Essays on the Organizational Consequences of On-line Behavior of Audiences

by

Rebeka Johnson

Thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Institute of Management (IMA)

Università della Svizzera Italiana (USI), Lugano, Switzerland

August 2015

SUPERVISOR: Prof. Balázs Kovács

to my family

ABSTRACT

Over the past 2 decades, internet use has become increasingly more a part of our every-day lives. We communicate with our friends and colleagues using the internet, we work using the internet, we also shop using the internet. We learn and increase our knowledge from information available on the internet. While on the one hand, we advance from the instant access to online contexts individualistically, on the other hand we participate as members of a community for example when we share our experiences online.

The ever growing use of the internet and its flourish in new segments of our daily life brings significant changes not only to us, the individuals, but also to the organizations. In the past decade, there has been a shift in the field of organizational theory considering the environment of organizations. Current approaches extend the horizon of the classical view proclaiming that organizational environment is not only constituted by rival organizations but also their audience members. Several studies found evidence that audience members' perceptions and behavior influence organizational success. For example, category-spanning organizations on average suffer from social and economic disadvantages in markets because they cannot meet the expectations of their audiences. This shift towards understanding the effect of audience responses on the organizational outcome motivates my dissertation. More specifically, I study how individuals on-line behavior affects organizations. I analyze three aspects of internet mediated communication and their consequences to the organization.

Firstly, I address the need to compare how traditional face-to-face communication compares to the modern email communication (Chapter 2). Studies tend to take it for granted, that on-line information exchange mirrors it's off-line counterpart at the work place. Although, there are great advantages in the availability of email data, as it retains communication in its completeness, it does not fully correspond to previously studied relations, like friendship or advice seeking. The characteristics of on-line communications also differ from off-line information exchange. Employees respect divisional and hierarchical boundaries in face-to-face conversations while these boundaries are blurred out within the email exchange .

Secondly, I analyze a special type of on-line behavior, the on-line word-of-mouth communication among audience members (Chapter 3). Online reviews play an increasingly important role in shaping organizational performance. Drawing conclusions on how customers perceive quality and typicality of a producer and how it manifests in on-line ratings increase the predictability of producer success.

Thirdly, I approach audience behavior from a collective behavior perspective (Chapter 4). Specifically, I analyze audience dynamics with threshold models. Doing so I address the micro level mechanism of how audience behavior creates certain macro level patterns of producer success rather than assuming that they are simple aggregates of individual characteristics.

Contents

List of Tables

List of Figures

Acknowledgement

Although I am the one who wrote the sentences of this dissertation, it embraces the work of many others. I would like to attribute it to the people, whose work remains invisible behind the lines but without which, I wouldn't have been able to have written this thesis.

Firstly, I would like to express how grateful I am for the values I have learned from my family throughout my upbringing. They taught me the value of knowledge, the belief in one's self, the values of hard work, and the ability to turn challenges into victories, among others. I believe my devotion towards research is in a big part due to the very inspiring conversations I had with two of my relatives, Mami and Papa, my grandparents, whom I miss a lot.

Secondly, I would like to say thank you to my graduate school professor, Prof. Tamás Rudas and to my former boss and co-author of one of my papers, András Vicsek, who both encouraged me to continue my studies in a form of a Ph.D. It meant a lot to me, that they expressed their opinion that I would be a good candidate for this career line. When times were hard, I was able to draw a lot of energy from their belief in me.

Thirdly, I am very grateful to my supervisor, Prof. Balázs Kovács, who put his trust in me. I feel that I am very fortunate to have such an excellent supervisor, who has continued to stand as a role model for me. Besides his extensive knowledge, experiencing his devotion to research and his commitment to teaching has had a very positive influence on me. He also never kept me waiting when I had a question - work related or otherwise, he was always available for me, which I highly appreciate.

Fourthly, I am thankful to Prof. Filippo Carlo Wezel who organized the Ph.D. program. Prof. Wezel spent much time on reaching out to top scholars to invite them to share their knowledge with the students. I believe that the quality of the program and his inspiring and insightful feedback helped me grow into a better scholar.

Fifthly, I would like to mention the role of my friends and fellow Ph.D's. in my dissertation. I am grateful for the support I received from them, for the long discussions we had on our work or on other life matters, and also for showing me the fun side of the otherwise hard-working Ph.D.

life. Thank you Judit, Parvaz, Martina, Zsófi, Chanchal, Sayed, Ivona, Zsuzsi, Manos, Margarita and Lívia.

Next, focusing on the work after I became a mother was increasingly difficult, but our amazing team of grandparents did everything in their reach to help me achieve my goal.

Lastly, I am greatly indebted to Andrew, my husband. He supported me throughout the years not only with his unconditional love, but also helping me to find the time for my work when it was needed (even if it meant that he has to provide the dinner, help with the housework, or tend to our children). He took an active part in my thesis too. He helped me immensely with the data collection of the third paper. He thought me how to use perl programming language on a more advanced level, and helped me with tailoring my script according to the research question.

Thank you very much for all of you!

Chapter 1

Introduction

1.1 Background

1.1.1 Attention shift to the audience members

The question as to why a producer becomes successful trails a long history in organizational research. Typically scholars argued that the producer's choices contain the key to success. Examples of such choices are, which industry segment to enter (Schmalensee, 1985; Rumelt, 1991), whether to enter into a highly competitive market or into a less competitive one (Carroll, 1985a), producing single or multiple products (Hoskisson and Hitt, 1988; Teece et al., 1994), becoming a specialist or generalist (Hannan et al., 2007b) and which other organizations to choose to enter a into relationship with (Gulati, 1995; Powell et al., 1996; Uzzi and Lancaster, 2004).

In recent years this focus on the producer shifted towards their audience. Studies of legitimacy showed that audience members play a key role in shaping organizations such that audience members put pressure on the producer to conform and fulfill their shared beliefs of the product (Suchman, 1995; Zuckerman, 1999a; McKendrick and Carroll, 2001b; Hsu, 2006c). Hsu found evidence that producers who meet the expectations of their audience are more successful (Hsu, 2006c).

This shift in the literature led to further, more detailed research on how audience perception of the producers' identity influences the producers' success . These studies however mainly focus on a certain segment of the audience - on the critics claiming that experts as third parties of “mediated markets” (Zuckerman, 1999a) through their distinguished role affect other audience members’ attention and evaluation patterns (Hsu, 2006c).

Critics as third parties of “mediated markets” (Zuckerman, 1999a) through their distinguished role affect other audience members’ attention and evaluation patterns (Hsu, 2006c). Zuckerman argues that the role of a professional critic is to account for the magnitude of mismatch between the product and its social identity. The critic perceives a product illegitimate and devalues it if the product doesn’t fit well into the expectations of the audience. Zuckerman refers to the critics as the primary audience whose behavior influences the secondary audience, the actual buyer’s behavior. There has been research conducted on other domains such as the film industry (Zuckerman and Kim, 2003; Hsu, 2006c) or French cuisine (Rao et al., 2003b, 2005) documenting similarly to Zuckerman’s study on stock prices (Zuckerman, 1999a) that intermediaries can serve as good estimates of producer success.

There are a few reasons however why information purely on intermediaries cannot serve as perfect proxies of producer success. Firstly, Zuckerman and colleagues points it out that these studies disregard the structure of the primary audience (Zuckerman et al., 2003b), the structure of the actual buyers, which could lead to different effects. A product has to meet the expectations of all its audience, who are coming from different backgrounds with different tastes, holding different opinions of the product in question and towards the category to which the product belongs. These expectations can be highly similar among the audience members leading to consolidation, and, at the other extreme, each member might hold diverse expectations (Zuckerman et al., 2003b) leading to different evaluation. Secondly, Critics may be pressured by both producer and audience side (Zuckerman and Kim, 2003) and by their personal preferences or assigned medium (Zuckerman and Kim, 2003), which further influences their evaluation. Thirdly, researches relying on data of critics so far have not shed light on the diverging influence intermediaries have on their audience. Zuckerman and Kim argue for example that for lowbrow movies, critics have less influence on the audience than for highbrow films. Fourthly, there might be specific audience segments, which do not follow intermediaries opinion and there are markets where intermediaries do not exist, for example on the crowdsourcing market. Finally, studies prove that intermediaries often tend to specialize in certain categories, ignoring products in multiple categories (Zuckerman, 1999a) and leaving the innovating attempts solely to the secondary audiences' judgment.

Recent studies responding to this critic extend their scope to follow full audience populations. Studies on the role of intermediaries in product success not only brought the secondary audience in the spotlight, but also emphasized the interconnectedness of audience members' decisions.

1.1.2 On-line behavior

A few years ago The Economist magazine (July 1999)[1] wrote that "the internet (will) change everything - the way we work, the way we learn and play, even maybe, the way we sleep." By now, nobody would disagree that this prediction has come true. What the Economist magazine did not predict in 1999, however, was that computer mediated communication shapes not only everyday behavior but it also fundamentally influences organizations.

Individual's on-line behavior influences organizations in two aspects. Not only that internet eliminated the geography and time constrained boundaries of communication (Duan et al., 2008), but it also organizes individuals into loosely linked communities. Individuals communicate not only on a personal basis, but they exchange their opinions and experience through the world wide web.

Organizations offering on-line opinion sharing services to their audience on the one hand receive direct feedback from them, on the other hand use it as a marketing opportunity where audience members may influence each other through word-of-mouth . Since Amazon.com customer review system became exceedingly popular the number of business with similar services have since risen vastly (Chevalier and Mayzlin, 2006). Websites appeared in recent years, with the sole reason to publish reviews, for example TripAdvisor, Yelp, or Booking.com.

Access to users and buyers opinion before product purchase changed the dynamics of product evaluation (Putsis Jr et al., 1997; Godes and Mayzlin, 2004; Salganik et al., 2006; Chevalier and Mayzlin, 2006; Dellarocas et al., 2007; Luca, 2011b). Luca (2011b) demonstrates for example that

[1] http://www.economist.com/node/215657

one star increase on the Yelp reviewer site may increase 5-9% the restaurant's income. Chevalier and Mayzlin (2006) finds evidence that review content influences future purchases of prospective buyers altering the overall sale indexes. Liu (2006) shows the positive relationship between the size of TV show audiences and online reviews. "Online user reviews have become an important source of information to consumers, substituting and complementing other forms of business-to-consumer and offline word-of-mouth communication about product quality" argues Chevalier and Mayzlin (2006, page 1.). The number of consumers, who pay attention to online reviews prior to purchasing products is increasing every day. Kee (2008) finds that 68% of Internet users read at least four reviews before product purchase (Godes and Mayzlin, 2004). BizRate questioned 5,500 on-line consumers and found that, 44% of respondents consult reviewer sites before buying a product and even more, 59% of the respondents considered reviews written by their peers more valuable than the ones by experts (Piller, 1999). "Online product reviews provided by consumers who previously purchased products have become a major information source for consumers and marketers regarding product quality", states Hu et al. (2008, page 201.).

Although, researchers undoubtedly find correlation between performance and online reviews, the underlying mechanism between the two is not yet clarified. On the one hand audience valuations influence product performance, while on the other hand product performance influences audience valuations too. Researchers studying only one side of the mechanism cannot differentiate between weather an increase in the product performance is due to higher quality or due to the effect of social influence (Duan et al., 2008).

The appearance of reviewer sites provides the possibility on the one hand to study audience populations (for example Hsu, 2006c; Kovacs and Johnson, 2013) to a more complete extent and highlights the challenge and on the other hand to understand the interconnectedness of audience behavior (Godes and Mayzlin, 2004; Chevalier and Mayzlin, 2006; Salganik et al., 2006; Salganik and Watts, 2008; Lee and Bradlow, 2011; Zuckerman, 2012).

1.2 Motivation

In the above section I intended to highlight the areas where audience members' internet-mediated communication has a direct influence on the organization. In the next part of my dissertation I aim to answer three research questions which motivated my empirical investigations.

In the first study I argue that the rapid growth in on-line communication provides scholars with a goldmine of available data, which allows them to map the structure of communities. However, data from on-line sources does not reflect perfectly previously studied "off-line" relationships. I compare off-line to on-line interactions using exponential random graph (ERG) models to understand what aspect face-to-face communication approximates it's on-line counterpart. I analyze the roles of the two different communication channels in reflecting organizational boundaries, organizational hierarchies and gender.

Secondly, studying online reviews I investigate why producers with multiple categories receive lower ratings from reviewers than single-category producers. In the empirical investigation I directly test and disentangle typicality-based and quality-based explanations for the negative consequences of category spanning.

Thirdly, I analyze the interconnectedness of audience decisions of an online crowdsourcing platform. I address the micro level mechanism of how audience behavior creates certain macro level patterns of producer success rather than assuming that they are simple aggregates of individual characteristics. I approach audience members' behavior from a collective behavioral perspective. Specifically, I analyze audience dynamics with threshold models.

1.3 Outline of the thesis

In the following sections I am going to introduce the above detailed studies in depth. Chapter 2 and Chapter 3 are based on the papers that I had published with my co-authors:

Johnson, R., Kovács, B., & Vicsek, A. (2012). A comparison of email networks and off-line social networks: A study of a medium-sized bank. *Social Networks, 34(4)*, 462-469.

Kovács, B., & Johnson, R. (2013). Contrasting alternative explanations for the consequences of category spanning: A study of restaurant reviews and menus in San Francisco. *Strategic Organization, 12*, 7-37.

Chapter 4 is a working paper constituted by unpublished material. In Chapter 5 I draw general theoretical conclusions based on the three empirical studies, and propose arenas for further research.

References

Carroll, G. R. (1985a). Concentration and specialization: Dynamics of niche width in populations of organizations. *American Journal of Sociology*, 90:1262–1283.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354.

Dellarocas, C., Zhang, X. M., and Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive marketing*, 21(4):23–45.

Duan, W., Gu, B., and Whinston, A. B. (2008). Do online reviews matter?—an empirical investigation of panel data. *Decision support systems*, 45(4):1007–1016.

Godes, D. and Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing science*, 23(4):545–560.

Gulati, R. (1995). Does familiarity breed trust? the implications of repeated ties for contractual choice in alliances. *Academy of management journal*, 38(1):85–112.

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

Hoskisson, R. E. and Hitt, M. A. (1988). Strategic control systems and relative r&d investment in large multiproduct firms. *Strategic Management Journal*, 9(6):605–621.

Hsu, G. (2006c). Jacks of all trades and masters of none: Audiences reactions to spanning genres in feature film production. *Administrative Science Quarterly*, 51(3):420–450.

Hu, N., Liu, L., and Zhang, J. J. (2008). Do online reviews affect product sales? the role of reviewer characteristics and temporal effects. *Information Technology and Management*, 9(3):201–214.

Kovacs, B. and Johnson, R. (2013). Contrasting alternative explanations for the consequences of category spanning: A study of restaurant reviews and menus in san francisco. *Strategic Organization*, 12(7-37).

Lee, T. Y. and Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, 48(5):881–894.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of marketing*, 70(3):74–89.

Luca, M. (2011b). Reviews, reputation, and revenue: The case of Yelp.com. Technical report, Harvard Business School Working Paper, No. 12-016.

McKendrick, D. G. and Carroll, G. R. (2001b). On the genesis of organizational forms: Evidence from the market for disk drive arrays. *Organization Science*, 12:661–683.

Piller, C. (1999). Everyone is a critic in cyberspace. *Los Angeles Times*.

Powell, W. W., Koput, K. W., and Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative science quarterly*, pages 116–145.

Putsis Jr, W. P., Balasubramanian, S., Kaplan, E. H., and Sen, S. K. (1997). Mixing behavior in cross-country diffusion. *Marketing Science*, 16(4):354–369.

Rao, H., Monin, P., and Durand, R. (2003b). Institutional change in toque ville: Nouvelle cuisine as an identity movement in french gastronomy1. *American journal of sociology*, 108(4):795–843.

Rao, H., Monin, P., and Durand, R. (2005). Border crossing: Bricolage and the erosion of categorical boundaries in French gastronomy. *American Sociological Review*, 70(6):968–991.

Rumelt, R. (1991). How much does industry matter? *Strategic Management Journal*, (12):167–185.

Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856.

Salganik, M. J. and Watts, D. J. (2008). Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. *Social Psychology Quarterly*, 71(4):338–355.

Schmalensee, R. (1985). Do markets differ much? *The American Economic Review*, pages 341–351.

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of management review*, 20(3):571–610.

Teece, D. J., Rumelt, R., Dosi, G., and Winter, S. (1994). Understanding corporate coherence: Theory and evidence. *Journal of Economic Behavior & Organization*, 23(1):1–30.

Uzzi, B. and Lancaster, R. (2004). Embeddedness and price formation in the corporate law market. *American sociological review*, 69(3):319–344.

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

Zuckerman, E. W. (2012). Construction, concentration, and (dis) continuities in social valuations. *Annual Review of Sociology*, 38:223–245.

Zuckerman, E. W. and Kim, T.-Y. (2003). The critical trade-off: identity assignment and box-office success in the feature film industry. *Industrial and Corporate Change*, 12(1):27–67.

Zuckerman, E. W., Kim, T.-Y., Ukanwa, K., and von Rittmann, J. (2003b). Robust identities or nonentities? typecasting in the feature-film labor market1. *American Journal of Sociology*, 108(5):1018–1073.

Chapter 2

A comparison of email networks and off-line social networks: A study of a medium-sized bank*

Abstract

Recently there has been a surge in the availability of online data concerning the connections between people, and these online data are now widely used to map the social structure of communities. There has been little research, however, on how these new types of relational data correspond to classical measures of social networks. To fill this gap, we contrast the structure of an email network with the underlying friendship, communication, and advice seeking networks. Our study is an explorative case study of a bank, and our data contains emails among employees and a survey of the ego networks of the employees. Through calculating correlations with QAP standard errors and estimating exponential random graph (ERG) models, we find that although the email network is related to the survey-based social networks, emails networks are also significantly different: while off-line social networks are strongly shaped by gender, tenure, and hierarchical boundaries, the role of these boundaries are much weaker in the email network.

*with Balázs Kovács and András Vicsek

2.1 Introduction

The rapid growth in online communication provides social networks scholars with a goldmine of data. Previously, to obtain information on the social structure of communities, researchers had to conduct survey-based, observational, or archival research. Today, however, data on electronic networks are readily available. This surge in data availability allows network scholars to explore the structure of communities through emails (Grippa et al., 2006; Kleinbaum et al., 2008; Quintane and Kleinbaum, 2008, 2011), phone logs (Eagle et al., 2010), or online collaborations (Stewart, 2005).

Relying on online data to map the structure of communities is not necessarily a straightforward exercise. Researchers now use email networks as if they were perfect mirrors of communication (or other) networks (e.g., Grippa et al., 2006; Kleinbaum et al., 2008). But this is assumption is questionable. Researchers need to understand how these online connections relate to previous "off-line" measures of relations and networks (Menchik and Tian, 2008; Matzat and Snijders, 2010). This paper provides an explorative case study, in which we compare the email network of a bank with the corresponding self reported friendship networks, information networks, and advice networks.

This paper adds to a few existing studies on the topic. Here, we briefly overview these studies in order to indicate how our current research extends them. Grippa et al. (2006) question the use of email data as a proxy for traditional social network analysis. In order to understand whether email parsing provides similar network structures as traditional social network analysis surveys do, Grippa et al. (2006) study the communication among employees of a small graduate school and compare network properties such as density, betweenness centrality, and core/periphery structure of various forms of communication. They find that email networks overestimate the communication of peers with technological expertise, but underestimate communication between the core and the periphery. In this article, we extend Grippa and colleagues' analyses with the inclusion of attribute data and multiple survey-based social network data. Besides the descriptive analysis of the network, we use exponential random graph (ERG) models to estimate when and how email networks and networks based on traditional sociometric methods coincide or diverge.

Quintane and Kleinbaum (2011) compare email data with survey reports on social networks, and find that the email network corresponds to the communication network as recalled by the participants. Although their research question is similar to ours, this paper extends their results in two important aspects. First, the survey-based ego-network they analyze is a communication network, while we collected network data on a wider range of relations: based on a more extensive set of survey questions (which we shall describe later), we are able to differentiate three main types of connections between employees: friendship, communication, and advice. We believe it is important to have a multidimensional measure of social networks because the extent to which emails correspond to "off-line" ties might depend largely on what kind of ties we are interested in. Second, we would like to test their results on a larger dataset. Quintane and Kleinbaum's analysis of surveys is a small sample (n=23), which might be one reason why many of their ERGM coefficients are insignificant. Our dataset contains a significantly larger number of observations.

The structure of the paper is as follows. First, we describe the empirical setting and the network data in detail, with emphasis on how we collected and transformed the variables and with a detailed discussion on the issue of missing observations. Second, we provide descriptive statistics on the networks and compare the networks using network correlations. Third, we estimate exponential random graph models on the networks, and compare what mechanisms drive the email, friendship, advice, and communication networks. Finally, we discuss the findings and their implications for research in social networks, and communications.

2.2 Data

We collected our data from a Central European bank. The dataset contains attribute information about the employees, email messages among the employees, and a sociometric survey. The company provided us attribute information about all employees who worked at the headquarters of the firm at the time of data collection (N=1,311), such as gender, job function, tenure, age and hierarchical level. The company also provided us the email activity of these employees for a three-week period. We conducted a sociometric survey, which was completed by 761 employees (overall response rate:

57%).

For purposes of this analysis, we select a subsample of the data. In order to minimize the possible biases from missing data (Holland and Leinhardt, 1973; Kossinets, 2006), we selected the three organizational departments with the highest response rates on the sociometric survey[1]. There is no missing data on the attributes of the employees and on the emails. The average response rate in these three departments is 67% (out of 168 employees in the three departments, 113 completed the survey). Table 2.1 contains the descriptive statistics of the sample, and compares the set of respondents with all the employees of these three departments. As the table shows, the characteristics of the respondents closely match the characteristics of the non-respondents (this and following results are based on chi-squared tests of proportion under multinomial sampling assumptions). We also check whether the three chosen departments are typical departments of the bank in terms of age, gender, tenure, and hierarchical composition. Table 2.1 illustrates that the three chosen departments do not differ significantly along most attributes of the employees from the population of all departments. The only exception is gender distribution: males are overrepresented in one of the three highest response-rate departments (the department with the second highest response rate happens to be the IT department, and males are traditionally overrepresented in IT departments). There is no gender bias in response rate within the departments, however. Therefore, although the response rate is the selected sample is only 67%, due to the fact that the sample matches the population on the main attributes, we have no reason to believe that non-response biases our model estimates[2].

The mean age of employees in our sample is 32 years; the mean tenure is 3.7 years; 34 employees were hired within one year; the longest tenure is 15 years. The employees in the analyzed departments are all located in the same building (which is important in the light of Marmaros and

[1]When deciding how many of the highest response rate departments to include into the analysis, we had to balance two opposing criteria: on one hand, by including more departments in the sample we would increase the number of observations; on the other hand, including more departments decreases the overall response rate, which we wanted to avoid. Eventually, we decided to include the three departments with the highest response rates because this sample provides a relatively large number of observations and at the same time keeps the response rate relatively high.

[2]We also investigated whether ERG estimates are biased, but our analyses (not shown here) do not indicate the presence of a strong bias.

Table 2.1: Demographic composition of the three departments included in the analysis, and their comparison with the whole organization.

		Employees who completed the sociometric survey		All employees in the selected departments		All employees in the company	
		Count	Percent	Count	Percent	Count	Percent
Gender:	Male	85	75%	120	71%	599	49%
	Female	28	25%	48	29%	612	51%
Hierarchy level:	Head of department	4	4%	4	2%	36	3%
	Supervisor	18	16%	22	13%	120	10%
	Employee	91	81%	143	85%	1054	87%
Departments:	Department 1	35	31%	49	29%	35 different	departments,
	Department 2	37	33%	59	35%	with 36	employees
	Department 3	41	36%	60	36%	on average	
Age:	18-24	3	3%	5	3%	146	12%
	25-29	43	38%	58	35%	436	36%
	30-34	39	35%	55	33%	298	25%
	35-39	14	12%	23	14%	168	14%
	40+	14	12%	27	16%	162	13%
Tenure:	Within 1 year	38	34%	52	31%	421	35%
	2-4 years	37	33%	56	33%	405	33%
	More than 4 years	38	34%	60	36%	384	32%
	Total	113		168		1,211	

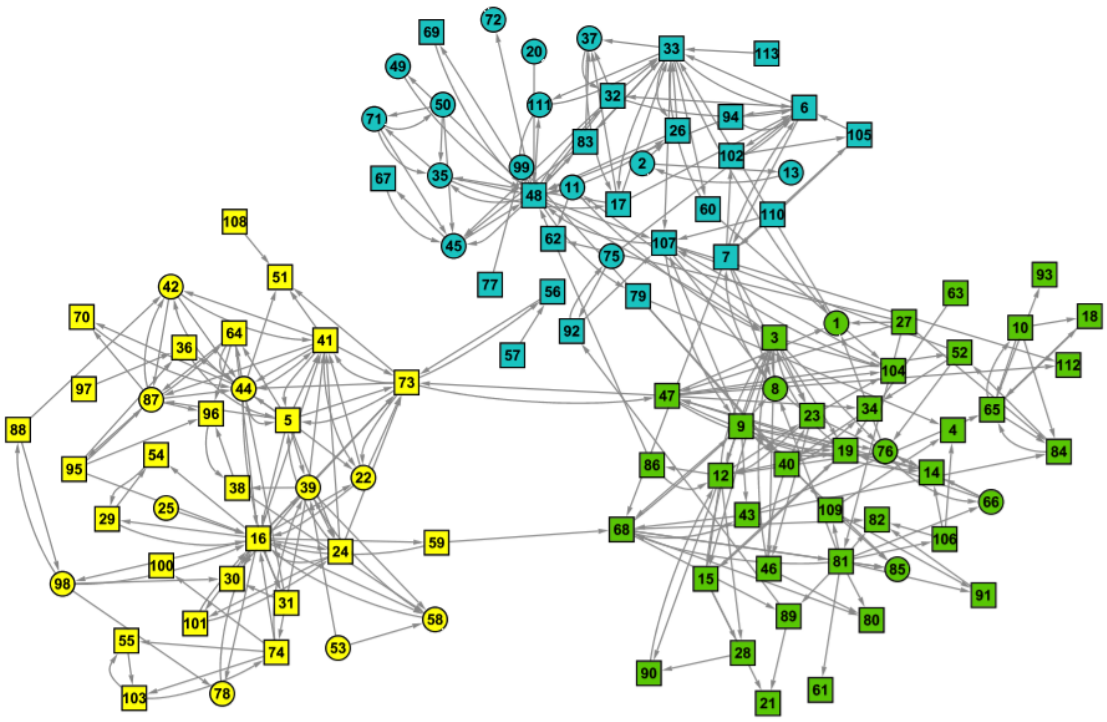
Sacerdote (2006) and Kleinbaum et al. (2008) who both find that geographical distance plays a crucial role in emailing behavior).

Sociometric data. We conducted the sociometric survey in July 2007. To map the egocentric networks of employees, we sent all the employees an email inquiring as to whether they would be interested in participating in our study. Included in the email was a direct link to the on-line survey. Respondents were presented with a list of all the members of the company, and were asked to select the names of their contacts. Respondents could select among the names in two ways: either by choosing directly from all the names using a drop-down list or by first selecting the department of the employee and then selecting the name from a list which contains only the employees of the selected department.

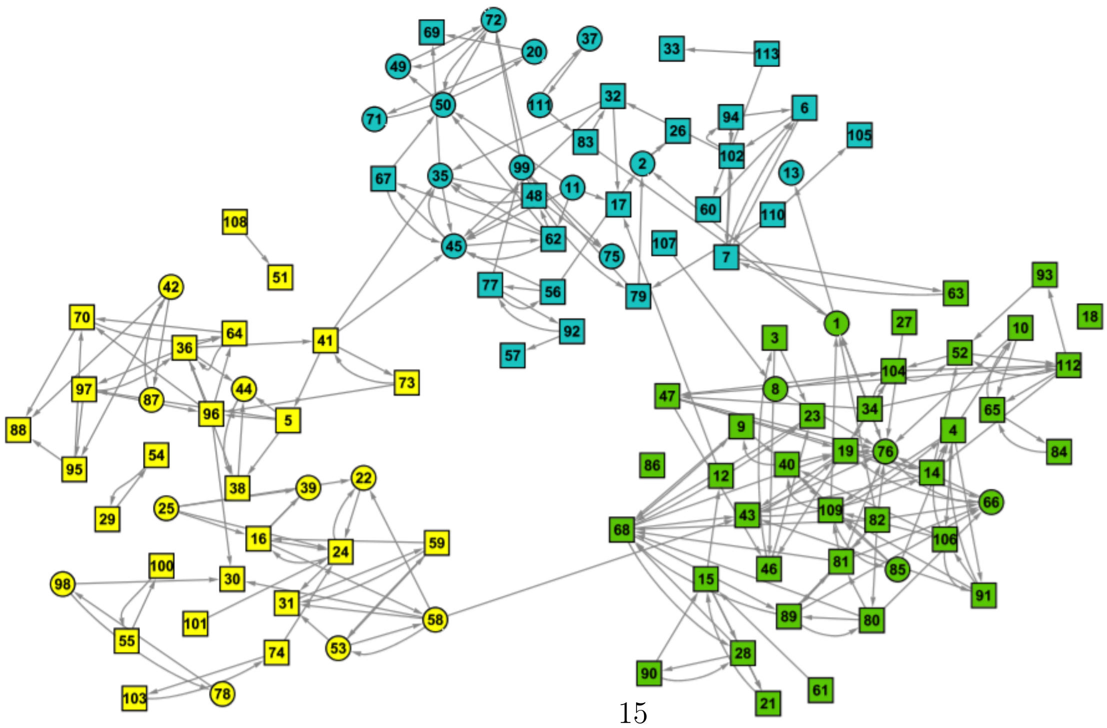
In the survey, we used traditional egocentric questions to map the *friendship*, *advice*, and *information flow* networks. This technique has a long history in network studies, communication research, and organizational studies (Moreno, 1933; Sampson, 1988; Podolny and Baron, 1997; Morrison, 2002; Goodreau, 2007). For each of the friendship, advice, and information flow networks we asked two questions, and we created a tie between ego and alter if ego named alter in any of the questions along the respective dimension. For example, if ego listed alter in either Question 3 ("Which of your colleagues do you believe have the most professional knowledge or expertise?") or Question 4 ("Which of your colleagues do you turn to for professional advice?"), then the advice seeking network is set to have a tie between ego and alter. See Table 2.2 for the survey questions (the survey contained other questions as well, but these questions are not directly relevant to our research question so we do not analyze them here). Note that these networks are directed. Figure 2.1 illustrates the friendship, advice seeking and information flow networks. One might

Figure 2.1: Visualization of the all email, friendship, advice seeking, and information flow networks. For easier comparison, we forced the same layout on the networks. Different colors represent the different organizational departments; squares denote male employees and circles denote female employees.

