}(S) = \frac{1}{m} \sum_{i,j} \left[X_{ij} - \frac{d_i^{in} d_j^{out}}{m} \right] \delta(p_i, p_j)$$

where $d_i^{in} = \sum_j X_{ji}$ is node i 's in-degree and $d_i^{out} = \sum_j X_{ij}$ is node i 's out-degree. For the object co-occurrence network W , which is an undirected weighted network, Q is computed as in equation (4) with m equaling the total weight of all edges in the network. For all network types, network modularity of 0 suggests that a given partition is no stronger than would be at random and therefore that it is insignificant. Modularity monotonically grows with growth in the strength of a partition.

Measuring network segmentation can be thought of as a problem of finding a partition that maximizes modularity. Solving such a problem is a computationally intensive challenge, and there are a variety of algorithms available for making it tractable. The spectral algorithm using eigenvectors (Newman 2006) is considered one of the most effective algorithms for maximizing modularity and is used in this study (Danon 2005; Porter et al 2009). Thus the comparison of maximal modularity for the same network over time can be used as a means to measure trends in its level of segmentation.³³

Segmentation is illustrated in two hypothetical interaction networks depicted in Figure 3.3. Both networks are comprised of the same number of nodes, but the networks differ in their levels of modularity: the network on the left exhibits a low level of modularity, whereas that on the right exhibits a high level. The interaction network on the left exhibits a disorganized topology whereby actors are not interacting exclusively within densely connected groups; the assignment of nodes to exclusive groups is

³³ Despite being commonly used as a means for detecting community structure in social, biological, chemical, and other types of networks, modularity suffers from various degeneracies (Good et al 2010). In particular, it often lacks a clear global maximum and therefore different heuristics for maximizing modularity might produce significantly different partitions with only modest differences in modularity. These degeneracies should be of concern when attempting to find an optimal partition and assigning nodes to different groups, as different partitions might result in different patterns of assignment. Nevertheless, as Good et al (2010, Fig. 2) demonstrate, modularity is fairly robust for the purpose of quantifying a network's potential divisibility into groups, irrespective of the actual composition, number, or sizes of different groups. Hence for the purpose of measuring segmentation, it is a useful tool. For more on modularity's sensitivity as a segmentation measurement tool, see the online supplement for Shwed and Bearman (2010).

not readily apparent in this network. The one on the right, on the other hand, clearly exhibits a segmented structure such that practically all nodes can be easily assigned to one, exclusive group. Imagine that these two networks represented co-occurrences between objects, ranging from an unstructured to a highly modular topology. In the latter case, the different groups of co-occurring objects delineate different categories with clear and pronounced boundaries. In the former, these boundaries are difficult to discern. Such a network would represent a semantic structure whereby objects are not associated with clearly marked categorical identities.

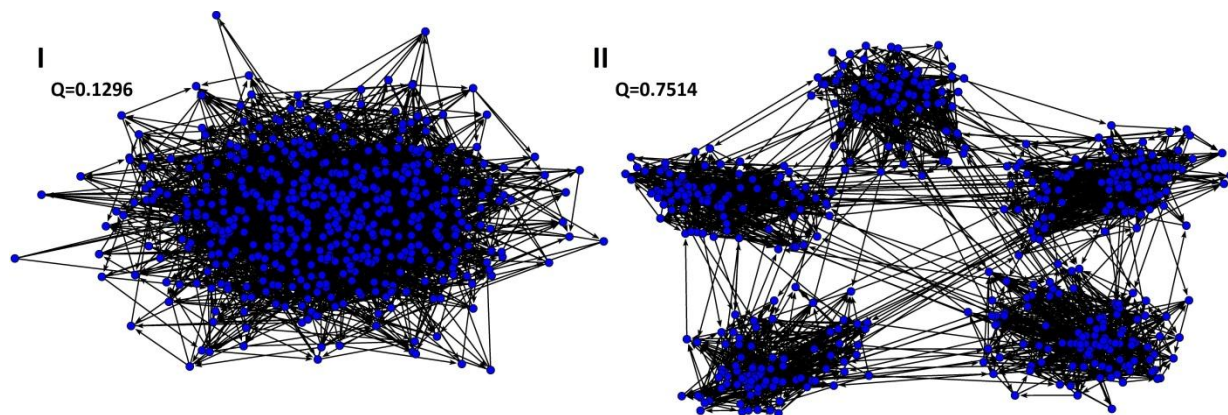


Figure 3.3: Two hypothetical interaction networks of size $N=500$. Panel I illustrates a network with low segmentation (Q) and Panel II a network with high segmentation. Network visualization is based on the Fruchterman-Reingold force-directed algorithm.

An interaction-categorization duality does not necessarily assume a strong division into clearly delineated groups. As discussed above, because of the co-constitutive nature of this dependence and its operation through cognitive mechanisms that are prone to error and drift, it can produce equilibriums of varying degrees of segmentation. Nevertheless, this duality implies that levels of segmentation in both network co-vary, and that when actors interact in dense cliques, categorical boundaries between the object they consume become more pronounced, and vice versa.

Hypothesis 3C: *Interaction and object co-occurrence levels of segmentation are correlated over time.*

3.4.2 Analysis

Do these hypotheses hold in the MOC and FOC data? Note that neither hypothesis is phrased in explicit causal terms. Rather, what each expects to find is evidence for a structural duality between interaction and categorization, whereby neither side of this duality is causally precedent. Hence, the statistical models that follow look for associations between two types of variables, one relating to categories, the other to interpersonal interaction. To do so I use a conventional approach: regression modeling. Though the choice of a 'dependent' variable in a regression model often implies casual dependence on an 'independent' variable of interest, this interpretation should be avoided here. What these models seek to demonstrate is merely a statistically significant relationship between two types of variables. In all of these models, the dependent variable – somewhat unconventionally – is the one relating to interpersonal interaction. This choice is tactical: as discussed in more detail below, it allows for including necessary controls for categorical proximity. It should not be understood as an implicit argument about categorization having causal precedence over interaction.

3.4.2.1 The Dyad

We begin our analysis by examining categorical proximity, which measures the extent to which two actors' collections of objects are similar in terms of their underlying categories. Recall that categorical proximity can be high even if the two collections include no overlapping objects, as it depends on these objects' categorical identities. In this respect, categorical proximity differs substantially from traditional measures for quantifying the similarity between two sets of objects, and that depend on the extent of compositional overlap between the two sets. The cosine similarity is one

such measure. It is commonly used for comparing similarity between datasets, especially in textual data-mining (Tan et al 2007).³⁴

Figure 3D plots (Panel I) the relationship in the FOC data between categorical proximity, P , and cosine similarity, $\cos$. Both measures are calculated per time-slot t , such that each data point represents the cosine similarity and categorical proximity for a pair of actors in a given time-slot. The diagram clearly illustrates that the two measures are unrelated: two collections that include no overlapping objects can still have high categorical proximity.

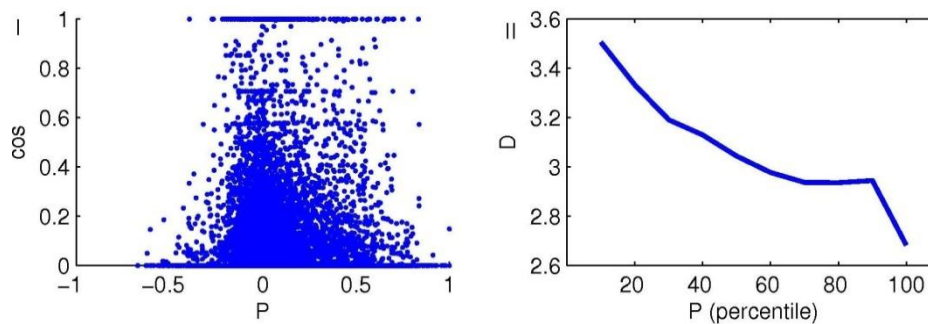


Figure 3.4: Scatter plot of categorical proximity (P) vs. cosine similarity ($\cos$) (panel I) and plot of average interaction distance (D) as a function of categorical proximity scaled in deciles (panel II). Both diagrams are based on FOC data.

The diagram on panel II of Figure 3.4 plots mean interaction distance (D) – which is defined as the shortest path between two actors in network S – as a function of categorical proximity (P). It demonstrates that as categorical proximity between two actors increases, the distance between them in the interaction network decreases. The relationship between the two distances appears S-shaped: the interaction distance decreases more precipitously toward the upper and lower ends of the categorical proximity distribution. In other words, proximity becomes more consequential for interaction distance as it nears either extreme.

³⁴ The cosine similarity is particularly useful for measuring distance in frequency data of the type analyzed here, as, unlike other measures such as Euclidean distance, it is not sensitive to zero-frequency coordinates. See Chapter 2 for details on how it is calculated.

The results of regression models testing the significance of the relationship between P and D are reported in Table 3.1. Model 1 estimates the relationship between categorical proximity as the independent variable of interest, and interaction distance as the dependent variable. Because interaction distance is a count variable, I use the following Poisson model to estimate it:

$$(3.6) \quad \ln(D_{ijt}) = \beta_0 + \beta_1 P_{ijt} + \beta_2 P_{ijt}^2 + \beta_3 P_{ijt}^3 + \beta_4 \cos(i, j, t) + \alpha X$$

where D stands for interactional distance and P for categorical proximity (note that each data point corresponds to the dyad ij at time t).³⁵ In order to model the S-shaped relationship between the two measures, I use a cubic function for categorical proximity. An alternative is to investigate whether actor i is likely to nominate actor j as a friend as a function of their categorical proximity. Model 2 therefore estimates the existence of dyad $i \rightarrow j$ in network S at time t , S_{ijt} , using the following logit model:

$$(3.7) \quad \ln\left(\frac{P(S_{ijt})}{1-P(S_{ijt})}\right) = \beta_0 + \beta_1 P_{ijt} + \beta_2 \cos(i, j, t) + \alpha X$$

where $P(S_{ijt})$ is the probability the dyad exists.

Both models allow adding cosine similarity as a control variable, as a means for testing whether categorical proximity is related to interaction distance above and beyond the overlap between two actors' object collections. The models also includes dummy variables (which are not reported in Table 3.1) for each time-slot, to control for time effects, represented by X in equations (3.6) and (3.7).

³⁵ The model is estimated using data that includes dyads between actors who are reachable from one another in the interaction network. Non-reachable dyads were removed.

Table 3.1
Poisson and logit models estimating interaction distance/dyad between pair $i \rightarrow j$ in S

Variable	Model 1 (Poisson) [†]		Model 2 (Logit) [‡]	
	MOC	FOC	MOC	FOC
P_{ij}	-0.0736 *** (0.0029)	-0.2910 *** (0.0194)	1.4542 *** (0.0049)	1.35896 *** (0.0173)
P_{ij}^2	1.9977 *** (0.0570)	0.0404 (0.0679)		
P_{ij}^3	-5.5246 *** (0.3098)	-0.2006 * (0.0971)		
$\cos(i, j)$	0.0736 *** (0.0025)	0.0354 (0.0195)	0.9829 *** (0.0051)	0.9869 (0.0133)
Constant	1.4046 *** (0.0022)	0.7085 *** (0.0176)		
χ^2	69184.98 ***	3575.98 ***	28773.98 ***	2255.07 ***

Robust standard errors in parentheses. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

[†] Dependent variable is interaction distance. [‡] Dependent variable is dyad dummy. Coefficients reported as standardized odds ratio.

Figure 3.5 plots the relationship between the two types of distances in the MOC and FOC, based on the Poisson model estimates. It also plots the distribution of categorical proximity in both communities (see caption for details). In support of hypothesis 3A, it demonstrates that the interaction distance between two actors decreases as their categorical proximity increases (net of the similarity in their two collections). This effect is particularly pronounced in the MOC: a transition over the full range of categorical proximity corresponds to a decrease of nine degrees in interaction distance. Given that the maximal distance in the network is 14, and that 99.92% of pairs have a 9 or smaller degree of separation, this is a substantial decrease. Note that the slope of the line intensifies toward the tails of the distribution of categorical proximity: a change in proximity is reflected in a substantial change in interaction distance when proximity is significantly high or low; in the mid-range, where two actors are neither significantly similar nor dissimilar in their categorical identities, changes in proximity correspond to little differences in interaction distance. Hypothesis 3A is similarly supported by model 2's results. An increase in one standard deviation in categorical proximity between actors i and j increases the odds of a tie between them by 45.4 percent in the MOC and 35.9 percent in the FOC.³⁶

³⁶ The results in Model 1, especially for the MOC data, may seem overly sensitive to outliers through the inclusion of second and third level polynomial transformations for P . To account for this possibility I also ran the same

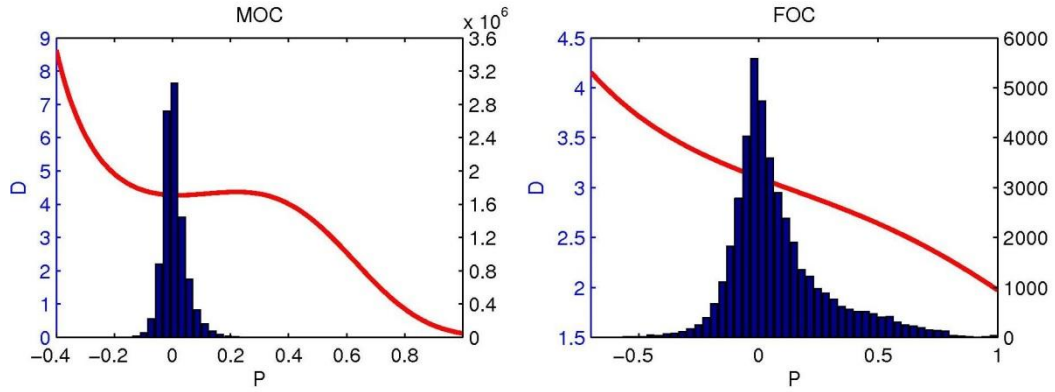


Figure 3.5: Poisson model of interactional distance (D) as a function of categorical proximity (P) in MOC and FOC (red lines). The function plotted corresponds to mean control variable values (because the Poisson model has a logarithmic link function, the magnitude of the curve is not independent of other control variables). The histogram plots P 's distribution in each community.

3.4.2.2 The Actor

Next we turn to the actor level, to explore the relationship between categorical cohesion and interactional clustering. Hypothesis 3B expects to find a positive relationship between the two, such that actors with categorically cohesive object collections also tend to have clustered ego networks in network S . This relationship is not entirely straightforward, however, as the two measures depend on other variables. Categorical cohesion, φ , is calculated as the mean distance between objects in the actor's collection and their geometric mean. Hence it depends on the actors' collection size, $|R|$. For $|R| = 1$, cohesion will always be 1, whereas as $|R|$ increases, the likelihood of cohesion equaling 1 decreases. Similarly, the interactional clustering coefficient measures the fraction of connected neighbors in ego's network, and is therefore dependent on the size of this network, or the actor's total node degree, d^T . Clustering is meaningless for $d^T < 2$ and is likely to decrease as d^T increases above 2. Both variables have heavy-tailed distributions, with few actors having large collections or high node

regression model with a simple linear function for P . In both the MOC and FOC data, P remains significant and negative, though the coefficient for P decreases in magnitude. Moreover, I calculated variance inflation factors (VIF) for an OLS version of Model 1 for the MOC data, as a means to estimate the degree of multicollinearity produced by the cubic function of P . VIF does not exceed 2.94 for all variables, which is generally considered a reasonable level of multicollinearity (Kutner et al 2004).

degrees. They are therefore included in the following models as logarithmically transformed control variables.

Table 3.2 reports the results of a regression where the dependent variable is interactional clustering, C . Because C is a proportion that is bounded by 0 and 1, I use a general linear model with a logit link function and a negative binomial distribution family with robust standard errors (Papke and Wooldridge 1996).³⁷ The model is defined as follows:

$$(3.8) \quad C_i = \frac{1}{1+e^{-z}} ; \quad z = \beta_0 + \beta_1 \varphi_{it} + \beta_2 \ln(d_{it}^T) + \beta_3 \ln(R_{it}) + \alpha X$$

where X are a set of time-slot dummy variables. Each data point represents an actor at time t . The estimation excludes data points for which $|R| < 2$ or for which $d^T < 2$.

Table 3.2
GLM estimating actor interactional clustering in S

Variable	MOC	FOC
φ	1.1357 *** (-0.1828)	0.7119 ** (0.2696)
$\ln(d^T)$	-0.4452 *** (0.0228)	-0.6721 *** (0.0613)
$\ln(R)$	-0.2395 *** (0.0188)	0.0142 (0.0292)
Constant	-1.7068 *** (0.1865)	-3.3018 (0.3476)
Pseudo loglikelihood	-2395.28	-477.24

Logit link function, negative binomial distribution. Robust standard errors in parentheses. ** p<0.01 *** p<0.000

In support of hypothesis 3B, in both the MOC and FOC data, an actor's interactional clustering significantly increases as a function of that actor's categorical cohesion. Because of the logit link function, the coefficients are not easily interpretable. Figure 3.6 therefore plots the relationship between φ and C for the average actor, as estimated by the model (see caption). The effect is visibly

³⁷ The logit link function constrains the dependent variable to the range [0,1], and the negative binomial distribution family is appropriate for the dependent variable's distribution. Robust standard errors are used to correct for distribution family misspecification.

bigger in the MOC: an increase along the full range of categorical proximity translates to more than a 0.2 increase in clustering. The effect in the FOC may seem small, however a change across the full range of categorical proximity translates to a change in clustering which is larger than its interquartile range. Overall, in both communities actors are characterized by categorical and interactional positions that are homologous with one another. Those whose collections are cohesive – who consume music or financial products of the same kind – also tend to interact in cliques. On the other hand, actors who occupy bridging positions in the interaction network also tend to consume a categorically diverse array of objects.

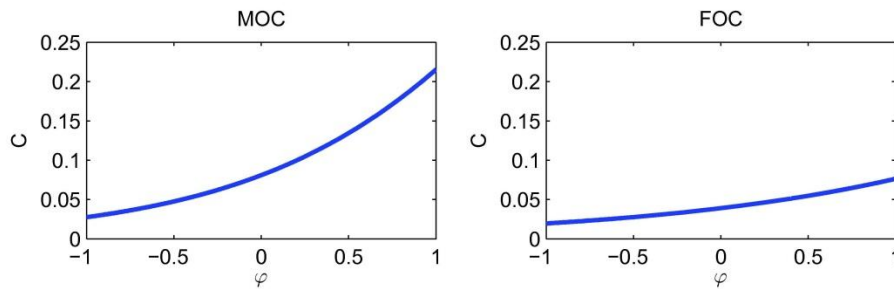


Figure 3.6: General linear model of interactional clustering (C) as a function of categorical cohesion (ϕ). The function plotted corresponds to mean control variable values (because the model has a logit link function, the magnitude of the curve is not independent of other control variables).

3.4.2.3 The Network Topology

Finally, we examine the topological relationship between both networks. If the interaction-categorization duality inheres in the emergence of groups in both the interaction and object co-occurrence networks, then, following Hypothesis 3C, we should expect to find a positive correlation between both networks' levels of segmentation (measured as network modularity) as they evolve over time. Figure 3.7 plots the segmentation of the interaction network S , Q^S , in both the MOC and FOC, as a function of time. Interestingly, the two communities follow opposing trends in their levels of segmentation. Whereas the MOC becomes decreasingly segmented over time, in the FOC, with the exception of a few initial time-slots, segmentation gradually increases. Granted, in the FOC the trend is relatively weak and subject to stronger deviations, nevertheless both trends are statistically significant at the $\alpha=0.05$ level.³⁸ In other words, the interaction pattern in the MOC is such that actors initially interact in relatively small clusters, and gradually begin interacting across clusters such that eventually the overall structure of the network becomes significantly less segmented. The opposite occurs in the

³⁸ In both communities, maximal modularity reaches 0.45. Such a level of modularity is generally considered very high for social networks. Initial time-slots were excluded for the FOC calculation.

FOC: over time, actors increasingly interact in clearly delineated groups.

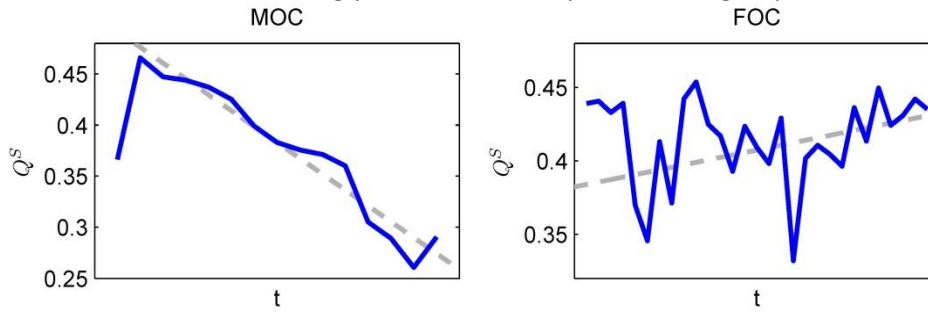


Figure 3.7: Network segmentation as function of time for the MOC and FOC data. Each data point represents modularity (Q^S) for network S at a given time-slot (t). Fitted linear models for the lines' slopes are represented with a dashed gray line.

The reasons for these opposite trends are difficult to investigate, as they depend on actors' motivations and other dispositional data that are not available. Nevertheless, this difference is congruent with what we can assume about these motivations. Musicians and music aficionados join the MOC with the intention to interact with other music enthusiasts and, in the words of the website's founder³⁹, to discover new music by listening to others' mixes. With this open-mindedness as a motivating impetus, it is not surprising that actors initially interact with those whose music they recognize as categorically similar, and gradually explore beyond their initial categorical comfort zone. In the FOC however, actors are not motivated by discovery for discovery's sake, but by profit maximization. Joining the FOC at the height of the 2008 financial crisis, they presumably seek to find clear strategies in others' financial behaviors during a time of unprecedented market instability. It is not surprising therefore, that rather than exploring in a discovery mode, they gradually converge toward interaction in closed groups.

Regardless of its underlying cause, the difference in the two trends provides a strong indication that modularity is a construct-valid measure for operationalizing segmentation. In both communities, network S grows monotonically over time in roughly the same manner,⁴⁰ yet segmentation follows

³⁹ In an interview with the author as well as in the mission statement on the website's homepage.

⁴⁰ Daily network sizes grow linearly in both communities with a slope of 28.64 and 20.8 in the MOC and FOC communities respectively.

opposite trends. This suggests that these changes in modularity are not merely artifacts of changes in network size, but rather correspond to significant structural differences over time. Indeed in theory, modularity should be insensitive to network size, as it is standardized by the overall number of edges. Yet some studies have argued that it can be affected by network size (Kreimer et al 2008). Shwed and Bearman (2010, online supplement⁴¹) conduct a series of preferential attachment-based simulations and find that in directed networks modularity is correlated with the natural log of network size. The citation networks they explore are structurally very similar to the ones under investigation in this study, with a scale-free distribution of in-degrees and a practical limitation on the number of out-degrees per node. Consequently I follow their conclusions and standardize Q^s by the log of the network's size, $\ln(|S|)$.

Calculating modularity for the standardized object co-occurrence network, Q^w , presents other challenges. Unlike traditional networks for which modularity was developed, the standardized co-occurrence network $\tilde{W}$ is a highly dense weighted network where each dyad corresponds to a weight representing the similarity between the two nodes. The majority of such weights is close to the sample mean, and therefore carries insignificant information: such values are neither significantly high, suggesting strong similarity, nor significantly low, suggesting strong dissimilarity. Consequently, I use a simple bootstrapping method to determine which weights are significant, and remove weights that are insignificant at the $\alpha=0.05$ level.⁴² Modularity is calculated on the sparse network that is produced after removing insignificant edges. To overcome problems relating to the scaling of modularity with network size, I limit the network to a fixed size of the n most popular objects during time t .⁴³ Because co-occurrence networks are significantly different from the citation networks Shwed and Bearman explore, standardizing by the natural log of the network's size is not necessarily prudent. Moreover, because

⁴¹ Found at <http://asr.sagepub.com/content/75/6/817/suppl/DC1>

⁴² This method is identical to the one used in Goldberg (2011). Though the metric used to calculate edge weights is different in this study, in both cases it is a normally distributed distance metric. See Goldberg (2011) for more details (Appendix A).

⁴³ In the MOC $n=1000$ and in the FOC $n=500$.

object popularity follows a power-law distribution in both communities, removing objects beneath a fixed level of popularity has a marginal effect in terms of aggregate collection composition. Using a fixed upper limit to network size allows avoiding problems of modularity's potential sensitivity to network size.

Figure 3.8 presents scatter plots of the levels of segmentation in the object co-occurrence network as a function of segmentation in the interaction network. Even though segmentation follows opposite trends in the interaction networks in the two communities, in both its relationship with segmentation of the object co-occurrence network is significantly positive: as interaction becomes more segmented, object co-occurrence also becomes more segmented. In other words, regardless of whether interactional segmentation is intensifying or weakening, in both communities categorical boundaries between objects are more pronounced when actors interact with one another in strongly knit cliques. In support of hypothesis 3C, the interaction-categorization duality manifests itself in a correlation between the segmentation of the two types of networks, irrespective of the types of objects being consumed, and the potentially different underlying motivations for actors to join each community.

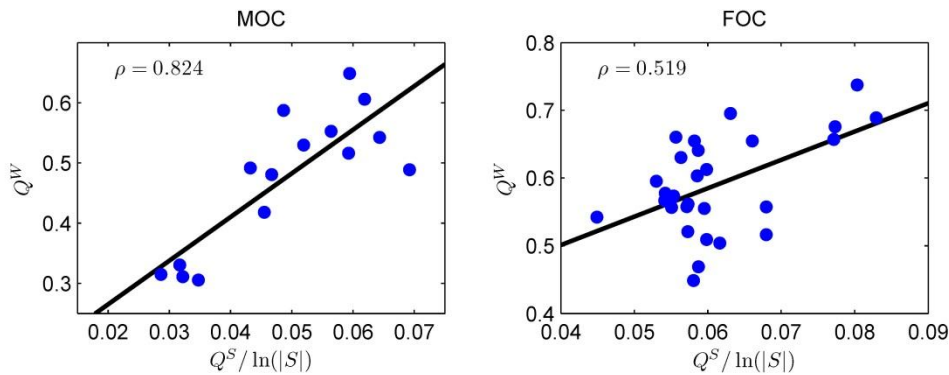


Figure 3.8: Segmentation in interaction (Q^S) and object co-occurrence (Q^W) networks for MOC and FOC data. Each dot represents the level of segmentation in both networks for a given time-slot. Q^S is standardized by network size. Black lines correspond to a fitted linear model. Pearson correlation coefficients are reported in the top left corner. Both are statistically significant: for the MOC $p < 0.001$, for the FOC $p = 0.004$.

3.5 Discussion

This chapter began with a seemingly simple question: where do classifications come from? The answer, so it emerged, is rather complicated. As discussed in the previous chapter, the human tendency to represent knowledge about the world using categories is built in to our cognitive processing systems. Yet we are not born with a prefabricated and cognitively iron-cast taxonomic guide. Rather, our categorical knowledge is built up incrementally through our daily experiences. When this sedimented knowledge depends on information we receive socially, sociologists refer to it as a ‘social construction’. In this chapter I have suggested thinking about the process of social construction as the emergence of a structural homology between two types of networks: networks of interaction between people, and networks of association between symbols and objects.

With the cultural turn in American sociology in recent decades, the idea of a duality between structure and culture has gained increasing currency. Even the staunchest of yesteryear’s structuralists have come to acknowledge the important role that cultural categories play. In an influential book that has quickly become a cornerstone of sociological literature on stratification, Charles Tilly (1999), for example, explains the durability of social inequality as a process of ‘categorical pairing’ between cultural boundaries in different settings. Yet most sociological explanations treat categories as external variables, with fixed meanings and boundaries (Friedland and Mohr 2004). When they do not, they often conceptualize the mutual constitution between cultural and material aspects of social life in very abstract terms. Here, drawing on Berger and Luckmann’s constructivist approach, I propose an analytical translation for this duality, and empirically test it. To my knowledge, such empirical evidence has yet to be presented in sociological literature.

The empirical evidence consists of three levels of structural homology between interaction and object co-occurrence networks: the dyad, the actor, and the network. On the dyadic level, we have seen

that there exists a strong and significant connection between pairs of actors' categorical proximity and their interactional distance. This may seem like an almost trivial finding; after all, it is hardly surprising that the networks under investigation are homophilous. Yet the homophily investigated in this study is not of the trivial kind. Homophilous actors in both the MOC and FOC are not identical in the makeup of their object collections. Rather, they tend to consume the same categories of objects, but not necessarily the same artists or financial products.

In fact, the results reported in Table 3A indicate that, when controlling for categorical proximity, actors whose object collections overlap do not necessarily tend to be closer to one another: in both communities, the coefficient for the cosine similarity in Model 1 is positive, and the odds ratio in Model 2 is smaller than 1. Moreover, in the MOC, both coefficients are statistically significant, suggesting that, net of categorical proximity, those whose collections are comprised of the same artists are, in fact, farther away from one another than those whose collections are different. In other words, people who interact with one another are categorically similar but tend to vary in the actual composition of their object collections. Traditional studies of homophily compare individuals on the basis of predefined and socially recognizable categories: they find, for example, that people of the similar race, gender, or occupational status, tend to have ties with one another (McPherson et al 2001). This study tells a more complex and nuanced story of coevolution. It shows that people tend to interact with those who are categorically closer to them not vis-à-vis exogenous and fixed categorical identities but rather on the basis of endogenously emergent categories as they evolve over time.

The ultimate manifestation of this coevolution is apparent in a second level of analysis – the topological level – in the co-covariance between network segmentation in the interaction and object co-occurrence networks over time. Collective patterns of interpersonal interaction, and object associations inherent in collective patterns of consumption, are strongly related. As actors follow one another in

increasingly clustered groups, their musical mixes or investment portfolios also tend to be structured by crisper categorical boundaries, which distinguish between clearly demarcated types of artists and financial assets. These categories come to mark, and circumscribe, the groups that consume them.

Examining the data at a third level of analysis – that of the individual actor – demonstrates that the different clusters in the two networks are coterminous with one another. The relationship between interactional clustering and categorical cohesion reported in table 3B implies that actors occupy structurally similar interaction and categorical positions: those who interact within closed cliques also tend to consume categorically homogenous objects, whereas those who occupy bridging interaction positions tend to consume categorically heterogeneous types of music or financial assets. Put differently, actors located on the borders separating different clusters of actors are also located on the borders interfacing between different emergent categories.

