The Dissertation Committee for Hsuan-Wei Chen Certifies that this is the approved version of the following dissertation:

ESSAYS ON NETWORK DYNAMICS AND INFORMATIONAL VALUE OF VIRTUAL COMMUNITIES

Committee:

Prabhudev C. Konana, Supervisor

Bin Gu, Co-Supervisor

Rajagopal Raghunathan

Maytal Saar-Tsechansky

Andrew B. Whinston

ESSAYS ON NETWORK DYNAMICS AND INFORMATIONAL VALUE OF VIRTUAL COMMUNITIES

by

Hsuan-Wei Chen, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2008

Dedication

To my loving family

Acknowledgements

I would like to thank the many people who have helped to make this dissertation possible. First, I would like to thank my advisors, Prabhudev C. Konana and Bin Gu, who played an indispensable role in my intellectual development. I am truly grateful for their constant guidance, encouragement and patience along the difficult trail of finishing my doctoral study. I would also like to thank my great mentor and dear friend in my committee, Maytal Saar-Tsechansky. Maytal has stood by me warmly for four years and never hesitated to offer advice on research and life. I would also like to thank the rest of my committee, Andrew B. Whinston and Rajagopal Raghunathan, for their commitment and support. This dissertation could not have been made a reality without them. Aside from my committee, Kyle Lewis also offered tremendous help through the dark days and, for that, I am very appreciative.

I also offer sincere thanks to all of my fellow students in Information, Risk, and Operations Management department, for understanding and supporting me as they climbed this mountain with me. My gratitude also extends to my lovely friends in Texas and in Taiwan for their invaluable friendships. I would especially like to thank the knowledgeable econometrician, Yu-Chin (A-chin) Hsu, for always being helpful in good and bad times.

I would like to thank my loving family—my dear parents, my brother-in-law Edward, Shu-Hui, and the God-blessed Lee family, for their constant care and belief in

my achievement, and especially my most adorable sister Grace, for sharing all the laughter and tears with me in Austin. Finally, I dedicate my deepest gratitude and love to my fabulous fiancé, Shun-Chuan, who is the meaning of my life. His love, support, encouragement and humor have made me alive throughout these rough and ready years.

Hook ‘em Horns!

Hsuan-Wei (Michelle)

Essays on Network Dynamics and Informational Value of Virtual Communities

Publication No._____

Hsuan-Wei Chen, Ph.D.

The University of Texas at Austin, 2008

Supervisors: Prabhudev C. Konana, Bin Gu

Public press and companies have increasingly strong interests in the impact on businesses brought about by virtual communities. In recent years, virtual communities have become significant sources of information for consumers and businesses by offering unprecedented opportunities for information sharing. Scholars recognize that information posted in virtual communities has important implications for the behaviors of community members and subsequent economic decisions and market performance. However, relatively less is explored about how the informational value of virtual communities results from an aggregated or fragmented community of information. In particular, the underlying motives and mechanisms of user interactions in virtual communities are challenging to understand because of the amount of information

available and the potential noises. To investigate user dynamics and the resulting informational value in virtual communities, I explore three major issues in my dissertation. First, I empirically examine whether community fragmentation or aggregation prevails in the context of virtual investment communities. Results indicate that instead of the common belief of virtual communities serving as melting pots that comprise opinions, online investors, in particular, show strong homophily behavior in virtual investment communities. Second, using data from virtual investment communities, I investigate the interactions among online investors that drive homophily and community fragmentation. I find that psychological needs for supportive opinions mainly drive the information seeking and interaction behaviors of online investors as compared to economic rationales. Following this line of exploration, I also identify the informational impact of virtual communities on user behaviors in the context of electronic markets. With data from online retailers, I examine the possible shrinkage of consumer product consideration that is reinforced by online recommendations. A resultant change of consumer consideration leads to a landscape shift of product competition for online retailers, suggesting strategic implications to manufacturers. All in all, my dissertation contributes to an understanding of the value of virtual communities as informational media, how virtual communities shape online user opinions, and how online user preferences impact businesses and markets in a networked economy. My research pushes the frontier toward understanding virtual communities and sheds light on the insights into exploring online network dynamics.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Motivation

In recent years, explosive growth of virtual communities is shown to significantly impact human behavior in terms of social networking. There are various reasons for which users join virtual communities. A large number of people jump into virtual communities to seek and share information. For instance, online investors browse daily financial news sites, join message board discussions, and exchange stock investment tips on Yahoo! Finance; movie viewers share opinions and comments on Netflix.com; healthcare professionals provide institutional and physician-related knowledge to patients on HealthBoards.com.