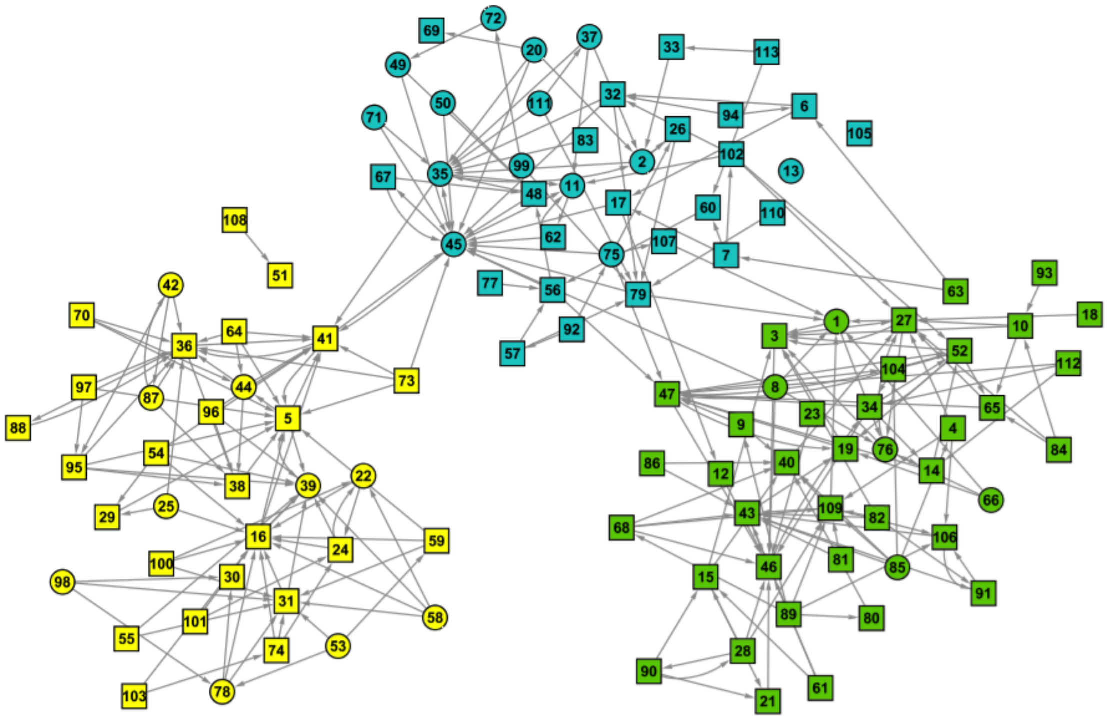
a) friendship network



b) advice network



c) information flow network



d) email network

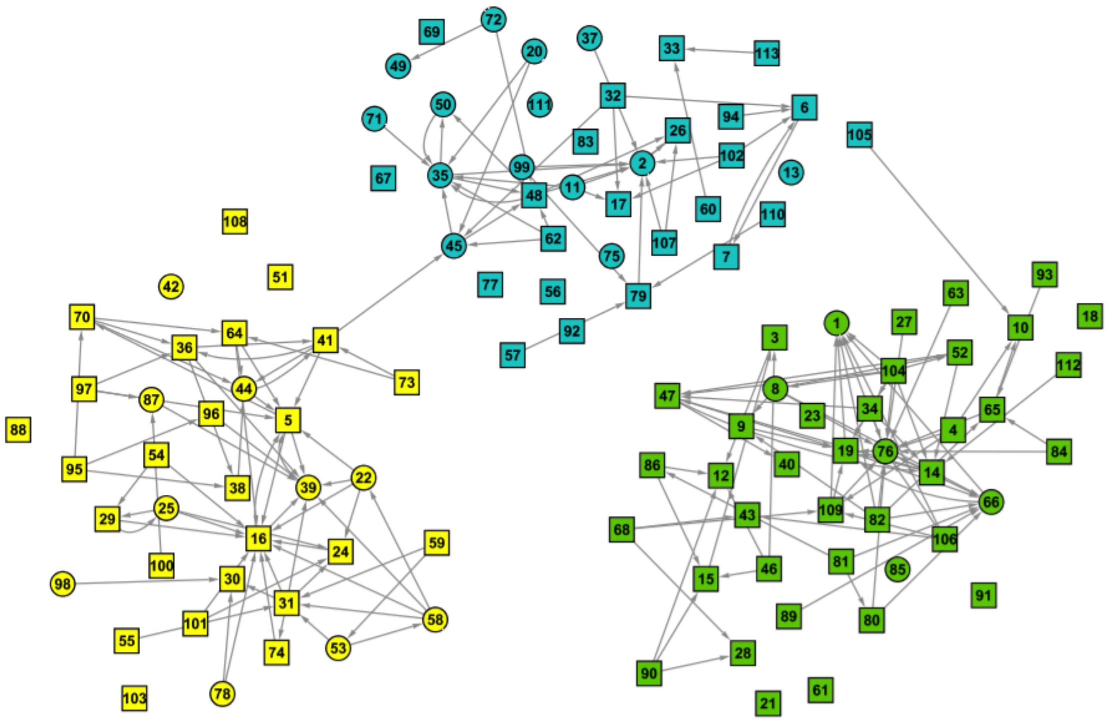


Table 2.2: Survey questions

	Survey questions
Friendship network	**1.** *Which of your colleagues would you turn to for advice with personal problems?*
	2. *Which of your colleagues do you like to have lunch with?*
Advice network	**3.** *Which of your colleagues do you believe has the most professional knowledge or expertise?*
	4. *Which of your colleagues do you turn to for professional advice?*
Information flow network	**5.** *Through informal channels, from which of your colleagues do you most quickly gain information concerning company news or organizational changes?*
	6. *Through formal channels, from which of your colleagues do you most quickly gain information about the company?*

argue that the research design, which, as described above, is a fixed-choice design, might introduce selection bias into our analyses as it censors the data (Holland and Leinhardt, 1973; Kossinets, 2006). While we agree that the research design is not optimal, we believe that in this case censoring does not alter the results significantly: For each sociometric question the employees could have listed up to four names, but they rarely listed the maximum four. In 42% of the answers only one alter was listed, in 37% of the answers two alters were listed, in 17% three alters, and only in 4% of the answers were all the four possible slots used. As only in these latter cases might censoring matter, we argue that censoring does not pose a serious problem.

Email data. The subsample we analyze includes the complete record of email activity among the 113 employees during an observation period of roughly three weeks. To protect the privacy of employees, the bank omitted the subject and the content of the messages, providing us only with information about the senders and recipients, the length of the email, the time and the date the message was sent. The names of employees were replaced with identifiers. Our sample contains 6,551 emails among the 113 employees.

Operationalizing the email networks. How should one build a network from data on email communications? The answer to this question is less than obvious. In short, one needs to operationalize what a tie is: should we consider two people A and B connected if there exists an email from A to B, or should we require say at least five emails? Should we include emails with multiple recipients (just think about whether a mass-email with 50 recipients represent a tie)? Should we consider a tie to exist between two persons if person A has cc-d person B in a single email but there is no other email from A to B? Because the literature provides no clear guidance on these issues, we investigate multiple operationalizations. Exploring multiple operationalizations on one hand ensures the robustness of our analyses, and on the other hand provides a reference for further researchers who aim at extracting networks from email communications.

To tackle the issue of operationalization of emails, we take an approach that mixes conceptual insights and sensitivity analyses. The first operationalization, which we call all-email network, defines a tie between ego and alter if there had been any email communications from alter to ego. This operationalization, we think, is a natural choice. In the second operationalization, we aimed to exclude ties between employees who only have mass-email communication to each other – the reason for this was the intuition that if only mass-emails exist between two employees then they are probably not connected. To identify mass-emails, Quintane and Kleinbaum (2011) selected emails with four or more recipients. As we find their cut-off at four recipients arbitrary, we have conducted sensitivity analyses: we explored how the networks change if we vary the cut-offs (for brevity, results are not shown here). These sensitivity analyses show that the natural cut-off is not at four recipients but at two: whether there exists an individual email between ego and alter (in which alter is the only recipient). Thus, we analyze the network of individual emails. This network, although emerged from sensitivity analyses, can also be defended conceptually, as person-to-person emails are qualitatively different from emails that have more than one recipient.

In the third operationalization, we create an email network in which a tie is defined to exist between alter an ego if ego sent at least four emails to alter. This operationalization emerged from the consideration to create an email-based network that has similar density to the survey-based networks (this is important in view of Anderson et al. (1999), who show how estimates of network parameters interact with the density of the network). After calculating the average degree for networks emerging from different cut-off values, we found that the network with four email cut-off has an average degree of 2.06 which is closest to the average degree of the friendship-, advice-, and information flow networks (see Table 2.3).

Table 2.3: Density of the networks and network correlations with QAP standard errors

	Number of edges	Average degree
All emails network	2,091	18.50
Individual emails network	1,221	10.81
Individual emails network (weight > 4)	233	2.06
Friendship network	251	2.21
Advice network	272	2.41
Information network	182	1.61
All possible edges	12,656	

Network Correlations	2	3	4	5	6
1 All emails network	0.735	0.308	0.272	0.248	0.230
2 Individual emails network		0.419	0.261	0.249	0.253
3 Individual emails network (weight > 4)			0.280	0.280	0.309
4 Friendship network				0.354	0.345
5 Advice network					0.348
6 Information flow network					

All correlations are significant at the 0.01 level (standard errors calculated with QAP).

2.3 Analyses: Comparison of survey-base and email-based social networks

2.3.1 Network correlations and descriptive statistics of the networks

Before turning to the analysis of the networks with exponential random graph models, we provide some descriptive statistics about the macro structure of the networks and the overall correspondence between the networks. First, we compare the networks using network correlations with standard errors calculated with Krackardt (1987)'s Quadratic Assignment Procedure (QAP). The network correlation coefficients, all of which are significant at the 0.01 level, are shown in Table 2.3. The networks correlate positively, although the strengths of the correlations are moderate: The survey-based networks have a correlation of around 0.35 with each other, and a 0.25 correlation with the email-based networks.

2.3.2 Comparison of the networks along centrality measures

After analyzing the networks with network correlations, in this section we compare the networks along structural properties such as degree centrality, betweenness centrality and closeness centrality (Wasserman and Faust, 1994; Grippa et al., 2006). Here we investigate whether centrality in email-based networks correspond to centrality and brokerage in the friendship, advice seeking, and information flow networks.

We calculate actors' degree centrality (the number of connections the individual has) for each network, as well as closeness and betweenness centrality (Wasserman and Faust, 1994). Similarly to Kossinets and Watts (2006), we calculate Spearman‘s rank order correlation to compare the structural position of actors in the different networks.

Table 2.4 shows a number of interesting correlations, but here we focus on whether the email-based networks correspond to the friendship, advice seeking, and information flow networks in terms of actors' centrality positions. Two general patterns are apparent. First, email networks have stronger correlations with the advice seeking and the information flow network in terms of centrality than with friendship networks. Indeed, some of the correlations are rather high (e.g., 0.642), indicating that being central in the email network corresponds to being central in the advice and information flow networks. Second, while centrality in email networks corresponds to centrality in off-line network in terms of betweenness- and degree-centrality, this does not hold for closeness centrality. Unfortunately we do not have any explanation for this finding.

2.3.3 Analyzing the networks with Exponential Random Graphs

The network correlations with QAP standard errors demonstrate that although the email-based and the survey-based networks are related, they differ substantially. Correlation is, however, just a crude measure of correspondence. To get at a more detailed comparison of the various networks, we turn to Exponential Random Graph models (Handcock, 2002; Hunter, 2007; Goodreau, 2007; Pattison and Robins, 2007). We choose this model family because it enables the comparison of the mechanisms and factors that explain the various social network data, and it allows for

Table 2.4: Spearman Correlations among Centrality Measures of the Employees

Betweenness centrality Spearman Correlation	2	3	4	5	6
1 All emails network	0.883***	0.454***	0.341***	0.551***	0.455***
2 Individual emails network		0.441***	0.331***	0.508***	0.426***
3 Individual emails network (weight>4)			0.171*	0.450***	0.472***
4 Friendship network				0.387***	0.247***
5 Advice network					0.384***
6 Information flow network					

****.Correlation is significant at the 0.01 level (2-tailed).*
***.Correlation is significant at the 0.05 level (2-tailed).*
**.Correlation is significant at the 0.1 level (2-tailed).*
(N=113; max.num.edge=12,656)

Closeness centrality Spearman Correlation	2	3	4	5	6
1 All emails network	0.278***	0.132	-0.034	0.296***	0.126
2 Individual emails network		0.097	-0.068	0.308***	0.046
3 Individual emails network (weight>4)			-0.012	0.369***	0.378***
4 Friendship network				-0.107	0.335***
5 Advice network					0.324***
6 Information flow network					

****.Correlation is significant at the 0.01 level (2-tailed).*
***.Correlation is significant at the 0.05 level (2-tailed).*
**.Correlation is significant at the 0.1 level (2-tailed).*
(N=113; max.num.edge=12,656)

Degree centrality Spearman Correlation	2	3	4	5	6
1 All emails network	0.888***	0.595***	0.400***	0.518***	0.642***
2 Individual emails network		0.583***	0.346***	0.448***	0.597***
3 Individual emails network (weight>4)			0.345***	0.393***	0.574***
4 Friendship network				0.450***	0.534***
5 Advice network					0.572***
6 Information flow network					

****.Correlation is significant at the 0.01 level (2-tailed).*
***.Correlation is significant at the 0.05 level (2-tailed).*
**.Correlation is significant at the 0.1 level (2-tailed).*
(N=113; max.num.edge=12,656)

simultaneously investigating the effects of node attributes, dyad attributes, and local network configurations (Handcock, 2002; Hunter, 2007; Goodreau, 2007; Pattison and Robins, 2007). Our research strategy entails modeling each network individually, and comparing the modeling results across the networks.

Exponential random graph models have the following form:

$Pr\left(X = x\right) = exp\left(\sum_{A} \gamma_{A} z_{A}\left(x\right)\right)/\kappa,$

where A is a subset of tie variables, defining a potential network configuration, γ_A is a model parameter corresponding to the configuration A (and is non-zero only if all pairs of variables in A are assumed to be conditionally dependent); $z_A(x)$ is the sufficient statistic corresponding to the parameter γ_A and indicates the number of configurations A occurring in the network x; and κ is a normalizing quantity (Pattison and Robins, 2007).

In the models we present, we incorporate both endogenous local network structure variables and attribute variables. The local network structure variables include: density, reciprocity, indegree and outdegree distribution[3], and transitivity. As for attributes, we incorporate hierarchy, gender, tenure, age, and information on whether the two employees in the dyad are in the same department. As the network is directed, we include a dummy of whether the sender is at higher hierarchy level than the receiver, or they are at the same hierarchy level (therefore, the baseline is when the receiver is at higher hierarchy level). Similarly, we add dummies for the dyadic gender effect, where the baseline is the female-female dyad. As both tenure and age are continuous variables, we included the absolute difference along tenure and age into the model (tenure and age are coded in decades).

Before discussing the specific models, we note that we have calculated a wide range of models with alternative specifications, but here for brevity we only show and discuss a few of them. We selected those models that are parsimonious but also have good fit: the models simulated based on the estimates fit the empirical networks well in terms of indegree-, outdegree, edge-wise shared

[3]Note that to model some of the networks, we needed to include two geometrically-weighted indegree and outdegree terms with different parameters. This is standard practice in modeling networks with bimodal degree distributions. Note that this explains the seemingly high coefficients of gwidegree, gwodegree, and edges in these models.

Table 2.5: Result of the ERG models for the survey-based networks

	Friendship network		Advice network		Information flow network	
Edges	-5.829	(0.117)	-578.800	(0.087)	-3.100	(0.095)
Reciprocity	2.935	(0.230)	1.733	(0.404)	1.474	(0.314)
Geom. w. in-degree, α=ln(2)	-0.497	(0.106)	-3.933	(0.064)	-3.616	(0.646)
Geom. w. out-degree, α=ln(2)	0.227	(0.046)				
Geom. w. out-degree, α=ln(2.5)			-53.130	(0.041)		
Geom. w. out-degree, α=ln(3)					-3.137	(0.118)
Geom. w. out-degree, α=ln(5)			628.300	(0.037)		
Cyclic Triples	-0.230	(0.010)				
Transitive Triples	0.911	(0.044)				
Hierarchy (high->low)	0.942	(0.059)	0.923	(0.046)	1.293	(0.066)
Hierarchy same	0.498	(0.035)	0.764	(0.038)	0.827	(0.020)
Male to Male	-0.303	(0.038)	-0.683	(0.040)	-0.503	(0.022)
Female to Male	-0.604	(0.080)	-0.469	(0.115)	-0.518	(0.070)
Male to Female	-0.414	(0.079)	-0.986	(0.077)	-0.516	(0.054)
AbsdiffTenure	-0.018	(0.007)	-0.001	(0.007)	-0.010	(0.003)
AbsdiffAge	-0.010	(0.004)	-0.006	(0.003)	-0.014	(0.002)
Same department	2.889	(0.032)	3.715	(0.040)	5.322	(0.015)
Isolates					0.662	(2.122)

Parameter values of the erg models; (N=113; max.num.edges=12,656) Standard errors are in parentheses.

partner, and triad-census distributions (Hunter et al., 2008). It is important to stress that our main findings hold across alternative model specifications.

2.3.4 Results of the exponential random graph models

Before discussing the individual parameter estimates, we have to note that in accordance with the network correlation results in the previous section, there is an overall positive correspondence between the email and survey-based networks in the sense that the signs of the coefficients are mostly the same. Besides an overall similarity, however, there are important differences as well.

Tables 2.5 and 2.6 show the model estimates[4]. First, we discuss the results for reciprocity. While reciprocity is present in all the networks, its strength varies. Reciprocity is very important in the email and the friendship networks. For example, if ego sent an email to alter, then the probability that alter sends an email to ego increases 24 times (exp(3.158)). Similarly, if an ego

[4]We estimated the ERG models with the Statnet package in R (Handcock et al., 2011)

Table 2.6: Result of the ERG models for three different operationalizations of the email data

	Network of all emails		Network of individual emails		Network of individual emails (weight>=4)	
Edges	13.410	(0.125)	191.700	(0.236)	-3.942	(0.124)
Reciprocity	2.825	(0.327)	3.158	(0.233)	4.414	(0.306)
Geom. w. in-degree, α=ln(2)	0.175	(0.537)			-2.241	(0.044)
Geom. w. in-degree, α=ln(3)			1.004	(0.280)		
Geom. w. in-degree, α=ln(5)	-8.737	(0.002)				
Geom. w. in-degree, α=ln(7)			-115.400	(0.003)		
Geom. w. out-degree, α=ln(2)	-1.351	(0.102)			-2.018	(0.046)
Geom. w. out-degree, α=ln(3)			-1.298	(0.004)		
Geom. w. out-degree, α=ln(5)	-10.930	(0.003)				
Geom. w. out-degree, α=ln(7)			-83.120	(0.003)		
Hierarchy (high->low)	0.817	(0.007)	0.401	(0.014)	0.986	(0.048)
Hierarchy same	0.546	(0.002)	0.396	(0.005)	0.416	(0.035)
Male to Male	0.039	(0.003)	0.111	(0.006)	-0.295	(0.038)
Female to Male	-0.102	(0.007)	-0.062	(0.014)	-0.008	(0.064)
Male to Female	0.200	(0.006)	0.222	(0.013)	-0.094	(0.060)
AbsdiffTenure	-0.023	(0.001)	-0.018	(0.001)	0.039	(0.004)
AbsdiffAge	-0.005	(0.000)	-0.002	(0.001)	-0.040	(0.003)
Same department	2.040	(0.004)	2.354	(0.003)	2.788	(0.021)
Isolates					-0.870	(0.099)

Parameter values of the erg models (N=113; max.num.edges=12,656) Standard errors are in parentheses.

nominates alter as friend, then the chance that alter nominates ego as friend is 19 times higher (exp(2.935)). Weaker is the effect of reciprocity in the information flow network and in the advice network (which is not surprising, as advice ties are often asymmetric).

Now we turn to analyze the role of organizational boundaries in shaping the interpersonal networks. It is apparent that being in the same organizational department dramatically increases the probability that two employees will interact by email or nominate each other in the friendship, advice seeking, or information flow networks (this is also apparent from Figure 2.1). The strongest is the effect for the information flow network, in which being in the same department increases the chance of a tie between two employees by almost two hundred times (exp(5.322)). More moderate but still strong is the effect of being in the same department for the advice seeking and friendship networks (the estimated coefficients are 3.72 and 2.89). Much weaker (but still significant) is the effect of departmental boundaries on emailing behavior. This finding indicates that studies that purely rely on email networks to investigate the effects of organizational boundaries on interpersonal networks and communications, such as Kleinbaum et al. (2008), underestimate the role of boundaries.

Organizational hierarchy also shapes off-line social networks stronger than it shapes email networks. While both emails and friendship, advice seeking and information flow nominations tend to be directed toward the same and lower levels of hierarchy (as viewed from ego's location), the effect size of hierarchy variables is higher for off-line networks (except in the case of email ties in which ego sent at least four individual emails to alter, but as this network signals stronger ties, this result is not surprising).

Gender plays a different role in the email and off-line social networks. While in the offline networks the female-female ties are the most likely, followed by male-male ties, the overall effect of gender on email networks is rather weak and inconsistent across operationalizations.

In summary, we have found that on one hand reciprocity plays a stronger role in email networks than off-line social network; on the other hand, off-line social networks are more influenced by organizational, gender, and hierarchical differences (we provide possible explanation for these findings later).

Table 2.7: Result of the ERG models for the email networks, with the friendship, advice, and information flow networks as covariates

	Network of all emails		Network of individual emails		Network of individual emails (weight>=4)	
Edges	12.970	(0.103)	25.650	(0.021)	28.590	(0.004)
Reciprocity	3.161	(0.255)	3.023	(0.053)	3.873	(0.007)
Geom. w. in-degree, α=ln(2)	0.377	(0.597)	0.879	(0.057)	-0.978	(0.017)
Geom. w. in-degree, α=ln(5)	-8.796	(0.004)	-16.310	(0.001)	-13.880	(0.000)
Geom. w. out-degree, α=ln(2)	-1.668	(0.060)	-0.318	(0.013)	0.083	(0.005)
Geom. w. out-degree, α=ln(5)	-10.460	(0.001)	-16.800	(0.001)	-21.290	(0.000)
Hierarchy (high->low)	0.736	(0.005)	0.252	(0.003)	0.524	(0.002)
Hierarchy same	0.489	(0.001)	0.343	(0.001)	0.229	(0.000)
Male to Male	0.079	(0.002)	0.143	(0.001)	-0.415	(0.001)
Female to Male	-0.012	(0.005)	0.023	(0.003)	0.042	(0.002)
Male to Female	0.279	(0.005)	0.307	(0.002)	0.001	(0.002)
AbsdiffTenure	-0.032	(0.000)	-0.019	(0.000)	0.038	(0.000)
AbsdiffAge	-0.015	(0.000)	-0.006	(0.000)	-0.066	(0.000)
Same department	1.984	(0.003)	2.259	(0.002)	2.416	(0.001)
Friendship Network	1.609	(0.173)	1.012	(0.045)	1.527	(0.019)
Advice Network	0.537	(0.177)	0.663	(0.037)	1.411	(0.019)
Information flow Network	1.030	(0.322)	0.852	(0.055)	1.308	(0.030)

Parameter values of the erg models (N=113; max.num.edges=12,656) Standard errors are in parentheses.

2.3.5 Can emailing behavior be explained with off-line social networks?

Why do people write emails? Although carefully answering this question would require the analysis of the content of the emails (which we do not have), the data at hand allows us to take a few steps toward understanding how off-line relations explain who writes email to whom. To keep consistency with our previous modeling approaches, we take the simplest approach to investigating how off-line networks explain emailing behavior: we include friendship, advice, and information flow ties as explanatory variables into the ERG models of Table 2.6. We expect these variables to have positive coefficients: an employee is more likely to write email to a friend, more likely to write email to whom she turns for advice, etc. We are interested in two things: first, which of these off-line relationships is the most important in explaining emailing behavior; and second, to what extent does including off-line relationships into the models help explaining emailing behavior.

Table 2.7 shows the ERG estimates for all previous email network operationalizations. First, as expected, the existence of a friendship, advice, or information flow relationship between ego and alter increases the probability that ego will write an email to alter. This finding is consistent across all three email-network operationalizations. Also consistent across operationalizations is that friendship relations are the most important in explaining the existence of emails from ego to alter. Second most important in two out of the three networks is the information flow relationship, except in the network of four or more individual emails, in which case advice is the second most important (this finding is not surprising as this is the network that mirrors the most frequent emails). These findings, we believe, are novel to the email and communication literatures.

Second, the results in Table 2.7 show that off-line relationships have strong explanatory power. On one hand, existence of a friendship tie increases the chance of an email, ceteris paribus, by 5 times (exp (1.609)). This magnitude makes it the most important determinant, together with reciprocity and being in the same department. Also, by incorporating friendship, advice, and information flow relationships as covariates, the fit of the models improves significantly. Note, however, that the strong structural effects remain, which indicating that there are social mechanisms at work on top of (the main effect of) friendship, advice, and information networks.

In summary, we can say that consistent with the previous results in the paper, the results of Table 2.7 demonstrate that while emailing behavior is strongly influenced by off-line relations, off-line relations do not explain emailing behavior perfectly. (This latter statement should be taken with caution, as we do not claim that there exists no model that explains emails. For example, we could include interactions between the variables etc. These we leave for future research.) What these analyses add, however, is that they demonstrate that friendship is the factor that most strongly drives the overall patterns of emailing, as opposed to advice seeking, or information flow (even though these factors are present as well).

2.3.6 Possible explanations for the findings

What might be the explanations for our empirical findings, namely that while there is a general correspondence between the email networks and off-line social networks, this correspondence is not perfect, and social, hierarchical, and gender boundaries are less important in email networks?

In general, there could be two answers. The first possible explanation for the divergence between survey-based and email-based social networks concerns a measurement issue: surveys are self-reported, while emails are observational. There is extensive literature in the fields of social networks and communications regarding the differences between self-reported and observed ties (Killworth and Bernard, 1976; Bernard et al., 1982; Brewer, 2000; Marin, 2004; Marsden, 2005). This literature mostly focuses on "recall-bias," i.e., that people cannot perfectly recall their ties to others. A general finding in the recall literature is that people are better in recalling stable and frequent relations (Freeman et al., 1987). Also, Brewer (2000) finds that people are better in recalling alters who are more centrally located in the network. This literature, however, tends to agree that there exists no good general predictors of whom ego will forget to recall, in terms of gender, status, or age differences (for example, Bernard et al. (1982) did not find any strong effect of these variables on whom people recall as a communication partner). These findings, thus, do not help in understanding the systematic differences we found between the email and the survey-based networks.

A second possible explanation for our findings is that online and offline relations are just simply different. As Susan Herring writes, "the Internet is said to be inherently democratic, leveling traditional distinctions of social status" (Herring, 2003, page 202.) While this is a slightly utopian statement, scholars have reported that perceived hierarchies, organizational boundaries, gender and age differences are less important in online settings (Wellman and Hampton, 1999). Interestingly, while there exists a sizable literature in the field of communication on how people use email, relatively few studies analyze empirically the effect of gender or status differences in emailing behavior (but see Boneva et al., 2001; Herring, 2003). These studies, using very different research approaches from ours (Boneva et al. (2001) uses survey data on internet usage), find similar results: females write more emails than males (Herring, 2003). Although these scholars observe a general tendency of online behavior to decrease the importance of off-line differences in gender, power, or age, a general agreement in the field is that these differences albeit decrease but do not go away.

In general, it is hard to come to a strong conclusion for the observed patterns, especially because survey networks and the email networks in our paper differ in both the recall/observational and the on/off-line dimensions. It would be an interesting way for future research to compare networks that differ in only one of these dimensions. An important step toward that direction is Matzat and Snijders (2010), who experimentally compare off- and on-line collected recall data.

2.4 Discussion

In this paper, we analyzed the structure of an email network and the underlying friendship, communication, and advice seeking networks. Our sample contains detailed email communications of 113 employees of three departments of a bank, and of a survey of the ego networks of these employees. Our goal was to understand the extent to which email patterns coincide with sociometric measures of the friendship, advice seeking, and communication networks. To study these questions, we used exponential random graph models along with structural analyses.

In general, our results demonstrate that although the structure of the email networks in our dataset is related to the structure of survey-based networks, emails networks and survey-based social networks do not perfectly correspond to each other. On one hand, email-based and survey-based social networks do coincide in many aspects. For example, email networks are relatively good in predicting the betweenness and degree centrality of employees, especially in the advice seeking and communication networks.

In many other aspects, however, email-based and survey-based social networks diverge. For example, if one is interested in the role organizational boundaries play in forming social networks, the two data types paint markedly different pictures. In the email-based networks, organizational boundaries play a weaker role (Namata et al., 2006) and email communication often moves across boundaries. The sociometric data, however, indicate that these boundaries do play a role, a role that is much stronger than pictured by email networks. This result questions the validity of Kleinbaum et al. (2008)'s finding that gender and hierarchical boundaries are not as strong as geographical boundaries - while these results may be valid, our results indicate that emails underestimate the role of organizational boundaries. Also, email networks do not capture the role organizational hierarchies and gender play in interpersonal networks.

Finally, we have to note that this paper is an explorative case-study of a single organization. Although we do not see any reason why this setting would be atypical, we believe that more work needs to be done to proper understand the relationship between email networks and off-line social networks. First, for the sake of generalizability, other organizations and empirical settings should be studied. Second, although in order to minimize the amount of missing data we chose the three departments with the highest response rates, we still have 33% of missing observations, which might bias our findings (but see our arguments about why we think that non-response bias is not strong in our sample). Third, future research is needed to analyze the content of emails: this would enable researchers to differentiate between friendship, advice seeking, or communication emails.

References

Anderson, B. S., Butts, C., and Carley, K. (1999). The interaction of size and density with graph-level indices. *Social Networks*, 21(3):239–267.

Bernard, H. R., Killworth, P. D., and Sailer, L. (1982). Informant accuracy in social-network data v. an experimental attempt to predict actual communication from recall data. *Social Science Research*, 11(1):30–66.

Boneva, B., Kraut, R., and Frohlich, D. (2001). Using e-mail for personal relationships the difference gender makes. *American behavioral scientist*, 45(3):530–549.

Brewer, D. D. (2000). Forgetting in the recall-based elicitation of personal and social networks. *Social networks*, 22(1):29–43.