These empirical findings reject the null-hypothesis: that there is no connection between who people interact with, and the categories they use to make sense of their actions. If that were the case, we would have seen no relationship between the structures of the interaction and object co-occurrence networks. Rather, like Victorians' understandings of the animal kingdom, actors in both communities use socially constructed categories to distinguish between different types of music or financial products; these categorical identities cohere, or dissolve, in tandem with interpersonal interaction. Critics might argue that there is not much new in these findings, as we already know that people mimic one another, and pursue interactions with similar others. Yet these findings provide empirical evidence for the co-constitutive interrelationship between interaction and categorization. They are consistent with a coevolutionary model whereby the categorical identities that draw people to interact with one another are shaped through the process of interaction. Moreover, they demonstrate that social influence and

choice homophily do not necessarily lead to a stable, balkanized equilibrium: whereas the FOC moved toward increased segmentation, in the MOC segmentation gradually diminishes.

The results are replicated in both communities, despite the obvious qualitative differences between consuming music and financial products. As discussed in Chapter 1, in both domains categories mitigate uncertainty and provide structures of meaning. That categories are coextensive with social interaction in these two different settings strongly suggests that these coevolutionary dynamics are not merely characteristic of particular types of action. Skeptics might nevertheless raise an eyebrow as to whether the same underlying mechanisms are in operation in both cases. They may argue that while there is extensive evidence that suggests that music functions as a symbolic identity marker (Bourdieu 184; Peterson and Kern 1996), the underlying motivations for financial activity are substantially different and that, because stocks are not conspicuously consumed, they do not imbue their investors with identity in the same way that music does. Yet these differences do not negate the theoretical implications of these findings. To the contrary, the fact that both domains are inherently different implies that people learn from one another in an automated reflexive manner that need not necessarily be consciously motivated by identity seeking. Like the Victorian distinction between whippets and greyhounds, people draw boundaries between different artists or stocks not necessarily on the basis of their inherent and objectively measurable qualities. Rather, they look for cues at others' behaviors in order to make sense of the world. Herein lays the apparatus of social construction.

Chapter 4

How Object Meanings Change

4.1 Introduction: How Apple and Jazz Straddle Different Meanings

The categorical structures guiding actors' consumption behaviors are, as we have seen in the previous chapter, strongly related to these actors' patterns of interaction. The empirical results analyzed so far have focused on the unchanging aspects of this relationship: on the structural homology between interaction and object co-occurrence that persists throughout the period of time under investigation. Yet beneath this veneer of stability operate dynamics that are in continuous flux. Both communities, despite following opposing trends, evolve over time. This evolution entails not only that categorical and group boundaries erode or become crisper, but also that the meanings that people associate with different objects are highly contingent. If we think of the categorical landscape as a Venn diagram, then not only do the locations and sizes of circles change over time, but individual objects' positions – and the meanings they entail – are also unstable.

Mainstream sociological work often conceptualizes categorical meaning as fixed (Abbott 1988). Fundamental social categories such as race or gender are assumed to have consensual interpretations in and of themselves and people's or objects' categorical identities are concomitantly assumed to be stable. But when categorical boundaries change, so do the identities of the objects they classify (Rao et al 2005; Negro et al 2011). Yet very little has been written by sociologists about how social dynamics can lead actors to rethink the meanings of objects.

Consider Apple Inc. as an example. Famously founded in 1976 in a Los Altos garage, Apple's profound impact on the computer, software and mobile phone industries is indisputable. Today, Apple is widely considered to be one of the most innovative and influential companies in the world.⁴⁴ Yet this multinational corporation was not always consensually revered. Following its initial meteoric success in the early nineteen eighties, Apple's personal computer market share began to decline by the second half of the decade, and the company's image as a maverick innovator continued to erode for the ten years that followed; by the mid nineteen nineties, Apple was perceived as a chronically unsuccessful niche producer in a market controlled by Microsoft and competing hardware manufacturers (Linzmayr 2004). That decade was characterized by Apple's successive botched attempts to reinvent its products and image. Its iconic CEO, Steve Jobs, at the time on forced exile from the company, similarly embarked on failed endeavors, including founding NeXT Computer which never managed to reach commercial success. By the late-2000s, Apple became synonymous with stock market out-performance. Yet in the period between 1987 and 1998 its stock consistently declined in value, during a time of almost uninterrupted financial market prosperity.

Jazz similarly experienced profound transformations in its public perception. Originating in African-American communities in the southern US at the turn of the previous century, during its first decades jazz had to battle relentless bad publicity. Partly owing to its black roots, partly to its popularization by 1920s middle-class counterculture youths, jazz was often vilified in the mainstream media for its morally debasing characteristics, and was dismissed by critics and intellectuals alike for its alleged lack of musical sophistication. Influential artists such as Dizzy Gillespie or Duke Ellington – today widely considered musical giants whose names adorn buildings and streets – resented their portrayal as lowbrow musicians. By the nineteen sixties, however, jazz had been transformed from popular music to art (Lopes 2002, 2005). Emphatically adopted and legitimized by critics and musicologists, it came to

⁴⁴ Fortune Magazine has named it the most admired company in the world three years in a row from 2008 to 2010.

occupy a public image of a high art form; in 1987 it was designated a “National American Treasure” in a resolution passed by the House of Representatives.

How and why are these meanings transformed? One might argue that this has to do with substantive changes in the objects themselves. Indeed, throughout its relatively short history, Apple had turned to a variety of markets; its recent success is largely attributable to its expansion into the music player and mobile communication markets, beyond its traditional focus on personal computers.⁴⁵ Similarly, jazz music evolved throughout the twentieth century. The emergence of Bebop in the 1940s, for example, introduced an emphasis on virtuoso features such as improvisation and dissonance which highlighted the music’s artistic foundations.

But these changes become meaningful only if interpreted by audiences as such. Apple had always strived – often unsuccessfully – to innovate and branch outside the personal computer market; yet its resurgence in the late 1990s was, in fact, related first and foremost to the introduction of a new generation of personal computers, its core and hitherto most identifiable market. Jazz was similarly characterized from the outset by features regarded today by enthusiasts as its artistic essence, such as its rhythmic structures and improvisational nature. In other words, change in object attributes does not necessarily translate to change in audience perception.

In and of themselves Apple’s and jazz’s evolutionary trajectories were inconsequential as long as audiences did not apprehend them as new. And these perceptions were always strongly tied to the social identities of those producing and consuming Apple’s products and jazz music. The computer industry of the early nineteen eighties, with Apple as one of its most prominent players, was constitutively inspired by the contrarian and individualistic counterculture of the nineteen sixties (Turner 2006). Apple’s iconic 1984 super-bowl ad, in which it was portrayed as a liberating force acting

⁴⁵ In fact in 2007 Apple changed its name from Apple Computers to Apple Inc. as a symbol of its expansion beyond traditional computer consumer markets.

against the tyranny of a big-brother-like technocracy, captured the essence of this spirit. And no one personified this essence more than Steve Jobs, an avid Beatles fan and motorcycle enthusiast who is known for his anti-business-attire sneakers and black turtleneck shirts. Jobs' departure from Apple between 1985 and 1997 coincided with its fading renegade image; his return signaled Apple's re-categorization as the epitome of 'cool' in the public imagination. The introduction of the iMac computer line in 1998 heralded a new phase in the company's image and public perception.

Jazz's fortunes are similarly strongly tied to the identities of its players. Its adoption by white middle-class urban youths disenchanted by the horrors of the First World War and rebelling against the rigidity of their parents' generation transformed jazz from an exclusively African-American art to a popular and socially crosscutting form of counterculture music. Yet its transformation into an art form was only accomplished several decades later with the institutionalization of a jazz art-world centered on dedicated clubs, magazines and books. Ultimately, this sacralization was only fully accomplishable when actors in academia and the art establishment provided it with their socially recognizable seal of approval (Lopes 2005).

This chapter traces the processes through which audiences, and the social positions their members occupy, shape the meanings of different objects over time. Granted, the limited timescale available in the data at hand does not allow for the kind of fundamental transformations experienced by Apple or jazz music and that transpire over decades. Institutions, such as those legitimating a new music as an art form, take many years to emerge and become influential. But this limitation serves as an advantage insofar as we can assume that over a period of several months profound exogenous effects – such as transformations that make an industry obsolete or a musical genre anachronistic – are unlikely to have occurred. Changes observed in the online communities under investigation are therefore most likely to have arisen endogenously.

In order to understand the social underpinnings of processes of reinterpretation I draw on two different sociological literatures. Boundary theory traces back to the origins of sociological thinking, and is concerned with how inter-group differences are sustained through the symbolic production of boundaries (Lamont and Molnár 2002). Motivated by understanding the durability of social inequalities, this literature has paid little attention to exploring why symbols change their meanings. Organizational ecology literature explores how multi-category membership affects processes of evaluation (Hannan 2010). Though this work has mostly addressed cases where the classifications in question are fixed, recent studies have explored how the effects of categorical incoherence depend on these categories' evolution (Ruef and Patterson 2009).

Yet, current work provides very little insight as to why objects – like Apple or jazz - are collectively reinterpreted. In this chapter I focus on two object attributes: (1) *ambiguity*, or the extent to which an object is perceived by a relevant audience to have an indistinct categorical identity, and (2) *boundedness*, or the extent to which an object is associated with singular social group. I demonstrate that ambiguous objects are more likely to undergo reinterpretation, and that the stability of an object's meaning is strongly related to its boundedness. Bounded objects maintain high levels of stability, but when members of different social groups begin to consume it, an object's meaning is more likely to change.

In the following paragraphs, I review existing literature on social boundaries and categorical ambiguity. I offer hypotheses as to how ambiguity and boundedness affect the stability of an object's meaning. In the next section, I describe how meaning stability, ambiguity and boundedness are measured, and use these measures in the analysis that follows. This chapter concludes with a discussion on the implications of these findings on theory.

4.2 Boundaries and How They Are Traversed

In order to explore under what circumstances categorical meanings change, it is important to ask, first, how they are ordinarily maintained. As we have seen in Chapter 2, in practice people rely on fuzzy and weakly defined categories; not only do these categories overlap but, in essence, they need not be cognitively demarcated in order to function as semantic anchors. Consequently, cognitively represented categories are brittle and context dependent. But, in order to coordinate social action, collectively shared categories must remain resilient to individual perceptual fluctuations. To achieve such stability, the boundaries separating them from one another need to be constantly reproduced.

This is precisely what social institutions do (Douglas 1986; Zerubavel 1997). Formal bureaucracies are a quintessentially modern manifestation of such institutions. They devise criteria for distinguishing between, to name but a few examples, citizens and aliens, work-related and personal expenses, credit-worthy or non-worthy national economies, or criminal and legal behavior. Yet these clear-cut definitions are alien to how we use categories in practice. Consequently, we are often exasperated by the seeming Kafkaesque arbitrariness of bureaucratic procedures. Because we normally do not consult taxonomic codebooks to accomplish every-day tasks, categorical distinctions need to be embedded in our daily practices. This is why sociologists have been focusing on how boundaries delineating social categories are erected and sustained.

Two different strands of sociological research are particularly relevant. The first, which I refer to as 'boundary theory', traces back to Durkheimian notions on the distinction between sacredness and profanity and how it maintains group identity (Durkheim 1965; Douglas 1966). This literature has been particularly concerned with the intersection between symbolic boundaries and social inequality, and how the two reproduce one another (Lamont and Molnár 2002). Organizational ecologists, on the other hand, have – especially in recent years – been focusing attention on how categorical distinctions

structure markets by providing legitimacy and coordinating actors' expectations (Hannan 2010). The following paragraphs fuse insights from both literatures to examine what categorical boundaries are, how they are traversed, and under what social circumstances such bridging results in changes in collective meaning.

4.2.1 Symbolic and Social Boundedness

The majority of the sociological boundary literature – whether explicitly or not – rests on an analytical distinction between symbolic and social boundaries (Lamont and Molnár 2002). Symbolic boundaries are conceptual distinctions that form the basis for a consensual understanding of a given social reality. The fundamental distinction, for example, between men and women, is obviously rooted in biology. But its social implications, namely the shared understanding that gender differences carry different role entailments, are conceptual; they are buttressed through symbolic displays of gender-appropriate behaviors (West and Zimmerman 1987).

Social boundaries, on the other hand, are objectified forms of social difference. Ultimately, they are manifest in how people are divided into different groups. Some social boundaries are highly institutionalized and consequently practically impermeable: the border dividing between North and South Korea is, perhaps, the most extreme of such examples. But most boundaries are traversable to some extent. And more often than not the different groups they circumscribe also vary in their access to a variety of resources. Sociologists have therefore been particularly interested in the processes through which symbolic and social boundaries become co-extensive and reproduce one another.

The collective erection of symbolic boundaries is what Gieryn (1983) calls 'boundary-work': the attribution of characteristics to an institution for the purpose of delineating and reinforcing a social boundary. While Gieryn discusses how boundaries are created to demarcate and legitimize science from

non-science, boundary-work is pervasive in a variety of social arenas. Symbolic boundaries demarcating national identities, for example, are central to the process of nation building, and are essential for maintaining these identities regardless of whether or not political borders are in force (Anderson 1991); national currency, for instance, is as much a symbolic tool as it is a monetary instrument (Carruthers and Ariovich 2010). Similarly, boundaries between people of different ethnicities or races are maintained when visible differences between individual actors are perceived to be socially meaningful (Brubaker et al 2004).

Whereas some markers – such as skin color or sex – are embodied and therefore difficult or impossible to change, others are exclusively symbolic. Their meanings are therefore potentially more transient. Yet some non-embodied symbols still function as strong and highly inflexible markers. The yarmulke, for example, continues to distinguish religious Jewish communities in contemporary Los Angeles as it did for centuries in Europe (Tavory 2010). It carries this meaning because yarmulkes, and other forms of male Jewish-orthodox traditional clothing, are exclusively associated with the orthodox Jewish community. Imagine that other non-Jewish-orthodox Angelinos would begin walking around the city adorning yarmulkes. If that were the case, its function as a boundary marker would begin to erode. Indeed while certain forms of New Age reinterpretations of Jewish practice have been adopted by non-Jews, especially in Los Angeles in recent decades (eg. Kabbalah, Jewish mysticism and Hassidic traditions, see Huss 2007), these emergent communities' symbolic markers are very distinct from those identifying traditional orthodox Jewish communities.

Like religious tokens, music, especially in contemporary Western societies, functions as a symbolic marker (Bourdieu 1984, 1993). Through their association with different social communities, different musical genres become aesthetic boundaries (Pachucki et al 2007). In the United States, musical and racial boundaries are often strongly correlated. Hip-hop, for example, has historically been associated

with its African-American roots. Like Jazz in the early twentieth century, it continues to be negatively perceived and portrayed by mainstream audiences and media. These connotations also serve to debase its primary and racially identified audiences (Lopes 2005; Binder 1993; Bryson 1996). In a working-class Puerto Rican enclave in Boston, for example, hip-hop and R&B music add to other aesthetic symbols in marking the neighborhood's cultural, and by extension social, contours (Small 2004).

But music is also semantically very malleable. What constitutes a musical boundary, and the meanings associated with the category a boundary demarcates, are historically and contextually contingent (Lena and Peterson 2008). Though musical consumption in the US is segregated by race, unlike yarmulkes, hip-hop is not, by any means, exclusively consumed by African-Americans or a racially homogenous audience. Indeed different musical styles, like Jazz, have over the years come to carry different meanings. To this day, some Americans continue to associate Jazz predominantly with black identity, whereas others conceive of it mostly as a sophisticated art form (Goldberg 2011). Because cultural consumption is elastic, it is prone to such changes. In fact, Peterson and colleagues have persuasively demonstrated that since the nineteen sixties there has been a significant shift not only in the demarcation of musical genres but in the underlying logic of artistic and cultural boundary construction. Cultural omnivorousness, they argue, has supplanted traditional distinctions between high and low-brow music such that one's breadth of musical preferences has become a dominant marker of social status (Peterson and Kern 1996).

The connection between symbolic and social boundaries can therefore be conceptualized as their mutual *boundedness*. Yarmulkes are exclusively bounded to a highly segregated social group – orthodox Jews – and therefore very effectively mark this group's boundaries. Jazz, on the other hand, is no longer predominantly consumed and performed by black audiences and artists, and is therefore amenable to multiple and contextual interpretations. Boundedness can therefore be defined as the extent to which a

symbol is exclusively associated with a socially distinctive group. As Tavory (2010) demonstrates, yarmulkes function as identity markers regardless of whether or not those wearing them are consciously displaying their Jewishness. Because they are highly visible and universal, and because Hassidic Jews congregate in very tight-knit communities, orthodox Jewish attire erects a highly impermeable symbolic boundary.

4.2.2 Ambiguity

Whereas boundary theory is predominantly preoccupied with how boundaries are constructed and maintained, organizational ecologists have mostly studied boundaries as sites of meaning straddling. As opposed to the former, the latter are concerned less with questions of inter-group inequalities and more with individual variability in outcomes.

Central to this line of work is the concept of *ambiguity*: an object that spans one – or multiple – categorical boundaries. A burgeoning literature on categorical spanning and its consequences has found, mostly, that ambiguous products and producers tend to fare relatively badly in markets (Hsu et al 2009; Zuckerman 1999; Hsu 2006). More relevant to our discussion, these findings also suggest that in order to understand how object meanings' are collectively transformed one needs to examine how, and why, ambiguity arises (Hannan et al 2007; Rao et al 2005; Ruef and Patterson 2009).

According to this theoretic approach, categories are understood as social agreements about the meanings of labels that people assign to objects (Negro et al 2010). Categories are important for the understanding of markets because labels confer identity to organizations. These identities consist of sets of codes that specify the features that an organization is expected to possess (Hsu and Hannan 2005). As such, categorical identities structure audiences' expectations and evaluations of producers and product

in a market. Consequently, they also orient producers' engagement in markets in pursuit of these audiences (Hsu et al 2009).

But organizational identity is, more often than not, difficult to discern. It may relate to the attributes of the products an organization produces, for example how a movie conforms to genre conventions (Hsu 2006) or how a car brand conforms to conventional attributes of its automobile class (Rosa et al 1999). It may also relate to organizational characteristics such as size or modes of production (Negro et al 2010; Carroll and Swaminathan 2000). Thus markets can be understood as a matching process between organizational attributes and audiences' evaluations of these attributes' identity entailments, be the audiences comprised of consumers, analysts or other organizations.

From an ecological perspective this matching process can be seen as tension between specialists, who adopt and conform to a singular categorical identity, and generalists, who display features of myriad categories (Zuckerman et al 2003). A variety of studies suggest that because audience perceptions are strongly shaped by collective classifications, organizations that fail to specialize are not easily classifiable, and are consequently less effective in targeting an audience; ultimately, ambiguous organizations are devalued by the market (Zuckerman 1999; Rao et al 2005).

The negative effects of categorical ambiguity relate to actors' inability to reliably ascertain other actors' qualities (Hsu et al 2009). Because quality is not easily observable, audiences often interpret accessible signals as quality indicators in order to classify producers and their products. When an object emits dissonant signals, audiences find it difficult to assign it a label, and therefore generate expectations about it. This is also why institutionalized forms of classification play prominent roles in markets. Faced with uncertainty, actors – whether investors, consumers, or producers – rely on mediated knowledge produced by experts to make sense of market activity. These experts' classification

schemes reify certain identity signals as official classification criteria, which in turn construct actors' evaluations and behaviors in a manner that reproduces these reifications (Ruef and Patterson 2009).

A prominent classification system of this sort is the US government's Standard Industrial Classification (SIC) which has been used by various government agencies since the early twentieth century. Though gradually being replaced by the North American Industry Classification System since 1997, the SIC continues to be used by the US Securities and Exchange Commission. SIC codes therefore remain influential indicators in securities markets, and concomitantly in scientific studies of such markets. Zuckerman (1999), for example, very persuasively demonstrates that firms that are not covered by analysts that specialize in their SIC-classified industrial segment, under-perform in the stock market.

Indeed, FOC users often describe their investment decisions in industrial sector terms. For example, on October 12 2008, almost one month into the 2008 financial meltdown, one user writes:

"Taking positions in **energy** ... I understand the mob psychology behind the market crash but can't for the life of me see how this affects energy ... Those guys are tied down with oil contracts for a decade. Money will keep flowing down those pipes for a long time ... " [my emphasis]

Twenty four users replied to this comment, mostly concurring with its analysis. Though various companies were mentioned in these comments – corresponding to various energy-related SIC codes - all these individual respondents seemed to implicitly agree that 'energy' is a meaningful category. Similarly, a different thread started in May 2009 discusses the merits of investing in solar energy producers and manufacturers. Altogether 444 different comments included the word 'energy'. Other categorical labels such as 'technology' (1031 comments) or 'banking' (826 comments) were also commonly used.

But SIC is not the only classificatory prism through which investors understand the market. On January 2009 another user posted the following opinion piece:

“Looking for ex-Giants. My theory is that 3 years from now the worries of toxic money will have evaporated. Uncertainty is what sent so many companies down the value chain, some nearly whipped out ... I suppose my rule of thumb would be companies which have slid over 70% without losing significant market or customer share”

In the discussion thread that followed,⁴⁶ a variety of such companies are mentioned, spanning a range of sectors: from United States Gypsum Corporation (SIC-classified as ‘Gypsum Products’), through Ford and GM (‘Motor Vehicles’), to Cemex (‘Ready-Mixed Concrete’) and Ebay (‘Business Services’). The logic informing this discussion seems to be less about the audiences these various firms target, and more about these organizations’ performance and size characteristics.

These different ways of lumping firms together are also apparent in users’ aggregate behaviors. Let us look again at Apple, and one of its major competitors, Dell. Like Apple, Dell is a leading US multinational corporation that manufactures and markets computers as well as servers, networking devices, smartphones, and other computational devices. Both firms are traded on NASDAQ and are classified by SIC as Electronic Computers manufacturers. Their target markets are very similar. But FOC users do not appear to perceive them as belonging to the same class of organizations. Figure 4.1 plots categorical similarity (τ) between Apple and Dell throughout the period under investigation. Recall that categorical similarity measures the extent to which two objects are perceived to belong to the same

⁴⁶ This thread comprised of 24 comments by 10 different users.

category.⁴⁷ As the figure illustrates, during most of the period the categorical similarity between the two companies was insignificantly different from 0, at times the two were even seen as having opposing categorical identities. In contrast, the company perceived to be most categorically identical to Apple was Google, a NASDAQ-traded internet search corporation classified by SIC as an Information Retrieval Service. Though both companies operate in the information technology space, they provide very different products that are traded in different markets.⁴⁸ Nevertheless, FOC users' collection compositions suggest that they consider them to be similar.

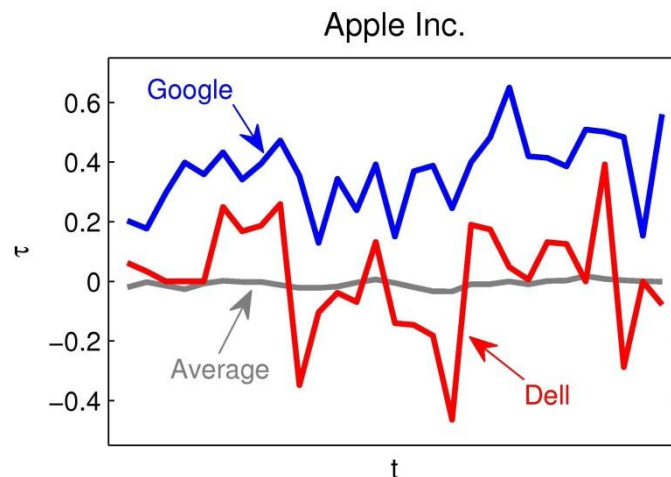


Figure 4.1: Categorical similarity (τ) as a function of time (t) between Apple, Dell and Google. The gray line corresponds to the average categorical similarity between Apple and all other stocks.

Of course, the similarity between Apple and Google would come as no surprise to anyone familiar with their histories. Both have successfully created highly recognizable IT brands and have become synonymous with the Silicon Valley from which they operate.⁴⁹ Like Jobs, Google's founders Larry Page and Sergei Brin are known for their non-conformist styles. Moreover, both companies are famous for

⁴⁷ Formally, categorical similarity is calculated as the structural equivalence (row correlation) between two objects in network $\tilde{W}$, which is the standardized co-occurrence network. It measures the extent to which two objects tend to co-occur with the same other objects in users collections. See Chapter 2, section 2.4.2 for details.

⁴⁸ According to SIC's nested structure, Apple and Google not only belong to different sectors but to different industry types altogether. Whereas Apple is classified as a manufacturer, Google is classified as a service provider.

⁴⁹ In their comments, FOC users often refer to both companies' strong brand identities. The word 'brand' appears in 14 comments that include the name Apple, and 14 comments that include the name Google. The two company names appear together in 27 different comments.

their informal and non-hierarchical corporate cultures. Indeed, ever since the nineteen seventies the Silicon Valley has been identified with a decentralized industrial system, characterized by inter-organizational cross-pollination and the abandonment of traditional symbols of organizational hierarchy. This communal identity is so strong that many employees identify more with Silicon Valley as an institution, than they do with their individual employers (Saxenian 1994). Both Apple and Google appear to be perceived by FOC users as Silicon Valley exemplars. Dell, on the other hand, seems to be understood as a more old-fashioned, traditional US corporation. Indeed amongst the firms it is categorically similar to are corporations such as Wal-Mart, GM, Ford and Cemex. Unlike its Silicon Valley competitors, Dell appears to be seen as one of the ex-giants referred to in the user comment cited above.

That Apple, Google and Dell's identities are not exclusively dependent on the products they sell is not unique. In fact, because quality is not easily observable, actors in markets often turn to socially-determined signals in order to infer it. Perhaps the most ubiquitous of such signals is price. Financial investors customarily rely on past market performance as proxies for future gains (Markowitz 1991), in the same way that music consumers interpret popularity as signals about song quality (Salganik et al 2006). Yet the collective judgments of investors or consumers are not always reliable, or readily available (Malkiel 2003). Consequently, as is the case with Apple and Dell, the social positions producers and consumers occupy are consequential for how others evaluate them (Granovetter 1985; White 1981). Such evaluations are shaped by organizations' locations in firm status hierarchies. In the securities market, for example, investment banks' status (as inferred from their name's position in security offering ads) has an impact on the price they charge for underwriting (Podolny 1993). Similarly, actors look at producers' relations with other actors to make judgments about quality. For example, biotechnology firms with prominent alliances issue stocks faster, and get better evaluations than do firms lacking such partnerships (Stuart et al 1999).

Actors occupying socially incoherent positions are sending mixed signals about their quality, and should therefore find themselves ambiguously evaluated by others. In the context of the online communities under investigation here, an object's social position relates to the positions occupied by the users that consume it. In consumer markets, goods and brands often acquire status through the people who consume them. Jewelry, fashion items, as well as cars and computers become symbols of wealth, success and cultural sophistication through their conspicuous consumption (Veblen 1991; Podolny 2005; Molotch 2003). Consumption objects come to carry social meanings when they are consumed by a socially bounded group (Pugh 2009; Muniz and O'Guinn 2001; Berger and Heath 2007). Hence ambiguity and boundedness are strongly associated concepts: if signals about quality are inferred by actors from social context, then market actors whose corresponding audiences are socially bound should have more coherent identities.

4.2.3 When Are Objects Reinterpreted?

As we have seen, sociological research has paid growing attention to questions of boundary maintenance and crossing. Yet most of this literature treats the categorical landscape itself as exogenous, and consequently as fixed. Relatively few studies ask why objects – such as firms traded on Wall Street or musicians – don new identities.

One strand of literature focuses on categorical emergence. Ruef (2000) finds that population-level dynamics can encourage, or hinder, the emergence of new organizational forms. When the number and size of existing forms reaches saturation, new forms become recognized as distinct categories. Yet as McKendrick and Carroll (2001) demonstrate, density alone cannot always explain emergence. The disk array industry, which is the focus of their study, failed to gain legitimacy because most of its producers derived their identities from other industrial categories. When new forms emerge they evolve through a

process of meaning negotiation. At first, meanings are in flux and there is little consensus among audiences about what attributes make a given object worthy of a categorical label. As time goes by, meanings congeal through institutionalization, and categorical boundaries solidify. Objects that fail to conform to these boundaries are penalized (Ruef and Patterson 2009; Rosa et al 1999).

These studies suggest that, with the exception of new forms emerging, categorical boundaries should become less permeable as time goes by. But existing categories and objects can undergo processes of reinterpretation (Negro et al 2011). Why does such reinterpretation occur? Of course a trivial answer is that, simply, the objects themselves change. IBM, for example, underwent a fundamental strategic reorientation in the nineteen nineties moving from computers as its core business to consulting; in 2005 it sold its personal computer division to Lenovo (O'Reilly et al 2009). Similarly, British singer Robbie Williams began his career in the highly successful yet critically unappreciated boy-band *Take That*. After embarking on a solo career in 1995 he turned to what is commonly considered more sophisticated music, earning accolades from critics and winning seventeen BRIT awards, more than any other single artist to date.⁵⁰

Yet, as we have seen, changes from within do not necessarily translate into new identities: Apple's attempts at reinvention were not always successful; nor was it easy for Robbie Williams to persuade new, more snobbish audiences, to listen to his music. An existing identity often functions as a strong barrier to change even when an actor actively pursues such change. In the feature-film labor market, for example, actors are often unable to escape a typecast trajectory established in the earlier stages of their careers (Zuckerman et al 2003). It is often the actions of others, and the social positions they occupy, that confer new meanings to existing objects. In the case of French gastronomy, for example, Rao and colleagues (2005) find that categorical boundaries weakened when high-status actors' borrowing from

⁵⁰ Retrieved from the BRIT Awards website's history section August 2 2010, <http://www.brits.co.uk/history>

other categories triggered emulation. Legitimacy played an important role when chefs associated with traditional French cuisine began borrowing dishes from the rival Nouvelle-Cuisine movement. At first, cross-categorical borrowing resulted in critics' disapproval. But such behaviors from highly acclaimed chefs were less likely to result in criticism, and therefore more likely to encourage others to do the same. Consequently, categorical borrowing became socially legitimized.