In addition to information support, virtual communities offer an unprecedented platform for building social relationships. Two of the most well-known examples of online social-networking communities include MySpace.com and Facebook.com, which link friends, and even strangers, in virtual worlds to share interests and activities or to explore the interests and activities of others. In virtual communities such as electronic marketplaces (e.g. Amazon.com and eBay.com), people join virtual communities to perform more efficient and profitable transactions. There are also a large number of people who participate in virtual communities for fantasy and fun. For example, Battle.net is famous for its virtual multi-player gaming environment for networked and

interactive entertainment. In addition, Xbox Live provides an integrated entertainment service for MSNBC, MSN Portal, Xbox gaming, etc. ESPN.com, ranked the worldwide leader in sports, provides a “Fantasy Sports” service to its audience for simulated gaming based on live scoring and player performance.

A great deal of evidence shows that the drastic progression and lapse in social dynamics from traditional networking experience brought on by virtual communities has significant impacts on subsequent economic decisions and businesses. Procter & Gamble (P&G) reports a 40% gain on research productivity after tapping into virtual communities (P&G Report 2007). This demonstrates the monetary value and prosperity of virtual communities. Further, MySpace.com and its parent company, Intermix Media sold for $580 million because of the potential for revenue growth for advertiser companies through its information sharing abilities (USA Today 2006). Stanford GSB News (2007; Nair et al. 2006) shows that pharmaceutical companies decide to spend approximately 32% of their total marketing dollars to influence opinion leaders (e.g. healthcare professionals, physicians, etc.), which will be rapidly disseminated among online medicine-related forums. In addition, Facebook.com obtained a $240 million investment from Microsoft in October 2007 for its unique audience growth by 1.5 million people in that month (Fortune 2007). Amazon.com, the largest online retailer, continued its upward trend in revenue growth in the first quarter of 2008, posting a 37% increase in revenue and 30% growth in earnings for the period (Portfolio.com 2008). Finally, Xbox expects combined revenue of $726 million from its TV and movie services by 2011

(Emerging Media Dynamics Executive Reports 2006).

The skyrocketing success of businesses as a result of virtual communities has drawn great attention not only from the public press and practice, but also from academia. In particular, researchers are becoming more and more interested in underlying user behaviors and the interaction dynamics that have created the value of virtual communities, which lead to the foresight of the potential impacts on businesses. Extant studies are the inspiration for this dissertation: to add to academia's current understanding of virtual communities by unraveling the underlying dynamics of interaction behavior in virtual communities. Specifically, I investigate what motives drive online users to seek and exchange information and what the consequence are for the informational value derived from such behavior in virtual communities. Examination of these issues is presented in Chapters 2 and 3. The value of virtual communities based on information sharing tightly ties in with the third area in my study, which is given in Chapter 4. In sum, the present study in Chapter 4 is concerned with the managerial implications of user interactions and consumption in virtual communities. A detailed conceptual description of my dissertation is given in the following subsection.

1.2 Conceptual Overview

The first section in Chapter 2 of my dissertation investigates the informational value of virtual communities that results from user interactions. While it is commonly believed that virtual communities serve as "melting pots" of opinions and thoughts, there

is little empirical evidence or a consistent theoretical basis for this assumption. Therefore, I examine whether a “melting-pot” or “homophily” of online individual opinions in fact prevails in virtual communities, particularly in the context of virtual investment communities.

To do so, I mainly draw upon literature that includes bounded rationality (e.g. Simon 1957), information economics for information acquisition (e.g. Stigler 1961), and homophily in social networks (e.g. McPherson et al. 2001) to investigate information seeking and interactions among online investors. I conduct an empirical investigation using stock investment sentiments of 72,019 online investors on 29 Yahoo! Finance stock message boards. The objective is to examine whether individuals have a stronger tendency to interact with like-minded people, as is homophily, or whether they are more likely to seek dissimilar interests and information, which indicates the melting pot idea. I find that, instead of information aggregation as might be expected from online networks, online investors reveal strong tendencies toward homophily due to psychological needs in virtual investment communities.

The mechanisms that drive homophily in virtual investment communities are then further explored in my second essay in Chapter 3. In particular, I examine the underlying factors that influence user interactions for homophily. My study draws heavily upon the psychological biases of online investors (e.g. Barber and Odean 2001a & 2001b; Kahneman and Tversky 1996). In addition, several major streams of literature are presented to provide insights into user interactions in virtual communities, as briefed

below.

Prior work investigates the incentives of individuals to participate in virtual communities—largely corresponding to the different purposes for which people join virtual communities in reality as mentioned above, such as searching for knowledge, building relationships, verifying identities, seeking emotional support, etc. (e.g. Butler 2001; Dholakia et al. 2004). Participation activities further reinforce some key issues for maintaining and creating the sustainability and competitiveness of virtual communities, including, for example, the appropriate membership size (e.g. Butler 2001) and online trust and reputation (e.g. Dellarocas 2003).

With the Yahoo! Finance data presented in Chapter 2, I use a discrete choice model to empirically examine the economic rationales and psychological needs that drive online investors to choose to participate in online discussion groups. I find that individuals' incentives in exchanging information with like-minded people are heavily affected by the nature of virtual investment communities and individual statuses. Specifically, availability of thread choices, increase in stock volatility and holding minority opinion are associated with homophily behavior, while membership size of discussion threads determines whether homophily or heterophily behavior occurs. I show that the first three results are consistent with psychological needs that drive homophily and the last result indicates interaction between economic and homophily behavior. The study shows the interactions of online investors are mostly motivated by homophily behavior due to psychological needs instead of economic concerns.