Eagle, N., Macy, M., and Claxton, R. (2010). Network diversity and economic development. *Science*, 328(5981):1029–1031.

Freeman, L. C., Romney, A. K., and Freeman, S. C. (1987). Cognitive structure and informant accuracy. *American anthropologist*, 89(2):310–325.

Goodreau, S. M. (2007). Advances in exponential random graph (p*) models applied to a large social network. *Social Networks*, 29(2):231–248.

Grippa, F., Zilli, A., Laubacher, R., and Gloor, P. A. (2006). E-mail may not reflect the social network. In *Proceedings of the North American Association for Computational Social and Organizational Science Conference*, pages 1–6.

Handcock, M., Hunter, D., Butts, C., Goodreau, S., and Morris, M. (2011). Statnet: software tools for the statistical modeling of network data. 2003. *URL http://statnetproject. org*.

Handcock, M. S. (2002). *Statistical models for social networks: Inference and degeneracy*. na.

Herring, S. C. (2003). Gender and power in online communication. *The handbook of language and gender*, pages 202–228.

Holland, P. W. and Leinhardt, S. (1973). The structural implications of measurement error in sociometryâĂă. *Journal of Mathematical Sociology*, 3(1):85–111.

Hunter, D. R. (2007). Curved exponential family models for social networks. *Social networks*, 29(2):216–230.

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481).

Killworth, P. and Bernard, H. (1976). Informant accuracy in social network data. *Human Organization*, 35(3):269–286.

Kleinbaum, A. M., Stuart, T., and Tushman, M. (2008). *Communication (and coordination?) in a modern, complex organization*. Harvard Business School Boston, MA.

Kossinets, G. (2006). Effects of missing data in social networks. *Social networks*, 28(3):247–268.

Kossinets, G. and Watts, D. J. (2006). Empirical analysis of an evolving social network. *Science*, 311(5757):88–90.

Krackardt, D. (1987). Qap partialling as a test of spuriousness. *Social networks*, 9(2):171–186.

Marin, A. (2004). Are respondents more likely to list alters with certain characteristics?: Implications for name generator data. *Social Networks*, 26(4):289–307.

Marmaros, D. and Sacerdote, B. (2006). How do friendships form? *The Quarterly Journal of Economics*, pages 79–119.

Marsden, P. V. (2005). Recent developments in network measurement. *Models and methods in social network analysis*, 8:30.

Matzat, U. and Snijders, C. (2010). Does the online collection of ego-centered network data reduce data quality? an experimental comparison. *Social networks*, 32(2):105–111.

Menchik, D. A. and Tian, X. (2008). Putting social context into text: The semiotics of e-mail interaction1. *American Journal of Sociology*, 114(2):332–370.

Moreno, J. (1933). Emotions mapped by new geography. *New York Times*, 3:17.

Morrison, E. W. (2002). Newcomers' relationships: The role of social network ties during socialization. *Academy of management Journal*, 45(6):1149–1160.

Namata, G., Getoor, L., and Diehl, C. (2006). Inferring formal titles in organizational email archives. In *Proc. of the ICML Workshop on Statistical Network Analysis*.

Pattison, P. and Robins, G. (2007). Handbook of probability theory with applications, chapter probabilistic network theory.

Podolny, J. M. and Baron, J. N. (1997). Resources and relationships: Social networks and mobility in the workplace. *American sociological review*, pages 673–693.

Quintane, E. and Kleinbaum, A. M. (2008). Mind over matter? the correspondence between e-mail and survey measures of social networks. *Under Review*.

Quintane, E. and Kleinbaum, A. M. (2011). Matter over mind? e-mail data and the measurement of social networks. *Connections*, 31(1):22–46.

Sampson, R. J. (1988). Local friendship ties and community attachment in mass society: A multilevel systemic model. *American Sociological Review*, pages 766–779.

Stewart, D. (2005). Social status in an open-source community. *American Sociological Review*, 70(5):823–842.

Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.

Wellman, B. and Hampton, K. (1999). Living networked on and offline. *Contemporary Sociology*, pages 648–654.

Chapter 3

Contrasting alternative explanations for the consequences of category spanning: A study of restaurant reviews and menus in San Francisco*

Abstract

Recent literature on organizational category spanning demonstrates that organizations that span multiple categories on average suffer social and economic disadvantages in markets. While multiple mechanisms have been proposed to explain this finding, most studies do not test directly nor contrast these mechanisms. In this paper, we contrast two of the main mechanisms proposed in the literature: the audience-side typicality-based explanation (category spanners are atypical of each categories spanned) and the producer-side quality-based explanation (category spanners produce lower quality output because they cannot develop expertise in any of the categories spanned). We find evidence for both mechanisms. Furthermore, we argue that quality and typicality interact such that high-quality organizations can benefit from being atypical. Finally, we contrast two kinds of spanning: "fusion" and "food court," and argue that their effects are different depending on the overall quality of the organization. Our empirical setting is the restaurant domain, and we analyze menus and reviews of 474 restaurants located in San Francisco.

---

*with Balázs Kovács

3.1 Introduction

An active line of contemporary research on organizations investigates the effects of category spanning on organizational outcomes. A now accepted empirical pattern shows that organizations that span multiple categories on average suffer from social and economic disadvantage in markets. Category spanners receive less attention and legitimacy and they have lower chances of success (Zuckerman, 1999a; Zuckerman and Kim, 2003; Hsu et al., 2009b). Organizations assigned to multiple categories tend to be either ignored (Zuckerman, 1999a) or devalued (Pólos et al., 2002b; Rao et al., 2003b; Hsu, 2006c; Kovacs and Hannan, 2010, 2011). This pattern is shown to hold in domains such as stock recommendation (Zuckerman, 1999a), films (Zuckerman and Kim, 2003; Hsu, 2006c), wine producers (Negro et al., 2010b, 2011), online auctions (Hsu et al., 2009b), and restaurant reviews (Kovacs and Hannan, 2010).

Although researchers studying category spanning have put forward an impressive set of empirical findings over the last decades, the theoretical underpinnings of these findings are not clarified, and a number of different theoretical processes have been advanced to explain the consequences of category spanning. The explanations proposed by organizational theorists mainly fall in two areas, which we term the "quality-based" and the "socio-cognitive-based" explanations. The quality-based explanations relate to the producer-side consequences of category spanning and explore the effects of spanning on the quality of the offerings of organizations. The socio-cognitive explanations focus on the audience-side consequences of spanning. One socio-cognitive explanation argues that category spanners violate institutionalized expectations and thus are viewed illegitimate (Zuckerman, 1999a). Another approach focuses on the confusion and ambiguity stemming from category spanning: category spanning confuses audience members who do not know what to expect from these organizations (Hsu, 2006c; Hsu et al., 2009b; Kovacs and Hannan, 2010, 2011). A mechanism emphasized here is typicality: organizations that span multiple categories are atypical to each of the categories spanned, thus audience members cannot rely on category schemata to form clear expectations about the offerings of the organizations (Hsu et al., 2009b; Negro et al., 2010b; Kovacs and Hannan, 2011).

While researchers are aware that quality-based and socio-cognitive-based explanations could be confounded (Zuckerman, 1999a; Zuckerman and Kim, 2003; Hsu, 2006c; Hsu et al., 2009b), there has been little attempt to test the mechanisms directly[1], nor have researchers put forward an empirical analysis in which multiple alternative mechanisms are contrasted. There is a need to open the "black box" of the consequences of spanning in order to identify the mechanisms that cause them. In most above-mentioned articles, one (or more) of the mechanisms is emphasized to derive macro level consequences, and the tests operate on the macro level consequences. In this paper, we contrast the typicality-based and the quality-based mechanisms. Getting at the mechanisms is important not merely for theoretical purposes but because different mechanisms have distinct implications. For example, the quality-based and the typicality-based explanations provide different predictions for quality-category organizations: the quality argument would not differentiate among single-category organizations, while the typicality-based argument predicts that single-category organizations that are atypical to their category have lower value than single-category organizations that are typical to the category they are in. The different mechanisms also provide diverging recommendations for organizational action: if the locus of punishment lies in the audience side (Zuckerman, 1999a; Hsu, 2006c), then organizations need to focus on audiences' perceptions to dodge the negative consequences of spanning. If the mechanism is on the producer side and involves reduction in skills and capabilities, then organizations need to assess the relatedness of skills and technologies before they decide to span categories.

Our empirical setting is restaurants and restaurant reviewing. In short, we investigate whether restaurants that span categories (such as "Japanese" and "Mexican") receive lower ratings from reviewers. The review data come from the online review website Yelp.com. We collected reviews between January 2010 and October 2011. The sample consists of 59,605 reviews written about 474 San Francisco-based restaurants by 32,624 unique reviewers. These data have been used previously in Kovacs and Hannan (2010) and Kovacs and Hannan (2011), who demonstrate that restaurants that span categories receive lower ratings. However, the two papers by Kovács and

[1]With the major exceptions of Hsu (2006c) and Negro and Leung (2013), which we discuss later. We shall also discuss research on diversification in strategy.

Hannan, as most other papers in the field, suffer from two shortcomings: they do not measure the typicality of restaurants but assume that multiple-category restaurants are less typical than single category restaurants; and they do not control for the alternative, skill- or quality-based mechanisms. To further their results, in this paper we collect and analyze two additional data sources: to assess typicality, we collected the menus of the restaurants in our sample. To assess quality, we collected the "food" quality scores of The Zagat Guide, which has previously been used as a measure of restaurant quality (Roberts et al., 2013). Combining these two additional data sources with the review data allows us to directly test and disentangle typicality-based and quality-based explanations for the negative consequences of category spanning.

A main novelty of this paper is a direct instrument of typicality. Previous research assessed typicality indirectly, and assumed that the more categories an organization populates, the lower its typicality in each of the categories populated (Hsu, 2006c; Kovacs and Hannan, 2010; Negro et al., 2010b). For example, a restaurant that is labeled both "Japanese" and "Mexican" is likely to be atypical of both categories. While we are sympathetic to this approach, here we argue that a more direct instrumentation of typicality is needed. First, approaches that use multiple category membership, in lack of better evidence, have to assume that single-category organizations are all typical to their category. This is clearly an oversimplification. For example, not all single-category "Italian" restaurants are 100% typical Italian. As single-category organizations are prevalent (for example, they constitute almost half of our sample), dealing with them is crucial to test the theory[2]. Second, not all multiple category organizations violate the "categorical imperative" to the same extent: for restaurants, an "Indian" and "Pakistani" combination is likely to be less detrimental than an "Indian" and "Japanese" combination. To account for such cases, ideally one needs to examine the actual offering of the organizations[3].

Our empirical strategy to assess the typicality of restaurants is to contrast the offering of the

[2]Note that in Zuckerman (1999:1418)'s Table 2, single-category organizations also constitute about half of the population (11 out of 20).

[3]An alternative approach is to take into account the structure of the category system. The only paper we know that attempts this is Kovacs and Hannan (2011), who use the organizational categories' co-occurrences to infer the structure of the category system. While this is a step in the right direction, it does not deal with the single-category organizations problem.

restaurant (items on the menu) with its labels (Italian, Japanese, etc.) and assess the extent to which the menu fits the label(s) the restaurant claims. We use a commonly utilized computational linguistics approach, word-category co-location mapping (Manning and Schütze, 1999b), to explore the schemata of organizational categories. We establish the typicality of the restaurants in the categories by comparing the restaurant's offering to the schemata of the categories. We assert that the category mismatch of a restaurant is high when the offerings are atypical of the label(s) to which it is assigned. To our knowledge, ours is the first paper in the category spanning literature that actually measures the offerings of the organizations and assesses typicality in such a way.

Besides contrasting the two main mechanisms for the effects of category spanning on the average effect of category spanning, we also contribute to current literature by theorizing about situations in which category spanning can be beneficial. We argue that the above-specified mechanisms, typicality and quality, interact in a way that typicality is advantageous for low- and mid-quality organizations but high-quality organizations can benefit from being atypical.

Finally, we distinguish between two kinds of category spanning. One, where the items of different categories appear side by side on the organization's profile, we call it "food court" type of spanning with respect to our empirical setting. The other type of spanning we label "fusion"; the elements of different categories combined within the organizational items. We argue that "fusion" type of category spanning is more beneficial of high-quality organizations, while "food court" type of category spanning is more beneficial to lower quality organizations.

3.2 Theoretical background and hypotheses

The interest in the implications of category spanning has a long tradition in organizational research, resulting in a wide range of seemingly contradictory findings. Many researchers emphasize the negative average effect of spanning. As we discuss in detail below, these researchers evoke explanations such as the negative average effect of spanning on quality or typicality (eg., Zuckerman, 1999a; Hsu, 2006c; Hsu et al., 2009b; Kovacs and Hannan, 2010; Negro and Leung, 2013).

Other studies document cases in which category spanning is beneficial. The literature on related diversification, for example, discusses cases in which category spanning promotes the evolution of new organizational capabilities (Markides and Williamson, 1994). Alvarez et al. (2005) develop a micro theory of creative action by examining how distinctive artists shield their idiosyncratic styles from the isomorphic pressures of a field. They show that auteur directors receive both critical and public acclaim because they span multiple movie genres. Importantly, most of the innovation literature emphasizes the positive effect of spanning ("recombination") on innovative output (Schumpeter, 1934; Fleming, 2001). Baker (1992) documents cases in which diversification can increase value. Villalonga (2001) argues that after taking selection effects into account, no overall negative effect of "diversification discount" can be found. Dobrev et al. (2001b) finds that larger niche width decreases the hazards for exit and disbanding.

While we do not aim here at fully reconciling all these findings, we would like to note a few points that help situate our theorizing and delineate the scope conditions of our theory. First, the literature that emphasizes the negative consequences of spanning typically focuses on the average effect, while the creativity and innovation literatures focus on the upper tails of the distribution by showcasing that highly successful organizations or innovators tend to be category spanners. So it might be the case that spanning leads to higher variance (as Fleming, 2001, shows) but at the same time spanning is detrimental on average. Second, as we demonstrate later, it can be the case that spanning is beneficial if the organization is of high quality or status (Phillips and Zuckerman, 2001a) but not otherwise, thus not controlling for quality or status might result in contradictory findings. Third, if the typicality argument is right, it might be the case that category spanning has a negative effect on organizational outcome only in settings where audience perceptions play a significant role. Fourth, the effect of spanning might differ by the outcome variable used. Fifth, the consequences of category spanning might depend on environmental conditions. As Freeman and Hannan (1983b) assert, specialists are preferred in stable environments but generalists have higher survival rates in uncertain environment.

Given these considerations, we emphasize that our theorizing below refers to a setting in which audience perceptions are important and that our main outcome variable is rating of the organizations' [products] by audiences. This is also a setting in which the environment is quite stable, at least in the few years our observational window encompasses. In the first three hypotheses we focus on the average effect of spanning, while in the last two hypotheses we explore how the effect of spanning might differ by organizations.

3.2.1 The producer-side: How category spanning affects average quality

Category spanning, or as it used to be referred to, the "generalists vs. specialists" issue, was a central topic in early population ecology (Hannan and Freeman, 1977b, 1989a; Freeman and Hannan, 1983b; Dobrev et al., 2001b). Building on the niche theory of Levins (1968), researchers argued that because an organization's level of resource, budget, and attention is finite, the more markets the organization engages, (ceteris paribus) the fewer resources it can spend on developing products, to attend to specific audiences, and to cumulate expertise in each of the categories spanned (Hannan and Freeman, 1989a; Hannan et al., 2007b). For example, a film actor has to decide whether to specialize in comedy or to prepare for both comedic and dramatic roles. Because of scale advantage, such a dispersion of attention and budget implies that organizations that span multiple categories cannot excel in each of the categories spanned. Organizations that attempt to develop skills and invest in quality in multiple categories run the risk of becoming a "jack of all trades but master of none" (Hannan and Freeman, 1989a; Hsu, 2006c). Building on this argument, researchers predicted that category spanners have lower value to audiences. This prediction has been confirmed in various settings, such as restaurants (Freeman and Hannan, 1983b), and feature films (Hsu, 2006c)[4]. (It is important to note that Freeman and Hannan (1983b) argue that generalists have higher survival chances in highly uncertain environments, where the benefits of hedging the risk against specializing in the wrong niches can outweigh the negative effects spanning has on average quality.)

The diversification literature in strategy also investigates the consequences of category spanning. For example, Lang and Stulz (1993) and Berger and Ofek (1995) find that diversified firms

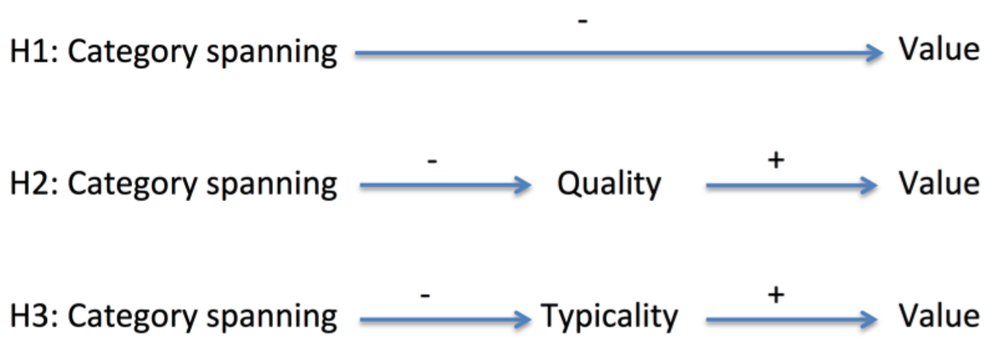
[4]Another main prediction of the principle of allocation theory (Hannan and Freeman, 1989a; Hsu, 2006c) is that category spanners (generalists) attract a wider set of audiences. Hsu (2006c), for example, demonstrates that movies that span genres tend to get larger viewership. Although preliminary results show that this pattern holds in our data as well, we do not pursue this line of investigation further.

trade at a discount relative to single-segment firms. While the validity of these finding is questioned (Villalonga, 2001), their argumentation mostly concerns producer-side consequences. For example, Berger and Ofek (1995) cite as potential benefits of diversification "greater operating efficiency, less incentive to forego positive net present value projects, greater debt capacity, and lower taxes;" as possible negative consequences of diversification they list "the use of increased discretionary resources to undertake value-decreasing investments, cross-subsidies that allow poor segments to drain resources from better-performing segments, and mis-alignment of incentives between central and divisional managers." (Berger and Ofek, 1995, page 40.)

A common shortcoming of the above research streams is that they rarely distinguish empirically the producer-side and audience-side effects of spanning and diversification. We think that aggregate level tests are not adequate to settle whether there exists a producer-side advantage or disadvantage of spanning, as audience-side effects might counteract producer-side effects. Indeed, this might possible explain the inconsistent findings in the diversification literature. As an important exception, Negro and colleagues (Negro et al., 2010b; Negro and Leung, 2013) recently analyzed the effect of spanning on quality using a unique empirical dataset: blind tasting of wines. Using blind taste ratings of wines to compare the quality of the wines produced by specialist and category-spanning wine producers, they find that wines of category spanning wine producers on average get lower ratings. This, they argue, is clear evidence of the quality effect, assuming that in blind tasting experiments the raters are unaware of the identity of the producers. Negro et al. (2010b) also compare the blind wine tasting results with non-blind wine tasting results, and hypothesize that the negative effects of category spanning should be stronger for the non-blind tasting setting because besides the quality-based effects, socio-cognitive effects are present as well.

We extend previous research by using a mediation framework to investigate whether quality mediates the negative consequences of category spanning. We aim to demonstrate three relationships. First, we test whether category spanners on average indeed confer lower quality offerings. Second, we test whether low quality offerings lead to diminished value for audiences. Third, we investigate whether quality mediates the effect of category spanning. Figure 3.1 demonstrates these relationships visually.

Figure 3.1: Visual representation of the three hypotheses



Hypothesis 1: Organizations that span multiple categories on average receive lower value ratings from audiences than single-category organizations.

Hypothesis 2a: Organizations that span multiple categories provide on average lower quality offerings than single-category organizations.

Hypothesis 2b: Organizations that provide lower quality offerings receive lower value ratings from audiences.

3.2.2 The audience-side: Category spanning, typicality, and audience evaluations

A distinct group of mechanisms used to understand the consequences of category spanning focus on the social and cognitive consequences of category spanning. Researchers in this tradition emphasize the role audiences play in shaping organizational outcomes. For example, Hsu and Hannan (2005b), when discussing organizational identity, assert that "[organizational identity is] not simply a list of observable properties," but resides in the "perceptions, beliefs, and actions of contemporaneous audiences." (p. 474)

Central to the perception of organizations by audiences are labels and categories used by organizations. Categories provide an interface between organizations and audiences (White, 1992; Ruef and Patterson, 2009). Labels evoke expectations in audience members, assisting them to navigate the organizational space. From this aspect, category spanning is detrimental because category spanners confuse audience members. Zuckerman (1999a) demonstrates that when financial analysts specialize, they tend to ignore firms that span categories. Hsu (2006c) argues that category spanning confuses audience members about the offerings of the organization, which hinder audience members in affiliating to organizations that fit their preferences. A main mechanism used to justify the negative effect of spanning on cognitive confusion is typicality. Because category-spanning organizations are atypical of each of the categories spanned (Hsu et al., 2009b; Kovacs and Hannan, 2011), audience members will not be able to make sense of them. This is especially the case when the categories spanned are distinct (Kovacs and Hannan, 2011).

The hypothesis that typicality has a positive effect on value builds on extant behavioral research. An illustrative study is presented in Fiske et al. (1987), who investigate how the consistency between labels and attributes (i.e., typicality) influences affect. They presented experiment participants with descriptions of hypothetical persons. Each description consisted of a label that describes the occupation of the person (such as doctor, artist) and personality attribute lists (e.g., obedient, productive, greedy). Fiske and her co-authors demonstrate that subjects automatically judge the consistency between the occupation schemata and the listed attributes (e.g., the doctor-reliable pair is consistent, while the artist-reliable pair is less consistent), and show that the more typical the presented person is to the category-schema, the more likely that the affect toward the category will apply to the presented person. For example, the higher the typicality of the presented doctor is to the "doctor" schema, the higher the affective rating will be. Thus, typicality in positively valued categories increases value.

Demonstrations of the positive effects of typicality on average value can be found in marketing as well. Ward et al. (1992) study how the typicality of the physical design of fast-food restaurants affects customer liking, and demonstrate that the more typical a fast-food restaurant is in terms of physical design (i.e., the more similar it is to the prototype, McDonald's), the more customers like it. Babin et al. (2004) show that the fit of the physical environment of a department store to the category prototype increases customers' affect, quality perceptions, and shopping value. These results are all congruent with the hypothesis that lowered typicality results in lower value.

The chain of explanation goes as follows: category spanning decreases typicality, and lowered typicality results in lowered value to the audience members. Because organizational research has not yet tested this mechanism directly, we state these two steps as the following two hypotheses.

Hypothesis 3a: Organizations that span multiple categories are less typical to the categories they populate than single-category organizations. Hypothesis 3b: Typicality on average has a positive effect on value ratings by audiences.

3.2.3 The interaction between category spanning and quality: How high-quality organizations can benefit from category spanning

The above arguments theorized about the average effect of category spanning. But could certain types of organizations benefit from category spanning? Here we build on the innovation literature and the literature on status and conformity (e.g., Phillips and Zuckerman, 2001a) to explore the interaction of quality, typicality, and spanning. We argue that organizations that are of high status or high quality might benefit from engaging in category-spanning activities. Our argument combines two theoretical streams.

On the one hand, the innovation literature argues that innovation and problem solving are enhanced when previously unrelated categories of knowledge or technology are brought together and integrated. According to this view, spanning distinct and taken-for-granted categories provides actors with a broader variety of information and perspectives, which prompts unexplored mental models (Holyoak and Thagard, 1996), leads to unusual insights (VanLehn and Jones, 1993) and, thereby, initiates innovation (Fleming, 2001; Rosenkopf and Nerkar, 2001; Burt, 2004). The innovation enhancing effect of category spanning has been systematically documented across levels of analyses, including individual inventors (Nerkar and Paruchuri, 2005), innovation projects (Fleming, 2001), business units (Rosenkopf and Nerkar, 2001), teams (Reagans and Zuckerman, 2001) and firms (Hargadon and Sutton, 1997). Fleming (2001) shows that category spanning (recombination) leads to higher variance in performance: recombinative patents are more likely to be a breakthrough success but also more likely to be a failure.

On the other hand, audiences' reaction to category spanners depend on the status of the organization that engages in category spanning. Phillips and Zuckerman (2001a) distinguish between three strata of status: low, middle, and high status, and argue that the consequences of "categorical imperative" differ across these levels. Low-status actors are deviants who do not conform to a minimal set of requirements to be considered as legitimate "players" in the market (Phillips and Zuckerman, 2001a, page 385.). Middle-status actors are only considered legitimate to the extent they conform to the rules and expectations of the relevant audience, therefore one could expect the highest level of conformity (typicality) for middle-status actors. High-status actors, however, are considered legitimate members of the market thus do not risk losing legitimacy by innovating. These actors, thus, are more likely to engage in behaviors and practices that are less typical to their category, and to use these behaviors and practices to further differentiate themselves from their middle-status peers.

Taken these two sets of arguments and applying them to the food facilities setting, we predict that category spanning and atypicality are more prevalent in high- and low-status food facilities than in middle-status food facilities. As Phillips and Zuckerman (2001a) would predict, low-status food facilities such as hot dog stands are not considered legitimate members of the restaurant domain. Our dataset does not contain very low status food facilities that have no chance to become legitimate member of the restaurant domain (we sampled on facilities that are listed as restaurants). Thus the lower and middle status and quality[5] restaurants in our sample have higher need to conform to their categorical schemata and thus stay typical; only restaurants that are high status and quality can successfully use atypicality as a way to differentiate themselves from other restaurants in their categories.

Hypothesis 4: Typicality is mostly beneficial for mid- and low-quality restaurants. Atypicality can be beneficial for restaurants that are of high quality.

[5]In this argument we assume that quality and status correlate (see Podolny, 1993). We acknowledge the possibility that status and quality are loosely coupled. Future research should address whether high quality or high status provides more freedom for restaurants to benefit from categorical differentiation.

3.2.4 Distinguishing the "food court" and the "fusion" type of category spanning

Category spanning can be of two kinds: one in which the elements of the spanned categories co-exist side-by-side but are not combined, and one in which the elements of the spanned categories are combined into the same products[6]. In the case of restaurants we call this "food court" and "fusion" type of spanning (Baron, 2004b). That is, a Mexican-French restaurant can be either such that on one side of its menu it offers Mexican dishes while on the other side it offers French dishes; or, it might mostly serve dishes that fuse elements of the two cuisines. These are qualitatively distinct cases of category spanning. Here we argue that these restaurants would attract different audiences. We build this argument around sociological theory of omnivores of audiences (Peterson and Kern, 1996). Traditionally, as Bourdieu (1984) documented, high-status individuals tended to differentiate themselves from others whom they viewed as lower status by engaging in cultural consumption patterns that were exclusive to them (such as going to the opera, see Bourdieu, 1984). Bourdieu argued that there exists a homology between social stratification and cultural consumption, whereby each social group consumes a set of cultural products that are typical to them. In the last decades, however, this cultural consumption pattern has undergone a substantial shift, and empirical research in the sociology of consumption has documented a shift towards omnivores in modern societies (e.g., Peterson and Kern, 1996; Vander Stichele and Laermans, 2006). High-status individuals have opened up to a wide-range of cultural preferences and consume cultural products from a wide variety of genres, be it traditionally defined high- or low-brow (Peterson and Kern, 1996; Vander Stichele and Laermans, 2006; Warde, 2005). Not only are omnivores open to a wide variety of genres but they are also more tolerant to, and are often in search of an interesting combination of categories and innovation. Thus, we predict that omnivores are more likely to appreciate the "fusion" type of spanning.

As omnivores tend to come from more well-to-do socioeconomic backgrounds (Warde, 2005), we predict that they can afford visiting better quality and more expensive restaurants. Therefore,

[6] The "fusion" vs. "food court" issue could be studied at other levels of analysis as well, such as brand or line of business. We return to this issue in the Discussion.

we predict that:

Hypothesis 5: "Fusion" type of category spanning is more beneficial for high-quality restaurants.

3.3 Empirical setting, data sources, and data operationalization

Our empirical setting is the restaurant domain of San Francisco. The restaurant domain provides an apt setting to study the consequences of category spanning. First, although the restaurant domain contains a variety of organizations, these organizations are similar enough to be compared meaningfully. Namely, they have easily comparable product structure (menus), and the same notion of quality applies to them. This would not be the case if we were to use a multi-domain setting: for example, comparing the product structure of a car manufacturer and a fast-food chain is unobvious, and so is comparing the quality of the offerings of such organizations. Second, analyzing the restaurant domain allows us to build on previous research in restaurants and categories (Freeman and Hannan, 1983b; Rao et al., 2003b, 2005; Carroll and Wheaton, 2009b; Kovacs and Hannan, 2010, 2011). Third, in the restaurant domain detailed and comprehensive records are available on restaurants, their menus, and customer evaluation. Fourth, restaurants are mostly of comparable (and small) size, which rules out the alternative quality-based explanation that larger organizations could develop expertise in multiple categories.

Below we discuss the three data sources we used: Yelp.com, Menupages.com, and the 2011 edition of the Zagat Guide. Our sample contains the 474 restaurants that were covered in all three data sources.

3.3.1 Restaurant ratings

We collected the reviews on the restaurants from the website Yelp.com. The website generates its reviews through a volunteer process in which customers can go online and write a review. Each review captures four pieces of information that is linked to the restaurant: (1) a unique identifier of the reviewer; (2) a star rating, ranging from one to five as an integer number; (3) a text review; and (4) the date of the review. We use the reviewer IDs to control for reviewer specific effects. The average of the ratings is 3.8; the median is four stars. In this paper, we do not utilize the text of the reviews. Note that Yelp.com encompasses a broader audience than many food and gourmet magazines and media outlets (cf. Johnston and Baumann, 2007). We downloaded the reviews for the restaurants starting from January 1st, 2010 until November 1st, 2011[7]. The emerging sample contains 59,605 reviews.