If perceptions about objects are structured by collectively shared categorical definitions, then ambiguity should play a central role in how such definitions change. Recall that ambiguity is not a property of the object; rather, it corresponds to the overall coherence in collective interpretations of an object. And audience members are influenced by others' interpretation in formulating their own understanding of an object. When these collective interpretations exhibit a strong consensus, then actors receive consistent and reinforcing signals from others about the identity of an object. But when there is disagreement between audience members on how a particular object is understood, it generates mixed signals about its identity; such ambiguity creates the potential for a transformation in how the object is collectively perceived (Hannan et al 2007).

Ambiguity should therefore be consequential for meaning stability only if people indeed look at others for deducing meaning. If, on the other hand, they devise their interpretations merely by examining these objects' attributes, then ambiguity – as defined here as the product of collective interpretations – should have no effect on an object's likelihood of reinterpretation. Hence:

Hypothesis 4A: *High ambiguity should result in subsequent reinterpretation*

As we have seen, in arenas of consumption, boundedness should be strongly related to ambiguity. Existing literature points to the social composition of actors labeled with a categorical identity as an important factor in the label's potential of being reinterpreted. Berger and Heath (2008) argue that when outsiders start consuming an object that is associated with a group, this object's identity is

transformed. They find that once members of a low-status dormitory began wearing a particular wristband, higher-status college students who were hitherto affiliated with that wristband stopped wearing it. Consequently its function as an identity signal changed. Negro et al (2011) similarly find that when certain Italian Barolo and Barbaresco winemakers began to challenge traditional winemaking practices, their fusion into a group labeled ‘modernist’ catalyzed a reinterpretation of the original winemaking category. In both cases, the boundedness of the objects in question diminished as different groups of actors made (explicit in the latter case and implicit in the former) claims for categorical identity. Consequently, these categories became more ambiguous.

Boundedness can help elucidate Apple’s fortunes. From the outset, Apple had a strongly identifiable brand. During the eighties and nineties, this brand was associated with Apple enthusiasts who often referred to their consumption of its products in almost religious terms (Kahney 2004). Ten years later, as Apple’s public image was being transformed, its products started to be consumed by a far more socially diverse crowd. More generally, when an object is being consumed by consumers from a hitherto unassociated social group, their consumption of this product begins sending mixed signals about this object’s identity. Such signals should spawn reinterpretation. Hence:

Hypothesis 4B: *A change in an object’s boundedness should result in subsequent reinterpretation.*

4.3 Measurements

4.3.1 Semantic Stability

Meaning, as discussed in Chapter 2, inheres in relationships between objects. Accordingly, the categorical similarity between two objects is defined (Formula 2.3) as the structural equivalence

between them in network $\tilde{W}$. Recall that $\tilde{W}$ is the standardized object co-occurrence network. A weight between two nodes in this network corresponds to the amount of collections in which the two objects co-occur, standardized by their popularities. Structural equivalence between nodes in $\tilde{W}$ measures the extent to which both objects co-occur with the same other objects.

A single object's meaning as it evolves over time can also be conceptualized in structural equivalence terms. Recall that using the sliding window technique, time is divided into T slots of equal size. T different snapshots of network $\tilde{W}$ are therefore created. Imagine that there is no change whatsoever in object meanings between two timeslots, t_1 and t_2 . If that were the case, we would expect all objects to maintain exactly the same standardized co-occurrence with one another; snapshots t_1 and t_2 of network $\tilde{W}$ would be identical. Consequently, the row in network $\tilde{W}$ corresponding to each object would retain the same values in both snapshots. Thus, we can think of the extent to which object i 's meaning changes between timeslots t_1 and t_2 as its structural equivalence with itself across the two timeslots. As is the case with categorical similarity, structural equivalence is measured using Pearson's correlation coefficient.

Recall that $\vec{\tilde{W}}_i$ denotes a vector of length K , corresponding to the i 'th row in matrix $\tilde{W}$. Object i 's semantic stability at time t is therefore defined as the correlation between $\vec{\tilde{W}}_i$ at times t and $t-1$.⁵¹ Formally:

$$(4.1) \quad \eta_{i,t} = \rho(\vec{\tilde{W}}_{i,t}, \vec{\tilde{W}}_{i,t-1})$$

Figure 4.2 plots the semantic stability of a few of the artists that appeared at least in two collections in all T timeslots. Unsurprisingly, of these artists, *The Beatles* enjoy the most stable meaning, averaging at 0.5214. Other sixties and seventies classic pop and rock artists such as The Rolling Stones,

⁵¹ The vectors include coordinates corresponding to co-occurrences with objects that were included at least in two collections at times t and $t-1$.

Led Zeppelin and Bob Dylan are also amongst the most stable artists. Four decades after releasing their most influential materials, these artists' genre identities appear to have been strongly solidified. In contrast, *Final Fantasy*, the stage name of the classically trained violinist-cum-Indie pop artist Owen Pallett, is highly unstable. Pallett's musical repertoire is very difficult to describe in conventional genre terms. Classified by AllMusic.com as Rock/Alternative/Indie Pop/Indie Rock, his music conforms to neither of these genres' conventional characteristics.⁵² His second album, for example, consists entirely of string quartet arrangements. Consequently, MOC users do not appear, for the most part, to be categorizing him in any coherently meaningful manner.

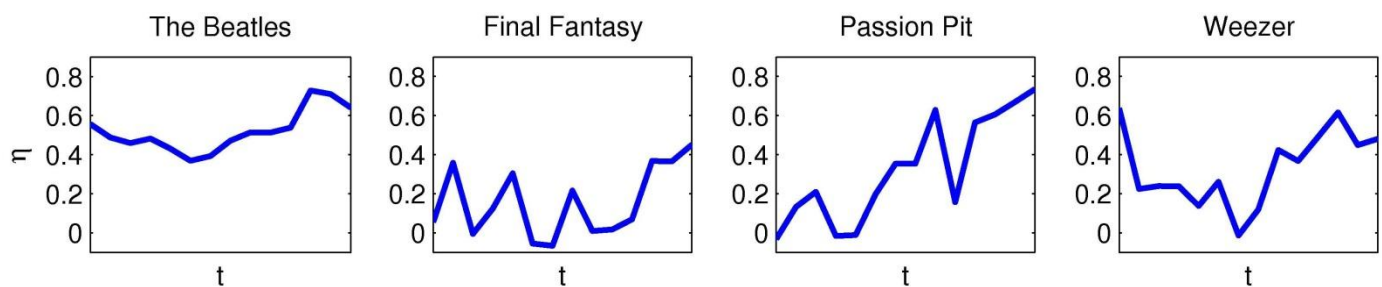


Figure 4.2: Semantic stability over time for four exemplary MOC artists

In both these cases – a highly established classical genre exemplar and a fringe experimentalist – semantic stability behaves as we would expect it to, yet ultimately reveals nothing surprising. But consider two other cases depicted in Figure 4.2. In the first, electro-pop band Passion Pit's stability gradually increases over time, peaking at almost 0.8 toward the end. The band's collective meaning appears to be cohering as time goes by. In contrast, alternative rock band Weezer's semantic stability begins quite high, but diminishes quickly only to re-stabilizes again toward the end of the time period. The band appears to be undergoing collective reinterpretation. Indeed unlike Passion Pit which only released its first single in 2008, Weezer had been on the alternative rock scene since 1992, and had gone through several reincarnations. During the first time-slots, Weezer is strongly associated with

⁵² AllMusic.com is a popular and comprehensive online music guide, run since 1995 by All Media Guide. The website employs experts to classify musicians.

contemporary late 2000s alternative rock bands such as Muse and Snow Patrol. Toward the end it increasingly becomes associated with bands that were prominent in its early days such as The Red Hot Chili Peppers, The Pixies and Nirvana, and even with classical acts such as The Beatles. Over time, it appears, Weezer's MOC interpretation morphs from that of a contemporary alternative rock band to that of a more established rock classic.

4.3.2 Ambiguity

Object ambiguity is defined as the extent to which an object is perceived to have more than a singular categorical identity. As discussed in Chapter 2, categories inhere in clusters of objects that are strongly connected to one another. An object that has more than one categorical identity is therefore an object that is categorically similar to objects belonging to more than one cluster.

The process of calculating categorical similarity, τ , results in a $K \times K$ matrix. Because τ is calculated using the Pearson correlation coefficient, the significance of this similarity can easily be tested. Thus, matrix τ can be transformed into a sparse network $\hat{\tau}$, such that only statistically significant positive values in τ are retained. Edges in $\hat{\tau}$ correspond to significant categorical similarity between the objects they connect. Picture the hypothetical network depicted in Figure 2.2 (Chapter 2) as such a network. Nodes, such as the one corresponding to Wal-Mart, who are embedded in strong clusters, have very low ambiguity. On the other hand, nodes like the one corresponding to Amazon, which span different network clusters, have high categorical ambiguity. In other words, nodes whose immediate neighbors are also strongly connected to one another have low ambiguity.

As we have seen in the previous chapter, the clustering coefficient is a common tool for measuring the density of a node's neighborhood. In simple binary networks, the clustering coefficient simply measures the fraction of ego's neighbors that are connected to one another. In an undirected,

weighted network such as $\hat{t}$, the calculation of the weighted clustering coefficient needs to take into account edge weights as follows (Onnela et al 2005):

$$(4.2) \quad C_i^W(w) = \frac{2}{d_i(d_i - 1)} \sum_{j,k \neq i} (\tilde{w}_{ij} \tilde{w}_{jk} \tilde{w}_{ki})^{\frac{1}{3}}$$

where d_i is node i 's degree and $\tilde{w}_{ij} = w_{ij}/\max(w)$ is the weight of the edge connecting nodes i and j , divided by the maximal edge weight in the network. In essence, the weighted clustering coefficient measures the extent to which ego's neighbors are connected to each other compared to the maximal weight in the network. Thus object ambiguity is defined as one minus the weighted clustering coefficient for a node in network $\hat{t}$:

$$(4.3) \quad a(i) = 1 - C_i^W(\hat{t})$$

Because node weights are standardized by the maximal weight in the network, the clustering coefficient rarely approaches 1 (in fact, in these data it rarely exceeds 0.85) and its distribution approximates a log-normal distribution. Ambiguity is accordingly strongly skewed toward 1. In the analyses that follow I consequently transform ambiguity by its natural logarithm.

4.3.3 Boundedness

Boundedness is defined as the extent to which an object is associated with a socially distinctive group. Thus, like ambiguity, it can be conceptualized in network terms as the density of a group of nodes. However, as opposed to ambiguity which is defined by the objects with which a focal object is associated, with boundedness the nodes in question correspond to users. As illustration, consider the hypothetical interaction network S depicted in Figure 4.3. Recall that each node in this network corresponds to an individual user, and edges to relationships between users. The network is visibly divided into two distinct clusters of individuals. Consider further that there are two objects that these

individuals consume. Nodes in the diagram are colored by the type of object that their corresponding individual consumes. If a user consumes neither object, its corresponding node is colored white.⁵³ As is clearly visible, whereas the black object is confined to the cluster of individuals on the left, the blue object is consumed by members of both clusters. In other words, the black object's boundedness is much higher than the blue object's.

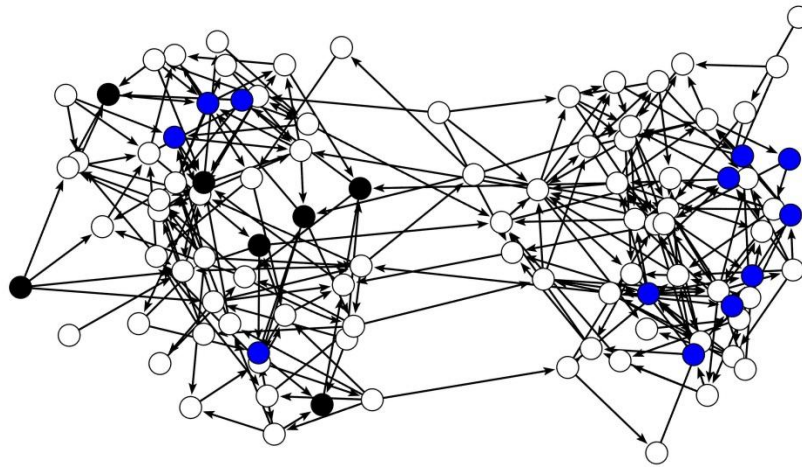


Figure 4.3: Illustration of boundedness. Node colors correspond to different objects. White nodes consume neither object.

Like ambiguity, boundedness can theoretically be calculated as the clustering of a sub-network, e.g. the density of ties between nodes comprising the sub-network formed by black nodes in Figure 4.3. But unlike network $\hat{t}$, network S is fairly sparse. Consequently, the clustering coefficient may produce misleading results. For example, whereas the six black nodes in Figure 4.3 have no edges connecting between them, the blue nodes are connected to one another through a handful of edges. As a result, the clustering coefficient for the blue sub-network is greater than the coefficient for the black network (which equals zero). If used as a measurement of boundedness, the clustering coefficient would erroneously suggest that the blue object is more bounded than the black object.

To overcome this problem, it is necessary to look for larger, underlying clusters in network S . Recall that in the previous chapter this was done using modularity maximization (Section 3.4.1.3).

⁵³ For the sake of simplicity and clarity, in this example individuals consume either object, but not both.

Modularity maximization divides a network into g mutually exclusive clusters, such that within-cluster edges are denser than between-cluster edges. Thus boundedness can be conceptualized as the cluster diversity of a subset of user nodes that consume a given object. A common measure of diversity over a set of discrete categories is Simpson's index of diversity (Hsu et al 2009). This measure is commonly known as the Herfindahl index when applied as a measurement of market concentration (Hirschman 1964).

Given a partition of a group of individuals, M , into g clusters, Simpson's diversity index is calculated as follows:

$$(4.4) \quad Div(M) = \sum_{i \in g} p_i^2$$

where p_i is the proportion of individuals that are assigned to cluster i . The greater $Div(M)$, the smaller M 's diversity. The boundedness of object i is therefore defined as the diversity of the users consuming it, with respect to network S 's division into discrete clusters. Let $[A_i] = \forall j, A_{ji} \geq 1$ denote the subgroup of users who include object i in their collection. Object's i boundedness is therefore defined as follows:

$$(4.5) \quad b_i = Div([A_i])$$

where diversity is calculated with respect to the partition of network S produced by modularity maximization.⁵⁴ Using this measure, the black object's boundedness appropriately equals 1, and the black object's equals 0.556.

⁵⁴ As mentioned in the previous chapter, modularity suffers from degeneracies. Please see Appendix B for further tests suggesting that its use for the purposes of boundedness measurement is not severely affected by these degeneracies.

4.4 Analysis

Our independent variable of interest is η , semantic stability. Semantic stability measures the extent to which an object – whether an artist or a financial asset – changes its meaning between one time slot and the next. The unit of analysis is therefore object-by-time. Figure 4.4 plots η 's distribution in both communities. As is apparent from both diagrams, in both cases these distributions approximate, though do not perfectly fit, a normal distribution. Semantic stability is rarely lower than 0; only on rare occasions an object's meaning changes so dramatically that it is negatively correlated with itself one time slot earlier. At the same time, semantic stability is equally rarely close to 1.⁵⁵ Because there are constant fluctuations in objects' meanings, an object rarely correlates perfectly with itself.

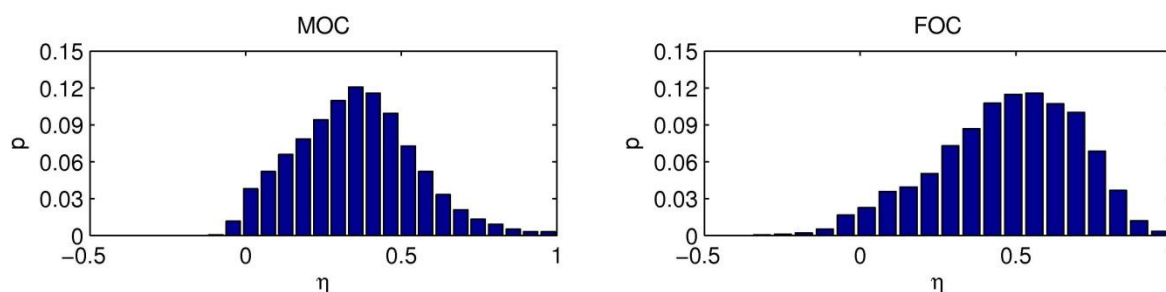


Figure 4.4: Distribution of semantic stability in both communities

4.4.1 Sources of Semantic Stability

4.4.1.1 Endogenous Sources of stability

Stability is sensitive to two types of parameters. Endogenous parameters relate to how η is calculated, and the effects of internal dynamics on this calculation. Recall that η is, in essence, a simple correlation between two vectors, relating to object i 's standardized co-occurrences with all other objects at times t and $t-1$. This measure is therefore sensitive to the overall number of cells in each vector. The greater the number of cells, the more η is resilient to changes in individual cells. The number of cells, in

⁵⁵ In fact it is never a perfect 1: the highest correlation is 0.99

turn, corresponds to the number of objects that appear in users' collections in both time slots, or to the objects that are active in both times. In a given time slot t , the set of active objects, K^t , includes all objects for which $d_{i,t}^A > 0$, or all objects whose degrees in network A are greater than zero. The number of active objects is therefore the size of the intersection of both sets:

$$(4.6) \quad |\eta_t| = |K^t \cap K^{t-1}|$$

As $|\eta|$ increases, η becomes more stable. It is therefore included in all the following regression models to account for stability that is an artifact of the number of active objects at a given time slot.

A second endogenous parameter that η is sensitive to is the number of users that include object i in their collections at times t and $t-1$, or the object's degree in network A , d_i^A , during both times. As $d_{i,t}^A$ increases, that is, as object i becomes more popular, η becomes less sensitive to idiosyncratic individual user behaviors. The difference between an object's degree at times t and $t-1$ captures the extent to which an object has changed in popularity over these two time slots. Because popularity has a heavy-tailed distribution, it is transformed by its natural log. Thus the change in an object's popularity, or its popularity delta, is defined as follows:

$$(4.7) \quad \Delta d_{i,t}^A = \ln(d_{i,t}^A) - \ln(d_{i,t-1}^A)$$

Note that popularity delta corresponds to an object's diffusion pattern: objects with high $\Delta d_{i,t}^A$ are those whose popularity increased between two consecutive time slots. Such an increase in popularity necessarily occurs when people mimic one another by including new objects in their collections; as new users adopt an object, its overall popularity increases. Thus, popularity delta can account for the effects of simple contagion. Moreover, as the number of users consuming an object increases, η becomes less sensitive to individual idiosyncrasies. In other words, as popularity increases due to simple contagion, so does semantic stability. Thus I include the popularity delta in order to control for changes in semantic stability that are attributable to simple mimetic diffusion patterns.

Finally, I include a control variable for the amount of time that has elapsed since an object first appeared in at least one collection, or its maturity, ΔT . Maturity is related to semantic stability in two ways. First, when a new object is introduced, simple contagion should translate into gradual diffusion, and consequently into increasing semantic stability. Second, maturity also confers legitimacy: the longer an object has been active, the more time users have had to observe it, and make sense of it.⁵⁶ When the rock band *Motion City Soundtrack*, for example, was first introduced, only 2 users included it in their collections, and its semantic stability in the following time-slot was very low at 0.094. Twelve time slots later its stability increased to 0.522. Controlling for time since first appearance therefore accounts for meaning stabilization that is the result maturation. To account for objects that were included from the first time slot, and for whom elapsed time is effectively left-censored, I also include a dummy variable that equals 1 if the object was active in the first time slot.

4.4.1.2 Exogenous Sources of Stability

Exogenous variables relate to object attributes that are not influenced by local dynamics. Because these attributes are inherent to the objects themselves, and not the result of user behaviors, they vary across the two online communities; after all, music and finance are very different domains.⁵⁷ Object attributes that warrant our attention are those that may have an influence on semantic stability. In the MOC, such attributes relate to artist characteristics that make these objects more or less likely to have stable musical identities.

One such attribute is an artist's recognizability. Artists that reach high levels of fame also enjoy greater influence on audiences' perceptions about music. Successful artists that fill stadiums and sell

⁵⁶ This argument is related to discussions on the 'liability of newness' in organization studies: new organizations are more likely to dissolve than established ones. As various studies have shown, this liability is also a product of low initial legitimacy (Hannan and Freeman 1984; Hannan et al 2007).

⁵⁷ Of course, these attributes may themselves be the result of social processes operating outside the online communities. For the sake of this analysis, the sources for these processes can be overlooked and treated as fixed externalities.

millions of albums presumably become, at least in the eyes of their fans, prototypical representations of the musical categories with which they are associated. Such artists should therefore exhibit greater semantic stability. Indeed *The Beatles* who, as we have seen, are the most stable artist in the MOC, are also the top selling music artist of all times (and the second most popular artists in the MOC).⁵⁸ Certified record sales reported by the Recording Industry Association of America (RIAA) are often used as reliable measures of artist success. But while these data refer exclusively to sales in the US, many MOC users reside outside the US, and include in their collections artists whose sales are not monitored through the RIAA.⁵⁹ As an alternative, I use data on artists popularity retrieved from the online music site Last.fm.⁶⁰ An artist's play-count on Last.fm serves as a measure of their success. Unsurprisingly, play-count has a heavy-tailed distribution ranging from 11 to 243,290,901 plays (the former is the play-count of 21 different MOC artists, and the latter is *The Beatles'* play-count). Consequently, I use the natural log of Last.fm play-count as a control variable.

A second attribute warranting attention is genre. While institutionalized categories structure individual perceptions in practically every domain of social activity, in music such categories are highly pronounced (Lena and Peterson 2008). Consider the difference between country music and jazz: one would be hard pressed to hear both played on the same radio program, or performed in the same venue. It is therefore likely that pre-existing institutionalized categorical identities play a role in structuring consumption patterns in the MOC. In fact, MOC members mostly, though not exclusively, listen to rock and electronic music. Other musical genres, especially those perceived to be very different from rock and electronic music, might consequently be included in more genre-homogenous collections,

⁵⁸ According to the Recording Industry Association of America's website, www.riaa.com, retrieved on June 18th 2011.

⁵⁹ In fact many of these artists distribute their music online and even if consumed by US audiences are invisible to the RIAA.

⁶⁰ Last.fm is an online radio and music consolidation website with tens of millions of users. Artist profile pages on the website also include information about the amount of times an artist was listened to by a user. For more information about Last.fm and how data retrieved from it was associated with MOC data, please see Appendix A.

and therefore subject to higher levels of stability. To account for such potential variability I include genre as a control variable. An artist's genre identity is retrieved from the online music guide AllMusic.com.⁶¹ Overall, MOC artists belong to twenty different genres.⁶² 62.92 percent of them are classified as 'Pop/Rock', and 11.16 as 'Electronic'.

Whereas the meanings associated with music artists are influenced by these artists' popularities and their adherence to genre conventions, what determines investors' perceptions of financial artists is, first and foremost, how these assets perform in the market. Two market performance metrics – rate of return (ROR) and volatility – are the cornerstones of modern portfolio theory (Markowitz 1991), and are highly institutionalized by analysts and the financial media as prisms through which to assess different financial assets (Bodie et al 2004). ROR, or the ratio of money gained or lost on an investment, is commonly reported by the press as an indicator of market success. Headlines following a crash on Wall-Street routinely refer to leading indexes' loss ratios as indicators of the US economy's health. FOC users are particularly primed to think of assets in ROR and volatility terms. When viewing others' profiles, FOC users are presented with a summary of those users' portfolio ROR and volatility. Other metrics such as high level (two-digit) SIC classification and market capitalization (share price multiplied by the number of outstanding shares) are also presented.

Rate of return is simply the ratio between a share's price at a given moment and its price when previously observed. This metric is conventionally expressed as a continuously compounded function in natural logarithm terms:

⁶¹ Each artist is associated with one primary genre on AllMusic.com. For more information about the website and how data was retrieved from it please see Appendix A.

⁶² These genres include (in descending order of popularity): Pop/Rock (63.93%), Electronic (11.16%), Rap (6.56%), R&B (5.56%), Jazz (2.70%), Reggae (1.56%), Country (1.53%), Folk (1.41%), International (1.38%), Latin (1.28%), Vocal (1.13%), Blues (0.69%), Stage & Screen (0.63%), Avant-Garde (0.33%), Comedy/Spoken (0.30%), New Age (0.24%), Religious (0.20%), Easy Listening (0.20%), Classical (0.13%) and Children's (0.07%).

$$(4.8) \quad ROR(i, t) = \ln \left(\frac{v_{i,t}}{v_{i,t-1}} \right)$$

where $v_{i,t}$ is share i 's value at time t (Bodie et al 2004). ROR measures the relative increase or decrease in a share's price over a certain period of time. Volatility relates to a financial asset's level of risk. It is commonly calculated as the standard deviation of this asset's continuous return over a certain period of time. Formally:

$$(4.9) \quad Vol(i) = \sigma_{ROR(i)}$$

where ROR includes a set of observations over a fixed period of time. High volatility means that the asset's price fluctuates excessively, and therefore suggests that this asset is risky (Bodie et al 2004).

Because FOC users are constantly exposed to ROR and volatility as descriptive asset metrics, I add them as control variables. Volatility in particular should, presumably, be related to semantic stability. Volatile assets are those which market actors are collectively undecided about their underlying values. Such uncertainty should translate into lower stability. Because asset metrics are measured during different points in time, I standardize ROR and volatility by the S&P 500 index (commonly used as an indicator of overall market performance) as observed during the corresponding time period, as follows:

$$(4.10) \quad ROR^s(i) = ROR(i) - ROR(sp500)$$

$$(4.11) \quad Vol^s(i) = \ln \left(\frac{Vol(i)}{Vol(sp500)} \right)$$

ROR for time t is calculated over a period of three months preceding t , and volatility is calculated for 36 months, using monthly observations as data-points. Additional standard metrics such as market capitalization and SIC classification were found to be unassociated with semantic stability and were therefore not included in the following models as control variables. All market performance data were retrieved from The Center for Research in Security Prices (CRSP) daily and monthly stock archives.

4.4.2 Results

The hypotheses examined in this chapter are concerned with the effects of endogenous social dynamics on fluctuations in object meanings. If these dynamics affect semantic stability then events occurring at time t should explain variability in semantic stability (η) at time $t+1$. The regression models used in this section are therefore structured such that independent variables are measured at time t , and the dependent variable η measures change in object stability occurring between times t and $t+1$.

We begin the analysis by examining sources of stability that are not the results of endogenous interaction dynamics: the control variables. Results are reported in Tables 4.1 and 4.2. Model 1 reports the effects of endogenous control variables on semantic stability. As expected, object popularity ($d_{i,t}^A$) and popularity delta ($\Delta d_{i,t}^A$) have a positive effect on stability in both communities. Popular objects tend to enjoy greater semantic stability, and as they diffuse to more users in the community, their stability continues to increase. Time elapsed from first appearance (ΔT) similarly has a positive effect: objects that appeared in users' collections stabilize as time goes by. Model 2 adds time-slot specific variables: time-slot (t) and active object count ($|\eta|$). In both communities, as the coefficient for time-slot suggests, stability gradually increases as time goes by. In other words, irrespective of the fact that both communities follow oppositional trends with respect to the strength of categorical boundaries – as we have seen in the previous chapter, these boundaries become less pronounced in the MOC and more pronounced in the FOC – in both communities meanings gradually stabilizes over time. This stabilization emerges above and beyond the effects of increases in popularity, or the overall number of objects. Though new users join the communities over time, and introduce new objects, overall meanings stabilize. Thus, controlling for variability in stability that is the result of how this measure is calculated, these results indicate that as both communities mature they gradually converge toward collective agreement on the meanings of objects.

Table 4.1
OLS regression estimating semantic stability η between times t and $t+1$ for MOC artist i

Variable	Endogenous		Exogenous	
	Model 1	Model 2	Model 3	
$d_{i,t}^A$	0.0646 *** (0.0019)	0.0644 *** (0.0019)		
$\Delta d_{i,t}^A$	0.0220 *** (0.0027)	0.0223 *** (0.0028)		
ΔT	0.0016 *** (0.0004)	-0.0044 *** (0.0012)		
$\min(t)$	-0.0294 *** (0.0037)	-0.0116 * (0.0050)		
T		0.0066 *** (0.0015)	0.0070 *** (0.0004)	
$ \eta $		0.0025 (0.2000)		
Success			-0.0424 *** (0.0073)	
success ²			0.0020 *** (0.0002)	
Classical			0.1057 * (0.0480)	
Rap			0.0191 ** (0.0066)	
Constant	0.2049 *** (0.0051)	0.1754 *** (0.0077)	0.4316 *** (0.0548)	
R^2	0.0913	0.0933	0.03	

Standard errors in parentheses. * p<0.05 ** p<0.01 *** p<0.001

Model 3 examines the effects of exogenous variables. In both communities, the two main exogenous sources of stability – artist success (measured as play-count) and share volatility at time t – have a curvilinear relationship with stability (see Figure 4.5). In the MOC, as expected, highly successful artists tend to have stable identities. But so do very marginal artists (those who draw a few thousand listeners at most). Presumably, because of their limited reach and recognizability, these artists are only known to a handful of enthusiasts who share a similar understanding of how they are categorically defined. Apparently such artists (who constitute less than 1 percent of the population of artists included in user collections) rarely diffuse beyond their core fan base.