3.3.2 Restaurant menus

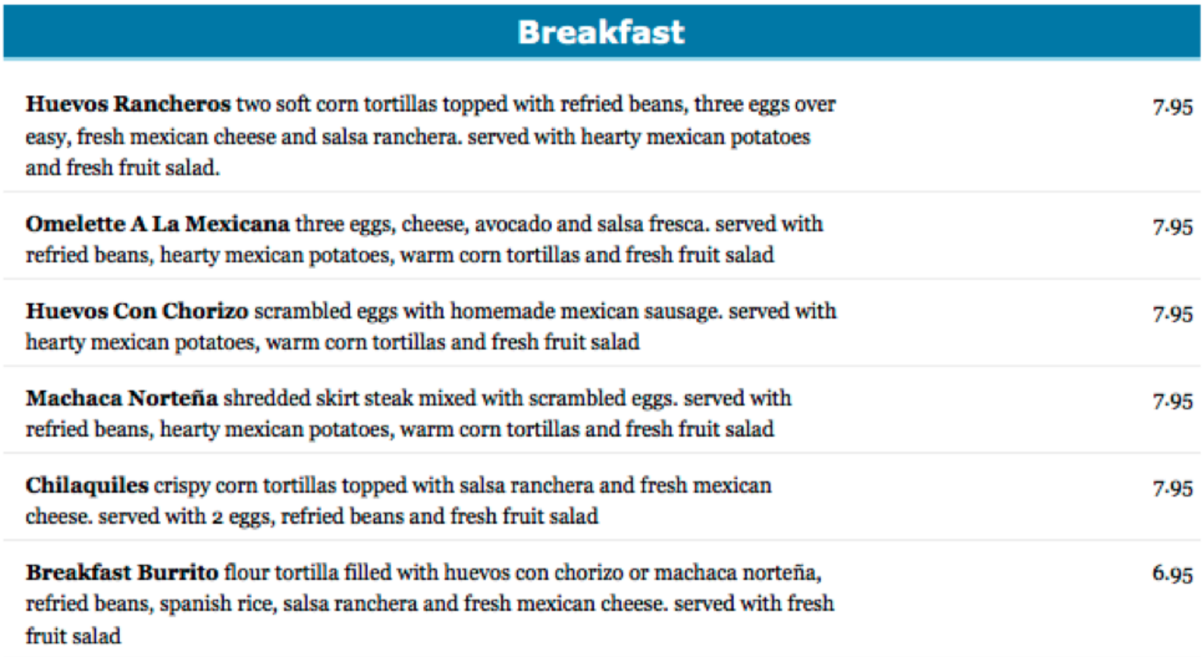
To compile a dataset on the menus of San Francisco restaurants, we used the website menupages.com. We downloaded the menus from menupages.com on October 18, 2011. The menus are sent by the restaurants to the management of the website (either by mail, fax, or by uploading electronically to menupages.com), where the menus are checked and formatted in a standard format. All font styles, pictures, colors and other stylistic items are removed, and all that appear on the menupages.com website are the names of the items offered in the restaurant, their descriptions, and the prices. Figure 3.2 illustrates this standard format, showing a snippet of the menu of "Andale", a Mexican restaurant in San Francisco. Note that in our analyses we use all items on the menu (i.e., not only food items), because we believe that drinks can be part of the schemata as well. For example, a typical Japanese restaurant is supposed to carry sake. As we explain later, our method of analysis ensures that items that appear indiscriminately on most menus (such as "coke" or "juice") drop out from the schemas.

The restaurants self-categorize themselves each into one or more cuisine categories. They can choose from 91 labels[8]. Most restaurants are in one category (44%), others are in two categories (40%), and some are in three or more categories (16%). The most popular categories are "Sand-

[7]One might argue that the menu of a restaurant or the restaurant quality might change substantially in a 22-month period. To check for the sensitivity of our results, we have rerun our regressions on the sample of reviews for the 12 months interval between November 1st 2010 and November 1st 2011, but our main findings did not change significantly (the results are available from the authors).

[8]Menupages.com actually uses a 92nd, "other" category. We excluded from our sample the six restaurants that are listed in the "other" category. On organizational consequences of being listed in the "other category," see Leung and Reschke (2011).

Figure 3.2: Illustration of the format of restaurant menus on menupages.com
This figure shows the first few items on the menu of "Andale", a Mexican restaurant in San Francisco.



Breakfast

Huevos Rancheros two soft corn tortillas topped with refried beans, three eggs over easy, fresh mexican cheese and salsa ranchera. served with hearty mexican potatoes and fresh fruit salad.	7.95
Omelette A La Mexicana three eggs, cheese, avocado and salsa fresca. served with refried beans, hearty mexican potatoes, warm corn tortillas and fresh fruit salad	7.95
Huevos Con Chorizo scrambled eggs with homemade mexican sausage. served with hearty mexican potatoes, warm corn tortillas and fresh fruit salad	7.95
Machaca Norteña shredded skirt steak mixed with scrambled eggs. served with refried beans, hearty mexican potatoes, warm corn tortillas and fresh fruit salad	7.95
Chilaquiles crispy corn tortillas topped with salsa ranchera and fresh mexican cheese. served with 2 eggs, refried beans and fresh fruit salad	7.95
Breakfast Burrito flour tortilla filled with huevos con chorizo or machaca norteña, refried beans, spanish rice, salsa ranchera and fresh mexican cheese. served with fresh fruit salad	6.95

wiches," "Chinese," "Italian," and "Japanese."

The menus vary in length. The shortest menu only lists seven items, while the restaurant with the longest menu offers 293 items (this is "Cheesecake factory"). The average number of menu items is 92.21. Not surprisingly, there is a significant positive correlation of 0.19 between the number of items on the menu and the number of labels assigned to the restaurant.

3.3.3 Using the Zagat Guide to assess quality

We obtained quality scores for the restaurants from a third source, the 2011 online edition of The Zagat Guide. The Zagat Guide, similar to Yelp.com, bases its restaurant ratings on the experience and satisfaction of restaurant goers, who voluntarily submit their ratings and reviews to Zagat. The scores along each of these dimensions can range from 0 (lowest quality) to 30 (highest quality). There are three major differences between Yelp.com and the Zagat Guide. First, and the reason for choosing the Zagat Guide, Zagat reviewers are asked to score the restaurant along three specific dimensions: food, décor, and service. That is, Zagat's reviewers supposedly do not take typicality into account when evaluating the restaurants (Schkade and Kahneman, 1998, see more about this later). Second, Zagat compiles (averages) the individual scores and only publishes the aggregated scores along these three dimensions but not the individual scores.

Third, Zagat provides lower coverage than Yelp.com and Menupages.com, covering 474 San Francisco restaurants in its 2011 edition. Zagat tends to cover "better" restaurants, so this is not a random subsample. The average Yelp.com star rating of restaurants covered in Zagat is 4.1, as opposed to the 3.72 average in the full sample. This might bias our estimates, but we do not see this as a primary source of concern as there is still much variance in the sample, both along the quality ratings and the Yelp star ratings.

3.3.4 Quality, typicality, and price as dimensions of value

The rating reviewers give to restaurants are a function of the perceived value of their visit to the restaurant. The value of a product or service to a consumer can be regarded as the person's overall assessment of the utility of a product, brand, service, or experience (Zeithaml, 1988, page 14.). This overall assessment often involves emotional, social, quality, and price dimensions (see Sweeney and Soutar, 2001). The main research question addressed here is how quality and typicality of a restaurant affect its perceived value and thus influence the choices consumers make. Clearly, our finding that quality increases value is not that surprising, so our main contribution in this sense is examining whether typicality confers value. One issue is worth discussing here. It is rather hard to obtain "objective" quality ratings, especially in the restaurant domain: many would argue that restaurant- and food quality is inherently perceptual and subjective. What we mean by "external quality assessment" is to separate the quality-related components of perceived value from the typicality-related components of perceived value. Thus, by using "food quality" scores from the Zagat Guide as a measure of quality, and also controlling for price, we separate various dimensions of value. We believe that Zagat, by asking to rate specific dimensions, is successful in priming the evaluation of specific quality criteria and hold typicality ratings in the (cognitive) background. In other words, while the typicality affects the perception of quality, elicitations of quality ratings in a dimension-specific way reduces the importance of typicality which it has in the overall evaluation question (which Yelp asks). This assumption builds on results in psychology. For example, research shows that by priming specific dimensions of evaluation, subject overweigh the primed dimension in their overall evaluation of options (this effect is known as the focusing effect, see Schkade and Kahneman, 1998). Schwarz (1996) reports an experiment in which subjects were asked two questions: one about the number of dates they had recently, and one about their general happiness. When the dating question was asked first, the correlation between the answers were 0.66; when the general happiness question was asked first, the correlation dropped to 0.12, proving that when a specific dimension is primed, subjects focus on that dimension of evaluation.

3.3.5 Assessing restaurant typicality

Hannan et al. (2007b) assert that established organizational categories build up of two main components: a label that denotes the category and a corresponding schema that describes the category. A schema usually consists of typical attributes that characterize the category. For example, the schema of the "Italian" category could include words such as "pizza, pasta, mascarpone, tiramisu, and cannelloni;" or the "Japanese" schema could include "sushi, teriyaki, udon, seaweed, nigiri." We note two properties of schemas: first, a schema is not simply a list of attributes but these attributes can vary in importance to the schema, some attributes being more core than others (Murphy, 2002); second, a schema could contain negative elements as well, i.e., elements that should not be in the attribute list, such as having "sushi" on the menu decreases the fit to the "Italian" schema.

The first step in our empirical strategy is to map the schemas underlying the restaurant categories. As Hannan et al. (2007b) write, a schema is a set of attributes that define a category, or more precisely, that describes the central tendency of a category (Hannan et al., 2007b; Murphy, 2002). The typicality of an item in the category is a function of the number of attributes the object shares with the prototype. In our empirical setting, the attributes are words in restaurant menus, and the schema of a restaurant category is a (weighted) set of these words.

Instead of defining what the category schemas are, we take a constructivist stance and learn the category schemas from the menus. We want to identify the words that tend to appear on menus of certain categories but not others, and recreate the category schemas from these word occurrences. In the parlance of computational linguistics, this means learning the category schemas from word-category associations (Church and Hanks, 1990; Manning and Schütze, 1999b).

Note that the approach we take here is a combination of the exemplar view and the prototype view of categories (Murphy, 2002). On the one hand, we follow the exemplar view because we take instances of the category and their descriptor and we learn the category schemas from these instances. Thus, the category schemas we map from the data describe the current schemas in San Francisco[9]. One the other hand, we follow the prototype approach because we identify categories with their central tendency and do not compare the restaurant to all other restaurants in that category individually to assess typicality.

Our approach to map category schemas is as follows. First, we calculate the typicality of each menu words in each of the 91 categories. We calculate the typicality of the word in the category by calculating the Jaccard-similarity of the word to the category. Formally, if $\#(word_i \& category_J)$ denotes the number of times the word i appears on menus in category J, $\#(word_i)$ denotes the total number of times the word i appears on the menus of the restaurants, and $\#(category_J)$ denotes the total number of items in category J, then:

$$Typicality(word_i, category_J) = \frac{\#(word_i \& category_J)}{\#(word_i) + \#(category_J) - \#(word_i \& category_J)} \tag{3.1}$$

See the Appendix for a detailed illustration of how word-category typicality is calculated.

Jaccard similarity is a commonly used similarity measure (Batagelj and Bren, 1995), and it satisfies three desiderata for the typicality measure: first, typicality of an item in a category

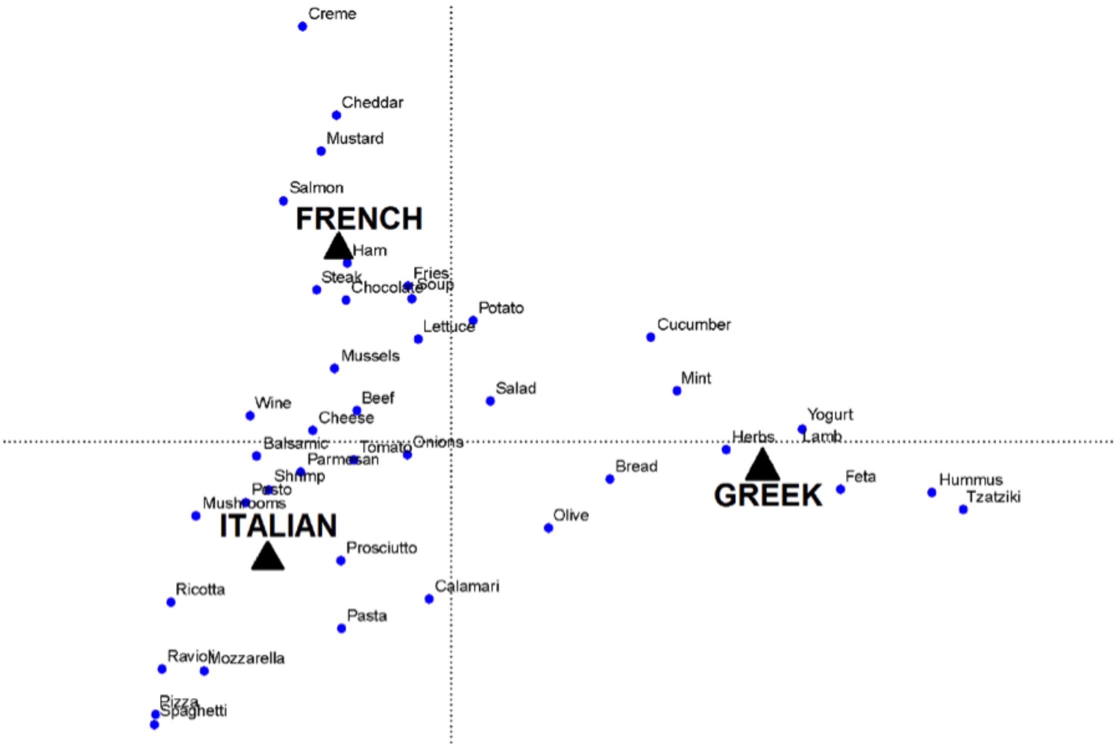
[9]Examining how category schemas vary across other cities and time periods would be an exciting project but it lies outside of the scope of the current paper. But we believe that focusing on restaurant schemata in San Francisco is the right approach to take, as we investigate the effect of category spanning on ratings from San Francisco, so whatever the local schemas are, they need to be used to assess category spanning.

increases with the number of co-occurrences; second, typicality of an item in a category decreases with the number of times the category appears. These two criteria ensure that items that tend to appear with a category have high typicality (e.g., the word "brie" tends to appear on menus of French restaurants). The third property of Jaccard-similarity (typicality of an item in a category decreases with the number of times the item appears), discounts items that appear on most menus indiscriminately (such as "juice" or "coke").

For computational simplicity, to assess category schemas we only include words that are mentioned at least five times in the whole dataset. We exclude all prepositions, conjunctions, and interjections. This leaves us with 12,323 unique words. We calculate the typicality of these 12,323 words in all the 91 categories.

In the next step, we calculate the typicality of each restaurant in each category by aggregating the individual word-category associations by taking the average of the category-word typicality for all words in the menu. Thus, each restaurant is assigned a typicality score in each category, where the typicality score ranges from 0 to 1; 0 denoting the lowest typicality and 1 denoting the highest typicality. Because of the typicality values are low in absolute number (due to the division by the count of words in the Jaccard formula), for better interpretability we rescale the typicality values so that the maximum observed value of typicality will be one. Note that as such a multiplicative rescaling only changes the unit of measurement. Before proceeding with the analysis, we provide three validations for our typicality measure. Figure 3.3, which illustrates category schema for three categories, provides the first validation: Italian, French, and Greek (for readability we only show a select set of words, the schemas contain many more words). In the figure, the centers of the categories are denoted by the triangle, and the distance of words from the category centers are inversely related to their typicality score. This graph was calculated with Correspondence Analysis, a commonly used method to plot dual item-category distance data (Greenacre, 1984). The figure shows that some words are close to the Italian category but not to other: e.g., "pasta," "prosciutto," "ricotta," or "mozzarella." Some words are close to the Italian category but are also close to the French category, such as "onion," "balsamic," or "wine". These are words that are typical to both the French and Italian schemas. Finally, some words such as "bread," "olive," and "salad" are at

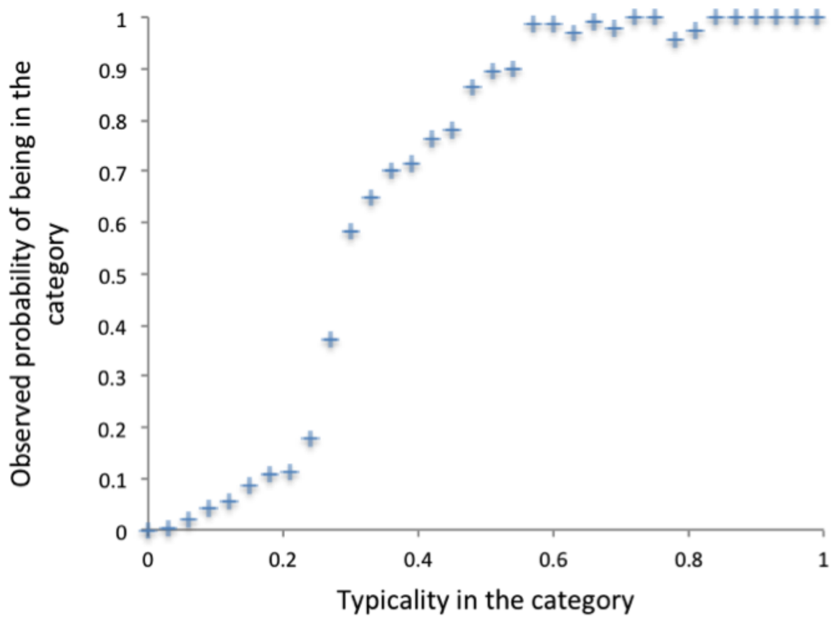
Figure 3.3: Category codes of three selected categories (Italian, French, Greek) Correspondence Analysis on a selected set of words.



equidistance from the three categories, indicating that they are present in all three categories but are not highly distinctive of each categories. The schemas on this figure do correspond to our expectation, providing face validity to our word-category typicality measure.

For a second validation, we build on previous results in cognitive psychology demonstrating that categorization is a positive function of typicality (Hampton, 1998). We expect that the more typical a restaurant is to a category, the more likely the restaurant is self-categorized in that category. To test this proposition, we ran logistic regressions using typicality to predict whether a restaurant categorizes itself in the category. Results (not shown here) demonstrate a strong correspondence between typicality and categorization: the coefficient is positive and significant, and typicality explains 62% of the variation in categorization. This relationship holds even when we do not use the same restaurants to estimate typicality and predict categorization: our analyses show that the typicality values have a strong predictive power even on restaurants outside the sample (taking a hold-out sample approach, see Stone, 1974). On Figure 3.4, we plot the probability of

Figure 3.4: Typicality and the probability that the restaurant will belong to that category



categorization as a function of typicality, and find a positive relationship (note that the shape of this curve is similar to that in Figure 2 of Hampton 1998).

Figure 3.5 provides a third validation, demonstrating that using menu words to explore the category structure of restaurants captures the macro category structure well. We calculated the Jaccard-similarity between the categories (Batagelj and Bren, 1995) in terms of menu word overlaps: two categories are similar if words that appear on menus of category A tend to appear on menus of category B and vice versa. The specific measure we use is

$$sim_{Jaccard}(catA, catB) = \frac{(count\,of\,words\,in\,both\,A\,and\,B)}{(count\,of\,words\,in\,either\,A\,or\,B)} \tag{3.2}$$

On Figure 3.5, we use these similarity values to create a hierarchical clustering of the restaurant categories (to avoid cluttering on the figure, we only plot the clustering for categories with at least five restaurants). As the figure shows, the category structure recovered by this method has high face validity, confirming the validity of mapping categories based on menu word occurrences.

Finally, to measure the typicality of the restaurants in the labels it claims, for each restaurant we

Figure 3.5: Hierarchical clustering of restaurant cuisines with more than five instances
Calculated based on the Jaccard similarity of the cuisines, based on the overlap in menu-words.

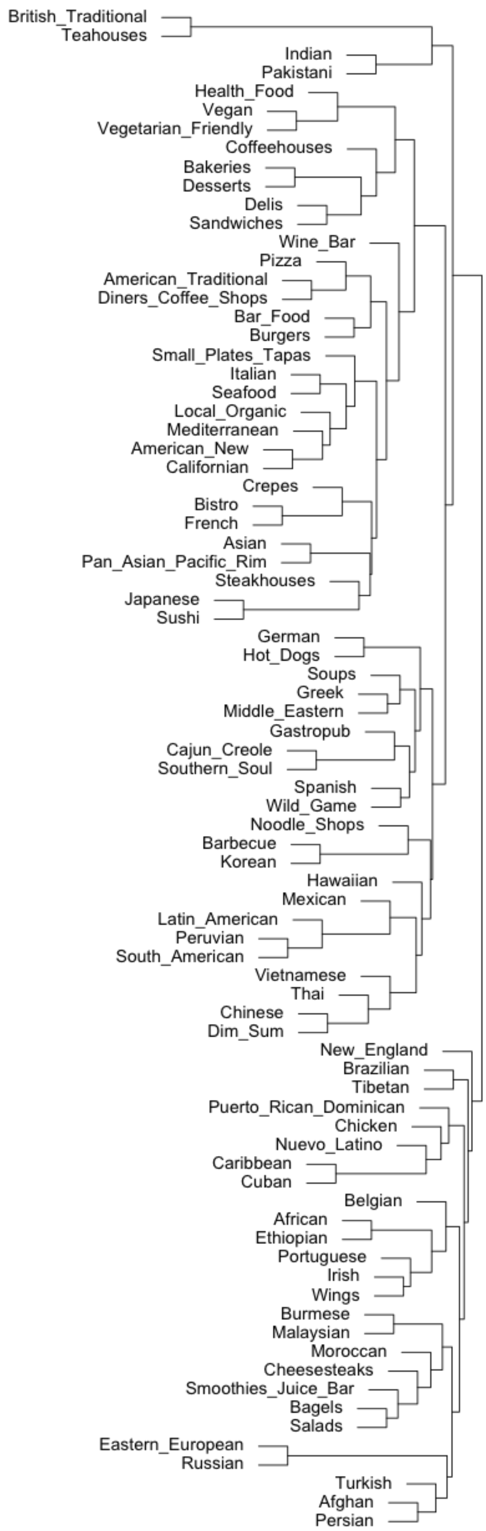
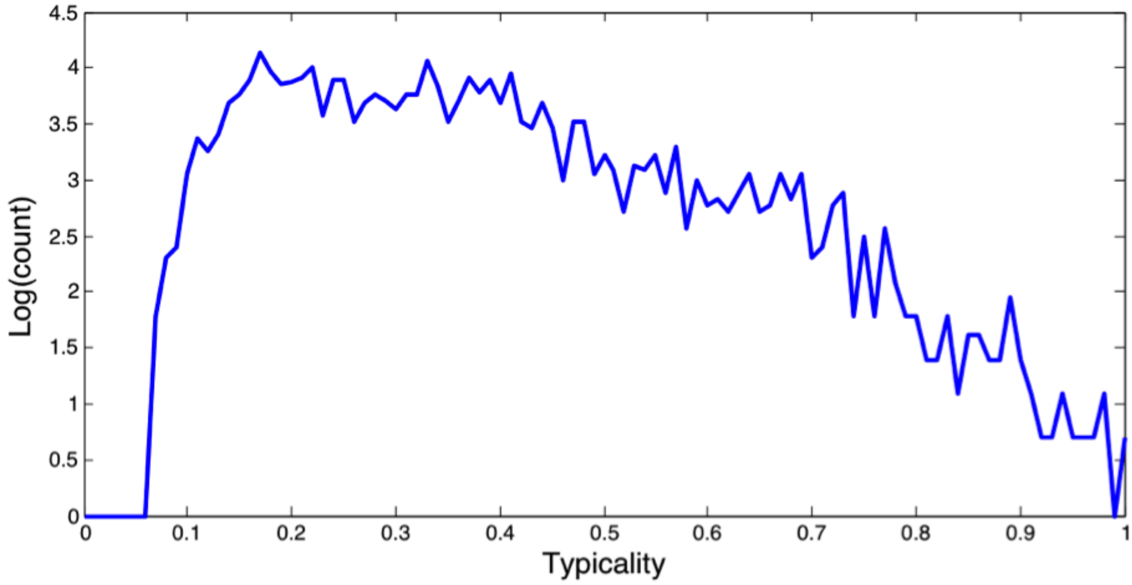


Figure 3.6: Distribution of the rescaled typicality values of restaurants in the categories they are classified in



aggregate the typicality of words it contains on its menu (calculations are shown in the Appendix). Figure 3.6 shows the distribution of the typicality values for restaurants in our sample. The figure shows a large variance in typicality of organizations in the categories they are in. We exploit this variance to understand the effects of typicality on audience value.

Our measure of typicality has an important assumption: we assume that schemas are shared by all actors. While it has been argued that category systems are more useful if there is a consensus about the category schemata (Hannan et al., 2007b), there is undoubtedly some variance in the schemata audience members use to evaluate organizations. As we do not have access to these schemata, we assume that the schemata are shared.

3.3.6 Distinguishing "food court" and "fusion" type of category spanning

As we discussed above, category spanning can be of two kinds: one in which the elements of the spanned categories co-exist side-by-side but are not combined ("food court"), and one in which the elements of the spanned categories are combined into the same products ("fusion"). To differentiate these two cases, we calculate the typicality of each of the dishes in the menus to each of the 91 categories. I.e., we follow the same procedure as for the menus above but now at the dish level. For food courts, we expect that each of the dishes will be only typical to a single cuisine, but that there will be dishes from multiple cuisines on the menu. For example, one half of the menu is Japanese, the other half is Italian. For fusion restaurants, however, we expect that the individual dishes will themselves belong to multiple categories. For example, the dishes on the menu combine ingredients and techniques from Japanese and Italian cuisines.

We measure the extent of "fusion" kind of category spanning by calculating the Herfindahl index of the grade-of-memberships of each dish in each cuisine, and then average these values for each restaurant. For example, if a menu contains two dishes, Dish 1 is 80% Italian and 20% Japanese, Dish 2 is 50% Italian 50% Japanese, then the measure for fusion is calculated as: ((.8^2+.2^2)+(.5^2+.5^2))/2=.59. Note that this fusion measure falls between 1/91 (pure fusion) and 1 (pure food court), controlling for the overall extent of spanning.

3.4 Results

3.4.1 Variables

Our main outcome variable is the rating provided by Yelp.com reviewers. Our first explanatory variable is continuous measure of category spanning: organizational niche width. Traditionally, organizational niche width has been measured with the count of market segments the organization targets (e.g., Hsu et al., 2009b). Recently, Kovacs and Hannan (2011) argued that such a simple counting of categories is not optimal because it does not take the similarity structure of categories into account. For example, while both in two categories, a "Mexican" and "Indian" restaurant has a wider niche than an "American (New)" and "Californian" restaurant. Kovacs and Hannan (2011) propose a novel measure of niche width that captures the similarity structure of the categories spanned.

$$NW = \begin{cases} 0 & if\, catnum = 1; \\ catnum\, x\bar{d_x} & if\, catnum > 1. \end{cases}$$

where catnum denotes the number of categories the organization claims, and $\bar{d_x}$ denotes the average Jaccard-distance between the categories spanned, as calculated in eqn (2). For example, as the distance between the "American (New)" and "Californian" categories is 0.56, the niche width of a restaurant that spans these two categories is 1.12, while the niche width of a "Mexican" and "Indian" restaurant is 1.94.

As we believe that this niche width measure is superior to simple category counting, this is the measure we use in the results presented throughout the paper. However, we note that additional analyses (not shown in the paper) revealed that the results hold even if we use the number of categories as a measure of niche width.

The other two main explanatory variables: (1) restaurant quality measured with Zagat food quality scores downloaded from the online version of the 2011 Zagat Guide and (2) restaurant typicality, have been described above.

We include a number of control variables. First, we control for the price level of restaurants. Yelp.com uses four categories to classify the price level of restaurant, indicating "the approximate cost per person for a meal, including one drink, tax and tips." (Yelp.com) A "$" restaurant denotes "cheap, under $10," "$$" denotes "moderate, $11-$30," "$$$" denotes "spendy, $30-$61, " and "$$$$" denotes "splurge, above $61". In our sample, 32% of the restaurants are in the lowest price range, 49% are in the "$$" range, 15% are in the "$$$" range, and the rest 4% are in the "$$$$" category. To allow for the non-linear effect of price, we included dummy variables for all level of the Yelp price rating. Second, we control for the overall popularity of the restaurants. We measure the popularity of a restaurant with the number of reviews it has received prior to the focal review. Third, we control for reviewers' activism, as previous studies show that activist reviewers tend to give lower ratings (Kovacs and Hannan, 2010). We measure reviewers' activism with the number of reviews they have written prior to the focal review (Kovacs and Hannan, 2010). To take into

Table 3.1: Descriptive statistics and Pearson correlations for the main variables (N=56,605)

Variable	Mean	S.D.	1	2	3	4	5	6	7
1. Rating	3.795	1.11							
2. Price	2.316	0.746	0.034						
3. Niche width	0.474	0.466	0.007	0.057					
4. Typicality	0.255	0.145	-0.048	-0.41	-0.411				
5. Ln(popularity)	5.705	0.777	0.098	0.095	0.082	-0.145			
6. Ln(activism)	1.346	1.276	-0.032	-0.021	-0.001	-0.016	-0.036		
7. Zagat's food quality score	21.988	2.871	0.209	0.383	-0.036	-0.148	0.271	-0.002	
8. Restaurant age at review (in days)	144.376	52.742	0.028	0.036	-0.037	0.118	0.459	-0.092	0.099

account the diminishing effects of popularity and activism, we used the log of the count of reviews written about the restaurant and the count of reviews written by the reviewer. Fourth, we control for the restaurant's age at the review, which we measure with the number of months since the first review of the restaurant. Fifth, to control for cuisine specific effects, we include dummy variables for the categories the restaurant is in. Finally, to control for geographical heterogeneity among restaurants, we include dummy variables for the ZIP code of the restaurant. See Table 3.1 for descriptive statistics and correlations for the main variables.

3.4.2 The main effect of category spanning

We start by investigating the main effect of category spanning on value. Because of the ordinal and bounded nature of the outcome variable (the rating is from one star to five stars), we analyze the effect of category spanning on reviews using an ordered logit modeling framework. This framework has been used previously to analyze restaurant ratings (Kovacs and Hannan, 2010, 2011)[10] . As some control variables are review specific, we conduct the analyses at the review level. To account for possible heterogeneity among reviewers and restaurants, we present robust standard errors, but the results we present hold with alternative approaches to standard error calculations as well, such as clustering on reviewers and clustering on restaurants.

Table 3.2 presents the results. Model 1 shows the main effect of niche width on value. Model

[10] We note that all the results presented below hold with linear regression specifications as well.

Table 3.2: Ordered logit regressions on Yelp ratings
The effect of category spanning (measured with niche width)

	(1)	(2)
Niche width	-0.095***	-0.139***
	(0.032)	(0.032)
Ln(popularity)		0.285***
		(0.014)
Ln(activism)		-0.105***
		(0.006)
Restaurant age at review		-0.006***
		(0.001)
Price=$$ (dummy)		-0.257***
		(0.032)
Price=$$$ (dummy)		-0.188***
		(0.037)
Price=$$$$ (dummy)		0.277***
		(0.056)
Cuisine dummies included	Yes	Yes
Zip code dummies included	Yes	Yes
Log-likelihood	-82,113	-81,665

*** p<0.01, ** p<0.05, * p<0.1.

Robust standard errors are in parentheses. N=59,605

Table 3.3: Linear regressions on Zagat food quality scores
The effect of category spanning (measured with niche width)

	Outcome: food quality score	
Niche width	-0.514***	-0.865***
	(0.042)	(0.04)
Ln(popularity)		0.970***
		(0.019)
Ln(activism)		-0.001
		(0.007)
Restaurant age at review		-0.309***
		(0.036)
Price=$$ (dummy)		1.206***
		(0.042)
Price=$$$ (dummy)		3.911***
		(0.069)
Price=$$$$ (dummy)		-0.019***
		(0.001)
Cuisine dummies included	Yes	Yes
Zip code dummies included	Yes	Yes
R-square	0.4	0.509

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Robust standard errors are in parentheses. N=59,605

2 includes the control variables. Both models confirm the hypothesis that restaurants with wider niches receive lower ratings. Table 3.2 shows that the effect of price is not linear: cheap ($) and very expensive ($$$$) restaurants are more highly rated than medium price restaurants. The control variables have the expected effect: more popular restaurants get higher ratings; and, consistent with the previous results of Kovacs and Hannan (2010), activists tend to give lower ratings.

3.4.3 Quality as a mediator of the negative effect of category spanning

Tables 3.3 and 3.4 analyze the mediating effect of quality. Table 3.3 contains two models investigating the effect of category spanning on quality. The results confirm Hypothesis 2a: category-spanning restaurants receive significantly lower food quality scores.

To test the second step of the mediation, we investigate how the food quality scores affect Yelp ratings. Table 3.4 shows the results. As expected, Zagat scores have a positive effect on Yelp

Table 3.4: The relationship between Zagat food quality scores and Yelp ratings (ordered logit regressions)

	(1)	(2)	(3)	(4)
Zagat's food quality score	0.139***	0.135***	0.135***	
	(0.003)	(0.004)	(0.004)	
Niche width			-0.025	-0.139***
			(0.033)	(0.032)
ln(popularity)		0.156***	0.157***	0.285***
		(0.015)	(0.015)	(0.014)
ln(activism)		-0.105***	-0.105***	-0.105***
		(0.006)	(0.006)	(0.006)
Restaurant age at review		-0.002***	-0.002***	-0.003***
		(0.000)	(0.000)	(0.000)
Price=$$ (dummy)		-0.213***	-0.213***	-0.257***
		(0.033)	(0.033)	(0.032)
Price=$$$ (dummy)		-0.350***	-0.349***	-0.188***
		(0.037)	(0.037)	(0.037)
Price=$$$$ (dummy)		-0.257***	-0.253***	0.277***
		(0.058)	(0.058)	(0.056)
Cuisine dummies included	Yes	Yes	Yes	Yes
Zip code dummies included	Yes	Yes	Yes	Yes
Log-likelihood	-81273	-81020	-81020	-81665

*** p<0.01, ** p<0.05, * p<0.1.

Robust standard errors are in parentheses. N=59,605

ratings: the better the quality of the food, the higher the average star rating on Yelp, even after including the control variables.

The results of Tables 3.3 and 3.4 confirm the mediating effect of quality in the negative relationship between category spanning and value. The comparison of models 3 and 4 in Table 3.4 shows that the mediation is complete: including the quality-related variables decreases the magnitude of the negative effect of category spanning to a level at which it becomes insignificant.

3.4.4 Typicality as a mediator of the negative effect of category spanning

In the next analyses, we investigate whether typicality serves as a mediator in the relationship between category spanning and value ratings. Table 3.5 shows the first step in the mediation analysis: the effect of category spanning on typicality. Because typicality is a continuous variable,

Table 3.5: The effect of category spanning on typicality (linear regression)

	Outcome: typicality	
Niche width	-0.024***	-0.017***
	(0.001)	(0.001)
ln(popularity)		0.001***
		(0.000)
ln(activism)		-0.010***
		(0.001)
Restaurant age at review		0.000**
		(0.000)
Price=$$ (dummy)		-0.054***
		(0.002)
Price=$$$ (dummy)		-0.102***
		(0.002)
Price=$$$$ (dummy)		-0.124***
		(0.002)
Cuisine dummies included	Yes	Yes
Zip code dummies included	Yes	Yes
R-square	0.819	0.846

*** p<0.01, ** p<0.05, * p<0.1.

Robust standard errors are in parentheses. N=59,605

we run linear regressions. Both models in the table confirm the hypothesis that category spanning decreases typicality. We note, however, that the effect size is rather small – this is partly due to the inclusion of cuisine- and ZIP code fixed effects but nevertheless casts shadows on the general practice of assuming that category spanners are clearly less typical to their categories than single-category organizations.

In Table 3.6, we test the second step of the mediation: the effect of typicality on ratings. In model 1, we only include typicality and the main control variables. In this model, typicality does not have a significant effect on ratings. However, as we demonstrated above, the quality scores strongly affect ratings. Thus, to model the effects of typicality, one needs to control for the quality-effects. In models 2-4, we include the food quality scores. To investigate whether the effect of typicality varies depending on the characteristics of restaurants, we include interaction variables with niche width, food quality, and price. As the comparison of the log-likelihoods shows, the full models provide significantly better fit. In models 3 and 4, the estimate of typicality becomes highly significant, and its effect size increases substantially. As models 3 and 4 are the best fitting models, we conclude that typicality indeed increases ratings. This finding confirms the second step of Hypothesis 3.

Hypothesis 4 posits that high quality restaurants can benefit from atypicality. The results in Models 3, 4, and 5 of Table 3.6 confirm Hypothesis 4: the main effect of typicality is positive, but the interaction with food quality is significant and negative. The Zagat quality scores range from 1 to 30, and model 4 suggests that atypicality is beneficial for restaurants that score 23 or higher (calculated as 4.157/.181). Thus, typicality is beneficial for lower quality restaurants but as the food quality of the restaurant increases, the benefits of typicality wane, and for high quality restaurants it turns negative. Figure 3.7 visualizes effect of typicality for restaurants at different levels of niche width and food quality. Finally, note that typicality increases ratings especially for middle-priced restaurant.

The negative interaction effect between quality and typicality also shows up in the negative pairwise correlation between quality and typicality. The restaurant-level -.127 correlation between Zagat food score and restaurant typicality (significant at $p<.01$) indicates that high quality restaurants are more likely to be atypical.

The strong negative interaction effect between niche width and typicality in Table 3.6 is worth further discussion. This finding, together with the positive main effect of niche width, shows that category spanning is detrimental to the extent that the restaurants actually engages in the multiple cuisines it claims in their labels. In other words, if a restaurant claims multiple labels and it tries to be typical to all the labels it claims then it is worse off than restaurants that claim multiple labels but concentrate their efforts to one cuisine and are atypical to the other labels. This finding is consistent with current understanding of category spanning. Kovacs and Hannan (2010), for example, assert that organizations that span multiple high contrast categories suffer most from spanning, because being member of multiple strong identity categories sends conflicting signals to audience members.

Hypothesis 5 posits that "Fusion" type of category spanning is more beneficial for high-quality restaurants. Model 5 in Table 3.6 investigates this hypothesis by adding to Model 4 "fusion" and its interaction with food quality. Recall that to assess whether a restaurant engages in "fusion" kind of spanning, for each dish on the menu we calculate the Herfindahl index of typicality values and then we take the average of these Herfindahl indices for each restaurant. The reason behind this measure is that fusion restaurants tend to have low average Herfindahl values as they combine ingredients and techniques within dishes. As these average values are not normally distributed, we use a binary measure here: a restaurant is tagged as fusion if its average within-dish Herfindahl index is below the population average. Model 5 shows that on average it is detrimental to be a fusion restaurant. However, high-quality restaurants can benefit from being a fusion. This result confirms Hypothesis 5.

Finally, in an additional set of analysis not shown here, we investigated whether reviewer activism moderates the effect of category spanning and typicality on ratings. In line with Kovacs and Hannan (2010), we found that category spanning and typicality have a weaker influence on ratings by activist reviewers. We share the explanation by Kovacs and Hannan (2010), who argue that activist reviewers and reviewers with domain expertise are less likely to use category cues to navigate the organizational space, so the possible confusion arising from category spanning are less likely to affect them.

3.4.5 Robustness checks

To investigate the robustness of the above findings, we conducted several sensitivity analyses. As none of the robustness checks lead to estimates that are substantially different from the results presented above, we do not show them here, just list the alternative specifications we tried. First, as an alternative specification for category spanning, we used a binary version of the niche width variable (0 if single-category organization, 1 if multiple-category organization), and also tried models in which niche width is measured with the number of categories the restaurant populates. Second, we reran all the above analyses on the organization level, aggregating ratings and organizational and reviewer characteristics by organizations. Although some coefficients have changed slightly and some estimated coefficients lost significance in these alternative specifications, all results are

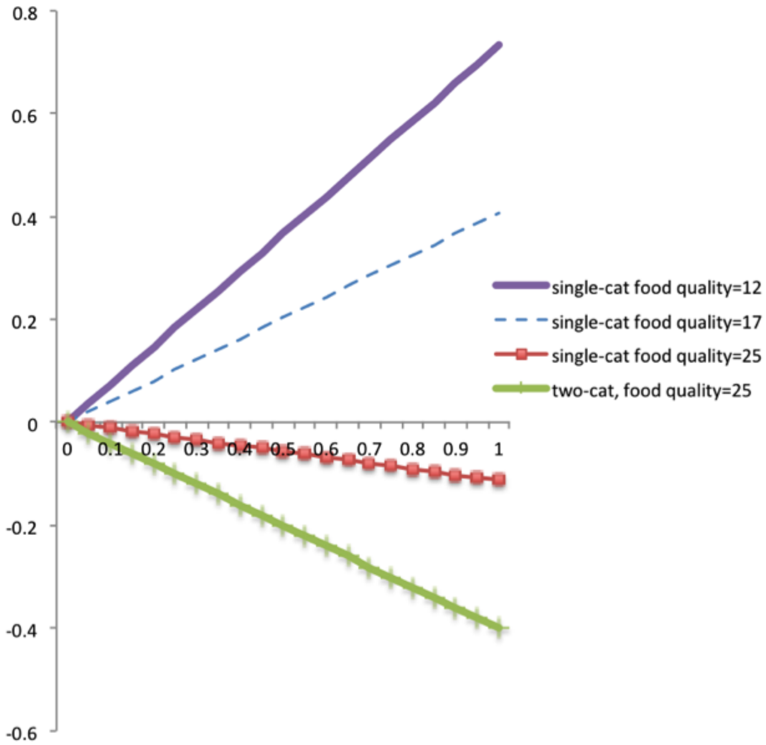
Table 3.6: The effect of typicality, quality, and category spanning on Yelp ratings (ordered logit regressions)

	(1)	(2)	(3)	(4)	(5)
Typicality	-0.040	0.231*	4.617***	4.157***	4.220***
	(0.132)	(0.133)	(0.632)	(0.651)	(0.652)
ln(popularity)	0.284***	0.158***	0.157***	0.154***	0.153***
	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)
ln(activism)	-0.105***	-0.105***	-0.105***	-0.104***	-0.104***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Restaurant age at review	-0.003***	-0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Price=$$ (dummy)	-0.257***	-0.200***	-0.202***	-0.475***	-0.456***
	(0.033)	(0.033)	(0.034)	(0.075)	(0.075)
Price=$$$ (dummy)	-0.197***	-0.327***	-0.342***	-0.633***	-0.619***
	(0.039)	(0.039)	(0.040)	(0.091)	(0.092)
Price=$$$$ (dummy)	0.258***	-0.229***	-0.301***	-0.188	-0.211
	(0.059)	(0.060)	(0.062)	(0.160)	(0.160)
Zagat's food quality score		0.136***	0.180***	0.181***	0.172***
		(0.004)	(0.008)	(0.008)	(0.010)
Niche width			0.178***	0.224***	0.237***
			(0.053)	(0.055)	(0.055)
Niche width x Typicality			-0.469***	-0.537***	-0.538***
			(0.112)	(0.117)	(0.118)
Zagat's food quality score x Typicality				-0.177***	-0.174***
				(0.029)	(0.029)
Price=$$ (dummy) x Typicality				0.773***	0.771***
				(0.189)	(0.191)
Price=$$$ (dummy) x Typicality				0.923***	0.941***
				(0.307)	(0.307)
Price=$$$$ (dummy) x Typicality				-1.696**	-1.533*
				(0.836)	(0.838)
Fusion restaurant					-0.372**
					(0.162)
Fusion X Zagat's food quality					0.013*
					(0.007)
Cuisine dummies included	Yes	Yes	Yes	Yes	Yes
Zip code dummies included	Yes	Yes	Yes	Yes	Yes
Log-likelihood	-81675	-81019	-80989	-80974	-80966

*** p<0.01, ** p<0.05, * p<0.1.

Robust standard errors are in parentheses. N=59,605

Figure 3.7: Predicted effect of typicality
(based on the estimates of Model 4 of Table 6)



consistent the general empirical patterns described above.

3.5 Discussion

In this paper, we set out to study the mechanisms that drive how category spanning affects audiences' value ratings of organizations. We focused on the two mechanisms that are most often evoked in current organizational literature: the quality-based explanation and the typicality-based explanation. Our empirical setting was restaurants and restaurant reviews in San Francisco. To assess the typicality of restaurants, we collected the menus of restaurants. To assess the quality of the restaurants, we collected the food quality scores from The Zagat Guide.

The first contribution of this paper is that we demonstrate that quality mediates the negative average effect of category spanning on value ratings: lower quality leads to lower ratings and restaurants in multiple categories on average confer lower quality than single-category restaurant. These findings provide a direct corroboration of the assumption behind the principle of allocation (Hannan and Freeman, 1989a). We also demonstrate the positive relationship between typicality and audience value ratings. While these two relationships have been widely hold true (Hannan et al., 2007b; Hsu et al., 2009b; Kovacs and Hannan, 2010, 2011), researchers have only provided indirect evidence by demonstrating the negative effect of category spanning on audience value (Hsu, 2006c; Kovacs and Hannan, 2010, 2011). We argued that a more direct test of typicality was needed. On the one hand, no previous research has shown that category spanning indeed diminishes typicality in the categories spanned. On the other hand, the multiple category approach cannot say anything about the typicality of single-category organizations. By collecting and analyzing restaurant menus, we were able to assess the offering of the restaurants and directly assess the fit to the schema of the categories claimed by the restaurants. We believe that such a test is unique in the organizational literature. Combining the effects of category spanning, typicality, and quality in a single model (Table 3.6, Model 4), we find that these mechanisms provide related yet distinct effects on value. Quality does not explain away the effects of typicality, nor does typicality account for the importance of quality.

Besides demonstrating that category spanning has a negative effect on value on average, we also argued that certain types of organizations could benefit from atypicality and category spanning. We found that being typical is especially important for low- and mid-quality restaurants. High-quality restaurants, on the other hand, may benefit from category spanning and atypicality. This pattern is consistent with multiple streams of previous research. Phillips and Zuckerman (2001a) demonstrate that high status actors bear lower costs of illegitimacy than middle or lower status actors. Waguespack and Sorenson (2010) show that high status of film producers increases their chance of getting favorable classification. Another explanation consistent with our findings could concern audience heterogeneity and the tendency toward novelty seeking. Ward and Loken (1988) demonstrate that the positive effects of product typicality are more pronounced if customers do not look for novelty or exclusiveness. Babin et al. (2004), although confirm an overall positive relationship between typicality and value, also note that some subjects are novelty seekers, and slight deviance from the prototypes could often result in higher value. High-quality restaurants can thus benefit from spanning if customers who visit high-quality restaurants are more likely to be novelty seekers than customers of lower quality restaurants.

Finally, we differentiate between two types of category spanning, "fusion" and "food court" (Baron, 2004a). We argued that high-quality restaurants could benefit from fusion type of spanning as they attract novelty-seeking patrons and the high quality or prestige of the restaurant might make these audiences less suspicious of innovative dishes. Restaurants that are of low- or mid-quality should not try to innovate, however, as atypicality will hurt them. Thus, if they are to engage in any kind of spanning, it should be the less innovative "food court" kind of spanning.

Here we would like to note that while we contrasted the two types of spanning at the restaurant menu level, this distinction has interesting parallels at other levels of analysis as well. For example, at the organization-brand level, it relates to the issue whether it is more beneficial for a firm to structure its products into separate but specialized brands or to combine the products under the umbrella of a unified brand. Or, at the organization structure level the "fusion" vs. "food court" differentiation relates to whether different lines of businesses should be integrated or kept separate. While further elaborating on these parallels is beyond the scope of the current paper, we note that, for example, the typicality-based argument would imply that keeping brands and lines of businesses separate is beneficial. This would be an argument favoring conglomerates that keep their brands focused and separate thus do not confuse their audiences. Combining the brands would be beneficial for brands with high quality (or status).

The corroboration of the positive average effect of typicality has implication for current research in organizational authenticity. Carroll and Wheaton (2009b) argue that restaurants that are viewed as authentic are viewed more favorable by patrons. One of their constructs of authenticity is type-authenticity: an organization is type-authentic if its attributes and practices are consistent with its type. This construct is similar to what we call typicality. Our findings indicate that type authenticity should be especially important for low- and mid-quality level restaurant. Future research could address how authenticity relates to quality and category spanning (see also Kovács, Carroll, and Lehman, forthcoming) and how the four types of authenticity proposed by Carroll and Wheaton (2009b) help explain the consequences of category spanning.

3.5.1 Limitations and future research

This study is not without limitations. First, we focused on the restaurant domain. A main advantage of the restaurant domain is that while it contains multiple categories, these categories and their schemas are highly commensurable as the offerings (menus) have the same structure. The question of generalizability to other contexts naturally arises. While we believe that the effects of category spanning, typicality, and quality are generalizable to other settings, we acknowledge that their relative importance might vary across settings. Take for example hypothesis 4, which argued that high-quality restaurants could benefit from being atypical but mid- and low-quality institutions benefit from being typical. This finding might be specific to domains that value creativity, innovation, and novelty seeking but are at the same time put emphasis on category fit. In domains where novelty seeking or artistry is not important (for example, traditional financial institutions) or very important (such as design or creative art), this pattern might not hold. We encourage future research in such domains.

Second, we focused on the effects of category spanning on value ratings by audiences, and one could ask whether our findings generalize to other outcome variables such as profitability or survival. Investigating the effect of category spanning, typicality, and quality on these outcome variables is not possible with our data, and given the intricate relationship between specialism, generalism, profit, and survival, we refrain from making firm predictions. One the one hand, as restaurant ratings positively influence revenues (Luca, 2011b), and value ratings negatively influence mortality (Hannan and Freeman, 1989a), one could argue that our findings would generalize to profitability and survival. On the other hand, however, as population ecologists have shown, the trade-off between specialism and generalism often depends on environmental factors such as the stability audience taste and demand (Hannan and Freeman, 1989a). When the environment is in flux, category spanning could be a more beneficial strategy in order to hedge against changes in taste. Similarly, atypicality might be beneficial if the organization can predict changes in taste and create their own market niches. We leave the disentangling of these effects for future work.

Third, for the categorization of restaurants we use the categories the restaurants self-declare on Menupages.com. This is possibly problematic on two accounts. First, a large proportion of consumers probably do not choose restaurants based on Menupages.com, so this categorization might not be to consumers. Consumers might decide based on alternative classifications for the restaurants or might not use category information at all. In this paper, we assumed that the category claims the restaurant makes on menupages.com coincide with the category claims it makes in other places such as on other consumer websites, on their menus, or during advertising. Second, the self-declared categorization of a restaurant might differ from the "real" categorization for strategic reasons: restaurant might intentionally try to misrepresent themselves by either claiming labels they shouldn't or by not claiming labels that they should. Although we do not think that this behavior is pervasive, with our data we cannot rule it out. We leave it for future research to explore and test the possible consequences of the above two self-reporting biases.

Admittedly, our tests suffer from survival bias: Our sample consists of restaurants that operated in October 2011, at the time of data collection. Unfortunately, data limitations prohibit us from addressing this selection problem. Future research could address this issue by prospectively downloading menus and coding the Zagat Guide in future years. Also, we focused on a single city, San Francisco, and mapped the restaurant schemas in that city. This is adequate for our empirical study as we use local schemas to predict local value, but it would be interesting to gather and analyze menus and reviews from other settings and understand how restaurant schemas differ across cities or countries.

We also assumed in this paper that the category system is stable. This is an especially sensitive assumption in the case of category spanning because in certain cases category spanners may create a new category by combining two previously distinct categories (for example, this is how the "minivan" category was created. See Rosa et al., 1999). We believe, however, that such cases are rare. In the domain of restaurant one could mention “minivan-like” new categories such as “Tex-Mex” or “Californian.” But these are rare cases and are very atypical. We [the authors] can personally recount numerous novel combinations from the San Francisco dining scene (Indian pizza, sushi taco, curry frozen yoghurt, just to mention a few) but very few of these caught on. Most category combinations never become successful and do not establish a new category. Second, not only are these new combinations unlikely to stick but they often take a longer time period to really become an established new category. We checked on menupages.com if any new category has been added during our observation period. No new categories have appeared on the San Francisco menupages.com website since January 1st, 2010. Slightly before our observation period started the “gastropub” category was added. As this example demonstrates, new categories do occasionally emerge. Given that only seven restaurants in our dataset are classified as gastropub, these cases do not threaten the validity our findings (our results hold after dropping these seven restaurants). One would need to take into account the changes in the category structure, however, if the study were to encompass a significantly longer time period, say 20-30 years.

A related limitation is that we analyze the data as cross-sectional and thus we cannot claim to show causality of the effect. Ideally, we would need panel data on both the Zagat scores and on the changes in the restaurant menus. Future research on category spanning is advised to collect panel data on the relevant variables in order to explore the causal relationships among category spanning, typicality, and quality.

Another limitation of the paper is that we could not account for the effect that labels (beyond the effect through typicality) might directly influence the perception of the restaurants and the food served. Research shows that customers change their perception of the products as a function of the labels applied to them. For example, Wansink and Park (2002) demonstrate that subjects rate the same (non-soy containing) snack less tasteful but more healthy if the label indicate that the snack contains soy. Wansink and Park (2002) show that subjects who ate foods with evocative menu names (such as "Succulent Italian Seafood Filet") generated a larger number of positive comments about the food and rated it as significantly more tasty than those eating regularly-named counterparts (such as "Seafood Filet"). Future research should address how these effects interact with typicality effects.

Some might also find it problematic that our measure for restaurant quality is perceptual. We would argue that quality is often ambiguous and in most domains of life quality is relative to viewpoints, tastes, preferences, or customs. Not only is this true for food and most cultural products but for most consumer goods as well (Shepard, 1987). If so, a researcher cannot do otherwise but measure quality as perceived by relevant audience members. Having said this, we encourage future studies of the effect of category spanning in domains where quality can be measured more objectively.

A final avenue for future research could be to use more advanced computational linguistic approaches to map category schemas, such as topic modeling (Griffiths and Steyvers, 2004). On a related point, future research could further scrutinize how category schemas are stored, specifically, whether categories should be represented with the prototype, the exemplar, or some other frameworks (Murphy, 2002). In this paper, we followed a combination of the prototype and exemplar views, but future research could investigate which representation describes audience members' behavior better.

References

Alvarez, J. L., Mazza, C., Pedersen, J. S., and Svejenova, S. (2005). Shielding idiosyncrasy from isomorphic pressures: Towards optimal distinctiveness in european filmmaking. *Organization*, 12(6):863–888.

Babin, B. J., Chebat, J.-C., and Michon, R. (2004). Perceived appropriateness and its effect on quality, affect and behavior. *Journal of Retailing and Consumer Services*, 11(5):287–298.

Baker, G. P. (1992). Beatrice: A study in the creation and destruction of value. *The Journal of Finance*, 47(3):1081–1119.

Baron, J. N. (2004a). Employing identities in organizational ecology. *Industrial and Corporate Change*, 13:3–32.

Baron, J. N. (2004b). Employing identities in organizational ecology. *Industrial and Corporate Change*, 13(1):3–32.

Batagelj, V. and Bren, M. (1995). Comparing resemblance measures. *Journal of Classification*, 12(1):73–90.

Berger, P. G. and Ofek, E. (1995). Diversification's effect on firm value. *Journal of financial economics*, 37(1):39–65.

Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgement of Taste*. Harvard University Press, Cambridge, Mass.

Burt, R. S. (2004). Structural holes and good ideas1. *American journal of sociology*, 110(2):349–399.

Carroll, G. R. and Wheaton, D. R. (2009b). The organizational construction of authenticity: An examination of contemporary food and dining in the us. *Research in Organizational Behavior*, 29:255–282.

Church, K. W. and Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29.

Dobrev, S. D., Kim, T.-Y., and Hannan, M. T. (2001b). Dynamics of niche width and resource partitioning1. *American Journal of Sociology*, 106(5):1299–1337.

Fiske, S. T., Neuberg, S. L., Beattie, A. E., and Milberg, S. J. (1987). Category-based and attribute-based reactions to others: Some informational conditions of stereotyping and individuating processes. *Journal of Experimental Social Psychology*, 23:399–427.

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

Freeman, J. and Hannan, M. T. (1983b). Niche width and the dynamics of organizational populations. *American Journal of Sociology*, pages 1116–1145.

Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. London: Academic Press.

Griffiths, T. L. and Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101:5228–5235.

Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition*, 65:137–165.

Hannan, M. T. and Freeman, J. (1977b). The population ecology of organizations. *American Journal of Sociology*, pages 929–964.

Hannan, M. T. and Freeman, J. (1989a). *Organization ecology*. Harvard University Press: Cambridge.

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

Hargadon, A. and Sutton, R. I. (1997). Technology brokering and innovation in a product development firm. *Administrative science quarterly*, pages 716–749.

Holyoak, K. J. and Thagard, P. (1996). *Mental leaps: Analogy in creative thought*. MIT press.

Hsu, G. (2006c). Jacks of all trades and masters of none: Audiences reactions to spanning genres in feature film production. *Administrative Science Quarterly*, 51(3):420–450.

Hsu, G. and Hannan, M. T. (2005b). Identities, genres, and organizational forms. *Organization Science*, 16(5):474–490.

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

Johnston, J. and Baumann, S. (2007). Democracy versus distinction: A study of omnivorousness in gourmet food writing. *American Journal of Sociology*, 113(1):165–204.

Kovacs, B. and Hannan, M. T. (2010). The consequences of category spanning depend on contrast. *Research in the Sociology of Organizations*, 31:175–201.

Kovacs, B. and Hannan, M. T. (2011). Category spanning, distance, and appeal. Technical report.

Lang, L. H. and Stulz, R. M. (1993). Tobin's q, corporate diversification and firm performance. Technical report, National Bureau of Economic Research.

Leung, M. D. and Reschke, B. P. (2011). Falling through the cracks: How naming assists object comprehensibility in ambiguous space. Technical report, University of California, Berkeley.

Levins, R. (1968). *Evolution in Changing Environments: Some Theoretical Explorations*. Princeton University Press.

Luca, M. (2011b). Reviews, reputation, and revenue: The case of Yelp.com. Technical report, Harvard Business School Working Paper, No. 12-016.

Manning, C. D. and Schütze, H. (1999b). *Foundations of statistical natural language processing*. MIT press.

Markides, C. C. and Williamson, P. J. (1994). Related diversification, core competences and corporate performance. *Strategic Management Journal*, 15(S2):149–165.

Murphy, G. L. (2002). *The Big Book of Concepts*. MIT Press.

Negro, G., Hannan, M. T., and Rao, H. (2010b). Categorical contrast and audience appeal: niche width and critical success in winemaking. *Industrial and Corporate Change*, 19(5):1397–1425.

Negro, G., Hannan, M. T., and Rao, H. (2011). Categorical challenge and response: Modernism and tradition in Italian wine production. *Organization Science*, 22(6):1449–1463.

Negro, G. and Leung, M. D. (2013). "actual" and perceptual effects of category spanning. *Organization Science*, 24(3):684–696.

Nerkar, A. and Paruchuri, S. (2005). Evolution of r&d capabilities: The role of knowledge networks within a firm. *Management Science*, 51(5):771–785.

Peterson, R. A. and Kern, R. M. (1996). Changing highbrow taste: From snob to omnivore. *American Sociological Review*, 61:900–907.

Phillips, D. and Zuckerman, E. (2001a). Middle-status conformity: Theoretical restatement and empirical demonstration in two markets1. *American Journal of Sociology*, 107(2):379–429.

Podolny, J. M. (1993). A status-based model of market competition. *American Journal of Sociology*, 98(4):829–872.

Pólos, L., Hannan, M. T., and Carroll, G. R. (2002b). Foundations of a theory of social forms. *Industrial and Corporate Change*, 11(1):85–115.

Rao, H., Monin, P., and Durand, R. (2003b). Institutional change in toque ville: Nouvelle cuisine as an identity movement in french gastronomy1. *American journal of sociology*, 108(4):795–843.

Rao, H., Monin, P., and Durand, R. (2005). Border crossing: Bricolage and the erosion of categorical boundaries in French gastronomy. *American Sociological Review*, 70(6):968–991.

Reagans, R. and Zuckerman, E. W. (2001). Networks, diversity, and productivity: The social capital of corporate r&d teams. *Organization science*, 12(4):502–517.

Roberts, P. W., Negro, G., and Swaminathan, A. (2013). Balancing the skill sets of founders: Implications for the quality of organizational outputs. *Strategic Organization*, 11(1):35–55.

Rosa, J. A., Porac, J. F., Jelena, R.-S., and Saxon, M. S. (1999). Sociocognitive dynamics in a product market. *Journal of Marketing*, 63:64–77.

Rosenkopf, L. and Nerkar, A. (2001). Beyond local search: Boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal*, 22(4):287–306.

Ruef, M. and Patterson, K. (2009). Credit and classification: Defining industry boundaries in 19th century America. *Administrative Science Quarterly*, 54:486–520.

Schkade, D. A. and Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, 9(5):340–346.

Schumpeter, J. A. (1934). *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle*, volume 55. Transaction publishers.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Erlbaum.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323.

Sweeney, J. C. and Soutar, G. N. (2001). Consumer perceived value: The development of a multiple item scale. *Journal of Retailing*, 77(2):203–220.

Vander Stichele, A. and Laermans, R. (2006). Cultural participation in flanders: Testing the cultural omnivore thesis with population data. *Poetics*, 34(1):45–64.