Table 4.2
OLS regression estimating semantic stability η between times t and $t+1$ for FOC asset i

Variable	Endogenous		Exogenous	
	Model 1	Model 2	Model 3	
$d_{i,t}^A$	0.0904 *** (0.0050)	0.0893 *** (0.0049)		
$\Delta d_{i,t}^A$	0.0624 *** (0.0078)	0.0590 *** (0.0077)		
ΔT	0.0014 * (0.0006)	-0.0064 *** (0.0014)		
$\min(t)$	-0.0802 *** (0.0076)	-0.0324 ** (0.0106)		
T		0.0110 *** (0.0015)	0.0027 *** (0.0006)	
$ \eta $		0.4415 *** (0.0540)		
$Vol^S(i)$			0.0319 *** (0.0089)	
$Vol^S(i)^2$			-0.0074 *** (0.0017)	
$ROR(i)$			-0.0112 (0.0068)	
Constant	0.3426 *** (0.0112)	0.0955 ** (0.0279)	0.3923 *** (0.0127)	
R^2	0.0944	0.1169	0.0116	

Standard errors in parentheses. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

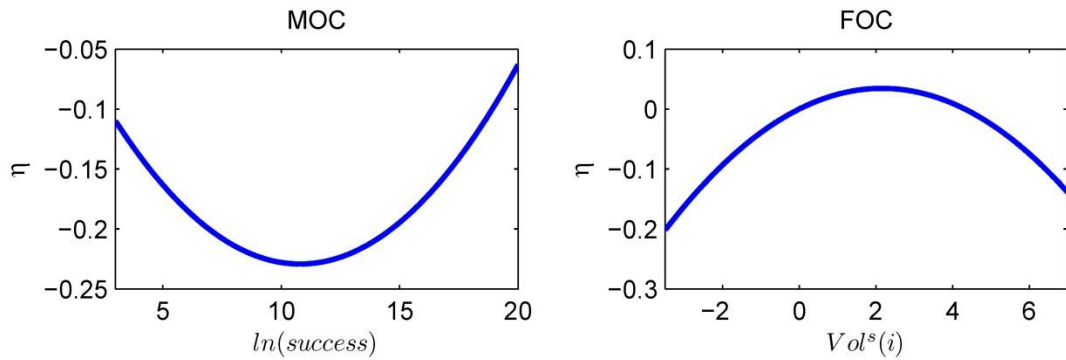


Figure 4.5: OLS estimations of the effects of artist success (left) and share volatility (right) on semantic stability

Volatility is similarly curvilinearly related with stability. As expected, highly volatile assets tend to exhibit low stability. For example, Visa Inc., the global credit card corporation, was traded with an average volatility several times greater than the S&P 500's. It's stability fluctuated accordingly, bottoming at -0.085. But assets with significantly low volatility also tend to suffer from low stability.

These assets appear to be those most related to the financial crisis that was unfolding in the second half of 2008. Funds related to Lehman Brothers, the investment bank whose collapse ignited the crisis, were amongst these highly nonvolatile assets. While this low volatility was presumably related to investors' uncompromising attempts to get rid of assets suddenly rebranded as 'toxic' – and consequently keeping their price consistently low – this toxicity appears to have also had a bearing on investors' ability to make sense of these assets; their meanings were, accordingly, highly unstable.

Two additional exogenous variable types are included in model 3. In the MOC, a series of genre dummy variables were added to the regression model. Of these, only classical music and rap are significantly more stable than other genres.⁶³ These two genres appear to have relatively distinctive identities in a community mostly focused on rock and electronic music. Overall, however, it seems that standard genre labels play a very minor role in structuring MOC users' understandings of music.⁶⁴ Similarly, an asset's ROR has an insignificant effect on its subsequent semantic stability. It appears that short-term market success has little to do with how users perceive different financial assets.

Table 4.3 reports the results of additional OLS models that include our two major variables of interest: ambiguity and boundedness. Model 4 examines the relationship between object ambiguity and its subsequent stability in both communities. The results strongly support hypothesis 4A: in both communities, as ambiguity increases, stability decreases. Figure 4.6 plots estimated object stability between times t and $t+1$ as a function of its ambiguity at time t . As it illustrates, the negative relationship between stability and ambiguity is particularly pronounced in the FOC: an increase along

⁶³ The omitted genre is 'Pop/Rock' which is the dominant genre in this community.

⁶⁴ To underscore this point, I examined the relationship between artists' genre identities and the tags with which users identified the mixes they appeared in. I listed the tags with which each genre was associated with by frequency. The top three most frequent tags associated with each genre included at least one of the following: 'rock', 'pop' or 'electronic'. In other words, regardless of artists' formal genre classifications, most users appear to perceive them as producing either rock, pop or electronic music.

the full range of ambiguity translates to a decrease of more than 0.4 in stability. As expected, ambiguous objects are more vulnerable to reinterpretation.

Table 4.3
OLS regression estimating semantic stability η between times t and $t+1$ for object i

Variable	Model 4				Model 5			
	MOC		FOC		MOC		FOC	
$\ln(a(i))$	-0.0682	*	-0.1852	***	-0.0603	*	-0.1710	***
	(0.0303)		(0.0216)		(0.0298)		(0.0215)	
$b(i, t)$					0.1408	***	0.2183	***
					(0.0138)		(0.0273)	
$\Delta b(i, t)$					-0.1164	***	-0.2798	***
					(0.0203)		(0.0454)	
$b(i, t) \cdot \Delta b(i, t - 1)$					0.0633	*	0.2439	***
					(0.0267)		(0.0587)	
$d_{i,t}^A$	0.0755	***	0.1016	***	0.0903	***	0.1252	***
	(0.0025)		(0.0051)		(0.0028)		(0.0060)	
$\Delta d_{i,t}^A$	0.0273	***	0.0582	***	0.0292	***	0.0594	***
	(0.0028)		(0.0076)		(0.0028)		(0.0076)	
ΔT	-0.0043	***	-0.0062	***	-0.0046	***	-0.0050	***
	(0.0012)		(0.0014)		(0.0012)		(0.0014)	
$\min(t)$	-0.0066		-0.0285	**	-0.0051		-0.0260	*
	(0.0051)		(0.0105)		(0.0051)		(0.0104)	
T	0.0063	***	0.0109	***	0.0057	**	0.0104	***
	(0.0015)		(0.0015)		(0.0013)		(0.0015)	
Constant	0.2474	***	0.0140		0.0728		-0.1247	***
	(0.0201)		(0.0297)		(0.0546)		(0.0348)	
R^2	0.0999		0.1373		0.1079		0.1531	

Standard errors in parentheses. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Additional community-specific control variables are included in the model but not reported.

Model 5 investigates the effects of boundedness on semantic stability. According to hypothesis 4B, a change in an object's boundedness should translate into semantic instability. Model 5 therefore adds boundedness delta, $\Delta b(i, t) = b(i, t) - b(i, t - 1)$, as an independent variable. If the boundedness of an object affects how it is perceived by actors, then we should expect that change in either direction – whether an increase or decrease in Δb – would result in lower stability. But the effects of boundedness delta on stability should not be straightforward. Bounded objects have, presumably, a more coherent categorical identity than weakly bounded objects. Thus, the greater the eventual boundedness at time t , the higher we would expect stability at time $t+1$, irrespective of the

change in boundedness that preceded it. I therefore add boundedness at time t as an additional variable. Moreover, the effect of Δb on stability depends on boundedness at time $t-1$: when boundedness is initially low, an increase should translate to instability; but when it is high, an increase should only contribute to stability. To account for this, I add an interaction term between boundedness at time $t-1$ and Δb .

The estimated relationship between boundedness at times $t-1$ and t , and between stability between times t and $t+1$, is plotted in Figure 4.6. The results support hypothesis 4B and are again more pronounced in the FOC. When an object diffuses to a diverse set of social groups – that is, when its boundedness drops from an initially high value – then its subsequent stability also drops; its adoption by members of new social groups translates to a change in its interpretation. At the same time, when an object's boundedness increases from an initial low – that is, when the diversity of the actors consuming it decreases – its stability remains low, because its identity is changing. The most semantically stable objects are those whose initial and subsequent boundedness are high.

How should we interpret this finding? One possibility is that the relationship between boundedness and stability is an artifact of diffusion dynamics. Consider again figure 4.3. Imagine that black colored nodes correspond to users including object A in their collections at time $t-1$, and that blue colored nodes correspond to users who included the same object at time t . Imagine further that all users in the left cluster include object B in their collection, and that all users in the right cluster include object C . Merely by diffusing to the right cluster, object A will no longer exclusively co-occur with object B , and its semantic stability will decrease. In order to address this possibility, I include ambiguity at time t as a control variable in model 5. Object A 's ambiguity in our hypothetical example would increase at time t after diffusing to the right cluster, as it would now be co-occurring with categorically dissimilar objects. Adding ambiguity as a control variable accounts for such a decrease in semantic stability. The results

reported in table 4C therefore correspond to the effects of boundedness on stability that are not the result of an object’s ambiguity. Rather, they relate to changes in stability that occur because users are reinterpreting an object’s meaning.

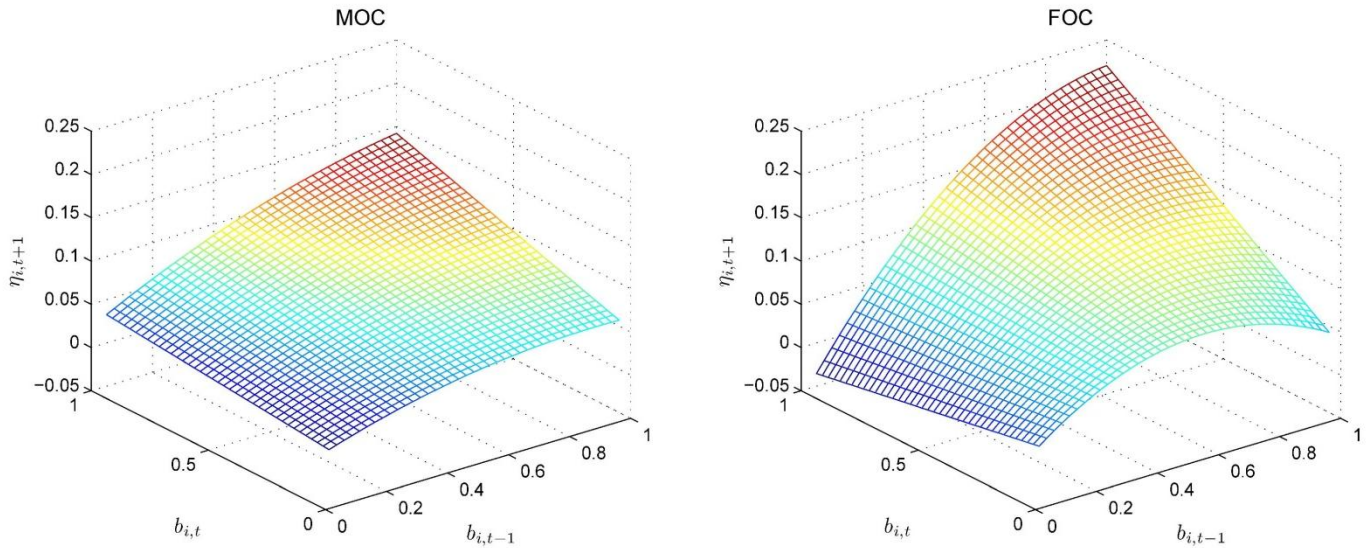


Figure 4.6: Estimated semantic stability between time t and $t+1$ (η), as a function of boundedness (b) at times $t-1$ and t , for both communities. Colors correspond to η , ranging from low (blue) to high (burgundy), and are provided for ease of interpretation.

4.5 Discussion

On September 14th 2008, one day before Lehman Brothers’ cataclysmic collapse which catalyzed a global financial crisis, Standard & Poor’s continued to rate the investment bank an ‘A,’ meaning that it considered it to have a “strong capacity to meet financial commitments”.⁶⁵ Retrospectively defending its position, the world’s largest and most influential rating agency argued that financial indicators suggested, up until the day before filing for bankruptcy, that the bank was capable of generating profit and withstanding liquidity pressures. “Yet,” Standard and Poor’s briefed in a report nine days later,

⁶⁵ Standard & Poor’s credit rating definition, retrieved from <http://www.standardandpoors.com/ratings/definitions-and-faqs/en/us>

“there are cases where negative market sentiment – whether or not grounded in fundamentals – can create significant difficulties for a company ... We believe the downfall of Lehman reflected escalating fears that led to a loss of confidence – ultimately becoming a real threat to Lehman's viability” (Sprinzen and Azarchs 2008). In hindsight, analysts pointed to systemic flaws in mechanisms for pricing and evaluating risk in mortgage-backed securities and credit default swap markets that served as the foundations upon which a bubble, and its unavoidable collapse, were building. At the time, however, only few market actors considered firms partaking in such activities to be at high risk of insolvency.

Stock market crashes provide exceptionally fast and violent examples of collective reinterpretation. In the weeks preceding and following Lehman's collapse, financial assets that were hitherto considered safe and profitable investments were suddenly redefined as 'toxic'. A change in their collective interpretation was precipitated when actors' behaviors in the aggregate began signaling to others the emergence of disagreement about these assets' riskiness. Such reinterpretation processes do not occur exclusively in markets; yet in other domains they are rarely as dramatic. Douglas Sirk's melodramatic films, for example, achieved great commercial success when released in the 1950s, but were very poorly received by his contemporaneous reviewers. Two decades later, a new generation of critics began reconsidering Sirk as a quintessential example of an auteur: a director whose films are characterized by a unified artistic vision despite the homogenizing pressures of the motion picture production process (Phillips 1991). By the 1990s his work had been entirely reinterpreted as a staple of artistic ingenuity, celebrated in homages by acclaimed art-house directors such as Quentin Tarantino and Pedro Almodóvar. Though his films remained exactly the same as they were when produced four decades earlier, their meaning was fundamentally transformed.

Sociological theory has, traditionally, focused on the durability of social structures; social change arising from within has consequently been relatively under-theorized (Sewell 1991). Indeed, the

literature on boundaries has been mostly concerned with explaining how symbolic and social boundaries reproduce one another. Similarly, though with increasingly more frequent exceptions, organizational literature has primarily conceived of inter-subjective schemas as a fixed categorical landscape in which organizations operate. The results presented in this chapter also point to the durability of such structures: in both the MOC and FOC, as time progresses, objects' overall semantic stability increases. But beneath this gradual trend lies a great deal of flux. When actors look at one another in order to make sense of the objects they are assessing, social dynamics can give rise to changes in their collective interpretations. In this chapter I demonstrated how change arises from within through ambiguity and boundedness.

Ambiguity arises from the interaction between an object's attributes, and an audience's perceptual prism. It is not a property exclusively inherent to the object; consequently, it can change. The results presented in this chapter demonstrate that when there is disagreement between actors on how an object is classified, this object is subsequently likely to change its collective meaning. In theory, this should not have necessarily been the case. Multivocal objects can mean different things to different people (Griswold 1987). From the point of view of the individual observer, such an object may appear to have a very coherent meaning. Ambiguity at the aggregate level should not translate into collective perceptual change as long as individuals do not find an object to have an incoherent meaning. But when actors observe each other's behaviors, they too become aware of an object's ambiguity. As the results reported in table 4C suggest, under such circumstances they are increasingly likely to assign new meanings to it.

An object's meaning is also related to the social identities of the individual actors consuming it. This identity is constitutively linked to how these actors' are located in social space: organizations positioned in intersections connecting disparate parts of a network, for example, also enjoy a

multiplicity of social identities (Padgett and Ansell 1993; Stark 1996; Powell et al 2005). This chapter demonstrates that the boundedness of an object – the extent of diversity in the identities of the individual actors who consume it – is related to its meaning. Reinterpretation occurs when the social composition of an object's consumer-base leads to change in its boundedness: when the object's boundedness decreases, its meaning destabilizes, and as it is consumed by an increasingly homogenous group of actors, its meaning stabilizes. This effect is significant in both communities above and beyond the effects of simple diffusion when an object's popularity increases, or of ambiguity in its overall perception.

Both in art and in financial markets, an object's underlying quality is very difficult to discern. In fact, what constitutes quality is often highly debatable (Yogev 2010). Under such circumstances, actors look at one another in order to infer quality. The results presented in this chapter shed new light on this process in two ways. First, because meaning is relational, actors do not merely examine objects independently. Rather, they look at others actors' behaviors in order to generate meanings from the patterns of co-occurrence that are manifest in such behaviors. Thus, when actors exhibit contradictory behaviors, ambiguity emerges. Such ambiguity leads actors to rethink objects' meanings. Second, actors are sensitive to the social identities of others actors. Consequently, these identities confer meanings to the objects people consume. Various studies have demonstrated that, especially with cultural consumption, people consciously use objects as vehicles for explicit identity-signaling (Berger and Heath 2008). These results point to the more subtle, implicit and presumably rudimentary ways through which group membership generates meaning.

Chapter 5

Performance, Status, and Non-conformity

5.1 Introduction: Status and Performance-Measurability in Markets

The existence of categorical deviance in markets – and elsewhere – poses a puzzle. Though popular culture often celebrates nonconformists as geniuses, in reality, there exist very strong pressures on actors to conform to existing behavioral norms (Meyer and Rowan 1977; DiMaggio and Powell 1983). Consider Michael Jackson, the King of Pop whose innovative fusion of rhythm & blues, soul and dance transformed pop music, making him the most successful entertainer of all times. Or Steve Jobs, the legendary Apple CEO and a pioneer of the computer and communications industries, who is widely considered today a technology visionary on a par with giants like Thomas Alva Edison. But for every posthumously glorified maverick, there are thousands of unrecognized novices struggling for survival. And as they painfully learn, behaviors that are at odds with how things are conventionally done, are almost always greeted with strong disapproval (Hargadon and Douglas 2001).

Why, then, are certain actors nevertheless motivated to take on unconventional behaviors? One strand of sociological literature ties an actor's proclivity to defy existing norms with his or her social position. Phillips and Zuckerman (2001) argue that conformity is especially strong amongst the mid-ranks of a status hierarchy. High-status actors feel sufficiently confident that their position will not be compromised by forays into uncharted terrains, whereas low-status actors have nothing to lose by

doing the same. Mid-status actors, caught between a rock and a hard place, have much to lose from nonconformity, and little latitude to explore it.

Central to this argument is the idea that unconventional behavior is interpreted differently by observers as a function of an actor's status. Indeed empirical studies find that in the semiconductor industry (Podolny and Stuart 1995) and the French haute-cuisine restaurant business (Rao et al 2005), when high-status actors defy existing categorical conventions their behaviors are more likely to be interpreted as innovation, and to be rewarded accordingly.

Absent from this explanation, however, is the question of performance: an actor's proven capacity to achieve a desirable goal. Performance and status are often correlated with one another; yet analytically, they are distinct. Whereas performance is inferred from past behavior, status is inferred from one's position in a set of social relationships. In some settings, performance is impossible to observe objectively, whereas in others distinguishing actors by performance is a relatively straightforward task. Ranking surgeons, for example, though not without its challenges, is easier than ranking novelists.⁶⁶

In this chapter I argue that the relationship between status and one's freedom to deviate from conventions depends on performance-measurability: the availability of a consensual and institutionalized metric for performance. In markets where performance-measurability is high, status becomes less consequential for how one's degree of conformity is evaluated. In settings where performance is difficult to appraise, status becomes an influential evaluative prism.

The two online communities investigated here provide an ideal setting for exploring this hypothesis. Whereas there is no consensual metric for appraising DJ performance, investors' strategies

⁶⁶ Presumably, this is why literary awards such as the Nobel Prize in Literature or the Man Booker Prize play more prominent roles in literary markets than best-surgeon prizes play in medical markets.

are conventionally evaluated in terms of their rate of return. And indeed return is built into the FOC website's architecture such that it serves as a salient and readily available indicator of user performance. In the empirical context of these two social network websites, nonconformity is conceptualized as the extent to which an actor's collection is categorically incoherent, meaning that it comprises uncommon object combinations. The longitudinal nature of the data at hand allows for a research design that tests for the impact of performance, status and nonconformity on subsequent success.

The remaining text is divided into four sections. In the first, I provide the theoretical motivation for this study, and offer several hypotheses. In the following section I define the main analytical constructs used in the analysis – success, coherence, status and performance – and explain how they are measured. The analytical strategy and the results are presented next, and are followed by concluding remarks in which I reintegrate the findings into a theory of performance-measurability and its implications for innovation.

5.2 Theory

5.2.1 The Contradictory Consequences of Incoherence

In practically every social context – though to varying degrees – the costs associated with nonconformist behavior serve as powerful constraints on individual actors. In organizational (Meyer and Rowan 1977; DiMaggio and Powell 1983, Ruef and Patterson 2009); financial (Lounsbury and Rao 2004) and artistic fields (Lena and Peterson 2008), category systems provide strong prescriptive principles for individual behaviors. Inter-subjective agreements about the labels that are attributed to actors, and the characteristics that warrant such labels, are therefore sources of order in social systems.

Indeed a variety of studies has consistently found that when actors defy these agreements they tend to suffer detrimental consequences.

This is because – as discussed in the previous chapter – when actors fail to conform to category systems they risk being misunderstood by others. The complexity of information required to make inferences about an actor's quality makes it necessary for others to rely on readily available and easily decipherable signals. When actors emit unconventional signals, they provide incoherent information about their identity. Consequently, other actors are more likely to ignore or, at worst, discount them. The negative impact of categorical noncompliance is particularly potent when activity is mediated through critics and gatekeepers. In an influential study, Zuckerman (1999) finds that firms that fail to garner the attention of reviewers specializing in their industry perform badly in the stock market.

Additional studies have similarly found that organizations or products that are not easily classifiable tend to be ignored and ultimately devalued by audiences. Feature films that do not conform to genre conventions (Hsu 2006), or products on the online bidding website eBay that are classified under multiple categories (Hsu et al 2009), suffer penalties due to their muddled identities. Whether in financial or artistic markets, atypicality makes it more difficult for observers to make sense of an object. Under such circumstances, we would expect the strong disincentives for ambiguity to weed out actors with incoherent identities.

But other studies find that ambiguity can, at times, be beneficial. Rather than seeing it as an obstacle that generates unintelligibility, these studies understand ambiguity as a manifestation of multivocality: the quality of having multiple meanings. Actors who have a singular coherent identity are easily classifiable, but have very limited room for maneuverability. Those who straddle different identities can more easily adapt. Padgett and Ansell's (1993) account of the rise the Medici in Renaissance Florence provides a classic example of the advantages of multivocality. Cosimo de Medici's

multivalent identity enabled him to take on different roles and provided him with the flexibility of 'robust action' that ultimately positioned him in an advantage vis-à-vis other Florentine actors.

How can this apparent contradiction between the advantages and disadvantages of ambiguity be reconciled? In market contexts, actors with multivocal identities are those who choose a generalist approach over specialization. Organizational ecology theory (Hannan and Freeman 1977) suggests that variation in market conditions can explain divergences in the consequences of ambiguity: generalism should be rewarded under conditions of volatility and flux; in contrast, when markets are stable, those with coherent identities will be more successful at targeting niche audiences (Ruef and Patterson 2009). Indeed Stark (1996) finds that in post-socialist Hungary, market actors responded to uncertainty by diversifying their assets. Powell and colleagues (2005) similarly find that in the emergent field of biotechnology, as new types of actors joined the nascent industry, organizations that adopted a 'multiconnectivity logic' benefited from an ability to take on multiple roles and interact with different actors.

In contrast, examining the US film industry, Zuckerman et al (2003) suggest that the varying benefits of generalism and specialization are attributable to difference in individual actors' career histories, rather than the labor market as a whole. They find that new entrants benefit from a genre-specialized identity as it allows them to be noticed by film producers and directors. Over time, however, such specialization begets typecasting which ultimately limits actors' ability to branch beyond their by-now solidified identity, setting a perimeter around their market potential. The film industry is not unique in this regard: similar pressures push people to specialize in academia (Leahey 2007) or information technology contracting markets (O'Mahony and Bechky 2006). In Hollywood, only few actors manage to escape this path and develop a truly diverse professional portfolio which they can successfully leverage. This variability cannot be explained by variation in market environment as all actors are operating under

roughly similar environmental conditions. If generalism thrives under conditions of volatility, we would expect to find all agents pursuing a broad skill-set under the same market conditions to succeed, or fail, more or less equally. But in the film industry, only a handful of actors are rewarded for cultivating a multifaceted cinematic persona. Why is that the case?

Zuckerman et al argue that only a select few who have reached high acclaim can afford deviating from role-specific behavior such that their deviation is interpreted by audiences as broadening, rather than sully, their identity. More broadly, this suggests a tradeoff between one's categorical identity and other sources of information available to others about that actor's quality. Actors with established reputations have more room for maneuverability, but those who are still struggling for recognition cannot afford appearing incoherent. In other words, the effects of ambiguity depend on how an actor is otherwise perceived by others (Podolny 1993, 2001, 2005; Gould 2002; Lynn et al 2009).

When low-reputation actors send ambiguous signals about their identity, these signals are interpreted as negative indications about their quality. But when established actors similarly follow unconventional behaviors, their seal of quality – inferred from their prestige – allows ambiguity to be interpreted as innovation and entrepreneurship. In scientific communities, for example, new theories and innovative research are more likely to occur on the margins of scientific respectability, and as a result are more often than not dismissed. If originating from or eventually championed by acclaimed scientists, their likelihood of diffusing and becoming legitimated increases substantially (Kuhn 1962; Frickel and Gross 2005; Evans and Foster 2011).

This is what Rao and colleagues (2005) argue happened in the field of French gastronomy in the final decades of the twentieth century. As the new category of Nouvelle Cuisine began challenging classical French cuisine in the late nineteen seventies, chefs who integrated classical and Nouvelle Cuisine elements on their menus were penalized by critics. The practice of gastronomical border-

crossing ultimately achieved legitimacy when highly acclaimed chefs began adopting it. The authors argue that these chefs' high status allowed them not only to escape critics' disapproval but to be seen as creative and original. In fact, if reputable actors are relatively immune to the detrimental effects of ambiguity, they have more incentives to cross categorical borders: behaving entrepreneurially differentiates them from others, while their acclaim mitigates the risk of delegitimization (Phillips and Zuckerman 2001). Indeed Kremp (2010) finds that established classical orchestras in the US in the late-nineteenth and early- to mid-twentieth centuries were more likely to include unknown composers in their repertoires. Their prestige mitigated the sanctions associated with deviating from the classical canon.

It is important to distinguish between two sources of information affecting how an actor is perceived by others: (1) performance-based information, and (2) status-based information. Performance-based information relates to expectations about an actor's behavior based on some sort of previously demonstrated capacity (Podolny 2005). If actors are evaluated by others based on a specific outcome of interest, then past ability to produce that outcome serves as indication about future ability. But there exists great variability in the availability of consensually agreed-upon tools for appraising performance across different contexts. In financial markets, for example, investment banks' performance is conventionally evaluated based on previous returns. While these evaluation criteria are far from perfect and are historically and socially contingent (Malkiel 2003; MacKenzie 2011; Millo and MacKenzie 2003), they are nevertheless more consensual than in the music market where measures such as commercial success and critical recognition are used as indications about quality and are often at odds with one another.

When quality is difficult to discern using past performance, we should expect status-based information to become increasingly important in shaping how an actor is perceived. In this context,

status refers to an actor's location in a social hierarchy that is not necessarily related to that actor's past performance. Podolny (1993, 2001, 2005) persuasively argues that an actor's status is inferred by others from the patterns of affiliations and relationships this actor engages in. A musician may be judged not solely on the basis of her commercial success or Grammy award count, for example, but also as a function of the identities of other musicians she has collaborated with in the past. In the remainder of this chapter I use status as shorthand for an actor's relative position in a network of relationships with other actors. Status-based information should be consequential for the effects of categorical ambiguity when there is little or no consensus on performance evaluation criteria. Under such conditions, high-status actors should benefit from ambiguity, whereas low-actors should be increasingly penalized by it. In contrast, when there is agreement on how performance is evaluated, status should have a marginal influence on the effects of ambiguity on success.

5.2.2 Status, Performance and Coherence in Online Music and Finance

How do the pressures of creativity and specialization manifest themselves in our two empirical settings? In both cases, the actors in question are consumers. Unlike the actors in the majority of the studies cited above, users in the two online communities do not produce services or goods. Their performance is therefore not rewarded monetarily; rather, the only measure of success in a social network site is the amount of attention that a user draws from other users. While the motivations for joining each community are different – as each is substantively focused on different types of activity – in both users join a group of people, rather than partaking in the activity in isolation, in order to share and receive information from others. One's overall popularity – defined as the number of followers a user has – functions as a tangible, visible and comparable measure of success in the community. The level of increase (or decrease) in the size of one's group of followers between two consecutive time periods – this user's *popularity delta* – is therefore a measure of that user's success over time.

In order to succeed, users need to be easily noticeable. Faced with the option of choosing from thousands of different profiles, users are presumably more inclined to follow other users whose collection compositions make sense to them. When browsing another user's profile, they need to determine whether that user is worthy of being followed without having to invest too much time or energy. And the cost of searching for new information decreases the more the information fits recognizable patterns.

Imagine an MOC user noticing a mix with the enigmatic title 'music that makes you feel good'. The ambiguous title makes it difficult for her to set clear expectations about what this mix might include. She is therefore less likely to sample it. But she nevertheless decides to give it try. Deterred by what she finds to be a hodgepodge including mellow chillout electronic music as well as Reggae and Ska, she decides to sample mixes created by other users. Indeed amongst the most popular mixes on the MOC are compilations with clear genre identities, such as those titled 'Indie Remixes', 'Classic Rock', or 'Classical Masterpieces'.⁶⁷ Users' comments provide a glimpse into the prisms through which they evaluate one another. Overall, the majority of comments posted in response to mixes are laudatory. But when users post a critical review of a mix, it is usually because they find its tracks to be incompatible with one another. "This doesn't seem to work well for some songs, I don't know why", writes one user on a mix titled "Chris's cool colour compilation" and which includes songs by British folk-rock band Mumford & Sons, alternative rock band Death Cab for Cutie and classic 1960s rock guitarist Jimi Hendrix. "Nice mix man", writes another user on a different mix, "but the MGMT doesn't go here". Unhappy with the psychedelic rock duo's fit into a mix otherwise dominated by experimental electronic music, this comment suggests replacing MGMT with more hard-core electronic bands. It appears that mixes with incoherent combinations of music draw disproportionate negative attention.

⁶⁷ Mix popularity is inferred by the number of users who have listened to it in full.

Search costs are similarly high in the FOC when investment portfolios appear to be unconventional. When browsing others' profiles, FOC users are displayed with a snapshot of these users' portfolio compositions. This composition breakdown provides an overview of its user's investment strategy. Combinations of stocks that do not correspond to an easily identifiable pattern are challenging for observers to categorize, and consequently comprehend. Some profiles include statements that elaborate on the user's investment strategy. But even in the presence of such statements, unconventional strategies are more difficult for potential followers to evaluate. A portfolio that does not correspond to a recognizable investment approach requires an observer to probe its merits more thoroughly; because studying a novel approach is cognitively taxing, observers are more likely to disregard such portfolios (Smith 2011).

Indeed amongst the most popular FOC users are investors whose strategies are conventional. Some investors' strategies – as self-described on their profiles – are based on established investment philosophies such as technical analysis or value investment.⁶⁸ Other investors state that they focus particularly on specific goods or industrial sectors. One such investor, for example, describes himself as an “unabashed gold bug”⁶⁹, whereas another declares that “a sustainable world is my game!” describing his target investments as “companies that deal in solar, wind and clean alternative energy”. Echoing Zuckerman's (1999) findings that market analysts and critics, and the classification systems they use, exert substantial influence on investors, one of the FOC's consistently most popular users writes about himself that he:

⁶⁸ Value investment is a popular strategy based on the identification of under-valued securities through fundamental analysis, or the analysis of firm's performance, financial health and market position.