VanLehn, K. and Jones, R. M. (1993). What mediates the self-explanation e ect? knowledge gaps, schemas or analogies? *Ann Arbor*, 1001:48109–2110.

Villalonga, B. (2001). Does diversification cause the'diversification discount'?

Waguespack, D. M. and Sorenson, O. (2010). The ratings game: Asymmetry in classification. *Organization Science*, 22:541–553.

Wansink, B. and Park, S.-B. (2002). Sensory suggestiveness and labeling: Do soy labels bias taste? *Journal of Sensory Studies*, 17(5):483–491.

Ward, J. C., Bitner, M.-J., and Barnes, J. (1992). Measuring the prototypicality and meaning of retail environments. *Journal of Retailing*, 68(2):194–220.

Ward, J. C. and Loken, B. (1988). The generality of typicality effects on preference and comparison: an exploratory test. *Advances in Consumer Research*, 15(1):55–61.

Warde, A. (2005). Consumption and theories of practice. *Journal of consumer culture*, 5(2):131–153.

White, H. C. (1992). *Identity and control: A structural theory of social action*. Princeton University Press.

Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. *Journal of Marketing*, 52:2–22.

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

Zuckerman, E. W. and Kim, T.-Y. (2003). The critical trade-off: identity assignment and box-office success in the feature film industry. *Industrial and Corporate Change*, 12(1):27–67.

3.6 Appendix

This Appendix provides an illustration how the typicality values are calculated for single and for multiple category restaurants. In this hypothetical example, there are four restaurants, each with only one or two items (see left panel). First, we transform this table to a category-word occurrence table (see right panel). In case of multiple category restaurants we divide the occurrence of the menu items with the number of categories the restaurant belongs to. For example in Restaurant C the “mushroom” item will get 0.5 value in the “French” and 0.5 value in the “Spanish” category.

Name	**Categories**	**Item**
Restaurant A	French	mushroom sandwich
Restaurant A	French	mushroom quiche
Restaurant B	Italian	mozzarella sandwich
Restaurant B	Italian	pizza
Restaurant C	French & Spanish	mushroom quiche
Restaurant C	French & Spanish	paella
Restaurant D	Tex-Mex	chili

Categories	**mushroom**	**sandwich**	**quiche**	**pizza**	**mozzarella**	**paella**	**chili**
French	2.5	1	1.5			0.5	
Italian		1		1	1		
Spanish	0.5		0.5			0.5	
Tex-Mex							1

Then we calculate the Jaccard similarity index for all word-category pairs. Each cell is calculated according to the following formula:

$$Typicality(word_i, category_J) = \frac{\#(word_i \& category_J)}{\#(word_i) + \#(category_J) - \#(word_i \& category_J)}$$

where $\#(word_i category_J)$ denotes the number of times the $word_i$ appears on menus in $category_J$, $\#(word_i)$ denotes the total number of times the $word_i$ appears on the menus, and $\#(category_J)$

denotes the total number of items in $category_J$. For example, the word “mushroom” appears 2.5 times in “French” menus, it appears three times in total, and there are 5.5 words in total in French menus: Typicality(“mushroom”,”French”)=2.5/(3+5.5-2.5)=0.42

Categories	**mushroom**	**sandwich**	**quiche**	**pizza**	**mozzarella**	**paella**	**chili**
French	0.42	0.15	0.25	0	0	0.08	0
Italian	0	0.25	0	0.33	0.33	0	0
Spanish	0.1	0	0.2	0	0	0.25	0
Tex-Mex	0	0	0	0	0	0	1

From this, we calculate the typicality of each restaurant in each category by taking the weighted average of the Jaccard-similarities of the menu words in that category. For example, the typicality of “Restaurant A” in the ”French” category is calculated as (2*0.42+0.15+0.25)/4 = 0.31, whereas the typicality of “Restaurant C” in the “French” category is calculated as (0.5*0.42+0.5*0.25+0.5*0.084)=0.13 Finally, to assess typicality of the restaurant in the labels it claims, we sum the typicality values for each organization in all categories it is in (these values are 0.31; 0.31; 0.13, 0.09 and 1).

Name	**French**	**Italian**	**Spanish**	**Tex-Mex**
Restaurant A	0.31	0.06	0.07	0
Restaurant B	0.05	0.31	0	0
Restaurant C	0.13	0	0.09	0
Restaurant D	0	0	0	1

Chapter 4

The dynamics of project popularity from a collective behavior prospective: A study of creative crowdfunding

Abstract

There are recent studies directing the attention towards the significance of peer influence in product valuation processes. Audience members rarely make their decision in isolation from others, but they infer quality from their peers' behavior. The ever-increasing use of reviewer sites boosts the number of markets where the dynamics of social influence significantly affects the predictability of product popularity. I analyze the dynamics of social valuations processes in the context of creative projects' crowdfunding activity. I approach audience members' decision from a collective behavior prospective. Building on the concept of threshold model, I hypothesize that early audience members draw their decisions relying more strongly on the objective information of project attributes, while audience members who donate to the project later will be influenced more strongly by their peers. This micro level mechanism leads to the macro level consequence, namely that the popularity of a project depends on its early evaluations. The analysis of more than five thousand projects on the Kickstarter.com website reinforces the above argument, demonstrating that objective attributes, such as high quality, does not determine but significantly constrain the dynamics of popularity. It further details attributes which preserve their strong role in the valuation process.

4.1 Introduction

That has puzzled researchers for many years, to be able to predict which products are going to become highly popular and which ones are going to fail. There have been studies comparing product popularity to their objective attributes to be able to respond to this challenge. However, it has been found in many cases, that product quality measures cannot be used as a perfect proxy of higher popularity. For example, Hsu's analysis of film ratings on the IMDB.com website found that many films associated with lower quality offerings through their multiple category membership gained higher popularity than their single category competitors (Hsu, 2006c).

Recent studies shed light on the root of such seemingly counterintuitive results by finding evidence that *"social valuations are not fully constructed but are significantly constrained by objective conditions"* (Zuckerman, 2012, page 227.). Audience members draw their decisions on market products rarely in isolation from others. Relying on previous buyers' judgments is especially amplified in the modern web-based consumer world. Since the foundation of Amazon.com's review system the number of reviewer sites rose vastly, covering a wide range of organizations offering their products and services. In sectors, where peer-influence is present and *"individual decisions are subject to social influence [...] there are inherent limits on the predictability of outcomes"* (Salganik et al., 2006, page 856.).

The MusicLab experiment of Salganik and colleague demonstrates that interpersonal influence among audience members may become as strong that it diminishes the effect of the product's true quality (Salganik et al., 2006; Salganik and Watts, 2008). Attention to the social influence among decision makers offers a new approach towards understanding the dynamics of product popularity. This approach is in line with the sociologist argument, claiming that one cannot fully understand any collective action without analyzing each participating member's behavior, linking the micro level inputs to the macro level outcome (Granovetter, 1978; Coleman, 1986). Outcomes of collective action cannot be inferred from the counting of preferences, because very small changes in them can lead to seemingly paradoxical outcomes.

Previous research has shown that once the effect of social influence takes off, it leads to further audience gain. This process has been referred to as cumulative advantage or Matthew effect (Merton et al., 1968; Merton, 1988; DiPrete and Eirich, 2006). These theories forecast that once a product has reached a certain degree of popularity, it will continue to become even more popular. The theory examines neither the role of product attributes in the process, nor it's relation to the ever-increasing popularity of certain products (Strang and Macy, 2001). The MusicLab experiment provides evidence that product attributes play a significant role, but that role may change dynamically in audience valuations with the increasing dominance of peer influence. Salganik and colleague found that the evaluation of songs were different when participants were shown the popularity ratings with comparison to an "unsocial" world, where they did not have information on how popular the songs were among other participants (Salganik et al., 2006). Moreover when the experimenters presented the inverse of the real popularity, they found that the new ratings did not affect the evaluation of outstanding songs, which continued to be received well in spite of their poor rating. These findings indicate that peer influence indeed alters how audiences evaluate products in some cases more heavily than in others. Salganik and Watts conclude that, *"which particular products turn out to be regarded as good or bad becomes increasingly unpredictable, whether unpredictability is measured directly or in terms of quality"* (Salganik et al., 2006, page 856.). I challenge this remark, which paints a rather negative picture of the predictability of product success. There have been studies demonstrating that in situations when there is no peer influence, product attributes provide a good estimation of success. I argue that the role of certain product attributes does not diminish fully, but changes in social valuation processes dependent on the strength of peer influence.

Peer influence may be of various natures and may not be active under different conditions (Zuckerman, 2012). I focus on social influence with complete anonymity, where audience members don't know each other, they only know the decision outcomes of other members. This condition is widely present especially in online markets[1]. Most of the review sites, such as Amazon, Yelp or TripAdvisor, recommendation systems such as Imdb and crowdfunding sites like Indiegogo,

[1]Zhang (2010) is introducing another very interesting market, where audience members base their decisions on the judgement of their unknown peers'. Analyzing the US kidney market she finds that refusion of a transplant kidney increases future refusals of the same donor kidney.

DonorsChoose and Kickstarter works on the basis of anonymity.

In particular my setting is creative projects on the crowdfunding market. I collected data from the Kickstarter.com crowdfunding website. On this website creators raise money for their creative projects through crowdfunding. Crowdfunding refers to the method where the producer collects small to medium size investment without any financial intermediaries (Ordanini, 2009). The special feature of this setting is that one can collect information on not only the producer side but also the audience side and of the product itself. The website publishes the project creator's previous funding and founder activities within the community; it also keeps track of every project backer, their geographical location, and their previous donations; and finally the website publishes each project in great detail about the project's intended goal, the project's special features, and about the risks and challenges involved in the project execution. On the Kickstarter website I followed the daily progress of more than 5000 projects, which started between the period of 1st and 31st March 2015. I continued the data collection until 1st March and 28th June 2015 - when the last project endedI draw conclusions on the role of project attributes in the dynamics of social valuations building on a dual analysis.

First, I compare how much early backers rely on information of the project when deciding about the donation in contrast to late backers. Borrowing from the threshold model I define early backers by their low threshold, while late backers will be defined by their high threshold (Granovetter, 1978; Valente, 1996; Granovetter and Soong, 1983, 1986). Threshold model is proposed to study situations where only binary choices are present such as leaving the neighborhood or not, buying a product or not, or funding a project or not. In the core of the model lies the interdependence among individual choices. Each individual makes their own choice depending on the number of others who choose before them. Everyone therefore has a threshold for the choice adoption. In short, one's choice of behavior is a function of others' (Schelling, 1971, 1972; Squazzoni, 2008; Granovetter, 1978; Granovetter and Soong, 1983, 1986). I use the sequence of backers to measure the individual thresholds. Low threshold individuals fund projects early. They are the ones who are willing to take higher risks. I expect that once high threshold individuals enter the funding it signals higher popularity, which in turn signals lowered risks and attracts further audience. Respectively, I define early backers, as those members who donate to the project when there is no or very small effect of social influence present while late backers enter their donation when the effect of peer influence is stronger. I argue that the decision-making situation of early audience is markedly different, as these members cannot infer quality information from others' previous behavior. I hypothesize that early backers decide about the project's worthiness of the donations more strongly based on its objective quality, rather then relying on information about the progress of the project. The regression analysis of the full profile of over 130 thousand backers who were donating money towards 5243 projects demonstrates that early backers choose projects with matching geographic location, with higher typicality, with more experienced project founders, with more detailed project introduction, and projects associated with less risk and projects which are open for a longer funding period. Further results shed light on the differences in audience compound between early and late audience. In accordance with the hypothesis, low threshold individuals are more strongly influenced by project attributes than by their peers in their valuations.

Second, I analyze how the role of project attributes change in shaping the overall valuations of creative projects as the effect of social influence increases. To describe the dynamics of the popularity of projects proposed for crowdfunding I use negative binomial regression where I predict the number of daily backers. To account for the magnitude of social influence I use the number of previous backers prior to the focal day. In accordance with previous research I also find evidence for peer influence. Overall, I find that the role of project attributes decrease or fully disappear when controlling for the effect of social influence. The results indicate that project location and it's typicality become less relevant information in the valuation process, while the project's current popularity becomes a more relevant basis of decision. Attributes, like the projects goal amount and it's duration, keep their significant role at a somewhat decreased level. Further, investigating the interaction between the effect of social influence and project attributes I find that under the condition of increased peer influence, the pure effect of project attributes diminish further. For example, the negative relation between higher project goal and popularity turns around with the growing group of backers.

Analyzing how constraints external to the product itself alter the dynamics of its popularity, I find that increased competition affects negatively the project's popularity even after accounting for peer influence. This negative effect however turns around as the project reaches a highly popular state.

This research circumscribes the relation between social influence and social valuations connecting the micro level mechanism with its macro level consequences. The backer level analysis demonstrates that objective conditions constrain more audience members' valuation early on, when peer influence is weak. While on the project-level tI show how increasing social influence affects product popularity. The results suggest that with a better understanding of the role that social influence plays in audience members' valuations, one can make predictions of popularity based on objective attributes with increased accuracy. This research also offers empirical evidence of the threshold model, which has been widely applied in simulation studies but much less on real life observational data.

4.2 Theory and hypotheses

4.2.1 Social valuation and the MusicLab experiment

The recent experiment of Salganik and colleague (Salganik et al., 2006; Salganik and Watts, 2008) draws the attention to the role of peer influence in product valuation.

In their study they created an internet based artificial "music market", where participants rated 48 previously unknown songs from unknown bands. In the independent condition participants did not have any information on other users' prior ratings, while in the social influence condition participants were aware of their peers valuations. They further specified the experimental conditions using different methods to transmit the information on previous ratings of other participants', such as introducing the most popular songs at the top of the list, sorting the songs randomly or presenting the available songs in a grid rather than in a list format.

The results reiterate that social influence indeed plays a significant role in the process of social evaluation. Further it decreases the predictability of outcomes. They find that the popularity ratings of the same set of songs were distinct when participants were rating them within the different conditions. They conclude therefore that *"unpredictability is inherent to the process and cannot be eliminated simply by knowing more about the songs or market participants"* (Salganik et al., 2006, page 856.)

There has been a growing amount of research pointing out the difficulty in forecasting popularity due to the lack of systematic results. Earlier studies hypothesized that this unpredictability stems from environmental changes (Peterson and Berger, 1971). Peterson and Berger label the environment of the music industry as turbulent, as the styles and preferences of their mainly young audience members follow rapid changes. This approach emphasizes the role audience members have in understanding product success. Studies following this approach based their conclusion on an aggregate level analysis of audience preferences. This approach is sufficient when audience members don't influence each others' decisions, or the magnitude of that influence is negligible compared to the magnitude of other, more controlled influences such as newspaper advertisements.

The sudden spread of new, internet based marketing techniques provide a less controlled environment for the emergence of social valuations. Amazon.com in 1995 introduced the possibility for it's users to write their personal opinion of books purchased on the website in the format of a review. This feature of the website quickly become very popular, and other online sellers started adopting the strategy. By now most sellers in product categories like books, electronics, games, videos, music, beverages and wine (Chen and Xie, 2008) allow their customers to help their fellow shoppers in the decision making process with sharing purchase experiences and quality valuations with them. Several studies document that audience members heavily rely on product information from their peers once it is available to them (Chevalier and Mayzlin, 2006; Liu, 2006). Given these conditions audience members participate in social interaction even in spite of not knowing each other personally, as they may reflect on each others previously made decisions. Individual preferences therefore may be considered as a function of others' previous decisions, where one acts - either making the purchase or not - based on others' previous actions. To conclude, once audience members are provided with information on the decisions of their peers, the knowledge of which may alter their own decision, they are participating in a collective action, even if the information is transmitted through the internet rather than through trusted acquaintances.

Sociologists have long argued that individual values and preferences cannot be concluded from the aggregate outcomes (Granovetter, 1978; Coleman, 1986). Social values are the product of social interaction (Merton, 1948, 1995). Individuals in social interaction may alter each others decisions in ways that it is not consistent with their initial intentions. For example, one may prefer film A over film B, but if film B receives more positive ratings the person may decided based on the higher popularity of film B to watch that film in contrast to the initially preferred film A. In order to understand why the given outcome developed, and not any other realization one has to analyze each participating member's behavior. Outcome of a collective action where people may influence each others' decision cannot be inferred from the counting of preferences, because very small changes in them can lead to seemingly paradoxical outcomes. For example, after an interesting and innovative presentation with a very good speaker everyone from the audience may think it deserves an applaud. In the first variation one might go ahead and starts clapping. Soon everyone else joins in. It is clear for everyone that the presentation was a great success. In the second variation of the same situation no one goes ahead to start clapping. In this variation, the success of the presentation is not implicit, where only one person changed their behavior from clapping to not to clapping. This example demonstrates that it may be risky to conclude individual attributes from the aggregate outcomes. Instead, following the sociologist tradition I attempt to connect the micro level inputs of audience members' behavior to the macro level outcome of product valuations (Coleman, 1986). Mapping micro-macro linkages may help explain seemingly counterintuitive results of low quality products achieving high popularity. Overall the MusicLab experiment pinpoints the difficulties of concluding the determinants of product popularity under the condition of social influence. It further demonstrates that certain product attributes carry a significant role in audience members' valuation processes also when peer influence is present. These product attributes may further indicate the outcome of social valuations and product popularity.

4.2.2 Social influence in audience members' choices

In organizational theory the mainstream approach to studying audience choices mostly relied on the analysis of differences between the individual preferences and attitudes towards the producer or the product. For example Zuckerman studies the determinants of why stock market analysts would decide for or against writing reviews about firms on the stock market based on the overall pattern of reviews. He claimed that multiple industry firms are less appealing to specialized analysts as they cannot transmit a clear identity (Zuckerman, 1999a). Hsu, accounting for the trade-off between niche width and fitness, analyzes aggregate audience preferences towards films (Hsu, 2006c), while Kovacs and Johnson claim that higher quality and more typical restaurants are more appealing to their audience (Kovacs and Johnson, 2013). Such an approach however disregards the strength of social forces that ones decision has on others[2] and assumes that audience members make their choices solely relying on information about the producer and in isolation from other members. Accounting for the interdependence of decisions leads to a more precise understanding of the role of product attributes in the emergence of popularity.

Social influence theory explains how people affect each other's decision by their action. Cumulative advantage processes and the theory of Matthew effect predict that once a high number of audience is attracted and the effect of social influence gets stronger it is more likely that further audience members will join in the same decision. One cannot conclude from these theories how audience members value products when there are no other members yet from whom they would be able to infer quality information? There are studies demonstrating that the effect of social influence may get as strong that it makes other effects of product attributes disappear (Salganik et al., 2006; Salganik and Watts, 2008; Koning and Model, 2013). Koning and Model (2013) found proof in a natural experiment of the outweighing effect of social influence. The authors invested $40 into randomly selected projects proposed for crowdfunding on the DonorsChoose website. Their results

[2]There has been a massive line of social network research studying the social influences among related individuals. Although, network theory is greatly concerned with the interconnectedness of individual decisions, the main condition of network analysis is that individuals are related to each other. In the present study I analyze members of a market, who have presumably no connection to each other, and who do not continue any communication between each other.

show that the success rate of those randomly selected projects into which they invested the initial $40 boosts, significantly increased. These experiments, where authors decouple the effect of social influence from the effect of objective product attributes are demonstrating that audience members' influence on each others' decision may be stronger than the effect of objective product attributes. Once we agree that audience perceptions are not pure functions of objective attributes the question arises naturally, what the relationship between objective attributes and audience perception is (Zuckerman, 2012). Why people make certain decisions even though no one made the same decision before them? Why people buy certain products without being able to draw conclusions from others' purchases? And how they decide about which product to choose? To answer these questions I use the threshold model.

4.2.3 Threshold model of interdependent decisions

Schelling described the phenomenon of neighborhood segregation by race, a question, which gained high visibility throughout the 60ies of America, through threshold model. He studied the problem of segregation from a collective behavior perspective. He found that every resident has a composition preference with respect to race of their neighborhood. This is what he called the threshold. Respectively, residents stay until the threshold is not reached. Once the the number of neighbors from the undesired race reach the threshold, residents decide to move to another location. The model demonstrates first, that residents base their action on other member's previous actions, and second, that small changes in the individual preferences may lead to full segregation in the local society (Schelling, 1971).

The model Schelling introduced, generally applies to those collective situations where actors are faced with binary decision. The model's main goal is to investigate the interdependence among decisions, namely how each member's decision depends on everyone else's decision (Granovetter, 1978). For example, one might buy a product only if others have already bought the same product. How many others the person needs to see before choosing to buy the same product is going to be described by the threshold. A person has a low threshold if they are amongst the first to buy the product. High threshold people are defined by those purchasing products only after many others have already done so, in other words, when the product has already established a certain level of popularity. Each individual in the analysis makes their own choice depending on the number of others who choose before them. Everyone therefore has a threshold for the choice adoption and the model predicts the aggregate outcome based on the initial distribution of individual thresholds. In short, one's choice of behavior is a function of others' behavior. The model's underlying condition is that audience members are uncertain about the relative quality of the various products, therefore they infer quality information from its popularity (Gould, 2002; Lynn et al., 2009).

The model has several criteria. Firstly, individuals are presented with mutually exclusive choice decisions, where the costs and benefits of the chosen outcome are a function of the number of others who made the same decision. This condition is present in internet based purchases and reviewer sites where audience members decide between buying a product or not, or between writing a review or not. In the context of creative project funding for example, the risk of investing in the project is decreasing by the increasing number of other investors. Secondly, the model presupposes that decision makers are behaving rationally in the sense of trying to maximize their utility and that they have complete information on the other's decision's outcome (Granovetter, 1978). Rational individuals have the motive to pay attention to others' decisions as they may have limited information at hand or may not be able to make sense of all the available information (Watts, 2002). Third, to be able to trace how one's action influences someone else's behavior the model assumes that decisions are sequential. Although, real life decisions do not necessarily fulfill this condition, on the one hand it allows us to study the interdependence among decisions (Oliver et al., 1985) and on the other hand the realization of internet based purchases, ratings and reviews, happening in well traceable time sequence.

One shortcoming of the model is that it takes the individual thresholds taken for granted and does not offer a clear understanding of why audience members have lower or higher thresholds. The model's main focus is on describing the effect of the individual thresholds on the aggregate level. From the perspective of understanding why some products become highly popular, one needs to analyze the basis of threshold choices. Individual thresholds signal ones attitude towards product purchase decision.

4.2.4 Relationship between product attributes and individual thresholds

There has been a great number of studies concerned with the understanding of the outcome of collective action since the seminal work of Olson (1965). Collective action theorists typically try to develop formal tools to describe the dynamics of collective action and try to identify the factors which determine the outcome (Oliver, 1993). Some of these studies are directly concerned with the effect of participant attributes on the collective behavior outcome. For example, critical mass theory emphasizes the significant role of those actors who behave differently from typical group members (Oliver et al., 1985; Oliver and Marwell, 1988; Marwell et al., 1988; Prahl et al., 1991; Marwell and Oliver, 1993). Marwell et al. (1988) find that group heterogeneity early on accelerates the dynamics of collective action and leads to success. However there is much less interest in understanding the effects of the subject of collective behavior in its dynamics. Social movement studies are reporting that stronger opinion about the subject matter leads to a more predictable action (Oliver et al., 1985; Opp, 1988; Walsh and Warland, 1983).

To describe audience members product valuations from a collective behavior perspective requires the extension to understand how product attributes influence audience members's decision.

Valente (1996), analyzing the adoption the new drug Tetracycline in Illinois (Original study: Coleman et al., 1966), the adoption of hybrid corn and the adoption of family planning, finds that the first adopters were most exposed to further information on the product attributes. In his study the group of early adopters consisted of those doctors who were subscribers to the most medical journals, the earliest adopters of planting hybrid corn were those farmers who visited most often the nearest large city and the first adopters of family planning were those Korean women who experienced the most national family planning campaigns. Conversely, later adopters with high threshold were exposed less to other information sources about the subject matter. Studies of innovation diffusion build on the threshold concept to categorize audience members into groups depending on their attitude towards innovation (Valente, 1995, 1996; Strang and Soule, 1998; Van Braak, 2001; Rogers, 2010). These studies are concerned with the role of personal attitudes in social influence where the participating members are directly connected to each other. However it helps formulating the hypothesis about the role of product attributes in personal thresholds in settings where peer influence is not direct.

Valente (1996) finds that early audience members are less prone to make their decisions based on other's decision, as is also indicated by their low threshold. Early participants face higher risks at the time of decision. As the number of participants is increasing the risk of participation is decreasing. The increased risk in early participation may force them to rely more on information about the product. This leads to the expectation that early audience members value products on the bases of their objective attributes. The effect of product attributes therefore will be a stronger indicator for low threshold participants. Conversely, the effect of product attributes decreases as the product becomes more popular.

Hypothesis 1: As the threshold of the backer increases the effect of product attributes diminishes.

4.2.5 Relation between social influence and objective conditions

While in the previous section I argue that audience members value products differently depending on the level of peer influence, here I theorize about the macro-level outcome of this difference in evaluation. On a micro-level, I argue that early audience members evaluate products based on its objective attributes. But as the group of audience members increases, the process of valuation changes, which makes it difficult to conclude the overall evaluation of the product. In the MusicLab and in the DonorsChoose experiments (I introduced above in more detail), the authors demonstrate that once audience members are aware of each others' decisions, it indeed led to different overall product evaluation, compared to audience members who have no information about each others' behavior (Salganik et al., 2006; Salganik and Watts, 2008; Koning and Model, 2013). The process unfolds in the following way: As more people participate in the group, it signals a stronger positive quality, which further increases the number of audience members participating with high threshold. In spite of the conclusion of Salganik et al. (2006), which states that it is impossible to predict the outcome of social valuation, their findings provide evidence, that in some cases, the outcome can be predicted on the basis of objectivity. In the MusicLab experiment, comparing songs which received outstandingly good or bad overall ratings, they find that these songs keep their ratings even in spite of an increasing level of peer influence. This result suggests, that objective conditions do indeed play a significant role in the process of social evaluations. However the level of significance may change due to the dynamics of social influence.

Proceeding from the first hypothesis I argue that the significance of objective conditions will decrease with the increasing force of social influence.

Hypothesis 2: As the number of backers increases the effect of product attributes decreases.

4.3 Empirical setting

4.3.1 Crowdfunding

My empirical setting is the domain of crowdfunding. One special feature of crowdfunding is to get the "ordinary" audience involved. Audience obtains community-based experiences through visibly participating in a project investment (Belleflamme et al., 2014) providing a unique setting, which allows us to study a two-sided market, the audience behavior and preferences; and matching those to the producer behavior. Another specificity of the crowdfunding domain is that audience does not rely on expert opinion, rather they make individual choices about which project they wish to invest in. I use data from the Kickstarter.com website where creators raise money for their creative projects through crowdfunding. Crowdfunding refers to the method where the producer collects small to medium size investments without any financial intermediaries (Ordanini, 2009).

The idea of crowdfunding stems from a broader concept of crowdsourcing, turning to the crowd with specific requests, such as collecting feedbacks or solutions to problems in order to help entrepreneurial activities (Kleemann et al., 2008). The crowd of people who wish to support the producer may also co-produce the output, select and sometimes develop the offers they deem to be most promising or interesting (Ordanini, 2009). Crowdfunders recruit their investors via the Internet, mostly through specialized websites like Kickstarter.com (Belleflamme et al., 2014). Although founding is possible in several different ways, such as loan or pre-ordering, the project owners on the Kickstarter website are asking for investments.

The significance of crowdfunding strongly increased in recent years both in terms of the numerous business sectors to which it is applied (Agrawal et al., 2011; Belleflamme et al., 2014) and in terms of value of the transactions (Lawton and Marom, 2010). The entrepreneurial sectors where crowdfunding as a source of capital entered were mostly related to creative work, for example the first American crowdfunding site ArtistShare was created in the year 2000. Musicians on the Sellaband website have been raising money to record their albums' since 2006 and three years later Kickstarter.com started fundraising for creative projects. By 2012 there had been 452 crowdfunding sites according to the Crowdsourcing.org's Directory of Sites, conquering other business sectors such as software, food and sport industry, becoming a more and more powerful source of funding for small ventures.

According to the 2014 summary of Kickstarter.com, in that year, over 3.3 million backers - people who donate money towards the project - from nearly every country pledged a total of \$529 million and successfully funded 22,252 projects. Within one year, creators of design projects, the category receiving the most donations, raised \$97 million. On the March 12 2014, backers donated almost \$400,000 within one hour (www.Kickstarter.com/year/2014/data). The significance of crowdfunding can't be measured just in terms of money. Two years earlier, in 2012, 19 movies at the Sundance film festival were funded through the Kickstarter website, four winning top prices and one nominated for an Oscar next to its numerous other successes. To better understand the empirical setting, below I summarize the specificities of the data.

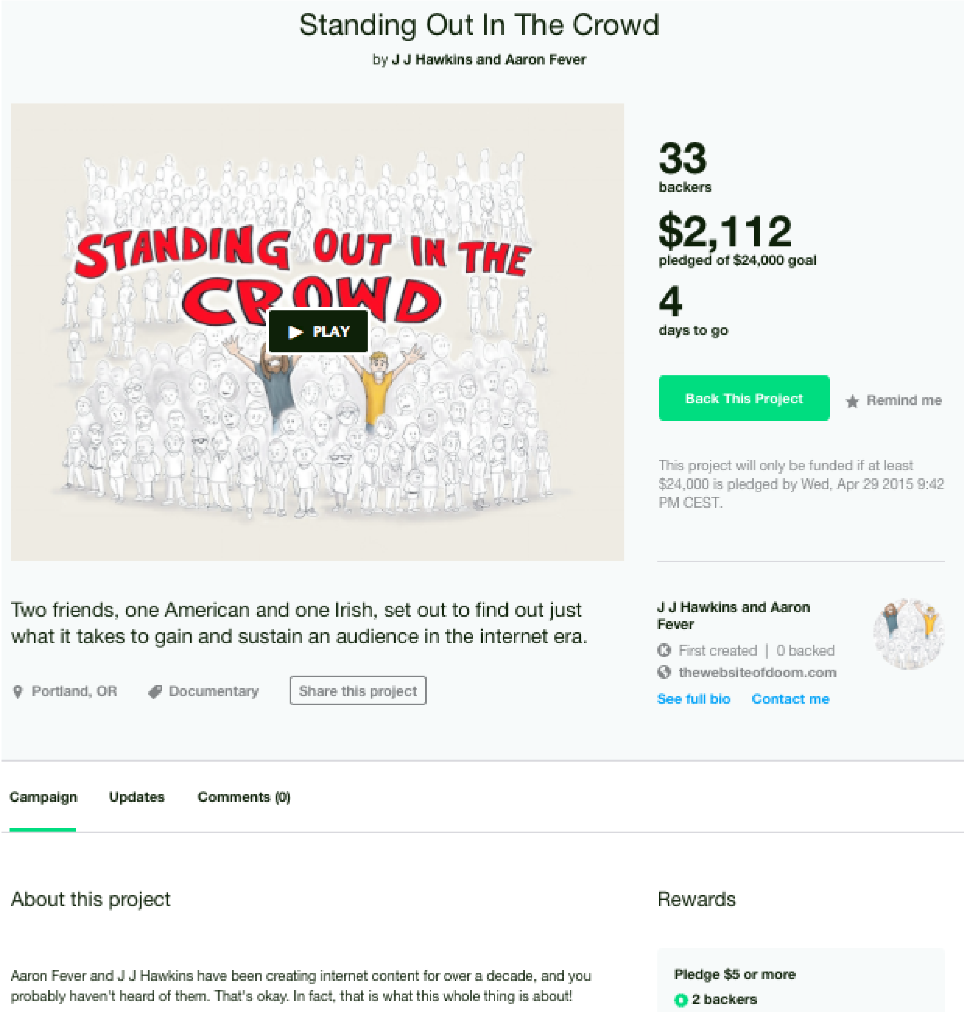
4.3.2 Kickstarter.com

On the Kickstarter.com website people propose creative projects for the purpose of collecting funding. Each project is assigned to one of the following fifteen categories: art; comics; crafts; dance; design; fashion; film and video; food; games; journalism; music; photography; publishing; technology; theater. Prospective backers can acquire further information from the projects introduction page. This page contains a series of information such as the title, detailed description of the project which may contain photos and videos and a short discussion of the possible risks and challenges. There is a possibility for the creator to publish the project on their Facebook page for higher publicity and to insert a link pointing to the official website of the project. Each project has a deadline, which is the final date when the project stops collecting donations. If the project reaches the target amount by the deadline, the creator will receive the total amount of funding. In case the project is not successful in raising its target amount, any money offered for donation will not be transferred to the project creator and the project fails to receive funding. The maximum allowed duration of any project on Kickstarter.com is 60 days.