⁶⁹ 'Gold bug' is a mostly pejorative term that refers to investors aggressively pursuing investment in commodity gold.

“looks for consensus among top research firms to validate selected trades. Patient, yet quick to reverse trades if position becomes wrong to hold. Top sectors for the current environment: Energy, Biotech, Infrastructure and financial.”

In both communities, the high cost of evaluating idiosyncratic object collections makes it less likely for such collections to draw other users' positive attention. Hence:

Hypothesis 5A: *The greater the coherence of an actor's object collection at time t , the more that actor should gain popularity at time $t+1$*

But whereas actors need to conform in order to be easily evaluated, they also seek to stand out in order to be noticed. And in the absence of analysts and critics of the kind Zuckerman (1999) and Hsu (2006) describe, there are no institutionally mandated gatekeepers whose function it is to explicitly penalize deviant behavior in the two online communities. There are two reasons why we should expect there to be incentives for MOC and FOC users to be entrepreneurial in their collection compositions. First, in both settings there exist intrinsic motivations to innovate. The MOC's stated purpose is to provide people with a platform to explore new music. And users often praise mixes they perceive to be original. ““Wow! What a strange collection of quirky gems!” writes one MOC user, “Glad to have heard this one - Très cool.” Both in art and in business, innovation is popularly celebrated as a fundamental characteristic of success: consider, again, the posthumous glorification of Michael Jackson and Steve Jobs as examples of the prominent role that innovation plays in the popular imagination of artistic and business ingenuity.

In financial markets, actors are fundamentally motivated by a desire to outsmart existing investment strategies. “I apply a common sense and moral attitude to investing” writes one of the FOC's most popular investors. He continues to spell out his investment philosophy:

"I haven't been convinced that rapid trading can beat a 'Buy & Hold' approach to the market and manage my portfolio keeping long term results in mind. I believe 'Brand' is the trickiest item to evaluate as it is, by definition, an asset not yet on the books. Because 'Brand Evaluation' is not a technical, but a philosophical process, I think truly strong brand names are usually undervalued - my job is to find those brands."

Being innovative in both settings translates to collections of conventionally unrelated objects.

Second, in both domains there are established norms that favor categorical mixing. The imperative of diversification championed by portfolio theory is a commonly employed strategy to reduce risk which, in the long run, is argued to produce overall maximal returns (Markowitz 1991; Bodie et al 2004). Diversified portfolios are, by definition, composed of non-related assets. While diversified portfolios are safer, they also have lower short term return potential. They do not cater to every investor's needs. But during times of heightened uncertainty – as was the case in the early months of the 2008 financial crisis – many investors seek risk-reducing strategies. Thus for some investors at least, strategies appearing to be following a diversified approach may have been particularly appealing.

Musical diversification, though frowned upon by artistic purists, is similarly a common practice. Inter-genre mixing and cross-pollination are, in fact, emblematic of the process of musical genre emergence and evolution (Peterson and Lena 2008; Lopes 2004). A variety of studies conducted in recent years suggests that a preference for a broad variety of musical genres is strongly associated with high status, especially in the US (Peterson and Kern 1996). Peterson (1992, 1997) has argued that the rise of cultural omnivorousness since the nineteen sixties is no less than a manifestation of a fundamental shift in the intellectual basis upon which societal status markers are established, from snobbism toward a cosmopolitanism that celebrates cultural diversity. The contemporary zeitgeist, he contends, associates omnivorousness with high social status because it symbolizes a political and

cultural openness that has become culturally valued. Diversified MOC collections therefore possess a cachet. Indeed, users with intentionally obscure, often ironic biographical statements rank among the most popular in the community. Though vague, they all appear to be attempting to signal their user's sophistication. Their inherent multivocality seems to be appealing to other MOC users.⁷⁰

In sum, in both the MOC and FOC there are contradicting incentives to conform and deviate from recognized categorical identities. At the same time, the information available to users in assessing each other's quality differs markedly across both communities. This difference should have a bearing on how these contradictory pressures are resolved in both communities. In the MOC, there is no consensual metric for evaluating user performance. In fact, when it comes to art, quality is a particularly evasive and debatable concept (Yogev 2010). Music is no exception. Methods for codifying and evaluating music are historically and socially contingent, even among those professionally specializing in musical production and analysis (Roy and Dowd 2010). What makes one MOC DJ better than the other is therefore an inherently subjective determination.

Consequently, status should become an important resource with which MOC users make inferences about one another. In the absence of a consensual evaluation metric, a user's status serves as a prominent available proxy for her quality. Low-status users vie for attention and recognition, and should therefore benefit from exhibiting easily recognizable musical combinations. High-status users, on the other hand, no longer need to be noticed, and should be less vulnerable to the detrimental effects of mixing music in unconventional ways. Having established their reputations, their musical entrepreneurship is more likely to be interpreted as innovation than as amateurism. Hence, we should expect to find an interaction between collection ambiguity and users status:

⁷⁰ Such statements are often curt and cryptic. Among the most popular ones are statements such as "I'm a writer. Take that as you will", "I have a useless degree and I'm underpaid. I sit with the cool kids at lunch" and "I'm always home. I'm uncool." The high frequency of the term 'cool' – in its various derivations – in users' comments is indicative of the extent to which they value mixes with an aura of fashionability. 2.04 percent of all comments included the word 'cool'.

Hypothesis 5B: *In the MOC, high-status users should gain popularity from exhibiting ambiguous collections, whereas low-status individuals should lose popularity.*

While there are no agreed-upon criteria for distinguishing between ‘good’ and ‘bad’ music, in financial markets there is a widely accepted standard for assessing investor performance: portfolio returns. There are, of course, disagreements about how exactly returns ought to be calculated. In particular, different time intervals may lead to significantly different performance evaluations. Moreover, different investors have different time horizons for capitalizing on their investments. Nevertheless, they all share a similar motivation: maximizing their profits. An investment’s rate of return is the most common and institutionalized evaluation metric used by analysts, finance professionals, and lay investors for quantifying profit (Bodie et al. 2004). And, when browsing each other’s profiles, FOC users are presented with other users’ performance metrics. These include an adjustable graph representation of that user’s returns over a period of one year, as well as a summary of returns and additional statistics (see Figure 5.1 for details). Because they figure so prominently on user homepages, FOC users are primed to treat returns as salient performance-based information.

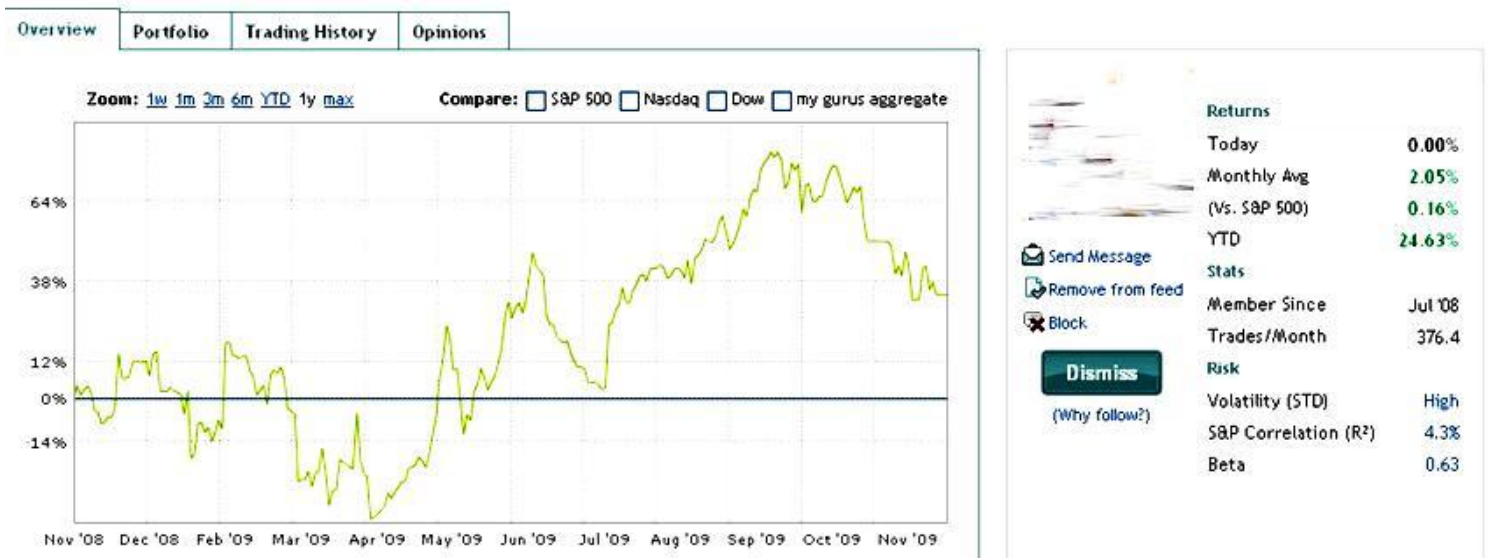


Figure 5.1: Snapshot of a typical FOC user homepage. Identifying user information is obscured.

Given the availability of performance-based information in the FOC, we should expect to find that status has little influence on how users infer each other's quality. This does not mean that status is inconsequential: high-status users are far more visible than low-status users, and we should expect to see a Matthew effect for status such that those with higher than average status levels are more likely to be noticed, and enjoy an increase in popularity. But once noticed, a user's status should not serve as a primary source of information about her quality because a more direct measure, which is widely considered a reliable one, is available. Status should therefore be insignificant in mitigating the effects of categorical ambiguity.

Hypothesis 5C: *In the FOC, high-status users should be as likely as low-status users to gain popularity from exhibiting incoherent collections.*

If status is overshadowed by performance as a quality signal, then we should expect the latter to have a bearing on the effects of ambiguity on success. Users with high-performance are those with proven qualifications. Thus, like high-status users in the MOC, when they exhibit unconventional portfolios these should be received as related to their success, rather than a signal of unreliability. Thus:

Hypothesis 5D: *In the FOC, high-performing users should gain popularity from exhibiting ambiguous collections, whereas low-performing users should lose popularity.*

5.3 Measurements

5.3.1 Success

The dependent variable in the analyses that follow is success. Because there is no monetary compensation for performance in the two social network sites in question, the only visible and

quantifiable reward for activity on the website is one's number of followers. Unlike popular general-purpose SNSs such as Facebook, in our two communities there are no complexities regarding the relationship between online and offline friendships, and the extent to which an online friendship extends to 'real-life' social capital (Ellison et al 2007): in the MOC and FOC, followship relationships are not assumed to relate to offline interaction. One's number of followers is therefore a straightforward measure of popularity. Previous studies demonstrate that in social network sites, one's number of friends is highly consequential for how others evaluate this actor's attractiveness (Tong et al 2008). Because it is perceived to relate to external validation, one's friend count is considered by others as a reliable measure of popularity (Utz 2010).⁷¹ Thus, I use popularity to measure of success.

In network terms, a user's popularity at time t is that user's node in-degree in the interaction network S at time t , or $d_{i,t}^{in} = \sum_j S_{ji,t}$, where $S_{ji,t} = 1$ if user j followed user i at time t or beforehand and has not ceased to follow user i since. Success is defined as user i 's popularity delta, or the change in the number of i 's followers over a given period of time. Because node in-degree has a heavy-tailed distribution, it is transformed by its natural log.⁷² Formally, popularity delta is defined as follows:

$$(5.1) \quad \Delta d_{i,t} = \ln(d_{i,t}^{in} + 1) - \ln(d_{i,t-1}^{in} + 1)$$

Popularity delta is sensitive to fluctuations in network size over time. An increase by 2 users when the overall network size is 100 is far more significant than a similar increase when there are 10,000 users.

Recall, moreover, that users do not require confirmation in order to follow another user. Nor does a relationship between two users imply personal acquaintance. One's overall number of followers is

therefore not restricted by structural or cognitive limits, as is often the case in other SNSs. Indeed, as we

⁷¹ As Tong et al (2008) demonstrate, in certain contexts, a large pool of friends is interpreted as popularity only up to point above which one's overabundance of friends is seen as indication of low desirability. Because followers in the MOC and FOC are not assumed to relate to offline friendships, overabundance does not present itself as a negative signal as it does on Facebook. It is therefore safe to assume a linear relationship between a user's number of followers and others' perceptions about her popularity.

⁷² Because the log function is undefined for values equal or smaller than 0, but some users have no followers and hence an in-degree equaling 0, I add 1 to the node degree.

have seen in Chapter 2, node degrees in network S follow a scale-free power-law distribution in both communities, and scales with network size. In order to correct for changes in network size, I standardize $\Delta d_{i,t}$ by its mean and standard deviation at time t , such that its variance equals 1 per each timeslot. This standardization makes $\Delta d_{i,t}$ comparable across different timeslots.

5.3.2 Coherence

Central to the hypotheses outlined earlier is the question of coherence: the extent to which a user's collection of objects conforms to a single and recognizable categorical identity. A coherent collection of objects is a collection comprised of objects that are similar to one another. For example, a collection of hip-hop music or an investment portfolio focused on technology stocks. Recall that in Chapter 2 we defined categorical similarity as the structural equivalence between two objects in the weighted co-occurrence network $\tilde{W}$ (equation 2.3). Structurally equivalent objects are those who tend to co-occur with more or less the same set of other objects. This similarity suggests that they carry similar meanings, or, more broadly, belong to the same category. The coherence of an object collection is therefore defined as its categorical cohesion⁷³, or the degree of categorical similarity between the objects comprising it.

Formally, the categorical cohesion of a set of objects is defined as the average categorical distance between all objects comprising the set, and the set's geometric center. Let L be a set of objects and $\vec{W}_k$ the vector of edge weights connecting node k to all other nodes in the weighted object co-occurrence network $\tilde{W}$. User i 's coherence, φ_i , or the categorical cohesion of his object collection, is defined as the average correlation between each object's vector of edge weights, $\vec{W}_k$, and the vector of mean edge weights in network $\tilde{W}$ for this collection (which is the set's geometric mean):

⁷³ Categorical cohesion was defined and discussed at greater length in Chapter 3.

$$(5.2) \quad \varphi_i = \frac{\sum_{\forall k \in K} A_{ik} \rho(\vec{\tilde{W}}_k, \vec{\bar{W}}^L)}{\sum_{\forall k \in K} A_{ik}}$$

where $\vec{\bar{W}}^L$ is the mean vector averaged over all $l \in L$.

Figure 5.2 illustrates how coherence is distributed in both communities. While both distributions seem to be roughly normal, FOC collections appear to be, on average, more coherent than those in the MOC. Note that coherence does not necessarily imply homogeneity vis-à-vis established classifications such as musical genres or industrial sectors: a user's investment portfolio is coherent as long as it comprises of securities that are often included together in other FOC users' portfolios, irrespective of how they are classified by the US government; such a portfolio is not necessarily 'coherent' by any objective criteria.

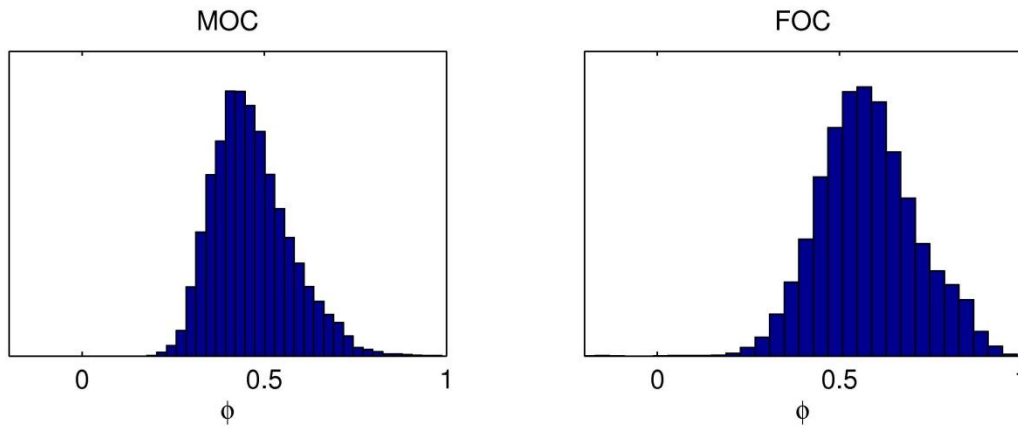


Figure 5.2: Distribution of coherence in both communities for all pairs of user-timeslot

This definition of coherence may seem problematic, particularly with regards to financial investment. Because portfolio diversification is such a common investment strategy, especially among lay investors, one would expect modal FOC users to invest in a variety of assets. If FOC users were diversifying in the way that modern portfolio theory expects them to – spreading risks by investing in a variety of unrelated assets – we should have found weak associations between different stocks. But, as Figure 5.3 demonstrates, this is not the case. Figure 5.3 illustrates a network of the 400 most traded securities throughout the entire period under investigation, such that edges correspond to categorical

similarity between stocks (measured over the entire period).⁷⁴ As is clearly visible, certain stocks are strongly associated with one another. Overall, the network appears to exhibit at least three different clusters of interrelated stocks. If FOC users were purely diversifying their portfolios we would not have seen such a segmented structure emerge.

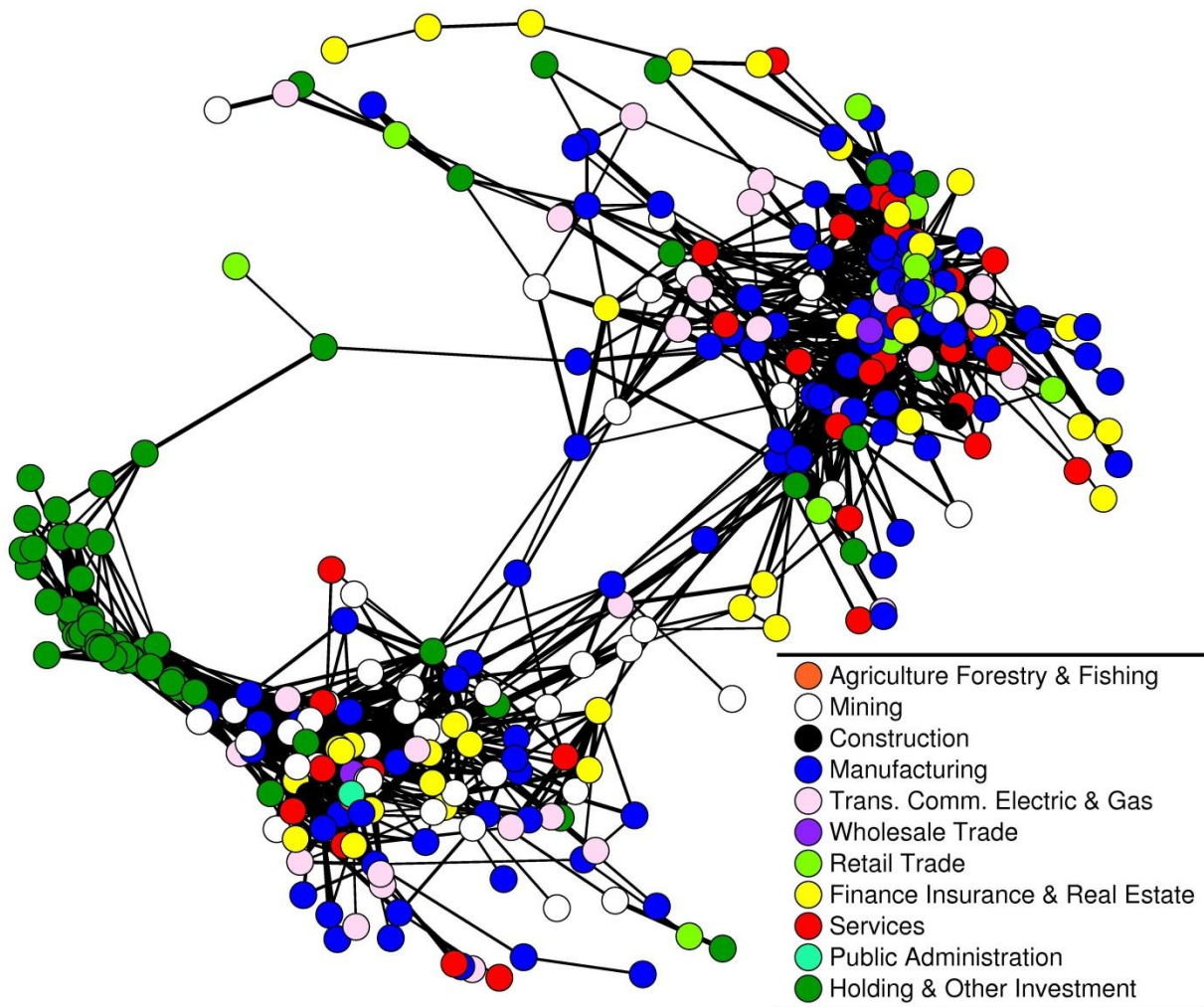


Figure 5.3: Network representation of categorical similarity between the 400 most traded securities in the FOC over the full time period. Nodes are color coded by SIC division. Edges correspond to statistically significant categorical similarity (line width corresponds to the magnitude of categorical similarity).

⁷⁴ Recall that categorical similarity is calculated as the Pearson correlation coefficient between two rows in the standardized co-occurrence network (see section 2.4.2 in Chapter 2 for details). Only significant correlations (for $\alpha = 0.05$) were used to produce this network diagram. For presentation purposes, only the strongest 10% edges are illustrated. The network is spatially laid out using the Fruchtmann-Reingold algorithm.

Node colors in Figure 5.3 correspond to one of ten different divisions by which SIC classifies stocks.⁷⁵ The different clusters are highly heterogeneous with respect to SIC: stocks that are categorically similar to one another are not likely to belong to the same SIC division. In other words, FOC users do not appear to follow SIC's categorical logic: they tend to trade in a variety of stock types, mixing, for example, manufacturing and financial stocks in their portfolios. The exception to this rule is the cluster on the left hand side of the diagram: it is exclusively composed of securities that belong to the 'Finance Insurance and Real Estate' division. In fact the nodes in this cluster are all classified by SIC as 'Holding and Other Investment Offices', a sub-category of the finance division (color-coded dark green). These types of securities include investment trusts and holding companies, but practically of these asset types in the FOC are exchange-traded funds (ETF). This suggests that certain FOC users tend to invest primarily in funds. These individuals appear to adhere to the diversification imperative, putting their eggs in several baskets, following the expert advice of directors in mutual funds.⁷⁶

The remaining FOC users invest directly in stocks. They do not specialize in particular industrial sectors, but this does not mean that they are adhering to the diversification imperative. Rather, the division of the network into clear clusters implies that, overall, these FOC users consider different stocks as making up different implicit categories. Central to the cluster on the bottom of the diagram are heavy industry stocks such as mining companies and steel manufacturers, as well as insurance companies. The cluster on the top comprises a diverse collection of firms ranging from electronics manufacturers and technology companies, to entertainment companies, business service providers, and health care providers. The rationale informing these investment patterns, though interesting in its own right, is

⁷⁵ The division is SIC's highest classification tier.

⁷⁶ These funds tend to be more specialized than generic index funds. Two of the most popular funds in this cluster are Direxion's 'Technology Bear' fund and ProShare's 'Ultra Short' fund.

beyond the focus of this chapter; what is important is that such a rationale seems to exist.⁷⁷ Coherent FOC users are those whose portfolios include securities that belong to either the clusters included in Figure 5.3: they may be diversifiers who invest primarily in mutual funds, or more specialized investors.

Table 5.1 provides a few examples of highly coherent and incoherent collections in both the music and finance communities. While coherent musical collections appear to adhere to conventional genre definitions (in the example in Table 5A, the high coherence collection is comprised exclusively of heavy metal musicians), coherent investment portfolios – as we just saw – exhibit greater industrial sector variation. The highly coherent FOC user presented as an example in Table 5A invests in a variety of stocks that seem consistent with the top cluster in Figure 5.3. Though difficult to label using conventional terms, this portfolio is nevertheless consistent with the overall investment pattern in the FOC. In contrast, the incoherent investor’s portfolio is comprised of technology stocks, aluminum and drug producers, as well as mutual funds. This investor diversifies her portfolio, but in an idiosyncratic way. Unlike conventional high-coherence diversifiers, she does not invest exclusively in mutual funds. FOC investors, even those with very coherent investment strategies, tend to invest in more than one industry sector. But overall they are less inclined than MOC DJs are to include unusual object combinations. The imperative of portfolio diversification notwithstanding, compared to MOC DJs, the average FOC user conforms to the investment strategies espoused by others in the community.

⁷⁷ Because these different clusters defy conventional classifications, they are difficult to label intuitively. It appears that the cluster on the bottom is focused on heavy industry (though why this cluster includes insurance companies is more difficult to make sense of) whereas the top cluster comprises firms that specialize in technologically advanced manufacturing, services and retail. In addition to visually inspecting the network diagram, I used the Newman (2006) modularity-based spectral partitioning algorithm to divide the network into different clusters. The partitioning algorithm divides the upper cluster in to two smaller clusters, one more focused on high-technology firms (such as Yahoo! and Google), the other on service firms, retailers and the entertainment industry.

TABLE 5.1		
Examples of high and low coherence object collections in the MOC and FOC		
MOC		
Artist	Genres	Styles
High coherence:		$\varphi = 0.9145$
Helloween	Pop/Rock	Heavy Metal, Power Metal, Speed Metal, Progressive Metal
DragonForce	Pop/Rock	Power Metal, Neo-Classical Metal, Heavy Metal
Stratovarius	Pop/Rock	Heavy Metal, Progressive Metal, Scandinavian Metal, Power Metal
Falconer	Pop/Rock	Power Metal, Heavy Metal
Firewind	Pop/Rock	Prog-Rock, Heavy Metal, Neo-Prog
Hammerfall	Pop/Rock	Heavy Metal, Scandinavian Metal, Power Metal
Gamma Ray	Pop/Rock	Heavy Metal, Power Metal, Hard Rock, Progressive Metal
Blind Guardian	Pop/Rock	Heavy Metal, Power Metal
Edguy	Pop/Rock	Hard Rock, Heavy Metal, Power Metal, Prog-Rock, Neo-Prog
Iron Savior	Pop/Rock	Heavy Metal, Power Metal
Low coherence:		$\varphi = 0.1779$
Radiohead	Pop/Rock	Alternative Pop/Rock, Britpop, Experimental Rock, Alternative/Indie Rock, Indie Electronic
Orbital	Electronic	Techno, Ambient Techno, Electronica
James Newton Howard	Film Score	
BT	Electronic	Progressive House, Progressive Trance, Nu Breaks, House, Club/Dance, Trance
Hans Zimmer	Film Score	
Hybrid	Electronic	Club/Dance, Trance, Progressive Trance, Nu Breaks
State Radio	Pop/Rock, Reggae	Contemporary Reggae, Alternative Pop/Rock, Alternative/Indie Rock, Jam Bands
<i>Genre and style classifications are based on Allmusic.com</i>		

TABLE 5.1 (continued)
Examples of high and low coherence object collections in the MOC and FOC

FOC				
Symbol	Name	SIC	Division	Industry
High coherence:		$\varphi = 0.9442$		
ATVI	Activision Blizzard, Inc	7370	Services	Business Services
CPRT	Copart, Inc.	5010	Wholesale Trade	Wholesale Trade-durable Goods
NTGR	NETGEAR, Inc.	3661	Manufacturing	Electronic And Other Electrical Equipment And Components
QSII	Quality Systems, Inc.	7372	Services	Business Services
BWA	BorgWarner Inc.	3714	Manufacturing	Transportation Equipment
GME	GameStop Corp.	5734	Retail Trade	Home Furniture, Furnishings, And Equipment Stores
LH	Laboratory Corporation of America Holdings	8071	Services	Health Services
MVL	Marvel Entertainment, Inc.	7829	Services	Motion Pictures
Low coherence:		$\varphi = 0.1161$		
SKF	ProShares UltraShort Financials	6726	Finance, Insurance and Realestate	Investment Offices
AAPL	Apple Inc.	3571	Manufacturing	Computer And Office Equipment
PALM	Palm, Inc.	3570	Manufacturing	Computer And Office Equipment
RMBS	Rambus, Inc.	3670	Manufacturing	Electronic Components And Accessories
WGNB	WGNB Corp.	6021	Finance, Insurance and Realestate	Commercial Banks
AA	Alcoa, Inc.	3334	Manufacturing	Primary Smelting And Refining Of Nonferrous
AOI	Alliance One International, Inc.	5159	Wholesale Trade	Farm-product Raw Materials
HRB	H&R Block Inc.	6159	Finance, Insurance and Realestate	Business Credit Institutions
LMT	Lockheed Martin	3764	Manufacturing	Guided Missiles And Space Vehicles And Parts
MMM	3M Company	3841	Manufacturing	Surgical, Medical, And Dental Instruments And Supplies
PFE	Pfizer Inc.	2834	Manufacturing	Drugs
S	Sprint Nextel Corporation	4813	Transportation, Communications, Electric, Gas, And Sanitary Services	Telephone Communications
TXN	Texas Instruments Inc.	3674	Manufacturing	Electronic Components And Accessories
USB	US Bancorp	6029	Finance, Insurance and Realestate	Commercial Banks
FAZ	Direxion Financial Bear 3x Shares	6726	Finance, Insurance and Realestate	Investment Offices

SIC classification definitions are based on data provided by the US Department of Labor

5.3.3 Status

Drawing on Gould (2002) I define status as prestige accorded to an actor because of an abstract position that actor occupies, rather than an immediately observable behavior. Key to this conceptualization is the insight that whatever the behavior of interest is – whether, as in our case, it is musical ability or investment prowess – high-status actors are assumed to be better at it. And this assumption is sustained by observers' recognition that other observers share a similar evaluation. Hence, this collective evaluation needs to be readily visible in order for it to confer status.

It therefore makes sense thinking of status in network terms: as a property attributable to an actor by virtue of a relationship of recognition or deference with other individuals. Indeed, Podolny (1993, 2005) argues that, at least in markets, an actor's status is contingent on the various relationships in which he engages. The computer-mediated communication afforded by the two online communities under investigation here significantly restricts users' ability to display status-related symbols they may have acquired elsewhere. Status signals embedded in speech, appearance or clothing, for example, are drastically muted when interaction is funneled through restricted website templates. Users' status within each community is therefore primarily a result of endogenous dynamics. And the most prominent and visible outcome of these dynamics, and one that can be associated with recognition by others, occurs when a user is followed by another user. One's number of followers, or network in-degree, is therefore an indicator of one's status. Indeed degree centrality is a common measure for status in network studies (Wasserman and Faust 1994).