The project page on Kickstarter also allows the creator to publish information about themselves, such as their name, the number of previously created projects, the location and the number projects they themselves previously sponsored. The website constantly updates how much money is donated so far to the project, the number of backers who made the donation, and the number of days left before the fund-raising project ends. There are ways in which the creator can try to engage more audience and show their own commitment towards the project. For example, they can write updates throughout the duration of the project as well as provide answers to questions and comments left about the project.

A project is considered successful when it raises the goal amount within the predetermined period. In this case the total amount of the donations (even if it exceeds the target amount) gets transferred to the creator. If the target amount is not raised by the end of the predetermined period, the backings are not executed into actual donations and the creator receives no funding for their project. In such a case the fundraising project is deemed unsuccessful and the money remains

Figure 4.1: Illustration of the project introduction page



with the backers. Therefore backers make the actual donation only if the project raises the target amount thus turning out to be successful (see Figure 4.1 for a snippet of a project introduction page).

Data from the Kickstarter website is suitable for testing the threshold model as the site publishes the number of backers instantly on the project page, allowing us to study the sequence of audience's decision. Each member can assess the number of other participants before deciding about their own participation.

Data

I have followed 6965 projects on the Kickstarter.com website between the period of 1st March 2015 and 31th May 2015, meeting the criteria that the project started between 1st March and 31st March. From the final dataset I have omitted several projects for the following reasons a) the project has been cancelled by the creator or the project has been terminated by the website for the reason that it is subject of an intellectual property dispute (5 projects); b) the creator of the project introduces the project purely with the use of images rather than text, making it impossible to compare how typical a project is to the other projects in the same category as there is no project textual description from which to gather this information (36 projects)[3]; or the project goal was defined in currencies other than US dollar (1681 projects)[4]. The final sample contains all project attributes, the daily changes in the number of backers and pledged amount, and the full text of the project descriptions of 5,243 projects. I also collected information on the backers' of these projects. The data file contains 133,484 backing instances, made by 132,778 individual backers. I collected information on the location, the backing sequence, and the project categories for which these backers were donating money previously.

To describe the micro and macro level processes and their relation to each other I build two separate analysis, one on the backer-level and another on the project-level.

4.4 Backer-level analysis

4.4.1 Variables

To analyze the relationship between product attributes and individual *threshold*s, I create a backer-level dataset. This dataset contains 3 types of information, a) backer specific information, b) project specific information and c) information on the project's progress. In case a backer has

[3]Comparing the distributions of the analyzed and omitted projects I found that the omitted projects do not differ from those, which are included in the final sample.

[4]The Kickstarter.com website is initiated from the US, New York, and started expanding to other countries only in recent years. The site mainly appeals to US-based projects and backers, however the number of projects based in other countries grows continuously. To reduce the effect of unobserved heterogeneity I omit projects with non-USD currencies.

multiple backing instances within the data collection period, I take the mean value of these instances across the different variables[5].

Backer threshold is the main outcome variable. I measure the individual threshold by the sequence of backing instances. The website lists all the backers who contributed to the project in descending order of the date and time of their contribution. I use this list to acquire the individual threshold. Further backer specific data are the backer location, the backing frequency, and the backing attitude. I distinguish between backers from the US and backers from outside of the US *(Backer from US)*. This information is readily available from the website. I separate backers who donated to more than one project within the 13 weeks data collection period. I define this group of backers as *frequent backers*.

I describe backing attitude with two variables. Firstly, I create the measure of *focused identity*, which differentiates between snobs and omnivores, the terms used by Peterson and Kern (1996) . I define backers with focused identity as those members of the community, who back projects mainly in one or a few categories in contrast to those backers who contribute to projects in many categories. To define how focused someone is, I count the number of categories to which the person had allocated donations previously and divide it by the total number of categories. The measure will indicate focused identity when the value is low (for example if a person donates to projects within the photography and art categories out of the 15 different categories, she will have a 2/15=0.13) in contrast to a value close to one. Secondly, I measure one's devotion to the categories. I name this variable as *category enthusiasm,* which shows the proportion of one's backings across the categories. For example, one might contribute to six projects in total, two photography and four art projects, which will indicate higher enthusiasm for art projects (4/6=0.66) and lower enthusiasm for photography projects (2/6=0.33). To create the above measures I used two pieces of information available on the backer profile: the total number of previous backings and distribution of these backing across the project categories. Although, these two attitude measures are similar to one another, they both contribute differently towards the understanding of backing behavior.

[5]Within the data collection period I observed 14,363 backers multiple times. Backers, who donated to more than one project constitute 11% of all the backers in the data.

While someone might be generally open to many different categories, the person may still show stronger preference towards one or few categories.

As the last backer specific variable, I include the number of previous backings.This is to account for the magnitude of one's general activity on the website (*previous backing count*).

To include variables accounting for project characteristics, I use information published on the project introduction page. I use two indicators of the magnitude of the project, a) the *goal* amount (measured in US dollars) and b) the project *duration* (measured in the number of days). I employ several quality indicators: a) the lengths of the project description (*description length*: measured by the logarithmic count of the number of words in the description text) and b) the length of the introduction of the possible risks (*risks length* - measured by the log count number of words in the introduction of the project's risks and challenges), c) whether the project has a video introduction, or d) a link to the official website (both dummy variables). These variables signal the time spent with creating the project. For example, writing a long description or creating a video is a lengthy task, which reflects the creator's invested energy. The official website may indicate the trueness of the project, while fewer risks also reflects higher quality.

I include data on the number of previously proposed projects by the same creator to control for the *creator's experience*. The creator has the possibility to use a Facebook link which automatically publishes her project on her Facebook account. I collect information on the number of *Facebook friends* who may be reached via this advertising method. Further, I control for geographic effects including a dummy variable with the value of one if the project was initiated from the US, and zero otherwise (*project from US*). The last project specific variable is the project *typicality*. Project typicality measures the distance between the project in question and the prototypical project of the category. The minimum value of the typicality is going to vary between zero and one, where the value is going to approach one when the project is a more typical member of the category, and it is going to approach zero, if the project is a less typical member (for more detailed discussion of the typicality measure, see Kovacs and Johnson, 2013). To illustrate the difference between high and low typicality I chose two projects from the art category. Project A with the title "Prolific NYC artist needs funding for upcoming exhibition" is created by a painter who is raising money to cover the costs of his exhibition. This project has a relatively high typicality score. Project B received the title “Play on words” from it’s creator, who is collecting money for making a picture book with illustrations of words and phrases in the english language. This project is a less typical one among all the art projects. Figure 4.2 shows the distribution of typicality among projects.

To differentiate between weather the backer donates to projects from a specific location or projects from local to them, I create two variables. The *same country* dummy variable indicates if the backer is from the same country as the project, while the *same city* dummy variable accounts for the similarity between the projects and the backers city.

The dataset contains project progress information. I include the *number of backers* who contributed to the project by the focal day, the percentage of the goal amount already raised *(percent raised)* and the number of days passed from the total project duration *(day sequence)*.

Finally, I include a set of dummy variables to control for unobserved heterogeneity across project categories.

See Table 4.1 and Table 4.2 for descriptive statistics and pairwise correlations.

4.4.2 Results

I study the relationship between product attributes and individual thresholds throughout the project duration using the linear regression model. The result in Table 4.3 presents two models. In both models the outcome variable is the threshold of the backer. As the variable is a continuous measure of the backing sequence, the coefficients are demonstrating how characteristics of the project relate to the time of the donation. Here I talk about the time of donation with respect to the number of previous adopters and not with respect to real time. Low threshold individuals donate early, in the sense that there are not many others who donated before them. In real time there still may be a big gap between the donations, or, on the contrary, the donations may follow each other in close succession.

The two models differ in their number of observations. While in the first model I include every backer who donated at least once throughout the data collection period, I restrict the second

Figure 4.2: Distribution of typicality (cumulative count)

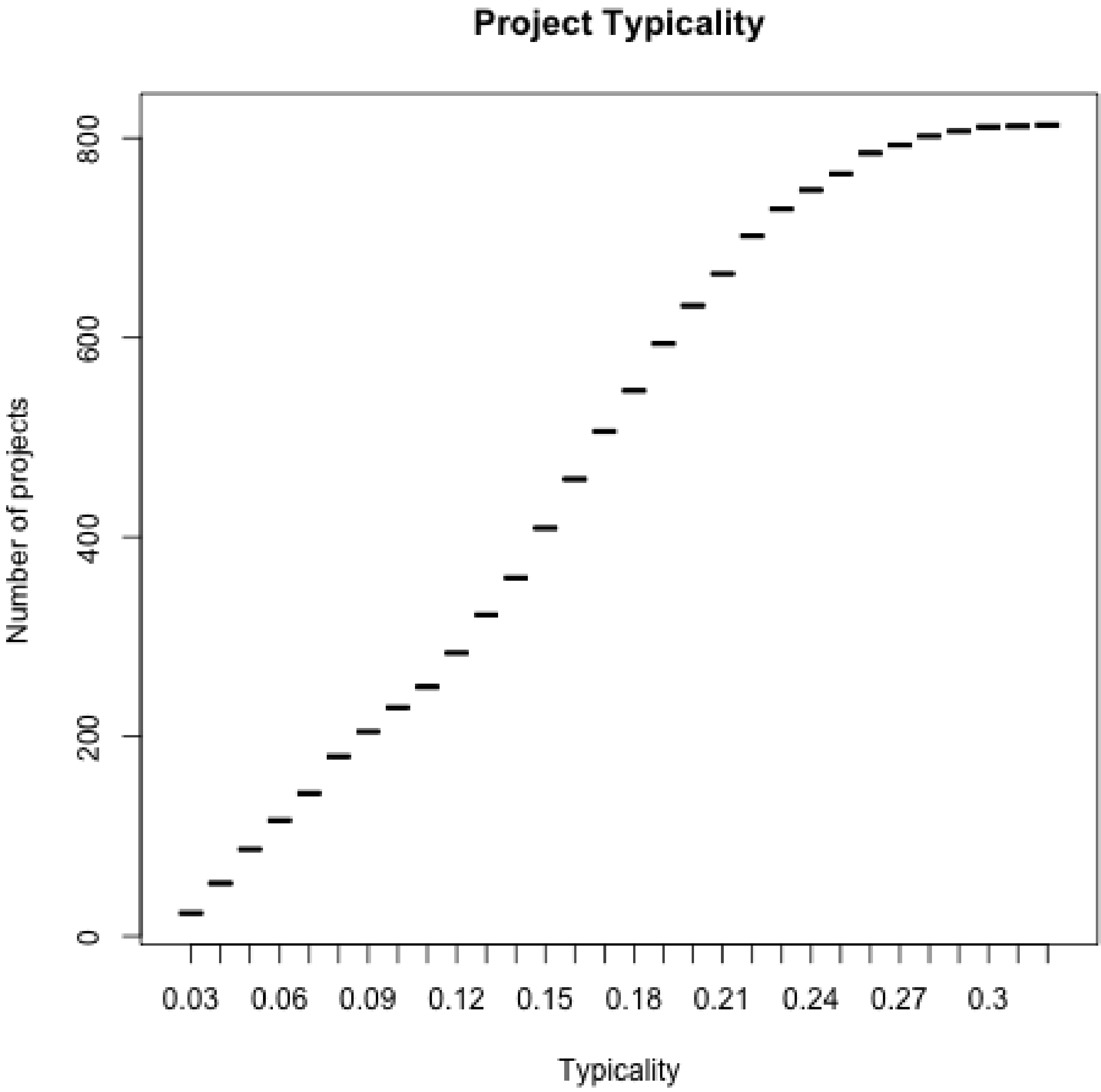


Table 4.1: Descriptive statistics of the backer profile

	Mean	StdDev
1) Threshold	485.306	950.73
2) Backer from US	0.314	0.463
3) Frequent backer	0.108	0.31
4) Focused identity	0.374	0.178
5) Category enthusiasm	0.26	0.211
6) Previous backing count	26.739	57.022
7) Goal (log)	10.109	1.401
8) Duration (days)	31.975	8.746
9) Description length (log)	6.627	0.795
10) Risks length (log)	4.779	0.774
11) Website	0.95	0.213
12) Video	0.959	0.194
13) Creator experience	1.841	3.183
14) Facebook friends (log)	2.272	2.478
15) Project from US	0.948	0.213
16) Typicality	0.199	0.066
17) Number of backers	582.574	1005.672
18) Percent raised	1.147	2.001
19) Day sequence	7.517	10.914
20) Same country	0.057	0.226
21) Same city	0.02	0.137

Table 4.2: Pairwise correlation of the backer profile

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
1) Threshold																				
2) BackerFromUS	-0.016																			
3) FrequentBacker	0.074	0.193																		
4) FocusedIdentity	-0.001	0.32	0.323																	
5) CategoryEnthusiasm	0.034	0.242	0.289	0.768																
6) PreviousBackingCount	-0.047	0.278	0.348	0.364	0.204															
7) Goal (log)	0.516	0.034	0.07	0.002	0.012	0.002														
8) Duration (days)	-0.072	-0.069	-0.067	-0.199	-0.234	-0.048	0.124													
9) DescriptionLength	0.321	0.093	0.109	0.145	0.175	0.067	0.524	0.026												
10) RisksLength	0.154	0.074	0.119	0.149	0.194	0.055	0.298	-0.003	0.425											
11) Website	0.107	0.042	0.004	0.028	0.026	0.011	0.206	-0.033	0.223	0.134										
12) Video	0.065	0.04	0.005	0.022	0.022	0.008	0.165	0.004	0.228	0.15	0.146									
13) CreatorExperience	0.1	0.094	0.123	0.238	0.268	0.088	-0.008	-0.283	0.028	0.158	0.043	-0.097								
14) FbFriends (log)	-0.05	0.059	0.096	0.003	-0.052	0.081	-0.056	0.077	0.029	-0.155	0.032	0.065	-0.08							
15) ProjectFromUS	0.071	0.022	-0.033	-0.066	-0.083	-0.02	0.04	0.013	-0.003	-0.042	-0.002	-0.019	0.009	0						
16) Typicality	0.144	0.11	0.128	0.227	0.299	0.073	0.22	-0.084	0.29	0.293	0.068	0.14	0.19	-0.049	-0.035					
17) NumberOfBackers	0.49	0.071	0.008	0.076	0.065	0.006	0.393	-0.151	0.234	-0.066	0.125	0.05	0.255	0.123	0.095	0.114				
18) PercentRaised	0.171	0.102	0.034	0.108	0.093	0.041	-0.005	-0.05	0.034	0.033	0.085	-0.026	0.173	0.077	-0.036	0.197	0.392			
19) DaySequence	0.075	-0.077	-0.017	-0.099	-0.108	-0.042	-0.024	0.276	-0.095	-0.03	-0.099	-0.054	-0.118	-0.057	-0.002	-0.105	-0.306	-0.164		
20) SameCountry	-0.056	0.357	0.007	0.048	0.01	0.048	-0.039	-0.01	-0.009	-0.024	-0.002	-0.003	-0.029	0.032	0.025	-0.021	-0.036	-0.027	0.009	
21) SameCity	-0.041	0.206	-0.003	0.02	-0.014	0.026	-0.035	0	-0.012	-0.022	-0.008	-0.001	-0.023	0.021	0.014	-0.035	-0.04	-0.036	0.019	0.584

Table 4.3: Linear regression analysis on the backer threshold

	(1)	(2)
BackerFromUS	-103.461***	-90.704***
	(4.752)	(14.558)
MultipleBackings	214.196***	
	(6.803)	
FocusedIdentity	-336.774***	-562.222***
	(17.762)	(56.538)
CategoryEnthusiasm	-387.008***	-639.617***
	(15.956)	(53.162)
PreviousBackingCount	-0.807***	-0.668***
	(0.038)	(0.053)
GoalLog	242.154***	473.661***
	(1.902)	(9.125)
DurationDays	-10.358***	-25.700***
	(0.252)	(1.365)
DescriptionLength	56.966***	178.989***
	(3.126)	(18.583)
RisksLength	46.024***	259.255***
	(3.123)	(16.849)
Video	-44.212***	193.933***
	(10.877)	(63.085)
WebSite	-64.864***	252.018***
	(9.299)	(51.993)
CreatorExperience	-11.304***	-16.574***
	(0.685)	(1.855)
FbFriendsLog	-21.401***	-44.781***
	(0.824)	(3.501)
ProjectFromUS	154.074***	594.498***
	(9.02)	(40.861)
Typicality	-97.923*	-354.891
	(57.945)	(304.809)
NumberOfBackers	0.322***	0.029*
	(0.003)	(0.015)
PercentRaised	46.053***	73.362***
	(1.132)	(5.137)
DaySequence	19.757***	65.335***
	(0.193)	(0.92)
SameCountry	-51.677***	1.525
	(10.816)	(46.999)
SameCity	68.958***	48.312
	(16.891)	(77.936)
Category dummies included	Yes	Yes
Constant	-2,613.695***	-7,665.316***
	(27.144)	(189.92)
Observations	133,484	14,363
R2	0.479	0.722
Adjusted	0.479	0.721
Residual Std. Error	686.414 (df = 133449)	782.333 (df = 14329)

model to those backers, whom I observe multiple times. To account for any possible differences between backers observed single or multiple times I include a control dummy variable in the second model, which has a value of one if the backer donated money to more than one project within the observation period. The coefficient estimates are similar through the two models, therefore I will focus on the results of the first model and I will highlight only the differences.

I start the investigation with describing the relationship between backer attributes and their threshold. As the negative and significant coefficient estimate demonstrates, backers from outside the US tend to donate earlier to projects. Analyzing the relationship between backing attitude and individual threshold I find that backers who allocate their contributions across many categories but show increased enthusiasm for the project's category, tend to donate earlier. Also, backers who donated to a high number of projects previously, tend to take higher risks.

Analyzing the relationship between project attributes and threshold I find that individual threshold increases with the magnitude of the project. Higher goal and longer duration attracts audience with high threshold. Similarly, projects with long description and many challenges are more appealing to high threshold backers.

Signaling higher quality with the inclusion of a video clip and link to the official website will increase the tendency to contributing earlier to the project. Experienced creators with a lot of Facebook friends have similar effect on backer threshold. Higher level of typicality is also more appealing to low threshold backers. Comparing the coefficient estimates in the two models I find that the effect of typicality has lost it's significance.

Backers tend to contribute later to US-located projects, which indicates the positive and significant coefficient estimate. While donating to projects which are in the same country is important for low threshold backers, donating to projects which are located in the same city is increasing the individual threshold. This relationship is also somewhat different when analyzing only frequent backers. Backers tend to donate later to projects located in the same country.

The three progress variables are showing the expected results, projects which attracted more audience before the focal day, which are closer to the end of the project period, and which have raised a higher percentage of the goal amount attract more audience members with higher thresholds.

Backers who contribute frequently to projects generally tend to contribute later, as it is shown in the first model. This might explain the differences between the two models.

4.5 Project-level analysis

4.5.1 Variables

I include three types of variable in the project-level analysis, where the outcome variable is the daily popularity of the project which I measure with the number of backers who donate to the project on the focal day (*Daily backers*).

Firstly, I include a set of attribute measures to be able to follow the change in the role objective attributes with increasing project popularity. Secondly, I include measures of the project dynamics. Thirdly, I include external controls of market competition.

I collect attribute information on each project from the project's page. As I use the same variables in the backer-level dataset, here I will only list them up in order to avoid repetition. I include the project goal, duration, the introduction lengths and the lengths of the description of possible risks, weather the project has a video introduction or an official website, the use of the Facebook link, project typicality, creator experience, and the creator's activity within the community. I further distinguish between projects initiated in the US, versus projects initiated outside of the US to control for possible differences due to the geographical location.

To follow the projects' progress, I include several variables in the dataset. I account for the day sequence, which has a value of one on the day of the project launch and the highest value per project is equal to the project's duration. To capture how close the project is to it's closing I calculate how much time has passed from the whole project duration (time passed %). I measure the strength of social influence by the total number of backers who had donated money up to the focal day. To exclude the possibility that the experienced popularity is due to other, non social but material effects I include a variable showing the pledged amount. One could argue that the pledged amount drives popularity in contrast to peer influence, the alternative argument of which needs to be falsified. An alternative measure of the effect of the total donated amount would be the percentage raised of the target amount.

I employ variables which are external to the project itself, namely measures of market competition. The number of new projects on the website accounts for the intensity of market competition, while I also control for category specific competition, counting the number of new projects within each category.

I include further control variables. Backing activity might be connected to the type of day, such that people are more likely to donate during weekdays, rather than at weekends. Similar results were published on the Kickstarter.com's yearly statistics webpage. Therefore, I include a dummy variable for separating weekdays from the weekends (*Weekend*). I also control for project success and finally, I include a set of dummy variables for each category. See Table 4.4 and Table 4.5 for descriptive statistics and pairwise correlations.

4.5.2 Results

I model the change in the daily popularity of creative projects proposed for funding over a thirteen week period. I analyze the dual effect of project characteristics and social influence to disentangle the dynamics of social evaluation. The form of my data is a time-series panel with a non-negative integer dependent variable. As the dispersion parameter of the daily backers exceeds one[6], the most suitable method of analysis is the negative binomial regression, where the unit of analysis is the daily changes in the funding activity. Figure 4.3 shows the frequency distribution of the number of daily backers.

[6]Following standard econometrics routine, I define dispersion as the ratio of the variance to the mean.

Table 4.4: Descriptive statistics of the project profile

	Mean	StdDev
1) DailyBackers	46.591	205.986
2) TotalBackers (log)	4.402	3.113
3) AmountDonated (log)	4.004	3.31
4) PercentRaised	0.387	0.079
5) Successful	0.173	0.378
6) DaySequence	19.222	13.306
7) TimePassed (%)	0.505	0.29
8) Weekend	0.285	0.451
9) ProjectLaunch-Weekend	0.16	0.366
10) Goal (log)	8.919	1.713
11) Duration (days)	38.031	12.781
12) DescriptionLength (log)	5.611	1.007
13) RisksLength (log)	4.258	0.844
14) Website	0.694	0.461
16) Video	0.775	0.418
17) FbFriends	0.466	0.499
18) Typicality	0.162	0.07
19) CreatorExperience	1.143	0.964
20) ProjectFromUS	0.946	0.226
21) CreatorActivity	7.336	16.623
22) NoNewProjects	276.864	84.86
23) NoNewProjectsWithinCat.	26.692	14.142

Table 4.5: Pairwise correlations of the project profile

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
1) DailyBackers																				
2) TotalBackers (log)	0.202																			
3) AmountDonated (log)	0.155	0.863																		
4) PercentRaised	0.297	0.389																		
5) Successful	0.154	0.376	0.483	0.447																
6) DaySequence	-0.181	0.001	-0.021	-0.022	-0.153															
7) TimePassed (%)	-0.203	0.027	0.02	0.027	-0.126	0.828														
8) Weekend	-0.036	-0.001	-0.001	-0.006	-0.001	-0.004	-0.006													
9) ProjectLaunch-Weekend	0.002	0.018	0.01	0.014	0.032	-0.001	-0.007	-0.008												
10) Goal (log)	-0.017	0.027	0.131	-0.106	-0.124	0.086	0.019	0.001	-0.019											
11) Duration (days)	-0.068	-0.064	-0.089	-0.096	-0.121	0.485	-0.002	0.001	0.011	0.146										
12) DescriptionLength (log)	0.087	0.398	0.479	0.163	0.21	-0.053	-0.029	-0.002	0.007	0.191	-0.065									
13) RisksLength (log)	0.055	0.275	0.333	0.088	0.11	-0.027	-0.014	-0.001	0.016	0.209	-0.034	0.48								
14) Website	0.059	0.258	0.295	0.111	0.163	-0.025	-0.022	-0.002	-0.005	0.052	-0.018	0.281	0.236							
15) Video	0.049	0.249	0.29	0.098	0.13	-0.045	-0.018	0.002	-0.026	0.08	-0.065	0.206	0.142	0.187						
16) FbFriends	0.021	0.096	0.098	0.046	0.068	-0.012	-0.011	-0.001	0	-0.047	-0.009	0.069	0.066	0.145	0.05					
17) Typicality	0.032	0.155	0.199	0.047	0.032	-0.011	-0.006	0	0.002	0.2	-0.015	0.195	0.109	0.072	0.142	-0.003				
18) CreatorExperience	0.025	0.03	0.041	0.054	0.037	-0.022	-0.009	0.001	-0.007	-0.036	-0.036	0.044	0.017	0.039	-0.068	0.035	0.04			
19) ProjectFromUS	-0.002	-0.025	-0.03	-0.005	0.003	-0.013	0	0	-0.018	-0.034	-0.028	-0.044	-0.023	-0.027	0.002	-0.008	-0.003	-0.004		
20) CreatorActivity	0.072	0.259	0.322	0.17	0.236	-0.073	-0.032	0	0.011	0.011	-0.107	0.262	0.141	0.166	0.123	0.138	0.096	0.123	0.008	
21) NoNewProjects	-0.004	-0.009	-0.004	-0.011	-0.024	0	0.002	-0.004	-0.717	0.013	-0.006	-0.005	-0.014	0.001	0.012	0	0.004	-0.001	0.035	0.004
22) NoNewProjectsWithinCat.	-0.003	0.009	0.028	-0.012	-0.026	0.005	0.004	-0.002	-0.392	0.1	0.002	0.025	0.003	0.028	0.077	0.01	0.504	-0.021	0.03	-0.012

Figure 4.3: Frequency distribution of number of daily backers

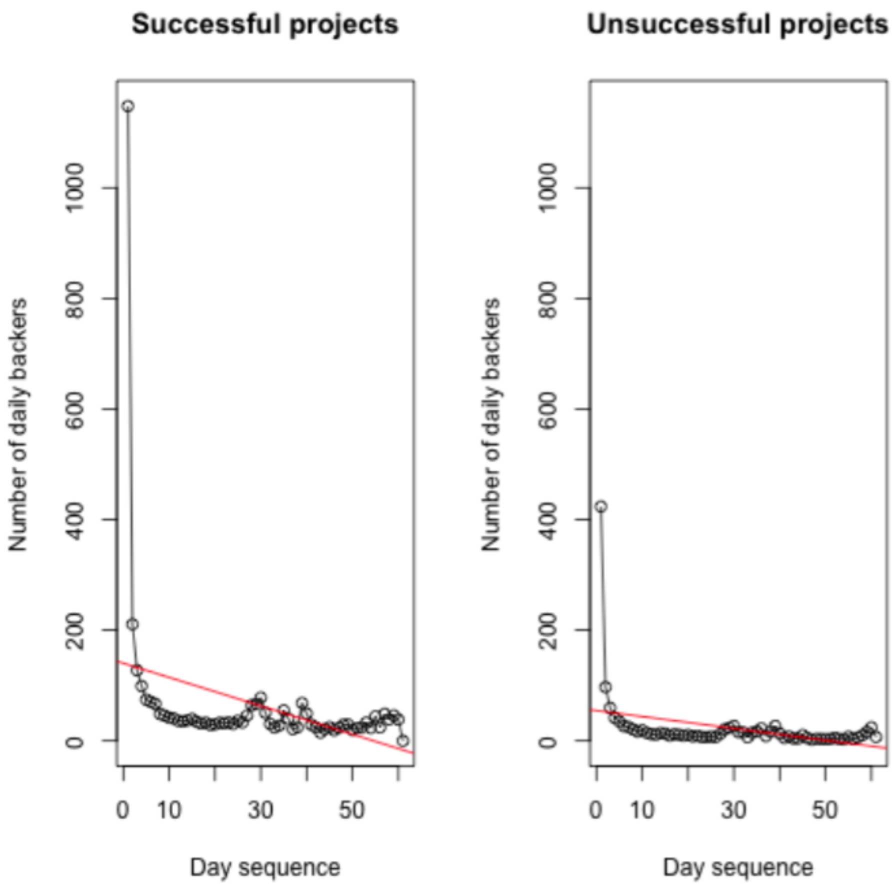


Table 4.6: Negative binomial regression on the project's daily popularity

	(1)	(2)	(3)	(4)	(5)
NumBackersLog		1.079***	1.131***	1.144***	0.993***
		(0.004)	(0.007)	(0.007)	(0.046)
AmmountPledgedLog			-0.079***	-0.104***	0.072***
			(0.007)	(0.007)	(0.015)
PercentRaised				0.414***	0.655***
				(0.027)	(0.027)
Akaike Inf. Crit.	612,864.100	575,300.500	575,152.200	575,093.100	574,354.400

	(1)	(2)	(3)	(4)	(5)
Successful	0.816***	0.178***	0.327***	0.198***	0.335***
	(0.035)	(0.028)	(0.03)	(0.031)	(0.032)
DaySequence	0.006*	-0.0004	-0.002	-0.002	0.018**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)
TimePassedPercent	-2.151***	-2.522***	-2.421***	-2.502***	-4.649***
	(0.13)	(0.114)	(0.114)	(0.114)	(0.29)
Weekend	-0.640***	-0.606***	-0.616***	-0.619***	-0.511***
	(0.027)	(0.023)	(0.023)	(0.023)	(0.056)
CountryUS	-0.137**	-0.021	(0.02	-0.02	(0.095
	(0.053)	(0.045)	-0.045)	(0.045)	(0.11)
ProjectStartWeekend	-0.160***	-0.246***	-0.249***	-0.254***	-0.339***
	(0.047)	(0.04)	(0.04)	(0.04)	(0.092)
GoalLog	-0.047***	0.019***	0.002	0.017**	-0.029**
	(0.008)	(0.007)	(0.007)	(0.007)	(0.013)
DurationDays	-0.021***	-0.017***	-0.017***	-0.017***	-0.019***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
DescriptionWordCountLog	0.231***	-0.025*	0.004	0.006	0.033
	(0.015)	(0.013)	(0.013)	(0.013)	(0.029)
RisksWordCountLog	0.124***	-0.007	0.006	0.008	0.039
	(0.016)	(0.014)	(0.014)	(0.014)	(0.032)
HasWebSite	0.168***	-0.042*	-0.022	-0.021	0.078
	(0.028)	(0.025)	(0.025)	(0.025)	(0.051)
HasVideo	0.162***	-0.252***	-0.219***	-0.219***	-0.430***
	(0.03)	(0.028)	(0.028)	(0.028)	(0.051)
HasFacebookLink	0.012	-0.034	-0.029	-0.027	0.011
Akaike Inf. Crit.	612,864.100	575,300.500	575,152.200	575,093.100	574,354.400

	(1)	(2)	(3)	(4)	(5)
	(0.025)	(0.021)	(0.021)	(0.021)	(0.049)
Typicality	1.920***	-0.128	0.089	0.07	-0.095
	(0.304)	(0.265)	(0.265)	(0.265)	(0.476)
CreatorExperience	0.039***	0.035***	0.037***	0.031***	0.03
	(0.013)	(0.01)	(0.01)	(0.01)	(0.046)
CreatorCommunityBelongingness	0.003***	-0.001	0.0004	0.0003	0.005*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
NoNewProjects	-0.001***	-0.001***	-0.001***	-0.001***	-0.003***
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0005)
NoNewProjectsInSameCategory	-0.004**	-0.0001	0.0002	0.0001	0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
NumBackersLog:AmmountPledgedLog					-0.039***
					(0.002)
NumBackersLog:DaySequence					-0.003***
					(0.001)
NumBackersLog:TimePassedPercent					0.356***
					(0.046)
NumBackersLog:Weekend					-0.023***
					(0.009)
NumBackersLog:CountryUS					0.014
					(0.017)
NumBackersLog:ProjectStartWeekend					0.01
					(0.015)
NumBackersLog:GoalLog					0.019***
					(0.002)
Akaike Inf. Crit.	612,864.100	575,300.500	575,152.200	575,093.100	574,354.400

	(1)	(2)	(3)	(4)	(5)
NumBackersLog:DurationDays					0.0001
					(0.001)
NumBackersLog:DescriptionWordCountLog					-0.003
					(0.005)
NumBackersLog:RisksWordCountLog					-0.005
					(0.005)
NumBackersLog:HasWebSite					-0.019**
					(0.009)
NumBackersLog:HasVideo					0.042***
					(0.009)
NumBackersLog:HasFacebookLink					-0.006
					(0.008)
NumBackersLog:Typicality					0.038
					(0.073)
NumBackersLog:CreatorExperience					0.001
					(0.007)
NumBackersLog:CreatorCommunityBelongingness					-0.001
					(0.0004)
NumBackersLog:NoNewProjects					0.0002**
					(0.0001)
NumBackersLog:NoNewProjectsInSameCategory					-0.001**
					(0.0004)
Category dummies included	Yes	Yes	Yes	Yes	Yes
Constant	3.524***	-0.662***	-1.097***	-1.191***	-0.512*
	(0.151)	(0.128)	(0.134)	(0.134)	(0.278)
Akaike Inf. Crit.	612,864.100	575,300.500	575,152.200	575,093.100	574,354.400

	(1)	(2)	(3)	(4)	(5)
Observations	161,052	161,052	161,052	161,052	161,052
Log Likelihood	-306,399.100	-287,616.300	-287,541.100	-287,510.500	-287,123.200
theta	0.044***	0.078***	0.078***	0.078***	0.079***
	(0.0002)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Note:*p<0.1;**p<0.05;***p<0.01					
Akaike Inf. Crit.	612,864.100	575,300.500	575,152.200	575,093.100	574,354.400

The first model in Table 4.6 introduces the effect of all variables on the number of daily backers apart from the main explanatory variables.

I start with analyzing the relationship between the project's daily appeal and it's characteristics. In Model 1 both variables which are controlling for the project's relative volume have a negative effect on the number of daily backers. Longer projects aiming to collect more money tend to attract less audience. Certain characteristics of the way the project is introduced also influence daily popularity. Longer, more detailed introduction increases the appeal, while audience members value positively if the creator proposes all the possible risks and challenges the project may have to overcome. Including a short video clip to the project introduction page to enrich the ways of presenting the main goal of the project or presenting a link to the project's official website both have a positive effect on the number of daily backers. Projects with high typicality value, also attract a higher number of daily backers.

Audience members may draw conclusions of the project by information about the creator. Creators advertising their project to Facebook friends bears no significant improvement on daily popularity. The creator's past experience in proposing projects and her activity on the webpage attracts more audience members, however the effect size remains modest.

The negative and significant coefficient estimate indicates, that projects, which are initiated in the US are less appealing to their audience.

Investigating the relationship between market competition and project popularity, I find that both increasing global and local competition decreases the project's popularity. The increasing number of new projects in the same category decreases the number of daily backers at a higher rate than the increasing number of projects on the website, without consideration of their category.

To investigate the effect of social influence on the number of daily backers, I include first of all the logarithmic count of attracted audience members, who backed the project before the focal day. Model 2 shows positive and significant coefficient estimates, demonstrating the effect of social influence. As the size of attracted audience increases, the more backers will donate the next day. After controlling for the effect of social influence,I argue that the effect size of project attributes decrease. The results of Model 2 demonstrate the decrease in many coefficients. The effect size of the project goal, the project duration and creator experience dropped. The significant effect of detailing the possible risks and the effect of typicality disappeared altogether. Introducing the project in a longer description seem to negatively affect the size of new backers once we control for peer influence. The significant and positive effect of the use of video clips and official websites also turned into a negative and significant coefficient estimate. Accounting for peer influence does not affect the negative effect of market competition when we measure it globally, while the negative effect of local competition disappears.

In Model 3 I include the donated amount in the model to test weather further backers are influenced in their decision by the previously offered amount or whether they decide based on the size of the attracted group. The variable has a negative and significant effect on the number of daily backers, while the effect size of social influence remains similarly strong. This finding demonstrates that audience members draw conclusions of the attracted group size rather than the attracted amount. To complement the previous argument, in Model 4 I include another measure of the size of donation. Namely, I include how much percentage of the target amount is offered for the project. Once the project starts reaching a higher percentage of the target amount, it will also become more popular. The effect size of reached target percentage remains behind the effect size of peer influence, which has the strongest relationship with popularity.

In Model 5 I further investigate the relationship between the increasing social influence and product attributes. Including a set of interactions, I find that projects with higher targets are less attractive, the negative effect of which decreases with the increasing social influence.

Figure 4.4: Interaction between project goal and backer size

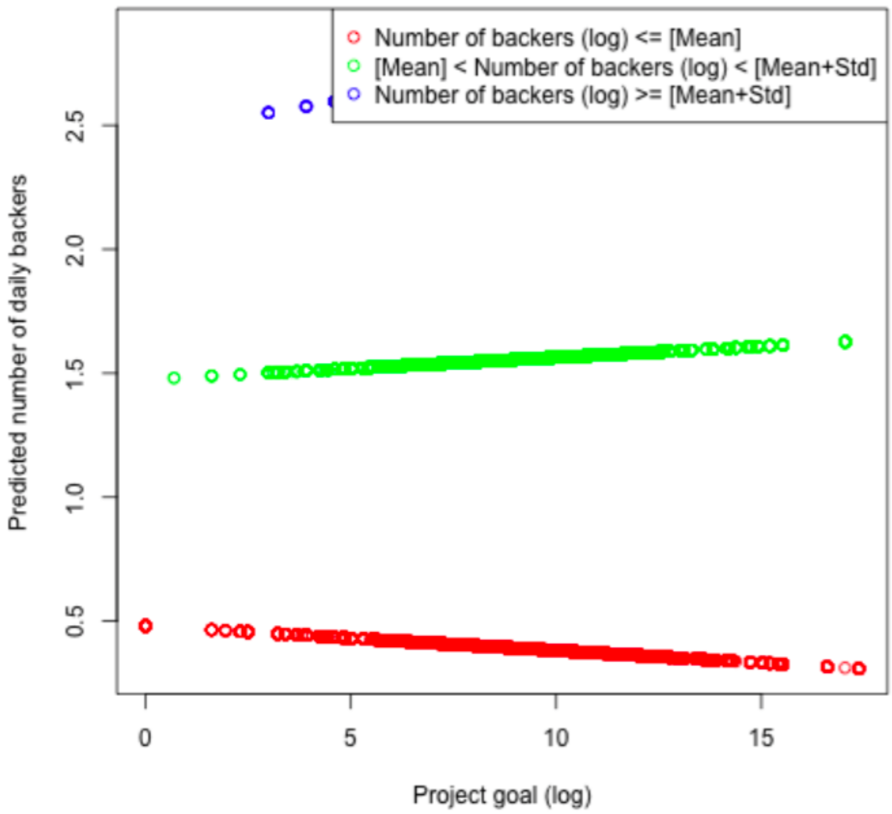


Figure 4.4 demonstrates the relationship between the number of daily backers and the project's goal amount. Similarly, I find that the negative effect of having a video clip attached to the project introduction is decreasing with the increasing peer influence. The existence of an official website is generally valued positively by audience members. Although, the coefficient remains positive when controlling for the the interaction, it's significance disappears. The effect of having an official website changes with the dynamics of peer influence. As peer influence increases, the positive effect of the website decreases.

Audience members tend to value shorter projects. Projects with longer duration have a negative effect on the daily popularity. This negative effect does not change when peer influence becomes stronger. The interaction term between backer size and project duration is not significant.

On the one hand, the negative effect of increasing global competition decreases as the project reaches higher popularity. On the other hand, including the interaction term illuminate the relationship between local competition and increasing popularity. Increasing local competition may increase the number of daily backers, but as the popularity increases the positive effect of competition decreases.

The comparison of the AIC values throughout the models underlines that Model 5 provides the best fit. Including interaction terms to describe how audience members' valuations change with respect to product attributes when popularity increases, improves the model.

4.6 Conclusions

Comparing the results of the backer and project-level analysis I conclude that the results support the hypotheses. Investigating the difference between low and high threshold backers I find that low threshold backers tend to donate to projects with a set of certain attributes, while high threshold backers differ in their project preferences. To understand this conclusion one has to keep in mind that the analysis was based on the assumption, that high threshold backers are contributing towards projects which already gained some level of popularity. This information is inherent in the fact that high threshold means someone who donated as the xth member in the backing sequence.

Those projects which are a) aiming to raise a relatively small amount of money, b) throughout a longer period, c) with a shorter description text, d) including a video clip and e) link to the official website, which are f) more typical and which are g) not located in the US are more attractive to individuals who cannot conclude about the likelihood of the project success from their peers but they have to rely on the available information on the project attributes. Projects with the above characteristics tend to gain popularity early on.

The project-level analysis reiterates the previous findings. Projects with similar characteristics are able to gain higher popularity. Information on the magnitude of the project has a significant effect on popularity even after controlling for the strength of social influence. Most of the quality indicators retain their original relationship to popularity when controlling for peer influence, although their effect size may decrease with the increasing number of attracted audience.

The general pattern shows that product attributes matter, however their role is changing with the dynamics of popularity. Individuals still rely predominantly on information on the product attributes when deciding to contribute to the project early on. Once the project gains some level of popularity, audience members start to rely more heavily on the project's current popularity, and less on the product information. However, there are certain attributes which may circumscribe the likelihood of project success.

4.7 Discussion

In this paper, I study endogenous audience dynamics. Firstly, I analyze the effect of the products' objective attributes on the audience valuation and its changes within the dynamics of social influence. Doing so I build on the concept of threshold model. Secondly, I investigate the consequences of the previous argument on the macro level. I find that audience members value products differently when social influence is week in contrast to when social influence is strong. Product attributes play a more significant role in the latter case. However, I demonstrate that certain product attributes retain their significant role in audience evaluations in spite of the strong effect of social influence. The empirical setting is creative projects proposed for crowdfunding on the Kickstarter website.

There have been studies analyzing the interconnectedness of audience members' decisions especially in the herding literature, where scholars analyze network externalities (Katz and Shapiro, 1985), the sanctioning of deviants (Akerlof, 1980) and taste for conformity (Becker, 1991) as sources of social influence. In the present study audience members are anonymous, therefore the main reason for following each other's decision is to infer product quality, which relates to the information-motivated herding literature (Banerjee, 1992; Bikhchandani et al., 1992).

Social influence among audience changes the dynamics of the overall producer evaluation. The pure effect of producer attributes changes as the interpersonal effect strengthens. According to previous findings, I also find that the dynamics among audience members diminishes the effects of product attributes (Salganik et al., 2006; Salganik and Watts, 2008; Koning and Model, 2013).

However my results demonstrate that there are aspects of the product information, which cannot be overwritten by the effect of social influence. When audience members are provided with the information on their peers' behavior, it may lead to a decrease in the predictability of outcomes, but it does not necessarily lead to complete unpredictability Salganik et al. (2006).

In previous studies of producer appeal, researchers conclude from aggregate results, the effects of certain producer attributes. The results of aggregate analysis in a setting where the information on peer decision is present may overestimate the effect size of the pure attribute and disregard the effect size of social influence. Kovacs and Johnson argue for example that high quality restaurants may benefit from being atypical members to their categories (Kovacs and Johnson, 2013). They provide the explanation of attracting customers who are more likely to be novelty seeking. Another possible explanation could be that the cause behind the positive aggregate effect of atypicality is the strong effect of social influence.

This paper contributes to the existing literature by demonstrating the applicability of threshold model in observational data. The theoretical background of the threshold model was created in the 1970's with demonstrations of the theory on mathematical simulation models (Schelling, 1971, 1972; Granovetter, 1978). Many more researchers followed earlier practices of using computer simulations to prove the nonexistence of a correlation between individual incentives and collective outcomes in different situations. Schelling applies threshold model to simulate neighborhood tipping (Schelling, 1971), Granowetter and Soong illustrate the dynamics of consumer demand (Granovetter and Soong, 1986) and public opinion (Granovetter and Soong, 1988) while Watts (2002) simulates global cascades in social and economic systems. One condition of the threshold model is that decisions are made in sequence, information on which is hard to obtain in real life data. The Kickstarter website publishes all backers in the order of their contribution, the data structure of which allows one to track the sequence of decisions. The analysis of more than 6000 projects strongly supports the core idea of the threshold model, that is, prospective contributors to the project make their decisions heavily based on the number of previous backers.

References

Agrawal, A. K., Catalini, C., and Goldfarb, A. (2011). The geography of crowdfunding. Technical report, National Bureau of Economic Research.

Akerlof, G. A. (1980). A theory of social custom, of which unemployment may be one consequence. *The quarterly journal of economics*, pages 749–775.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, pages 797–817.

Becker, G. S. (1991). A note on restaurant pricing and other examples of social influences on price. *Journal of Political Economy*, pages 1109–1116.

Belleflamme, P., Lambert, T., and Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. *Journal of Business Venturing*, 29(5):585–609.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, pages 992–1026.

Chen, Y. and Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54(3):477–491.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354.

Coleman, J. S. (1986). Social theory, social research, and a theory of action. *American journal of Sociology*, pages 1309–1335.

Coleman, J. S., Katz, E., Menzel, H., et al. (1966). *Medical innovation: A diffusion study*. Bobbs-Merrill Indianapolis.

DiPrete, T. A. and Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual review of sociology*, pages 271–297.

Gould, R. V. (2002). The origins of status hierarchies: A formal theory and empirical test1. *American journal of sociology*, 107(5):1143–1178.

Granovetter, M. (1978). Threshold models of collective behavior. *American journal of sociology*, pages 1420–1443.

Granovetter, M. and Soong, R. (1983). Threshold models of diffusion and collective behavior. *Journal of Mathematical sociology*, 9(3):165–179.

Granovetter, M. and Soong, R. (1986). Threshold models of interpersonal effects in consumer demand. *Journal of economic behavior and organization*, 7(1):83–99.

Granovetter, M. and Soong, R. (1988). Threshold models of diversity: Chinese restaurants, residential segregation, and the spiral of silence. *Sociological methodology*, 18(6):69–104.

Hsu, G. (2006c). Jacks of all trades and masters of none: Audiences reactions to spanning genres in feature film production. *Administrative Science Quarterly*, 51(3):420–450.

Katz, M. L. and Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American economic review*, pages 424–440.

Kleemann, F., Vos, G. G., and Rieder, K. (2008). Un (der) paid innovators. *Science, Technology and Innovation Studies*, 4(1).

Koning, R. and Model, J. (2013). Experimental study of crowdfunding cascades: When nothing is better than something. *Available at SSRN 2308161*.

Kovacs, B. and Johnson, R. (2013). Contrasting alternative explanations for the consequences of category spanning: A study of restaurant reviews and menus in san francisco. *Strategic Organization*, 12(7-37).

Lawton, K. and Marom, D. (2010). The crowdfunding revolution. *Social Networking Meets Venture Financing. Amazon Digital Services*.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of marketing*, 70(3):74–89.

Lynn, F. B., Podolny, J. M., and Tao, L. (2009). A sociological (de) construction of the relationship between status and quality1. *American Journal of Sociology*, 115(3):755–804.

Marwell, G. and Oliver, P. (1993). *The critical mass in collective action*. Cambridge University Press.

Marwell, G., Oliver, P. E., and Prahl, R. (1988). Social networks and collective action: A theory of the critical mass. iii. *American Journal of Sociology*, pages 502–534.

Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review*, pages 193–210.

Merton, R. K. (1988). The matthew effect in science, ii: Cumulative advantage and the symbolism of intellectual property. *Isis*, pages 606–623.

Merton, R. K. (1995). The thomas theorem and the matthew effect. *Social forces*, pages 379–422.

Merton, R. K. et al. (1968). The matthew effect in science. *Science*, 159(3810):56–63.

Oliver, P., Marwell, G., and Teixeira, R. (1985). A theory of the critical mass. i. interdependence, group heterogeneity, and the production of collective action. *American journal of Sociology*, pages 522–556.

Oliver, P. E. (1993). Formal models of collective action. *Annual Review of Sociology*, pages 271–300.

Oliver, P. E. and Marwell, G. (1988). The paradox of group size in collective action: A theory of the critical mass. ii. *American Sociological Review*, pages 1–8.

Olson, M. (1965). *Logic of collective action public goods and the theory of groups Rev. ed.*.

Opp, K.-D. (1988). Grievances and participation in social movements. *American sociological review*, pages 853–864.

Ordanini, A. (2009). Crowd funding: customers as investors. *The Wall Street Journal*, 23(3).

Peterson, R. A. and Berger, D. G. (1971). Entrepreneurship in organizations: Evidence from the popular music industry. *Administrative Science Quarterly*, pages 97–106.

Peterson, R. A. and Kern, R. M. (1996). Changing highbrow taste: From snob to omnivore. *American Sociological Review*, 61:900–907.

Prahl, R., Marwell, G., and Oliver, P. E. (1991). Reach and selectivity as strategies of recruitment for collective action: A theory of the critical mass, v*. *Journal of Mathematical Sociology*, 16(2):137–164.

Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.

Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856.

Salganik, M. J. and Watts, D. J. (2008). Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. *Social Psychology Quarterly*, 71(4):338–355.

Schelling, T. C. (1971). Dynamic models of segregationâĂă. *Journal of mathematical sociology*, 1(2):143–186.

Schelling, T. C. (1972). A process of residential segregation: neighborhood tipping. *Racial discrimination in economic life*, 157:174.

Squazzoni, F. (2008). The micro-macro link in social simulation. *Sociologica*, 2(1).

Strang, D. and Macy, M. W. (2001). In search of excellence: Fads, success stories, and adaptive emulation1. *American Journal of Sociology*, 107(1):147–182.

Strang, D. and Soule, S. A. (1998). Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual review of sociology*, pages 265–290.

Valente, T. W. (1995). *Network models of the diffusion of innovations*, volume 2. Hampton Press Cresskill, NJ.

Valente, T. W. (1996). Social network thresholds in the diffusion of innovations. *Social networks*, 18(1):69–89.

Van Braak, J. (2001). Individual characteristics influencing teachers' class use of computers. *Journal of educational computing research*, 25(2):141–157.

Walsh, E. J. and Warland, R. H. (1983). Social movement involvement in the wake of a nuclear accident: Activists and free riders in the tmi area. *American Sociological Review*, pages 764–780.

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771.

Zhang, J. (2010). The sound of silence: Observational learning in the us kidney market. *Marketing Science*, 29(2):315–335.

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

Zuckerman, E. W. (2012). Construction, concentration, and (dis) continuities in social valuations. *Annual Review of Sociology*, 38:223–245.

Chapter 5

Conclusions

5.1 Summaries of chapter results and contributions

The appearance of the internet, and its diffusion into many aspects of our daily life brought several implications to the organizations. The goal of my dissertation was to examine three areas where audience members' internet-mediated communication has a direct influence on the organization.

Firstly, I argue that the rapid growth in the use of emails as a channel of communication provides scholars with easily accessible data, which allows them to map the structure of communities. However, data from on-line sources does not reflect previously studied "off-line" relationships perfectly. I compare off-line interactions to on-line interactions using exponential random graph (ERG) models to understand what aspects of the face-to-face communication approximates it's on-line counterpart. I analyze the roles of the two different communication channels in reflecting organizational boundaries, organizational hierarchies and gender.

Secondly, studying online reviews written by audience members, I investigate why producers with multiple categories receive lower ratings from reviewers than their single-category competitors. In the empirical investigation I directly test and disentangle typicality-based and quality-based explanations for the negative consequences of category spanning.

Thirdly, I analyze the interconnectedness of audience decisions of an online crowdsourcing platform. I address the micro level mechanism of how audience behavior creates certain macro level patterns of producer success rather than assuming that they are simple aggregates of individual characteristics. I approach audience members' behavior from a collective behavioral perspective. Specifically, I analyze audience dynamics with threshold models.

5.1.1 Chapter 2

People communicating through online channels bring many advantages to research scholars who study interaction patterns. Relying on online data to map the structure of communities is not necessarily a straightforward exercise. Researchers now use email networks assuming that it mirrors communication (or other) networks (e.g., Grippa et al., 2006; Kleinbaum et al., 2008). In Chapter 2 questioning this assumption I investigate how these online connections relate to previous "off-line" measures of relationships and networks (Menchik and Tian, 2008; Matzat and Snijders, 2010). In Chapter 2 I provide an explorative case study, in which I compare the email network of a Central European Bank's employees with the corresponding self reported friendship networks, information networks, and advice networks. The dataset contains attribute information about the employees, email messages among the employees, and a sociometric survey for a three-week period. Besides the descriptive analysis of the network, I use exponential random graph (ERG) models to estimate when and how email networks and networks based on traditional sociometric methods coincide or diverge.

Chapter 2 has three main contributions to the research field. Firstly, The results of analyzing multiple operationalizations provides guidance on how a tie should be defined in the email correspondence. Secondly, I differentiate between three main types of employee connections: friendship, communication, and advice, in order to better understand the relationship between off-line and other types of communication networks. Having a multidimensional measure of social networks enables us to better understand the extent to which emails correspond to "off-line" ties. Thirdly, Chapter 2 demonstrates that although the structure of the email networks in the dataset is related to the structure of survey-based networks, emails networks and survey-based social networks do not perfectly correspond to each other. On the one hand, email-based and survey-based social networks do coincide in many aspects. For example, email networks are relatively good in predicting the betweenness and degree centrality of employees, especially in the advice seeking and communication networks. In many other aspects, however, email-based and survey-based social networks diverge. For example, if one is interested in the role organizational boundaries play in forming social networks, the two data types paint markedly different pictures. In the email-based networks, organizational boundaries play a weaker role (Diehl et al., 2006) and email communication often moves across boundaries. The sociometric data, however, indicate that these boundaries do play a role, a role that is much stronger than pictured by email networks. Also, email networks do not capture the role organizational hierarchies and gender play in interpersonal networks. Further, I enrich the analysis with detailed attribute data on the network participants and I test previous results on a larger dataset.

5.1.2 Chapter 3

In Chapter 3, I studied audience members' on-line reviews and their effect on producer success. Researchers studying category spanning and it's consequences to the producer, have put forward an impressive set of empirical findings over the past decades. I argue that the theoretical underpinnings of these findings are not well clarified. The number of different theoretical processes have been advanced to explain the consequences of category spanning which are not directly tested. The explanations proposed by organizational theorists mainly fall into two areas, which I term the "quality-based" and the "socio-cognitive-based" explanations.

In Chapter 3 I attempt to test and contrast the mechanisms of the quality-based and socio-cognitive-based explanations directly in the empirical setting of restaurant reviews.

One of the main novelties of Chapter 3 is a direct instrument of typicality. Previous research assessed typicality indirectly, and assumed that the more categories an organization populates, the lower its typicality in each of the categories populated (Negro et al., 2010b; Hsu et al., 2009b; Kovacs and Hannan, 2011). The empirical strategy to assess the typicality of restaurants is to contrast the offering of the restaurant (items on the menu) with its labels (Italian, Japanese, etc.)

and assess the extent to which the menu fits the label(s) the restaurant claims. I use a commonly utilized computational linguistics approach, word-category co-location mapping (Manning and Schütze, 1999b), to explore the schemata of organizational categories. I establish the typicality of the restaurants in the categories by comparing the restaurant's offering to the schemata of the categories.

In Chapter 3, I demonstrate that quality mediates the negative average effect of category spanning on value ratings: lower quality leads to lower ratings and restaurants in multiple categories on average were conferred lower quality than single-category restaurant. These findings provide a direct corroboration of the assumption behind the principle of allocation (Hannan and Freeman, 1989a). I also demonstrate the positive relationship between typicality and audience value ratings.

The evidence of the positive average effect of typicality has implication for current research in organizational authenticity. Carroll and Wheaton (2009b) argue that restaurants that are viewed as authentic are viewed more favorably by patrons. One of their constructs of authenticity is type-authenticity: an organization is type-authentic if its attributes and practices are consistent with its type. This construct is similar to what I call typicality. The findings indicate that type authenticity should be especially important for low- and mid-quality level restaurants.

5.1.3 Chapter 4

In Chapter 4, I study audience members' on-line behavior from a collective behavior perspective. Specifically, I conduct a multi-level investigation on how audience's evaluation changes with respect to the increasing level of social influence in the setting of creative projects' on-line crowdfunding activity. On the micro-level, I argue that audience members' value products based on the objective attributes when the effect of social influence is either small or non-existent. As peer influence get's stronger, the effect of product attributes decreases in the valuation process. On the macro-level, I study the changes of the role of product attributes in social valuation.

While previous studies of producer success mainly drew conclusions on product attributes, a novelty of Chapter 4 is that I study audience dynamics with threshold models.

As follows from the threshold theory, I argue that producers' ability to attract audiences hinges not solely on the producer's attributes but also on the social influence already attracted audience members have on the prospective audience. Accounting for interaction among audience members provides a more nuanced description of the role of certain producer attributes.

Analyzing more than 5000 projects, I find evidence that audience decisions are interrelated. The result supports the core idea of the threshold model, that is prospective contributors to the project make their decisions based on the number of previous investors and the higher the number of previous investors, the higher the positive influence on the number of investors on the focal day. The interpersonal effect among audience members have further implications for studies of producer success. My analysis supports, that the increasing effect of social influence will decrease the effects of other producer attributes. Specifically, I find that the positive effect of project typicality on gaining audience (where typicality is measured by the grade of membership of the project description) is going to decrease as more audience is gained. The positive effect of the producer's ability to attract members from the target population diminishes with the increasing size of audience. I define a producer's target population as those members of the total audience whose previous activity is higher than 50% in the given category.

I conclude that social influence among audience changes the dynamics of the overall producer appeal. In previous studies of producer appeal, researchers derived from aggregate results the effects of certain producer attributes. The results of aggregate analysis may overestimate the effect size of the pure attribute and disregard the effect size of social influence. The present analysis demonstrates that projects with mid-level typicality which are subject to strong social influence might be more likely to receive the target amount of funding than projects with initial high typicality but with weaker social influence. This study also demonstrates that product success highly depends on audience homogeneity. In line with the arguments put forward by population ecologists and opposite to the sociological arguments, appeal initially to one narrow segment of the audience will more likely lead to further audience gain. The study also demonstrates that successful projects were able to attract audience from broader niches. This result helps to understand why highly typical projects may not be as successful as their less typical rivals claiming that they may not be appealing enough to audiences of varied tastes.

References

Carroll, G. R. and Wheaton, D. R. (2009b). The organizational construction of authenticity: An examination of contemporary food and dining in the us. *Research in Organizational Behavior*, 29:255–282.

Diehl, C. P., Getoor, L., and Namata, G. (2006). Name reference resolution in organizational email archives. In *SDM*, pages 70–81. SIAM.

Grippa, F., Zilli, A., Laubacher, R., and Gloor, P. A. (2006). E-mail may not reflect the social network. In *Proceedings of the North American Association for Computational Social and Organizational Science Conference*, pages 1–6.

Hannan, M. T. and Freeman, J. (1989a). *Organization ecology*. Harvard University Press: Cambridge.

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

Kleinbaum, A. M., Stuart, T., and Tushman, M. (2008). *Communication (and coordination?) in a modern, complex organization*. Harvard Business School Boston, MA.

Kovacs, B. and Hannan, M. T. (2011). Category spanning, distance, and appeal. Technical report.

Manning, C. D. and Schütze, H. (1999b). *Foundations of statistical natural language processing*. MIT press.

Matzat, U. and Snijders, C. (2010). Does the online collection of ego-centered network data reduce data quality? an experimental comparison. *Social networks*, 32(2):105–111.

Menchik, D. A. and Tian, X. (2008). Putting social context into text: The semiotics of e-mail interaction1. *American Journal of Sociology*, 114(2):332–370.

Negro, G., Hannan, M. T., and Rao, H. (2010b). Categorical contrast and audience appeal: niche width and critical success in winemaking. *Industrial and Corporate Change*, 19(5):1397–1425.