In both the MOC and FOC users can easily observe other users' degrees; both websites prominently list a user's set of followers on that user's homepage. Node degree therefore becomes a very strong and visible differentiating symbol of prestige in both communities. But node degree is often

too simple a measure of status. A tie to a high-status individual confers more status than to a low-status individual (Podolny 1993). Bonacich (1972) therefore suggested using eigenvector centrality as a means to account for the cumulative status of a node's ego-network. Given an adjacency matrix A , eigenvector centrality satisfies the following equation:

$$(5.3) \quad \lambda x = Ax$$

such that x is an eigenvector and λ is matrix A 's leading eigenvalue. In the two communities investigated in this study, the hyperlink-based user interface not only clearly quantifies users' node degrees but also enables users to easily navigate from one user profile to another. Thus profiles of users who are followed by high-status users are more likely to be accessed by users browsing the website, and implicitly be associated with high prestige. Eigenvector centrality is therefore a more appropriate measure than mere node degree for inferring status.

But eigenvector centrality's mathematical properties make it inappropriate for directed networks in which certain nodes have zero in-degrees, because such nodes contribute nothing to the status of nodes with which they are connected. Thus, a node with high in-degree might be inferred to have zero status if all the nodes pointing to it have zero in-degrees. To overcome this problem, Bonacich and Lloyd (2001) introduce the alpha centrality measure, which is conceptually very similar to eigenvector centrality. This measure assumes that each node has some degree of external status which is independent of status conferred by network position. Let A denote an $N \times N$ directed adjacency matrix, and e an N -sized vector of external status, then alpha centrality c satisfies the following equation:

$$(5.4) \quad c = \alpha A^T x + e$$

where α is a scalar which reflects the relative importance of endogenous network position in determining status. Following Bonacich and Lloyd, I define $\alpha = 0.85/\lambda$ such that it is smaller than the leading eigenvalue. Because external factors are very limited in their ability to bear on the dynamics of

status emergence in both communities, I set e to a constant $\epsilon < \alpha$, such that all users have the same initial external status, and such that its relative contribution to overall status is small compared to network position.

Alpha centrality provides a reliable measure of relative status within a network, however, because eigenvector values are sensitive to network properties and size, the alpha centrality measure is difficult to compare over time. To overcome this problem I standardize c by its variance and mean for each timeslot t , such that it is expressed in terms of standard deviations. This makes it comparable across different times.

5.3.4 Performance

Whereas status is prestige dissociated from observable behavior, performance is prestige directly resulting from such behavior. With music, how this behavior is interpreted is inherently in the eyes of the beholder. The same musical mix may be greeted with enthusiasm by one observer and disappointment by another. But in the FOC portfolio return serves as a universal yardstick for investment performance.

Return is conventionally calculated as a percentage of profit on an investment, or the amount of capital gained minus the amount of capital invested as proportion of the amount of capital invested, over a given period of time. Formally:

$$(5.5) \quad \text{Return} = \frac{\text{gain} - \text{invested}}{\text{invested}}$$

where *gain* is the value of the investment at the end of the time period plus any capital withdrawn from the investment account over the time period, and *invested* is the sum total of capital put into the account over the time period.

The amount of time used for calculating return is consequential as different investment strategies vary in the timeframes for which they provide optimal returns. FOC users are presented with three different return timeframes: daily, monthly and year-to-date. Of these, preliminary analyses indicated that monthly returns are most strongly associated with popularity delta such that high-performing investors during the month preceding time t were rewarded with a consequent increase in popularity (OLS, $p < 0.001$). These analyses also find that the two variables are curvilinearly related such that the positive relationship between return and performance delta is attenuated at extremely high return rates.⁷⁸ To make return invariant to different times, it is standardized by the corresponding S&P 500 return.

5.4 Analysis

5.4.1 Model

The focus of this analysis is the individual user: how a user's collection composition and social position affect eventual success. The unit of analysis is therefore a user-time pair. The two datasets were restructured such that whenever a user made a change to her a collection, an observation was added to the dataset. This observation includes the time t on which this change was made, this user's collection cohesion at time t , the user's status as a function of her network position at time t , and, for FOC users, the return on their investment for the month preceding and up to time t . The dependent variable, Δd or popularity delta, represents the standardized increase in in-degree in network S during a time window following t . Because user activity follows a weekly circadian rhythm (in the MOC because users tend to be more active over the weekend, and in the FOC because trading is suspended during the weekend), I set the time window to seven days. A week is long enough to allow for a user's activity to

⁷⁸ These results are not reported in full here.

have been noticed by others, and sufficiently short for the attention generated by the activity not to have entirely waned.⁷⁹

In addition to the dependent and independent variables of interest, several control variables are added to the model. Because the dependent variable is popularity delta, popularity at time t was added to the model. Popularity at time t should be consequential for subsequent growth in popularity because popular actors are more visible and therefore more likely to receive the attention of other actors. As discussed in Chapter 2, the interaction network S in both communities is scale-free, with a handful of users reaching extremely high levels of popularity. Such networks are often characterized by preferential attachment such that the likelihood that a node will increase in in-degree is proportional to that node's current in-degree (Barabási and Albert 1999). To account for increases in popularity which are strictly the result of preferential attachment, popularity is added as a control variable. To correct for its heavy-tailed distribution and for network-size variation, it is logarithmically transformed and standardized by time. In both communities, popularity at time t alone explains roughly 15% of variance in subsequent popularity delta.

Four additional characteristics potentially add to users' visibility on both social network sites, and therefore have an effect on the likelihood of an increase in popularity. Whether or not a user has added a biographical (MOC) or philosophical (FOC) statement, irrespective of that statement's content, strongly shapes other users' ability to evaluate her. Without one, others have very little information about that user's profile and are less likely to pay attention to it. Some statements are short and enigmatic, whereas others are more informative. In order to account for statement verbosity I use its length in characters (logarithmically transformed) as a control variable. Whether a user posted an image on their website, and whether they point to an external personal website (a feature available only on

⁷⁹ Roughly similar results were produced when using slightly different time window specifications. Results ceased to be significant when the time window was significantly extended beyond a week.

the MOC), might also increase their visibility. Corresponding dummy variables were therefore added to the models. Finally, the amount of days that has passed since a user had joined the community also has a bearing on users' ability to be noticed. Other things being equal, as time goes by a user's novelty erodes and so does their attractiveness.⁸⁰ The number of days since joining the community (logarithmically transformed) was added as a control variable as a consequence.

Finally, I added the number of objects contained in the user's collection at time t as a control variable. Recall that coherence is calculated as the average categorical distance between the objects in the collection. This average is sensitive to the number of objects in the collection. On the one hand, as the number of objects grows, the likelihood that they are all proximal to one another decreases. On the other hand, the likelihood of high variance in distance decreases. Thus, in order to account for variation in coherence that is an artifact of collection size, I add collection size and its quadratic term as control variables.⁸¹

While user profile attributes such as biographical statement and photo explain some variability in popularity increase (8% in the MOC and 2% in the FOC), much of their effect is the result of unobserved heterogeneity. For example, two users may have both chosen to add photos to their profiles, but these photos may be distinctly different in the emotions they evoke and therefore how observers will interpret them. FOC photos, for example, range from unassuming personal portraits to evocative symbolic images. A tiger image (animal images are common on the FOC), for example, may suggest an aggressive investment strategy. Similarly, biographical statements can produce a variety of expectations. Some statements are very specific about a user's philosophy whereas others appear to be intentionally vague. These individual differences may lead to different effects of collection coherence on success. An MOC user, for example, who explicitly states that she is classical music fan, may be penalized if she

⁸⁰ This is particularly true in the MOC where new users are more prominently displayed by the website.

⁸¹ In analyses not reported here it was found that the number of objects contained in a collection (and transformed logarithmically) has a convex relationship with coherence.

includes non-classical music in her collection. In contrast, a user with a clever but ambiguous statement, and a snazzy profile photo, will likely be expected to produce unconventional music compilations.

The various ways in which these idiosyncratic user characteristics are interpreted by others are, in essence, unobservable differences between users.⁸² In order to account for this variability, I use a multi-level model where user-time observations are grouped by individual user. This model assumes that users differ in their overall level of attractiveness: some users, irrespective of their behaviors, are simply more attractive than others. It also assumes that users differ in the effects of collection coherence on their attractiveness: it is expected of some users to exhibit more experimental collections than others. The model is therefore specified as follows:

$$(5.6) \quad \Delta d_{it} = \beta_0 + \beta_1 \varphi_{it} + \beta_2 s_{it} + \beta_3 s_{it}^2 + \beta_4 \varphi_{it} \cdot s_{it} + v_{i0} + v_{i1} \varphi_{it} + \alpha X + \gamma Y + \varepsilon_{ij}$$

where β are fixed-effect coefficients, v are random-effect individual user coefficients, and α are fixed-effect control variable coefficients. γ are fixed-effect performance coefficients which are not included for MOC users, and which for FOC users equal to:

$$(5.7) \quad \gamma_1 r_{it} + \gamma_2 r_{it}^2 + \gamma_3 \varphi_{it} \cdot r_{it}$$

s_{it} is user i 's status (alpha centrality) at time t , φ_{it} is that user's collection coherence and r_{it} is her monthly return. The random effect terms are not assumed to be independent. Both status and returns include quadratic terms.

5.4.2 Results

Let us first examine how status and performance explain success, irrespective of one's collection coherence. Model 1 in tables 5.2 (MOC) and 5C (FOC) presents the analysis described in equation (5.6)

⁸² Because of the large number of users, it is practically impossible to analyze user photo and statements manually. Automatic analysis techniques such as natural language processing or photo recognition are insufficiently granular for the purposes of this study.

without including cohesion-related parameters.⁸³ In both communities, popularity at time t is a strong predictor of subsequent popularity delta. The effect is more than twice as large in the FOC than it is in the MOC: a one standard deviation increase in popularity at time t for FOC users translates to a subsequent 0.65 standard deviation increase in success. The Matthew effect in both communities is significant and substantial: users are more inclined to follow other users who are already popular.

Table 5.2
Fixed-effect estimates for MOC user success (standardized popularity delta)

Variable	Model 1		Model 2		Model 3	
Popularity	0.3219	***	0.3253	***	0.3273	***
	(0.0434)		(0.0436)		(0.0435)	
Status	0.0422		0.0086		0.2504	**
	(0.0460)		(0.0465)		(0.0832)	
Status ²	0.0585	***	0.0712	***	0.0694	***
	(0.0085)		(0.0088)		(0.0088)	
Statement	0.1248	***	0.1287	***	0.1277	***
	(0.0131)		(0.013)		(0.013)	
Image	0.8291	***	0.8392	***	0.8409	***
	(0.0749)		(0.0749)		(0.0748)	
Website	0.2337	***	0.2147	***	0.2127	***
	(0.0514)		(0.0514)		(0.0513)	
Days active	-0.2222	***	-0.2171	***	-0.2191	***
	(0.0134)		(0.0135)		(0.0135)	
Coherence			0.7219	**	0.9030	***
			(0.221)		(0.2314)	
Status* Coherence					-0.5127	***
					(0.144)	
Constant	0.7452	***	0.1549		0.0637	
	(0.1446)		(0.2228)		(0.2261)	
<i>Wald χ^2</i>	1223.05		1243.38		1268.79	

Standard errors in parentheses. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Popularity's effect on success is, presumably, mostly the result of structural dynamics: users attach preferentially to those who are already central, and hence visible. Status, on the other hand, also takes into account the popularity of alters following ego. Controlling for user popularity, status relates to the effect of a position conferred to an actor by virtue of being followed by high-status others. In both

⁸³ Collection size is included in all the analyses that include coherence as an independent variable, to account for coherence levels that result from the number of objects included in collection. For presentation purposes, it is not reported in these tables.

communities, status has a significant and curvilinear relationship with success. As Figure 5.4 illustrates, this curvilinearity is mostly related to the increasing effect of status on success in the upper range: extremely high-status individuals also enjoy an increase in popularity far greater than their mid-status counterparts.⁸⁴ This effect is particularly pronounced in the MOC: highest-status users are likely to increase in status by up to four standard deviations. In the absence of performance-based information, status is a strong predictor of subsequent success in the MOC.

Table 5.3
Fixed-effect estimates for FOC user success (standardized popularity delta)

Variable	Model 1	Model 2	Model 3
Popularity	0.6483 *** (0.0904)	0.5738 *** (0.0884)	0.5827 *** (0.0886)
Status	-0.3080 *** (0.0738)	-0.2009 ** (0.0721)	-0.2989 ** (0.1097)
Status ²	0.0659 *** (0.01)	0.0465 *** (0.0098)	0.0475 *** (0.01)
Performance	1.9775 *** (0.1627)	1.9787 *** (0.1649)	0.7487 (0.5653)
Performance ²	-0.9340 *** (0.139)	-0.9790 *** (0.14)	-0.9380 *** (0.141)
Statement	0.0531 * (0.025)	0.0604 ** (0.0216)	0.0618 ** (0.0218)
Image	0.2824 * (0.1264)	0.1872 (0.1126)	0.1881 (0.1131)
Days active	-0.0272 (0.0225)	-0.0262 (0.022)	-0.0259 (0.0221)
Coherence		0.9820 ** (0.3254)	0.9148 ** (0.3332)
Status*Coherence			0.1580 (0.1527)
Performance*Coherence			2.0191 * (0.8808)
Constant	0.4874 ** (0.1716)	-0.4200 (0.2934)	0.0637 (0.2261)
R ²	0.0944	0.1169	0.0116

Standard errors in parentheses. * p<0.05 ** p<0.01 *** p<0.001

In the FOC, performance is also a significant predictor of success: the difference in success between a user with a 100% gain on investment and a user with a 100% loss is estimated at roughly 4

⁸⁴ Note that status has a heavy-tailed distribution: high-status users in both communities have up to seven standard deviations more status than the mean user.

standard deviations. But the effect of performance on success begins to decline at extremely high return rates. It appears that a return rate above 100% is interpreted by others as an unreliable indication of future performance.⁸⁵ Additional control variables such as biographical statement length and whether a user posted an image on their profile have significant effects as expected. These attributes' effects are drastically more substantial in the MOC. In the absence of performance-based information, self-presentational information appears to be more consequential for DJ's success than it is for investors'.

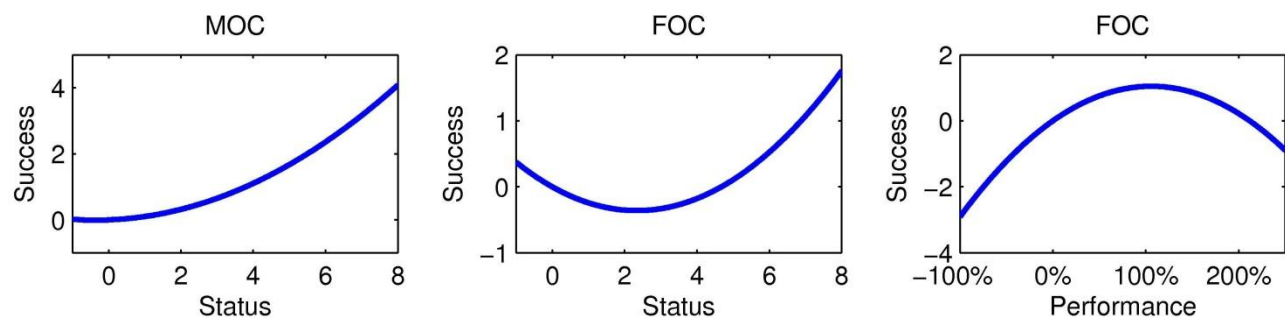


Figure 5.4: Success (popularity delta) as a function of Status (alpha centrality) and Performance (monthly return)

Model 2 includes coherence as an independent variable. In both communities, coherence is strongly and significantly associated with success. Consistent with hypothesis 5A those users whose collections are more coherent are also more likely to gain popularity. This finding is not especially surprising: in support of previous studies, this analysis reconfirms that, whether in music or finance, those actors who are more comprehensible to others, by virtue of exhibiting behavior that is consistent with others' behaviors, are more likely to be favorably noticed. The positive effect of coherence on subsequent success remains strong, and even increases in the MOC, when interaction effects are added in Model 3: in both communities an increase of one point in coherence translates to roughly a 0.9 standard deviation increase in success.

⁸⁵ Less than 1% of user-time observations have a return rate greater than or equal to 100%. It is likely that observers associate such a high return rate with excessive risk-taking, and are consequently deterred by it.

But the interaction between status and coherence has different effects on success in both communities. Consistent with Hypotheses 5B and 5C, this interaction is significant in the MOC, but insignificant in the FOC. As expected, the interaction term is negative in the music community. The relationship between status and coherence, and its implications on success, are illustrated in Figure 5.5. The diagram plots success as a function of coherence, such that each line corresponds to a fixed status level, ranging from very low (light gray) to very high (black) status and increasing in equal intervals. As we saw in Model 1, status has a strong and positive effect on success: high status individuals, irrespective of their coherence, are consistently more successful than their low-status counterparts. Note how the slope of the line changes as status increases: low-status actors benefit from exhibiting coherent collections, whereas high-status actors benefit from exhibiting unconventional musical compilations. At the highest extreme of status, an increase in one's coherence by one point translates to a drop of more than two standard deviations in success; moreover, a high-status actor with an incoherent collection is roughly six standard deviations more likely to gain subsequent popularity than a similar actor with low-status is. These findings suggest a tradeoff between status and the effects of coherence: incoherence is detrimental for low-status users who need to be noticed, but becomes a differentiating advantage for high-status users who are already visible.

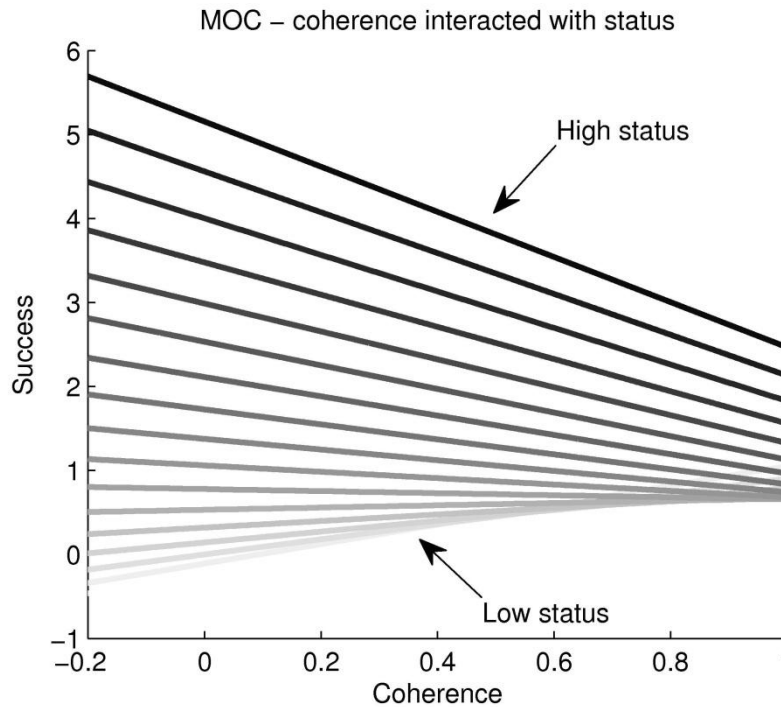


Figure 5.5: MOC success, estimated as a function of an interaction between coherence and status. Each line corresponds to a fixed level of status, ranging in constant intervals from low (light gray) to high (black).

Status is not similarly consequential in the FOC. As is the case in the MOC, users with coherent investment portfolios are likely to succeed more than those with incoherent portfolios irrespective of their status. Yet, unlike their MOC counterparts, high-status FOC investors do not have more latitude to behave unconventionally. As expected by Hypothesis 5C, given the availability of performance-based information, status does not serve to mitigate the effects of ambiguity in the FOC. But neither does performance. In fact, the interaction term between performance and coherence is, contra Hypothesis 5D, positive and significant: high-performing actors benefit more from coherence than do low-performing ones. As Figure 5.6 illustrates, the slope of the line representing the relationship between coherence and success changes from negative to positive as performance increases. Low-performing investors appear to benefit from exhibiting incoherent portfolios, whereas high-performing actors are penalized excessively for the same behavior. Not only does performance fail to mitigate the effects of ambiguity for high-performing actors, it appears to exacerbate them. In other words, while status

provides MOC DJs relative impunity from the risks of innovation, high-performance only increases expectations from FOC investors to behave conformingly. How can this discrepancy be explained?

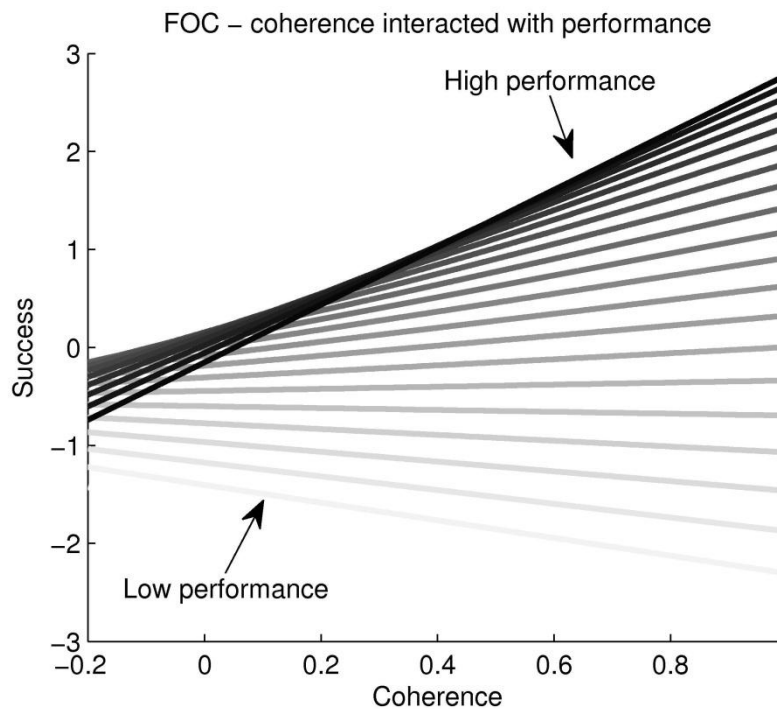


Figure 5.6: FOC success, estimated as a function of an interaction between coherence and performance. Each line corresponds to a fixed level of performance, ranging in constant intervals from low (light gray) to high (black).

5.4.3 Coherence as a Signal or as an Evaluative Lens

One possible explanation is that coherence is interpreted differently in both settings, such that it is understood as innovation by music aficionados, but predominantly as risk by investors. There are two possible reasons for these divergent interpretations. First, the penalty associated with failure in the FOC is far greater than in the MOC. Following the wrong users, and potentially adopting their failing strategies, can translate into tangible losses of capital. Investors therefore have strong incentives to steer away from nonconformist strategies. In the MOC, on the other hand, there is no clear definition for what constitutes a 'bad' mix. The absence of a consensual performance metric in the MOC makes

'failure' and 'success' inherently context-dependent – and therefore malleable – definitions. The line distinguishing risk from opportunity is consequently highly blurred. Second, DJs and investors partake in online social networking for different reasons, and under different circumstances. While DJs are there to explore new music, investors share information with the ultimate goal of maximizing profit. And during the tumultuous months of late 2008 and early 2009, as a financial crisis of historic proportions was unfolding, investors were presumably looking for stability and coherence in a financial environment increasingly characterized by unpredictability and disarray.

But the findings reported above undercut this explanation. If coherence is understood differently in our two empirical settings then it should also have different effects in each community. But the coefficient for coherence in Model 3 in both websites is practically identical, indicating that success increasing linearly with coherence. Moreover, if incoherence is associated with risk in the FOC, why are its negative effects particularly strong for high-performing actors? If anything, these actors have a proven record of success that should offset the riskiness inferred from their portfolios.

An alternative explanation is that coherence is utilized differently by users in each community. For MOC users, coherence serves as an indication about quality, and can either signal lack of clarity, or innovative potential. Status serves as a means to adjudicate these two competing interpretations. In the FOC, on the other hand, given the availability of performance-based information, cohesion is used – like status in the MOC – as a prism through which to assess other information: performance.

These different functions relate to a distinction proposed by Smith (2011) between identity as a signal, and identity as a lens. When functioning as a signal, identity is used by others as information about the actor. In contrast, when it functions as a lens, identity serves as a sense-making device with which other types of information are processed. Smith argues that when actors adopt nonconventional behaviors they amplify the potential drawbacks and rewards these behaviors carry: successful

nonconformism is celebrated, but when nonconformist behavior fails, it is reprimanded as imprudent. However, examining the global hedge fund industry, he finds, contra to his expectations, that nonconforming funds are penalized less severely for poor performance than conventional funds. The same applies for FOC users: rather than amplifying the consequences of performance, incoherence appears to be attenuating them.

To illustrate this point, let us revisit the interaction between coherence and performance plotted in Figure 5.6. Figure 5.7 plots the same results, yet here this interaction is illustrated through the relationship between performance and success. Each line corresponds to a fixed level of coherence, ranging from low (gray) to high (black). As is clearly visible, performance is far more consequential for investors with conventional portfolios. For such actors, the difference between very high and very low performance translates to more than five standard deviations in success. Yet for low-coherence investors, the effects of performance are far more nuanced. Moderate underperformance by incoherent actors reaps similar rewards as extreme over-performance. It appears that information about performance is discounted by observers when the actor in question is pursuing incoherent investment behaviors. When playing by the rules, investors are scrutinized by their performance, and rewarded accordingly. But when pursuing idiosyncratic strategies, their failure seems to be accepted with greater lenience, whereas their success with a great deal of suspicion.

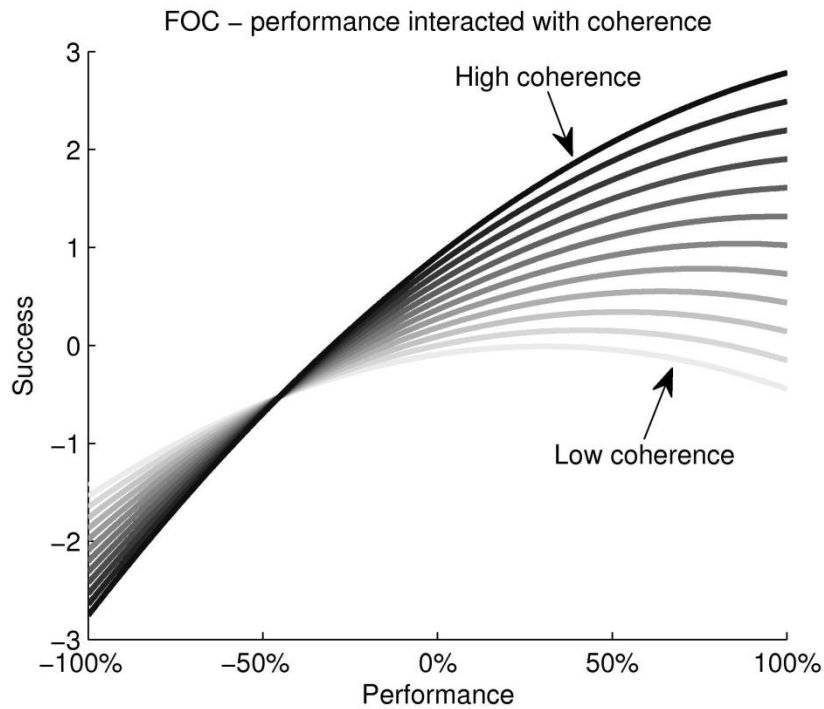


Figure 5.7: FOC success, estimated as a function of an interaction between performance and coherence. Each line corresponds to a fixed level of coherence, ranging in constant intervals from low (light gray) to high (black).

A broader conclusion can be drawn from these findings. Whether one's coherence is utilized as information in and of itself, or as a prism with which to evaluate other sources of information, depends on the availability of these different information sources. In the MOC, performance-based information is not available to users. In its absence, actors' collection compositions appear to function as primary sources of information for evaluating these actors' identities, and, by extension, their underlying qualities. But because identity is difficult to discern, users turn to status as a lens through which to evaluate this information. High-status actors therefore have more leeway for exploring new identities.

In the FOC, on the other hand, performance-based information is readily available. But investors are fully aware that previous market performance is far from a flawless indicator about future performance. They therefore turn to coherence as a lens through which to assess this performance. Coherent actors appear more predictable, and hence their previous performance is interpreted as a

reliable indication about the future. However, incoherent actors' performance is taken with a generous grain of salt: because their peculiarity is more difficult to decipher, their performance is not treated as a reliable signal. Moreover, given that coherence does not serve as a primary source for evaluation, status no longer functions as a lens through which to assess identity. Though high-status investors enjoy greater overall success, they do not have more maneuverability than others to pursue unconventional strategies.

5.5 Conclusion: Innovation and Performance-Measurability

The picture emerging from the analysis is one of a complex interplay between identity, status, and performance, resulting in three important insights. First, the manner by which coherence affects how an actor is perceived by others depends on the extent to which there is consensus among actors on how performance is measured. When such a measure is agreed upon, identity serves as lens through which performance-related information is evaluated. But when such a measure does not exist, identity is interpreted as a direct signal about underlying quality. In the latter case, status comes into play as an additional source of information which has a bearing on how identity is interpreted as a quality signal.

Second, contra our initial expectations, status and performance are not simply interchangeable proxies for an actor's intrinsic, but ultimately unobservable quality. As we have seen, whereas high-status actors have more latitude to explore unconventional behaviors, high-performing actors have less latitude to do so. This is because performance-measurability affects whether identity is used a signal, or an evaluative lens. Whereas in the latter case coherence is used in order to evaluate other information, in the former other information is used to evaluate the meaning of coherence.

Finally, these findings shed new light on the incentives that push certain actors, despite strong pressures for conformity, to defy categorical boundaries. These incentives, it appears, depend on the

extent to which performance is measurable. In contexts where performance-measurability is low – like French gastronomy – high-status actors are incentivized to innovate because such innovation differentiates them from others while their status mitigates the effects of appearing incoherent. But when performance is easily measurable, status no longer provides this protection. Rather, under such circumstances, poorly performing actors have stronger incentives to innovate, because continued bad performance will not result in penalties. Unlike high-performing unconventional actors, whose performance cannot be translated into success, poorly performing actors have nothing to lose by exploring unconventional behaviors.

More broadly, all other things being equal, performance-measurability should change the incentive structure for innovation, as depicted in Figure 5.8: when performance-measurability is low, status plays an important role pushing those with established reputations to innovate; when performance-measurability is high, however, performance supersedes status, incentivizing underperformers to think and behave outside the box. Thus, we should expect to find in domains such as academia, where methods for evaluating performance are highly debatable and contentious, that innovation emerges mostly from established figures. In competitive sports, on the other hand, where a baseball team's performance, for example, is fairly straightforwardly measured, status becomes less consequential. We should therefore expect to find that innovative techniques spring mostly on the margins of success, from low-performing teams.

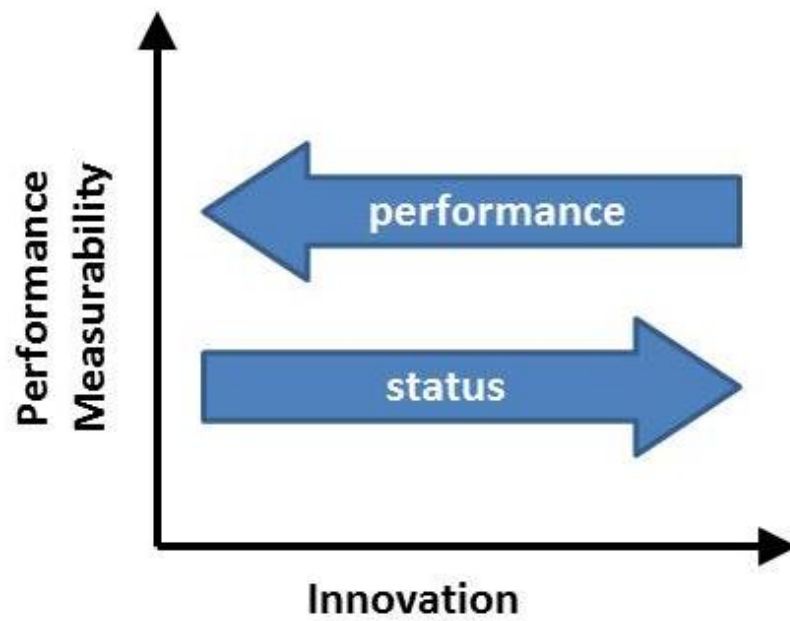


Figure 5.8: Who innovates? The implications of performance-measurability on innovation incentives.

Chapter 6

Conclusion: Change and Inter-subjective Synchronicity

The 2009 MTV Video Music Awards is probably best remembered for rapper Kanye West's hubristic and unselfconscious crashing of Taylor Swift's Best Female Video acceptance speech. Snatching the shocked country-pop star's microphone, West vocally protested the prize not having been awarded to R&B idol Beyoncé Knowles. The incident – receiving emphatic repudiations across the board, including from President Obama – distilled in one short moment the intermeshing of race relations and popular music in the US. The categorical work behind the scenes charged this altercation with social meaning: both West and Knowles are African-American, and are associated with music which has deep black roots. Swift, on the other hand, a white country singer-songwriter, is associated with a musical genre that lays claims to traditional white southern authenticity (Peterson 1997). Anyone watching the ceremony on TV would likely have interpreted the exchange as a metaphor for the tension between these two musical traditions, and the social identities they imply.

As much as Beyoncé and Swift are different, however, they also share many similarities. Both are accomplished female pop singers whose lyrics often draw on their own autobiographies when treating such teen-pop staple topics as failed relationships or unreciprocated love.⁸⁶ And though the two are consummate contemporary examples of 'pop',⁸⁷ the music industry consistently treats them as

⁸⁶ Two of their respective hits, *Irreplaceable* and *You Belong With Me*, for example, are harmonic pop ballads featuring soft guitars and love-related themes.

⁸⁷ Though often referred to as a genre, pop is best understood as commercial music intentionally obscuring genre characteristics with the aim of appealing to broad audiences (Lena and Peterson 2008). The online music catalogue Allmusic.com classifies both artists as pop singers.

operating in isolated, and in certain respects almost diametrically opposed musical worlds: R&B and country.⁸⁸ MTV Music Video Awards viewers likely follow suit. This is why sociologists traditionally turn to institutions when explaining the durability and persistence of social arrangements.⁸⁹ Institutions, they argue, through their formal classifications, rituals and regulations, produce behavioral conventions which sort people and things into different metaphorical boxes. But how are these conventions sustained in practice? Where do they come from in the first place? And under what circumstances do they change?

In this dissertation I have attempted to address these questions by tracing the social dynamics through which classifications emerge and take hold. Rather than examining the formal classification criteria used by critics and experts – which is often, though not exclusively, the case in sociological studies – I conceptualize a category system as an uncoordinated collective exploration of a latent semantic web. This conceptualization implies a mutually constitutive understanding of the relationship between individual actors and the inter-subjective structures that channel their behaviors: social categories are external to individual actors – insofar as they are apparent in these actors’ aggregate choices – but they are not strictly imposed from above; rather they are sustained and constantly reproduced through these actors’ behaviors.

Critical readers may point out that there is not much revelatory about this proposition. In fact, recent years have seen a proliferation of sociological accounts which rely on a conceptual duality between ‘culture’ and ‘structure’ (Mohr and White 2008; Sewel 1992; Giddens 1984). Yet this duality is often stated in the abstract and is very broadly conceived; empirical attempts to operationalize it are rare. There are good reasons for this dearth. First, high-quality data that simultaneously relate to two

⁸⁸ The two singers are listed on different Billboard charts and both have won several Grammy awards in different categories, Swift as best country singer and Beyoncé as best R&B singer.

⁸⁹ I use the term institution in its broad sociological sense as a cognitively widespread behavioral convention that has “rule-like status” (DiMaggio and Powell 1991).

different classes of social phenomena are very difficult to gain access to, let alone collect systematically over time. Hence most studies that attempt to demonstrate a mutual relationship of this kind tend to rely on snapshots taken in one given moment. They are therefore restricted in their ability to demonstrate a co-evolutionary and mutually constitutive process. A second difficulty relates to the social-network methodology very commonly employed in these studies. Steeped in a strong structuralist tradition, network analysis conventionally emphasizes the structural-interactional components of social reality, treating cultural categories as fixed entities which are external to the dynamics it tries to explain (Emirbayer and Goodwin 1994). Simulation models, for example, often think of ‘culture’ in very static terms as independent ideas or memes that travel on a socio-structural infrastructure. What is of interest is how this infrastructure facilitates or hinders the process of cultural diffusion (eg. Carley 1991; Mark 1998; Centola et al 2007).

This study tries to move this research agenda forward in several ways. First, it concretizes the “amorphous, indescribable mist” (Fine 1979: 733) of culture by focusing on one of its fundamental aspects: collectively shared category systems. As discussed in Chapter 2, recent research in cognition suggests that human neurophysiology predisposes people to perceive the world through a categorical lens. Categories are therefore central to the operation of culture: when shared inter-subjectively, they serve as cultural prisms through which our experiences of the world are interpreted, and function as blueprints that guide our behaviors (DiMaggio 1997). Sociological studies have traditionally explored category systems’ casual role by examining formal systems of classification. Yet these tidy taxonomies are very different from the fuzzy categorical divisions that people rely on in their daily sense-making. My approach is therefore to excavate collective systems of classification from people’s aggregate behaviors. This approach is more consistent with how categories are used in practice. Rather than searching for their synthetic reification in institutionalized arrangements, I trace collectively shared categories by

identifying recurrent patterns of associations between objects in different individuals' behaviors.⁹⁰ This allows for a flexible conceptualization of a category as a network of related concepts, as opposed to a fixed and clearly demarcated set, the evolution of which can be traced over time.

Though the methods I use are new, this dissertation is by no means the first empirical attempt to excavate folk taxonomies from people's practical behaviors (Bowker and Star 1999), nor is it groundbreaking in its conceptualization of meaning in network terms (Mohr 1998). Yet a second way in which this study is novel is its juxtaposition of practical classifications with network structures of interpersonal interaction. Drawing on network theory, social constructionist theory and cognitive sociology, it takes a position that Emirbayer and Goodwin (1994) call *structuralist constructionism*: it sees categories as emerging through social interaction but also possessing ontological qualities that feed back into the patterns of interaction on which they depend. The duality between structure and culture is here operationalized as a structural relationship between two types of networks: one connecting people, the other connecting concepts. In other words, networks are not merely a social infrastructure that connects people of different predefined categorical attributes. Rather, these categories change in the course of, and consequently affect, social interaction.

To empirically explore these theoretical assumptions I use data from two online social network sites. These data present two significant advantages. First, they provide unparalleled access to the complete set of behaviors of all members of these two online communities. Over a period of roughly one year, these data document all the various interpersonal interactions between the community members as well as these members' object collection compositions, whether comprising musicians or financial assets. Such fine-grained data is practically impossible to get hold of in non-computer-mediated

⁹⁰ The analytical focus on institutionalized systems of classification – by which I mean formal classifications produced by bureaucracies, analysts and researchers – is highly valuable in its own right, as a large volume of studies have demonstrated. My purpose is not to dismiss this research but to complement it by shifting the analytical lens to how categorization operates in practical everyday settings.

settings, especially when encompassing thousands of individual actors. As I discuss in Chapter 2, there are obvious limitations to these data.⁹¹ But these are offset by what the data afford: the ability to observe, as if being the proverbial fly on the wall, the unfolding of social dynamics in a setting which is not artificially produced for the exclusive purpose of social scientific research. The two website's user interfaces severely restrict the kinds of activities users can pursue online to those that are relevant for this study. In this respect, these data are ideal for exploring the questions this dissertation is concerned with.

A second advantage relates to the inherent differences between the two websites under investigation. Comparative analyses often choose as their test-cases comparable albeit non-identical empirical settings, either highlighting their differences or pointing to their overarching similarities. In this study I have mostly adopted the latter approach. My motivation was not to analytically superimpose the two communities in order to identify the various ways in which music and finance are distinctive; in many ways, to do so would be to state the obvious. Rather, my intention was to demonstrate – precisely because of their fundamental differences – how financial and musical consumption are driven by the same rudimentary social dynamics. This comparison is intuitively powerful because we normally imagine financial and artistic pursuits to be so diametrically opposed: whereas music, in Aldous Huxley's (1931) words, "comes nearest to expressing the inexpressible," money, as Ayn Rand's (1957) protagonist quips, is ultimately "only a tool." But in both domains socially constructed categories play important roles in creating and sustaining order, and are used by actors as means to ascertain quality. This study sought to use the comparison between music and finance in order to shed light on the similar, and arguably universal ways in which categories evolve through social interaction in both domains.

⁹¹ Of course, no dataset is entirely without limitations. Online data are relatively new in sociological research, and the onus is justifiably on their users to provide persuasive justifications for their use. I have attempted to do so in greater length in Chapter 2.

The payoff – so I hope the reader has by now been persuaded – has been substantial. Chapter 2 lays out a series of methodological tools – which are honed throughout the following chapters – for the application and operationalization of the basic concepts informing this dissertation. The three empirical studies that follow dissect the broad theoretical statements posed above into smaller, workable hypotheses, and develop the tools for testing them. Each chapter approaches the subject matter from a different angle. Chapter 3 examines the aspects of the interaction-categorization duality that are consistent over time. The two following chapters examine change from two different vantage points: Chapter 4 asks under what conditions objects’ categorical meanings change, whereas Chapter 5 explores what incentivizes actors to defy categorical boundaries. Each hinges on a distinctive analytical concept, and its operationalization: Chapter 3 defines a social construction as a structural homology between a network of interpersonal interaction and a network of object categorical similarity; Chapter 4 defines the concept of social boundedness and demonstrates how it is related to an object’s semantic stability; and Chapter 5 suggests that performance-measurability can explain how status relates differently to success in the music and financial communities.

Altogether, these three empirical viewpoints concatenate into a unified theoretical perspective which builds on Berger and Luckmann’s (1966) constructionism and the neo-institutionalist theory which it was central in catalyzing. Neo-intuitionalist accounts conceive of institutions as macro-level abstractions that inhere in taken-for-granted scripts and classifications. As DiMaggio and Powell (1991) concede, while focusing on the durability and persistence of institutional arrangements, these accounts largely neglect the micro-foundations of socio-cognitive order; consequently, they often fail to acknowledge endogenous sources of institutional change (Rao 2005).

The approach I take here is to conceive of institutions as brittle arrangements of *inter-subjective synchronicity* which are repeatedly sustained through social interaction: as long as actors reenact the

institutional order, it is maintained. But local divergences can gradually aggregate into the emergence of a new inter-subjectivity. Patterns of interpersonal interaction are therefore central to understanding how inter-subjectivity is maintained and under what circumstances it is transformed. The findings reported in this dissertation suggest that change can emerge through an undulating cycle whereby certain actors are incentivized to be creative in their behaviors, spawning diffusion through imitation, which results in a new collective equilibrium, and vice versa.

As we saw in Chapter 5, actors occupy different structural positions which provide varying opportunities and incentives for divergent behaviors. When these divergences diffuse into socially bounded groups – as we have seen in Chapter 4 – semantic reinterpretation is catalyzed. This change entails a transformation in the classificatory logic shared by the community as a whole, resulting in a new structural equilibrium between patterns of interaction and categorization (Chapter 3). Change can therefore emerge from within the delicate balance that holds an institutional order in place (Figure 6.1).

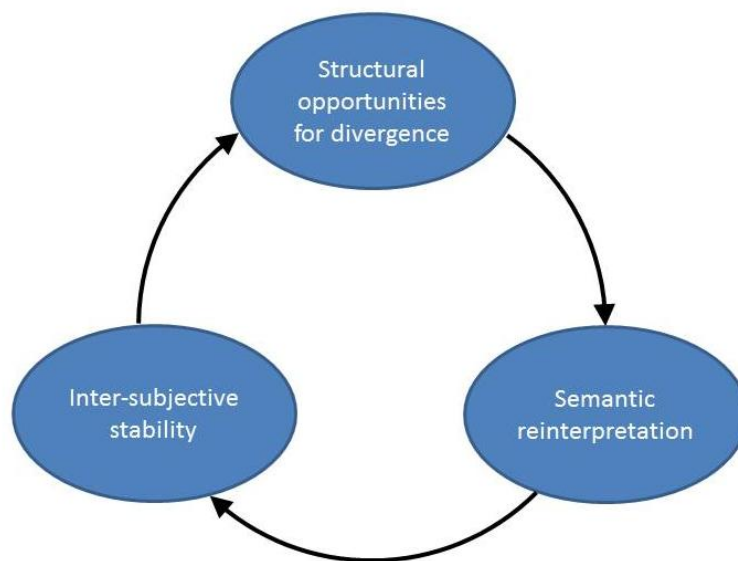


Figure 6.1: inter-subjective synchronicity and the cycle of change

The results presented in this dissertation are only initial steps in elaborating a coherent theory of inter-subjective synchronicity. Future research would need to address a variety of theoretical and

empirical challenges that are either only alluded to, or entirely neglected in this thesis. Two are particularly worthwhile mentioning. The first relates to the question of diffusion, which has been prominent in network studies (Burt 1987; Christakis and Fowler 2006; Watts 2003). Central to the model used here is the assumption that people do not simply mimic one another by adopting identical behaviors but rather are influenced to espouse similar categorical distinctions. And these categories inhere in semantic relationships between different concepts. In other words, the things that diffuse through the network are not independent ideas but associations between ideas which make up an implicit classificatory order.⁹² While the results presented in this dissertation, especially those in Chapters 3 and 4, are consistent with the idea of relational diffusion,⁹³ more evidence is necessary to substantiate its empirical existence.

A second question deserving future attention relates to the delicate equilibrium that sustains inter-subjective synchronicity. Chapter 3 provides evidence that such equilibrium exists, and that it changes over time. Yet further theorizing and empirical testing is necessary for conceptualizing, in detail, the dynamics through which relational diffusion occurs and how it leads to the emergence of a new stable balance. Agent-based simulations would likely prove particularly useful for this purpose. Drawing on insights derived from the studies included in this dissertation, these simulations will explore, from a systemic point of view, the structural conditions under which imbalances emerge.

An entirely different set of questions is substantively related to the two social domains in question in this study – music and finance – and the distinctive insights that can be gleaned about each. This dissertation has been primarily concerned with highlighting the universal similarities between both domains. With the exception of Chapter 5, the myriad and fundamental differences between them have

⁹² This is not to disavow the notion that certain behaviors do indeed diffuse like viruses. The behaviors of interest here, however, are those that originate from an intricate categorical logic.

⁹³ Relational in the sense of the diffusion of relationships between concepts, as opposed to the diffusion of concepts through interpersonal relations.

been essentially overlooked. Critics may rightly point out that this omission is potentially problematic as these differences, as Chapter 5 demonstrates, are consequential for how categories emerge through interaction. Moreover, the data at hand, and the methodologies developed to analyze them, can be useful for shedding light on the unique sociological features of music and finance.

Perhaps the most glaring difference between the two communities relates to the relationship between behavior and identity signaling. There exists a voluminous literature on how musical consumption functions as a symbolic identity marker which reinforces race-, class-, ethnicity- and gender-boundaries (Roy and Dowd 2010). But the idea that identity plays a fundamental role in structuring financial behaviors remains far more controversial.⁹⁴ When discussing how identity factors into financial and artistic behavior it is important to make an analytical distinction between *motivation* and *interpretation*, i.e., between the extent to which identity signaling motivates actors' behaviors, and how behaviors are interpreted by others as signals about an underlying categorical identity. In the studies included in this dissertation I have mostly assumed that actors' behaviors are interpreted as identity signals irrespective of these actors' motivations. This assumption is not farfetched, even in the context of financial investment: previous research suggests that when market exchange is not anonymous, actors develop reputations that are associated with their investment history (Abolafia 1996; Podolny 2008). It is nonetheless highly plausible that actors in both the music and finance communities are inspired by inherently different motivations and as a consequence respond differently to others' behaviors (eg. when mimicking one another). Such a difference may have a bearing on the overall dynamics, and deserves further attention.

Future research might also seek to address questions that are specific to the sociology of music and the sociology of finance. The apparent rise of the cultural omnivore, for example, has been debated

⁹⁴ Though, riding on the tide of behavioral economics' recent success, it is gaining increasing currency even amongst economists. See for example Akerlof and Kranton (2010).

extensively in sociological literature in recent years, especially with respect to musical consumption. The data used here may be instrumental in shedding further light on whether, and how, cultural omnivorousness is related to individuals' structural opportunities to become exposed to a wide variety of musical styles. These data may be equally useful for the sociological study of financial investment. Recent work in behavioral economics has potently demonstrated that different people exhibit varying degrees of tolerance to risk-taking, in a manner that profoundly affects their investment behaviors. Examining such behaviors in a network setting could potentially illuminate the social underpinnings of the distribution of risk-tolerance, and how responses to uncertainty propagate through social interaction.

The information technology revolution of recent years has opened up sociological research avenues which were up until recently almost impossible to imagine. The advent of the so-called Web 2.0 and its participatory online data-sharing has transformed the internet into a global repository of social data. And advances in computational algorithms and hardware have made the analysis of these unprecedentedly large datasets feasible. The potential, especially for the sociology of culture – traditionally crippled by low-quality data and conventional statistical tools ill-designed to address its complex theoretical assumptions – is immense. In this dissertation I have tried to tap this potential by approaching the study of culture through the prism of classification. Not incidentally, the problem of classification is also central to the burgeoning computer science field of data mining, which is concerned with the discovery of patterns in large datasets (Tan et al 2005). The methods computer scientists use to extract latent structures from messy data are highly applicable for uncovering the practical classifications informing the behaviors of the online users generating these data.

Mining culture from online data means discovering categories in people's behaviors rather than assuming them a-priori. We know that collectively shared categorical distinctions – like those assigning

Beyoncé and Taylor Swift to different musical traditions – play central roles in sustaining social order, and the inequalities it is founded on. Yet we still have a cursory understanding of the processes through which these categories take form and become meaningful. Using the online worlds of music and finance as its virtual laboratories, this dissertation has harnessed data mining techniques to dabble in some of the most fundamental questions inspiring sociological inquiry: what is ‘culture?’ how does it hold people together? and which of its properties should we measure? Far from providing a definitive answer to either of these questions I have nevertheless hopefully been able to show that exploring the micro- interactional foundations of social categories in online data is a sociologically worthwhile endeavor in the right direction.

Appendix A: Data Management

Data from both the MOC and FOC were provided by the respective site owners voluntarily. The data were provided in raw format as MySQL tables, an open-source relational database management system. The original data owners removed all identifying information (such as users' real names, addresses, passwords etc.), but did not make any additional changes to the data. Under the terms of agreement with both website owners, the data are kept securely on Princeton University-owned hardware and are password-protected. The data cannot be shared with other parties (whether within or outside Princeton) without explicit consent from the website owners.

The large data volume (totaling in several gigabytes, and comprising millions of records) is impossible to analyze using traditional social-scientific analysis tools. In order to process such a large dataset, the data were uploaded to MySQL database. The data were stored in a database catalog comprising a variety of tables, corresponding to users, objects, user-to-user relationships and user-to-object relationships. Additional tables were created to store variables produced for the purpose of the analysis, such as categorical similarity or ambiguity. Analyses were predominantly performed using MathWork's Matlab R2010b computing environment and programming language. Data were retrieved from MySQL using a free PHP-based Matlab-MySQL interface developed by Luigi Rosa, downloadable from <http://www.advancedsourcecode.com/matlabphpmysql.asp>. Some statistical analyses were performed using Stata 11 (after the data had been converted from MySQL to a Stata-readable format). Additional data structuring scripts were implemented using the Python 2.6 programming language.

Data relating to users in both websites were highly structured and well maintained, and required minimal processing. Data relating to objects, on the other hand, suffered from certain consistency and reliability problems. In the FOC, these inconsistencies were fairly limited. Stocks are uniquely identified

by their trading symbol. These identifiers were matched with data retrieved from the Center for Research in Security Prices (CRSP), operated by the University of Chicago's Booth School of Business, and which included historical securities performance data. A small fraction of these identifiers could not be matched with CRSP data (this was often the case when a firm filed for bankruptcy and its stock was traded on the Pink Sheets). These discrepancies were manually corrected.

In the MOC, data inconsistencies posed a greater challenge. Here, artists and tracks do not have unique identifiers. Instead, users provide artists names and track titles in freeform text. As a consequence, the same artist may be spelled in a variety of ways. Such duplications occur most often with artist names including prefixes (eg. 'The Beatles' v. 'Beatles'), special characters ('Beyoncé' v 'Beyonce'), hyphens and slashes ('AC/DC' v 'ACDC') or when artists are known by several names ('Beethoven' v 'Ludwig van Beethoven'). This poses a problem because it is necessary that records corresponding to the same artists – irrespective of spelling inconsistencies – be treated as such. Overall the data include 57,677 unique artist names.

Correcting for multiple spelling is a highly challenging problem. To overcome it, I performed several procedures. First, all names were transformed into lower case, and regular expressions were used to clean spaces, blanks, hyphens and other punctuation marks. I used two external online sources to match names: the online music website Last.fm and the online radio Pandora. Because some names are not unique to one artist, I used a combination of artist name and track to identify an artist. Each artist entry was matched with Last.fm and Pandora using both websites' name search engine. The search results were compared to the original name using the Levenshtein string distance metric (Gusfield 1997). The result with the shortest distance was chosen (in case of multiple minimal distance results the first one was chosen). When both Pandora and Last.fm returned the same result, the entry was assigned to that artist's name. In case of a discrepancy, if the Levenshtein distance was smaller than three, the

Last.fm name was chosen. Remaining discrepancies were settled manually. After cleaning the data and correcting duplicates, unique artists who were included only in one user's collection were removed.

Overall, this procedure reduced the number of unique artist names by roughly 70%.

References Cited

- Abbott, Andrew, 1988, "Transcending General Linear Reality", *Sociological Theory* 6: 169-186
- Aral, Sinan, Lev Muchnik, and Arun Sundararajana, 2009, "Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks", *Proceedings of the National Academy of Sciences*, 106(51): 21544–21549
- Anderson, Benedict R., 1991, *Imagined Communities: Reflections on The Origin And Spread Of Nationalism*, London: Verso
- Baldassarri, Delia, and Peter Bearman, 2007, "Dynamics of Political Polarization", *American Sociological Review* 72(5): 784-811
- Barabási, Albert-László and Réka Albert, 1999, "Emergence of Scaling in Random Networks", *Science* 286(5439): 509-512
- Baumann, Shyon, 2007, *Hollywood Highbrow: From Entertainment to Art*, Princeton: Princeton University Press
- Bearman, Peter S., and Katherine Stovel, 2000, "Becoming a Nazi: A Model for Narrative Networks", *Poetics* 27: 69-90
- Ben-Yehuda, Nachman, 1980, "The European Witch Craze of the 14th to 17th Centuries: A Sociologist's Perspective", *American Journal of Sociology* 86(1): 1-31
- Berger, Peter L. and Thomas Luckmann 1966, *The Social Construction of Reality*, Garden City, NY: Doubleday
- Berger, Jonah and Chip Heath, 2007, "Where Consumers Diverge from Others: Identity-Signaling and Product Domains", *Journal of Consumer Research* 34(2): 121-134
- Berger, Jonah and Chip Heath, 2008, "Who Drives Divergence? Identity Signaling, Outgroup Dissimilarity, and the Abandonment of Cultural Tastes", *Journal of Personality and Social Psychology*, 95(3): 593-607
- Bergesen, Albert J., 2004, "Durkheim's Theory of Mental Categories: A Review of the Evidence", *Annual Review of Sociology* 30: 395-408
- Binder, Amy, 1993, "Constructing Racial Rhetoric: Media Depictions of Harm in Heavy Metal and Rap Music", *American Sociological Review* 58(6): 753-767
- Bloor, David, 1982, "Durkheim and Mauss Revisited: Classification and the Sociology of Knowledge", *Studies in History and Philosophy of Science* 13(4): 267-297
- Bodie, Zvi, Alex Cane and Alan Marcus, 2004, *Investments*, New York: McGraw-Hill
- Boltanski, Luc, 1987, *The Making of a Class: Cadres in French Society*, Cambridge: Cambridge University Press

- Boltanski, Luc and Laurent Thévenot, 2006, *On Justification: Economies of Worth*, Princeton: Princeton University Press
- Bonacich, Phillip, 1972, "Factoring and Weighting Approaches to Status Scores and Clique Identification", *The Journal of Mathematical Sociology* 2(1): 113-120
- Bonacich, Phillip and Paulette Lloyd, 2001, "Eigenvector-Like Measures of Centrality for Asymmetric Relations", *Social Networks* 23(3): 191-201
- Bourdieu, Pierre, 1984, *Distinction: A Social Critique of the Judgment of Taste*, Cambridge, Mass.: Harvard University Press
- Bourdieu, Pierre, 1990, *The Logic of Practice*, Stanford: Stanford University Press
- Bourdieu, Pierre, 1993, *The Field of Cultural Production: Essays on Art and Literature*, Polity Press
- Bowker, Geoffrey C., and Susan Leigh Star, 1999, *Sorting Things Out: Classification and Its Consequences*, Cambridge: MIT Press
- boyd, Dana M., and Nicole B. Ellison, 2007, "Social Network Sites: Definition, History and Scholarship", *Journal of Computer Mediated Communication*, 13(1): 210-230
- Breiger, Ronald L., 1974, "The Duality of Persons and Groups", *Social Forces*, 53(2): 181-189
- Brubaker, Rogers and Frederick Cooper, 2000, Beyond 'Identity', *Theory and Society* 29(1): 1-47
- Brubaker, Rogers, Mara Loveman and Peter Stamatov, 2004, Ethnicity as Cognition, *Theory and Society* 33(1): 31-64
- Bryson, Bethany, 1996, "'Anything But Heavy Metal': Symbolic Exclusion and Musical Dislikes", *American Sociological Review* 61(5): 884-899
- Burt, Ronald S., 1987, "Social Contagion and Innovation: Cohesion Versus Structural Equivalence", *American Journal of Sociology* 92(6): 1287-1335
- Butler, Judith, 1993, *Bodies That Matter: On the Discursive Limits of Sex*, London: Routledge
- Callon, Michel, 2008, "What Does It Mean to Say That Economics Is Performative?", in Donald MacKenzie, Fabian Muniesa and Lucia Siu (eds.), *Do Economists Make Markets? On the Performativity of Economics*, Princeton: Princeton University Press
- Carley, Kathleen, 1991, "A Theory of Group Stability", *American Sociological Review* 56(3): 331-354
- Carruthers, Bruce G. and Laura Ariovich, 2010, *Money and Credit: A Sociological Approach*, Cambridge: Polity Press
- Cattuto, Ciro, Alain Barrata, Andrea Baldassarri, Gregory Schehr, and Vittorio Loreto, 2009, "Collective Dynamics of Social Annotation", *Proceedings of the National Academy of Science* 106(26): 10511-10515
- Centola, Damon, 2010, "The Spread of Behavior in an Online Social Network Experiment", *Science* 329(5996): 1194-1197

- Centola, Damon and Michael Macy, 2007, "Complex Contagions and the Weakness of Long Ties", *American Journal of Sociology* 113(3): 702-734
- Centola, Damon, Juan Carlos González-Avella, Víctor M. Eguíluz and Maxi San Miguel, 2007, "Homophily, Cultural Drift, and the Co-Evolution of Cultural Groups", *Journal of Conflict Resolution* 51(6): 905-929
- Chaiken, Shelly, and Yaacov Trope, 1999, *Dual-Process Theories in Social Psychology* New York: Guilford.
- Christakis, Nicholas A., and James H. Fowler, 2007, The Spread of Obesity in a Large Social Network over 32 Years, *The New England Journal of Medicine* 35:370-379
- Coleman, James S., Elihu Katz and Herbert Menzel, 1966, *Medical Innovation: A Diffusion Study*, Indianapolis: Bobbs-Merrill Co.
- Clauset, Aaron, Cosma R. Shalizi, and Mark E. Newman, "Power-law Distributions in Empirical Data", *SIAM Review* 51(4): 661-703
- Coval, Joshua D. and Tobias J. Moskowitz, 1999, "Home Bias at Home: Local Equity Preference in Domestic Portfolios", *The Journal of Finance*, 54(6): 2045-2073
- Custers, Ruud, and Henk Aarts, 2010, "The Unconscious Will: How the Pursuit of Goals Operates Outside of Conscious Awareness", *Science* 329(2): 47-50
- D'Andrade, Roy G., 1995, *The Development of Cognitive Anthropology*, Cambridge: Cambridge University Press
- Danon, Leon, Albert Diaz-Guilera, Jordi Duch and Alex Arenas, 2005, "Comparing Community Structure Identification", *Journal of Statistical Mechanics* P09008
- DeNora, Tia, 2000, *Music in Everyday Life*, Cambridge: Cambridge University Press
- Denrell, Jerker, 2008, "Indirect Social Influence", *Science* 321(5885): 47-48
- Denrell, Jerker, and Gaël Le Mens, 2007, "Interdependent Sampling and Social Influence", *Psychological Review*, 114(2): 398-422
- DiMaggio, Paul J., 1987, Classification in Art, *American Sociological Review* 52(4):440-455
- DiMaggio, Paul J., 1993, "Nadel's Paradox Revisited: Relational and Cultural Aspects of Organizational Structure" in N. Nohria and R.G. Eccles (eds.) *Networks and Organizations*, Harvard Business School Press: 118-42
- DiMaggio, Paul J., 1997, "Culture and Cognition", *Annual Review of Sociology* 23: 263-287
- DiMaggio, Paul and Hugh Louch, 1998, "Socially Embedded Consumer Transactions: For What Kinds of Purchases Do People Most Often use Networks", *American Sociological Review* 63(5): 619-637
- DiMaggio, Paul J. and Walter W. Powell, 1983, "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields", *American Sociological Review* 48(2): 147-160
- DiMaggio, Paul J. and Walter W. Powell, 1991, "Introduction" in: Walter W. Powell and Paul J. DiMaggio (eds.), *The New Institutionalism in Organizational Analysis*, Chicago: University of Chicago Press, pp. 1-38

- Dodds, Peter S., and Duncan Watts, 2004, "Universal Behavior in a Generalized Model of Contagion", *Physical Review Letters* 92(21): 218701
- Douglas, Mary, *Purity and Danger*, New York: Routledge
- Douglas, Mary, 1986, *How Institutions Think*, Syracuse: Syracuse University Press
- Durkheim, Emile, 1965, *The Elementary Form of Religious Life*, Oxford University Press
- Durkheim, Emile and Marcel Mauss, 1963, *Primitive Classification*, Chicago: The University of Chicago Press
- Edelson, Micah, Tali Sharot, Raymond J. Dolan and Yadin Dudai, 2011, "Following the Crowd: Brain Substrates of Long-Term Memory Conformity", *Science* 333(108): 108-111
- Eliasoph Nina, Paul Lichterman, 2003, "Culture in Interaction." *American Journal of Sociology* 108(4): 735-94
- Ellison, Nicole B., Charles Steinfield and Cliff Lampe, "The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites", *Journal of Computer-Mediated Communication* 12(4): 1143-1168
- Emirbayer, Mustafa, 1997, "Manifesto for a Relational Sociology", *American Journal of Sociology* 103(2): 281-317
- Emirbayer, Mustafa and Jeff Goodwin, 1994, "Network Analysis, Culture, and the Problem of Agency", *American Journal of Sociology* 99(6): 1411
- Evans, James A. and Jacob G. Foster, "Metaknowledge" *Science* 331(6018): 721-725
- Fagiolo, Giorgio, 2007, "Clustering in Complex Directed Networks", *Physical Review E* 76: 026107
- Fine, Gary Alan, 1979, "Small Groups and Culture Creation: The Idioculture of Little League Baseball Teams", *American Sociological Review* 44(5): 733-745
- Fiske, Susan T., and Patricia W. Linville, 1980, "What does the Schema Concept Buy us?", *Personality and Social Psychology Bulletin* 6(4): 543-557
- Fiske, Susan T., and Shelley E. Taylor, 2008, *Social Cognition: From Brains to Culture*, New York: McGraw-Hill
- Foucault, Michel, 1970, *The Order of Things: An Archeology of the Human Sciences*, Random House
- Foucault, Michel, 1972, *The Archaeology of Knowledge*, London: Tavistock Publications
- Fourcade, Marion and Kieran Healy, 2007, "Moral Views of Market Society", *Annual Review of Sociology* 33:285-311
- Frickel, Scott and Neil Gross, 2005, "A General Theory of Scientific/Intellectual Movements", *American Sociological Review* 70(2): 204-232
- Friedkin, Noah H., 1993, "Structural Bases of Interpersonal Influence in Groups: A Longitudinal Case Study", *American Sociological Review* 58: 861-872

- Friedland, Roger and John Mohr, 2004, "The Cultural Turn in American Sociology", in Friedland, Roger and John Mohr (eds.), *Matters of Culture: Cultural Sociology in Practice*, Cambridge University Press: 1:70
- Galdi, Silvia, Luciano Arcuri, and Bertram Gawronski, 2008, "Automatic Mental Associations Predict Future Choices of Undecided Decision-Makers" *Science* 22:1100–1102
- Ghaziani, Amin and Delia Baldassarri, 2011, "Cultural Anchors and the Organization of Differences: A Multi-method Analysis of LGBT Marches on Washington", *American Sociological Review* 76(2): 179-206
- Giddens, Anthony, 1984, *The Constitution of Society: Outline of the Theory of Structuration*, University of California Press
- Gieryn, Thomas F., 1983, Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists, *American Sociological Review* 48(6):781-795
- Girvan, Michelle and Mark E. J., 2002, "Community structure in social and biological networks" *PNAS* 99(12): 7821-7826
- Goetzmann, William N., and Alok Kumar, 2008, "Equity Portfolio Diversification", *Review of Finance* 12: 433–463
- Goldberg, Amir, 2011, "Mapping Shared Understandings Using Relational Class Analysis: The Case of the Cultural Omnivore Reexamined", *American Journal of Sociology* 116(5): 1397-1436
- Good, Benjamin H., Yves-Alexandre de Montjoye and Aaron Clauset, 2010, "Performance of Modularity Maximization in Practical Contexts", *Physical Review E* 81, 046106
- Gould, Roger V., 2002, "The Origins of Status Hierarchies: A Formal Theory and Empirical Test", *American Journal of Sociology* 107(5): 1143-1178
- Granovetter, Mark, 1978, "Threshold Models of Collective Behavior", *American Journal of Sociology*, 83(6): 1420-1443
- Griswold, Wendy, 1987, "A Methodological Framework for the Sociology of Culture", *Sociological Methodology* 17: 1-35
- Hacking, Ian, 1999, *The Social Construction of What?* Cambridge: Harvard University Press
- Hannan, Michael T., 2010, "Partiality of Memberships in Categories and Audiences", *Annual Review of Sociology* 36: 159-181
- Hannan, Michael T. and John Freeman, 1977, "The Population Ecology of Organizations", *American Journal of Sociology* 82(5): 929-964
- Hannan, Michael T., László Pólos, and Glenn R. Carroll, 2007, *Logics of Organization Theory: Audiences, Codes and Ecologies*, Princeton: Princeton University Press
- Hargadon, Andrew B. and Yellowlees Douglas, 2001, "When Innovations Meet Institutions: Edison and the Design of the Electric Light", *Administrative Science Quarterly* 46(3): 476-501

- Holland, Rob W., Merel Hendriks and Henk Aarts, 2005, "Smells Like Clean Spirit: Nonconscious Effects of Scent on Cognition and Behavior", *Psychological Science* 16(9): 689-693
- Hsu, Greta, Michael T. Hannan and Özgecan Koçak, 2009, "Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests", *American Sociological Review*, 74(1): 150-169
- Hsu, Greta, 2006, "Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production", *Administrative Science Quarterly* 51(3): 420-450
- Huss, Boaz, 2007, "The New Age of Kabbalah: Contemporary Kabbalah, the New Age and Postmodern Spirituality", *Journal of Modern Jewish Studies* 6(2): 107-125
- Huston, Ted L. and George Levinger, 1978, "Interpersonal Attraction and Relationships", *Annual Review of Psychology* 29:115-56
- Huxley, Aldous, 1931, *Music At Night and Other Essays*, London: Chatto and Windos
- Johnson, Victoria, 2008, *Backstage at the Revolution: How the Royal Paris Opera Survived the End of the Old Regime*, Chicago: The University of Chicago Press
- Kahneman, Daniel and Amos Tversky, 1979, "Prospect Theory: An Analysis of Decision under Risk", *Econometrica* 47(2): 263-292.
- Kahney, Leander, 2004, *The Cult of Mac*, San Francisco: No Starch Press
- Kalmijn, Matthias, 1998, "Intermarriage and Homogamy: Causes, Patterns, Trends", *Annual Review of Sociology* 24: 395-421
- Kay, Aaron C., S. Christian Wheeler, John A. Bargh and Lee Ross, 2004, "Material Priming: The Influence of Mundane Physical Objects on Situational Construal and Competitive Behavioral Choice", *Organizational Behavior and Human Decision Processes* 95(1): 83-96
- Kogut, Bruce and Gordon Walker, 2001, "The Small World of Germany and the Durability of National Networks", *American Sociological Review* 66(3): 317-335
- Kossinetz, Gueorgi and Duncan J. Watts, 2009, "Origins of Homophily in an Evolving Social Network", *American Journal of Sociology* 115(2): 405-450
- Kuhn, Thomas, 1962, *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press
- Kremp, Pierre-Antoine, 2010, "Innovation and Selection: Symphony Orchestras and the Construction of the Musical Canon in the United States (1879-1959)", *Social Forces* 88(3): 1051-1082
- Lazarsfeld, Paul, and Robert K. Merton, 1954, "Friendship as a Social Process: A Substantive and Methodological Analysis" in Morroe Berger, Theodore Abel, and Charles H. Page (eds.) *Freedom and control in modern society*, ed., New York: Van Nostrand, pp. 18-66
- Lazer, David, Brian Rubineau, Carol Chetkovich, Nancy Katz and Michael Neblo, 2010, "The Coevolution of Networks and Political Attitude", *Political Communication* 27: 248-274
- Lamont, Michéle and Virág Molnár, 2002, "The Study of Boundaries in the Social Sciences", *American Review of Sociology* 28: 167-195

- Landauer, Thomas K., Peter. W. Foltz, and Darrell Laham, 1998, "A Introduction to Latent Semantic Analysis", *Discourse Processes*, 25: 259-284
- Leahey, Erin, 2007, "Not by Productivity Alone: How Visibility and Specialization Contribute to Academic Earnings", *American Sociological Review* 72(4): 533-561
- Lena Jennifer C and Richard A. Peterson, 2008, Classification as Culture: Types and Trajectories of Music Genres, *American Sociological Review* 73(5): 697-718
- Lévi-Strauss, Claude, 1963, *Structural Anthropology*, New York: Basic Books
- Lieberman, Matthew D., Ruth Gaunt, Daniel T. Gilbert, and Yaccov Trope, 2002, "Reflection and Reflexion: A Social Cognitive Neuroscience Approach to Attributional Inference", *Advances in Experimental Social Psychology* 34:199–249
- Linzmayr, Owen W., 2004, *Apple Confidential 2.0: The Definitive History of the World's Most Colorful Company*, San Francisco: No Starch Press
- Liu, Ka-Yuet, Marissa King and Peter S. Bearman, 2010, "Social Influence and the Autism Epidemic", *American Journal of Sociology* 115(5): 1387-1434
- Lizardo, Omar, 2006, "How Cultural Tastes Shape Personal Networks", *American Sociological Review* 71(5): 778-807
- Lizardo, Omar and Michael Strand, 2010, "Skills, Toolkits, Contexts and Institutions: Clarifying the Relationship between Different Approaches to Cognition in Cultural Sociology", *Poetics* 38: 204-227
- Lopes, Paul, 2002, *The Rise of the Jazz Art World*, Cambridge: Cambridge University Press
- Lopes, Paul, 2005, "Signifying Deviance and Transgression: Jazz in the Popular Imagination", *American Behavioral Scientist* 48(11): 1468-1481
- Lynn, Freda B., Joel M. Podolny and Lin Tao, "A Sociological (De)Construction of the Relationship between Status and Quality", *American Journal of Sociology* 115(3): 755-804
- Lounsbury, Michael and Hayagreeva Rao, 2004, "Sources of Durability and Change in Market Classifications: A Study of the Reconstitution of Product Categories in the American Mutual Fund Industry, 1944–1985", *Social Forces* 82 (3): 969-999
- Macrae, C. Neil, and Galen V. Bodenhausen, 2001, "Social Cognition: Categorical Person Perception", *British Journal of Psychology* 92(1): 239-255
- MacKenzie, Donald, 2011, "The Credit Crisis as a Problem in the Sociology of Knowledge", *American Journal of Sociology* 116(6): 1778-1841
- MacKenzie, Donald and Yuval Millo, 2003, "Constructing a Market, Performing Theory: The Historical Sociology of a Financial Derivatives Exchange", *American Journal of Sociology* 109 (1):107-145.
- Malkiel, Burton G., 2003, "The Efficient Market Hypothesis and Its Critics", *Journal of Economic Perspectives*, 17(1): 59-82

- Mark, Noah P., 2002, "Cultural Transmission, Disproportionate Prior Exposure, and the Evolution of Cooperation", *American Sociological Review* 67(3): 323-344
- Marsden, Peter V., 1987, "Core Discussion Networks of Americans", *American Sociological review* 52(1): 122-131
- Martin, John L., 2000, "The Relation of Aggregate Statistics in Beliefs to Culture and Cognition", *Poetics*, 28: 5-20
- Martin, John L., 2002, "Power, Authority, and the Constraint of Belief Systems", *American Journal of Sociology*, 107(4): 861-904
- Martin, John L., 2003, "What is Field Theory?", *American Journal of Sociology*, 109(1): 1-49
- Mark, Noah P., 1998, "Beyond Individual Differences: Social Differentiation from First Principles", *American Sociological Review* 63(3): 309-330
- Mark, Noah P., 2003, "Culture and Competition: Homophily and Distancing Explanations for Cultural Niches", *American Sociological Review* 68(3): 319-345
- Mark, Noah P., Lynn Smith-Lovin and Cecilia L. Ridgeway, 2009, "Why Do Nominal Characteristics Acquire Status Value? A Minimal Explanation for Status Construction", *American Journal of Sociology* 115(3): 832-862
- Markowitz, Harry M., 1991, "Foundations of Portfolio Theory", *The Journal of Finance*, 46(2): 469-477
- Marx Ferree, Myra, 2003, "Resonance and Radicalism: Feminist Abortion Discourses in Germany and the United States." *American Journal of Sociology* 109 (2): 304-344
- McPherson, Miller and Lynn Smith-Lovin, 1987, Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups, *American Sociological Review* 52(3): 370-379
- McPherson, Miller, Lynn Smith-Lovin, and James M Cook, 2001, Birds of a Feather: Homophily in Social Networks, *Annual Review of Sociology* 27:415-444
- Mercken, Liesbeth, Tom A.B. Snijders, Christian E.G. Steglich, E. Vartiainen and H. de Vries, "Dynamics of Adolescent Friendship Networks and Smoking Behavior", *Social Networks* 32: 72-81
- Meyer, John W. and Brian Rowan, 1977, Institutionalized Organizations: Formal Structure as Myth and Ceremony, *American Journal of Sociology* 83(2): 340-363
- Mische, Ann and Philippa Pattison, 2000, "Composing a civic arena: Publics, Projects, and Social Settings", *Poetics* 27(2-3): 163-194
- Mohr, John W., 1998, "Measuring Meaning Structures", *Annual Review of Sociology* 24: 345-370
- Mohr, John W, and Harrison White, 2008, How to Model an Institution, *Theory and Society*, 37(5): 485-512
- Mohr, John W. and Vincent Duquenne, 1997, The Duality of Culture and Practice: Poverty Relief in New York City, 1888-1917, *Theory and Society* 26(2): 305-356

- Molotch, Harvey, 2003, *Where Stuff Comes From: How Toasters, Toilets, Cars, Computers, and Many Other Things Come to Be As They Are*, London: Routledge
- Montgomery, James D., 2011, "The Population Dynamics of Black-White-Mulatto Racial Systems", *American Journal of Sociology* 117(1): 46-89
- Moody, James, Daniel A. McFarland and Skye Bender-DeMoll, 2005, "Dynamic Network Visualization", *American Journal of Sociology* 110(4): 1206-1241
- Muniz, Albert M. and Thomas C. O'Guinn, 2001, "Brand Community", *Journal of Consumer Research* 27(4): 412-432
- Naccache, Lionel and Stanislas Dehaene, 2001, "Unconscious Semantic Priming Extends to Novel Unseen Stimuli", *Cognition* 80:215-229
- Negro, Giacomo, Michael T. Hannan and Hayagreeva Rao, 2011, "Category Reinterpretation and Defection: Modernism and Tradition in Italian Winemaking", *Organization Science* 22(6): 1449-1463
- Newman, Mark E. J., 2002, "Assortative Mixing in Networks", *Physical Review Letters* 89, 208701
- Newman, Mark E. J., 2004, "Analysis of Weighted Network", *Physical Review E* 70, 056131
- Newman, Mark E. J., 2006, "Finding Community Structure in Networks Using the Eigenvectors of Matrices", *Phys. Rev. E* 74, 036104
- Nickerson, David W., 2008, "Is Voting Contagious? Evidence from Two Field Experiments", *American Political Science Review* 102: 49-57
- O'Mahony, Siobhan and Beth A. Bechky, 2006, "Stretchwork: Managing the Career Progression Paradox in External Labor Markets", *The Academy of Management Journal* 49(5): 918-941
- O'Reilly, Charles A., J. Bruce Harreld, and Michael L. Tushman, "Organizational Ambidexterity: IBM and Emerging Business Opportunities," *California Management Review* 51(4): 75-99
- Pachucki, Mark A., and Ronald L. Breiger, 2010, "Cultural Holes: Beyond Relationality in Social Networks and Culture", *Annual Review of Sociology*, 36:205-224
- Padgett, John F. and Christopher K. Ansell, 1993, "Robust Action and the Rise of the Medici, 1400-1434", *American Journal of Sociology*, 98(6): 1259-1319
- Padgett, John F. and Paul D. McLean, 2006, "Organizational Invention and Elite Transformation: The Birth of Partnership Systems in Renaissance Florence", *American Journal of Sociology* 111(5): 1463-1568
- Peterson, Richard, 1992, "Understanding Audience Segmentation: From Elite and Mass to Omnivore and Univore", *Poetics* 21: 243-258
- Peterson, Richard, 1997, "The Rise and Fall of Highbrow Snobbery as Status Marker", *Poetics* 25: 75-92
- Peterson, Richard, 1997, *Creating Country Music: Fabricating Authenticity*, Chicago: University of Chicago Press
- Peterson, Richard and Roger Kern, 1996, "Changing Highbrow Taste: From Snob to Omnivore" *American Sociological Review*, 64(5): 900-907
- Phillips, Gene D, 1990, *Major Film Directors of the American and British Cinema*, London: Associated University Press

- Phillips, Damon J. and Ezra W. Zuckerman, "Middle-Status Conformity: Theoretical Restatement and Empirical Demonstration in Two Markets", *American Journal of Sociology* 107(2): 379–429
- Podolny, Joel M., 1993, "A Status-Based Model of Market Competition", *American Journal of Sociology* 98(4): 829-872
- Podolny, Joel M., 2001, "Networks as the Pipes and Prisms of the Market", *American Journal of Sociology* 107(1): 33-60
- Podolny, Joel M., 2005, *Status Signals: A Sociological Study of Market Competition*, Princeton University Press
- Podolny, Joel M. and Toby E. Stuart, 1995, "A Role-Based Ecology of Technological Change", *American Journal of Sociology* 100(5): 1224-1260
- Porter, Mason A., Jukka-Pekka Onnela, and Peter J. Mucha, 2009, "Communities in Networks", *Notices of the American Mathematical Society* 56(9): 1082-1097
- Powell, Walter W., Douglas R. White, Kenneth W. Koput and Jason Owen-Smith, 2005, "Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences", *American Journal of Sociology* 110(4): 1132-1205
- Pugh, Allison J., 2009, *Longing and Belonging: Parents, Children, and Consumer Culture*, Berkeley and Los Angeles: University of California Press
- Pullum, Geoffrey K., 1991, "The Great Eskimo Vocabulary Hoax", in Pullum, Geoffrey K., *The Great Eskimo Vocabulary Hoax and Other Irreverent Essays on the Study of Language*, University of Chicago Press: 159-171
- Rao, Hayagreeva, Philippe Monin and Rodolph Durand, 2005, "Border Crossing: Bricolage and the Erosion of Categorical Boundaries in French Gastronomy", *American Sociological Review* 70(6): 968-991
- Rand, Ayn, 1957, *Atlas Shrugged*, New York: Random House
- Ridgeway, Cecilia, 1991, "The Social Construction of Status Value: Gender and Other Nominal Characteristics", *Social Forces* 70(2): 367-386
- Ritvo, Harriet, 1989, *The Animal Estate: The English and Other Creatures in the Victorian Age*, Cambridge: Harvard University Press
- Ritvo, Harriet, 1997, *The Platypus and the Mermaid, and Other Figments of the Classifying Imagination*, Cambridge: Harvard University Press
- Rivera, Mark T., Sara B. Soderstrom and Brian Uzzi, 2010, "Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms", *Annual Review of Sociology* 36: 91-115
- Rosa, José Antonio, Joseph F. Porac, Jelena Runser-Spanjol and Michael S. Saxon, 1999, "Sociocognitive Dynamics in a Product Market", *The Journal of Marketing* 63: 64-77
- Rosch, Eleanor H., 1973, "Natural Categories", *Cognitive Psychology* 4: 328-350
- Rosch, Eleanor H., 1975, "Cognitive Representation of Semantic Categories", *Journal of Experimental Psychology* 104(3): 192-233

- Rosch, Eleanor H., 1978, "Principles of Categorization", in Rosch, Eleanor H. and Barbara B. Lloyd (eds), *Cognition and Categorization*, Hillsdale: Lawrence Erlbaum, pp. 27-48
- Roy, William G. and Timothy J. Dowd, 2010, "What is Sociological About Music?", *Annual Review of Sociology*, forthcoming
- Ruef, Martin, 2000, "The Emergence of Organizational Forms: A Community Ecology Approach", *American Journal of Sociology* 106(3): 658-714
- Ruef, Martin, Howard E. Aldrich and Nancy M. Carter, "The Structure of Founding Teams: Homophily, Strong Ties, and Isolation among U.S. Entrepreneurs", *American Review of Sociology* 68(2): 195-222
- Ruef, Martin and Kelly Pattreson, 2009, "Credit and Classification: The Impact of Industry Boundaries in 19th Century America," *Administrative Science Quarterly*, 54: 486-520
- Salganik, Matthew J., Peter S. Dodds, and Duncan J. Watts, 2006, "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market", *Science* 311:854-856
- Schatzki, Theodore R., 2001, "Introduction: Practice Theory", in, Theodore R. Schatzki, Karin Knorr-Cetina and Eike von Savigny (eds.), *The Practice Turn in Contemporary Theory*, London: Routledge, pp. 10-23
- Scheff, Thomas J., 2006, *Goffman Unbound!: A New Paradigm for Social Science*, Paradigm Publishers
- Schwartzman, Luisa Farah, 2007, "Does Money Whiten? Intergenerational Changes in Racial Classification in Brazil", *American Sociological Review*, 72(6): 940-963
- Sewel, William H. Jr., 1992, "A Theory of Structure: Duality, Agency, and Transformation", *American Journal of Sociology* 98(1): 1-29
- Shwed, Uri and Peter S. Bearman, 2010, "The Temporal Structure of Scientific Consensus Formation", *American Sociological Review* 75(6): 817-840
- Simmel, Georg, 1950, *The Sociology of Georg Simmel*, New York: Free Press
- Small, Mario, 2004, *Villa Victoria: The Transformation of Social Capital in a Boston Barrio*, Chicago: Chicago University Press
- Smith, Edward Bishop, 2010, "Identities as Lenses: Audiences' Evaluation of Organizational Performance", *Administrative Science Quarterly* 56(1): 61-94
- Snijders, Tom A. B., Gerhard G. van de Bunt, and Christian E. G. Steglich, 2010, "Introduction to Stochastic Actor-Based Models for Network Dynamics", *Social Networks*, 32(1): 44-60
- Stark, David, 1996, "Recombinant Property in East European Capitalism", *American Journal of Sociology* 101(4): 993-1027
- Strand, Michael, 2011, "Where Do Classifications Come From? The DSM-III, the Transformation of American Psychiatry, and the Problem of Origins in the Sociology of Knowledge", *Theory and Society* 40: 273-313
- Strauss, Claudia, and Naomi Quinn, 1997, *A Cognitive Theory of Cultural Meaning*, Cambridge: Cambridge University Press.

- Stuart, Toby E., 2007, "The Formation of Inter-organizational Networks" in James Rauch (ed) *The Missing Links: Formation and Decay of Economic Networks*, Russell Sage Foundation
- Stuart, Toby E., Ha Hoang and Ralph C. Hybels, 1999, "Interorganizational Endorsements and the Performance of Entrepreneurial Ventures", *Administrative Science Quarterly* 44(2): 315-349
- Swidler, Ann, 1986, "Culture in Action: Symbols and Strategies", *American Sociological Review* 51(2): 273-286
- Swidler, Ann, 2001, *Talk of Love: How Culture Matters*, Chicago: University of Chicago Press
- Tan, Pang-Ning, Michael Steinbach and Vipin Kumar, 2006, *Introduction to Data Mining*, Addison Wesley
- Tavory, Iddo, 2010, "Of Yarmulkes and Categories: Delegating Boundaries and the Phenomenology of Interactional Expectation", *Theory and Society*, 39(1): 49-68
- Tong, Stephanie Tom, Brandon Van Der Heide, Lindsey Langwell and Joseph B. Walther, 2008, "Too Much of a Good Thing? The Relationship Between Number of Friends and Interpersonal Impressions on Facebook", *Journal of Computer-Mediated Communication* 13(3): 531-549
- Tilly, Charles, 1999, *Durable Inequality*, Berkeley: University of California Press
- Turner, Fred, 2006, *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*, Chicago: Chicago University Press
- Utz, Sonja, 2010, "Show me your friends and I will tell you what type of person you are: How one's profile, number of friends, and type of friends influence impression formation on social network sites", *Journal of Computer-Mediated Communication* 15(2): 314-335
- Uzzi, Brian and Jarrett Spiro, 2005, Collaboration and creativity: The small world problem, *American Journal of Sociology* 111(2): 447-504
- Vaisey, Stephen, 2009, "Motivation and Justification: A Dual-Process Model of Culture in Action", *American Journal of Sociology* 114(6): 1675-1715
- Vaisey, Stephen and Omar Lizardo, 2010, "Can Cultural Worldviews Influence Network Composition?", *Social Forces* 88(4): 1595-1618
- Veblen, Thorstein, 1991, *The Theory of the Leisure Class*, Fairfield, NJ: A. M. Kelly
- Wasserman, Stanley, and Katherine Faust, 1994, *Social Network Analysis: Methods and Applications*, Cambridge: Cambridge University Press
- Watts, Duncan J., 1999, Networks, Dynamics, and the Small-World Phenomenon, *American Journal of Sociology* 105(2): 493-527
- Watts, Duncan J., 2002, "A Simple Model of Global Cascades on Random Networks", *Proceedings of the National Academy of Science* 99(9): 5766-5771
- Watts, Duncan J., 2003, *Six Degrees: The Science of a Connected Age*, New York: W. W. Norton & Company
- Watts, Duncan J. and Steven H. Strogatz, 1998, "Collective Dynamics of 'Small-World' Networks", *Nature* 393: 440-442

- West, Candace and Don H. Zimmerman, 1987, *Doing Gender*, *Gender and Society*, 1(2): 125-151
- White, Harrison C., 1992, *Identity and Control: A Structural Theory of Social Action*, Princeton: Princeton University Press
- White, Harrison C., 2008, 'Notes on the constituents of social structure – Soc. Rel. 10 – Spring '65', *Sociologica*, 1/2008
- White, Harrison C., Scott A. Boorman and Ronald L. Breiger, 1976, "Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions", *American Journal of Sociology* 81(4): 730-780
- Willer, Robb, Ko Kuwabara and Michael W. Macy, 2009, "The False Enforcement of Unpopular Norms", *American Journal of Sociology* 115(2): 451-490
- Wimmer, Andreas and Kevin Lewis, 2010, "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook", *American Journal of Sociology* 116(2): 583-642
- Wittgenstein, Ludwig, 1953, *Philosophical Investigations*, Oxford: Blackwell
- Yeung, King-To, 2005, "What Does Love Mean? Exploring Network Culture in Two Network Settings", *Social Forces* 84(1): 391-420
- Yogev, Tamar, 2010, "The Social Construction of Quality: Status Dynamics in the Market for Contemporary Art", *Socio-Economic Review* 8(3): 511-536
- Zelizer, Viviana, 1997, *The Social Meaning of Money: Pin Money, Paychecks, Poor Relief, and Other Currencies*, Princeton: Princeton University Press
- Zerubavel, Eviatar, 1997, *Social Mindscales: An Invitation to Cognitive Sociology*, Cambridge: Harvard University Press
- Zerubavel, Eviatar, 2003, *Time Maps: Collective Memory and the Social Shape of The Past*, Chicago: Chicago University Press
- Zhang, Yan, A. J. Friend, Amanda L. Traud, Mason A. Porter, James H. Fowler and Peter J. Mucha, 2008, "Community Structure in Congressional Cosponsorship Networks", *Physica A* 387(7): 1705-1712
- Zhao, Dingxin, 1998, "Ecologies of Social Movements: Student Mobilization during the 1989 Prodemocracy Movement in Beijing", *American Journal of Sociology* 103(6): 1493-1529
- Zuckerman, Ezra W., 1999, "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount", *American Journal of Sociology*, 104(5): 1398-1438
- Zuckerman, Ezra W., 2000, "Focusing the Corporate Product: Securities Analysts and De-Diversification", *Administrative Science Quarterly* 45(3): 591-619
- Zuckerman, Ezra W., 2004, "Structural Incoherence and Stock Market Activity", *American Sociological Review*, 69(3): 405-432
- Zuckerman, Ezra W., Tai-Young Kim, Kalinda Ukanwa and James von Rittmann, 2003, "Robust Identities or Nonentities? Typecasting in the Feature-Film Labor Market", *American Journal of Sociology* 108(5): 1018-1074