The interaction dynamics of information sharing and exchange among online users creates "word-of-mouth" activities, which are shown to have significant impacts on subsequent economic behaviors, such as purchase and investment decisions (e.g. Chevalier and Mayzlin 2006; Dellarocas et al. 2008; Duan et al. 2008). The third section in Chapter 4 offers an investigation of how information provided by virtual communities alters human behaviors, which in turn affects managerial strategies for businesses. Specifically, while much has been discussed that the Long Tail increases product variety in online retailers, prior work shows it could be the other way around. With the presence and reinforcement of online recommendations, product variety could, in contrast, decrease. Further, product alternatives considered by consumers for purchase decisions may become fewer. This possible shrinkage of consideration set of product alternatives has significant implications for product competition. In particular, from the perspective of manufactures, products need to not only compete for being purchased but also being first included in consideration by consumers. In this study, I show that in real online retailing environments, the size of consideration set is fairly small. I then investigate the resultant product competition for consideration and for choice by developing a consideration set choice model at the aggregate level. The model is applied to 38,400 unique products collected from Amazon's "What Do Customers Ultimately Buy After Viewing This Item?." The results show that despite the high level of product variety provided by online retailers, more than half (53.294%) of the purchase decisions are made by consumers considering two or fewer product

alternatives. In addition, the two levels of product competition are distinctive from each other: a product does well for consideration may not do so well for purchase choice, and vice versa. This study brings out two key implications. First, the results reveal that due to the small size of consideration set, being first considered by online consumers plays a dominant role in product competition. Second, the findings indicate that pricing strategies only influence product competition for choice. It suggests that despite overwhelming focus on price competition in prior work, it plays a rather limited role in product competition for consideration.

1.3 Contributions

My dissertation has three main contributions. First, I empirically demonstrate that versus economic rationales, online investors are driven principally by cognitive and psychological biases to interact in virtual investment communities. Thus, this dissertation is one of the first attempts to study individual interaction behavior in virtual communities and identifies the informational value of virtual investment communities. The community fragmentation and homophily explored in virtual investment communities shows how interactions in virtual investment communities affect user opinions and their subsequent economic decisions.

Second, the informational value of virtual investment communities suggests a number of insights for practitioners, such as virtual community providers and businesses that are interested in understanding the distribution of information and value of

advertising to virtual community members. For instance, virtual community members can be clustered based on their tendencies to expressing opinions, which is valuable information for marketing and pricing strategies. Incentives can also be provided to online members to encourage them to express opinions, which may foster diversity in virtual communities if desired.

Third, the investigation of product competition that results from consumer consideration and product choices that are suggested by online recommendations further explores the informational value of virtual communities for business practitioners. The findings of differentiated product competition for consumer consideration and choice in online retailers are insightful for product manufacturers. For instance, instead of pricing and product quality considerations, enhancing product awareness plays a more dominant role in a marketer's ability to increase product competitiveness in virtual communities.

In brief, exploring the issues in my dissertation advances the present understanding of the value of virtual communities as informational media and how they shape user opinions and economic impacts in today's networked economy. In the following chapters, I present detailed research for each of the topics of my dissertation.

Chapter 2: Melting-Pot or Homophily in Virtual Investment Communities?—An Empirical Investigation

2.1 Introduction

There is significant excitement in the popular press on the role of virtual communities in creating business value. *BusinessWeek*'s cover story on June 20, 2005, noted that "companies are using Internet-powered services [virtual communities] to tap into the collective intelligence of employees, customers, and outsiders, transforming their internal operations" (Hop 2005). The results are astounding: P&G reports a 40% gain on research productivity after tapping into virtual communities like MySpace.com (P&G report 2007). This potential has resulted in staggering evaluation of some virtual communities like MySpace.com (USA Today 2006). For example, News Corp acquired MySpace.com and its parent Intermix Media for $580 million and Microsoft gained 2.5% interest in Facebook.com for $250 million.

Virtual communities vary significantly in terms of compositions and topics. On one end individuals with prior ties come together to interact socially to the other end where complete strangers come together to discuss and share opinions and experience on a wide range of issues. These issues include politics, religion, culture, healthcare, entertainment, real estate, retail, auction, technology, financial markets, or purchasing decisions. Recent research shows that virtual communities for product/service

recommendations on websites like Amazon.com or tripadvisor.com[1] have significant impact on consumer purchase decisions and welfare (e.g., Chevalier and Mayzlin 2006; Godes and Mayzlin 2004; Li and Hitt 2007). Likewise, Ma and Agarwal (2007) suggest that loyal automobile owners discover new ways of exploring automobile features on discussion boards.

Early empirical evidences suggest that individuals are influenced by information posted on virtual communities. However, research to explicate the nature and dynamics that underlie the information generation process is still evolving. In this study, I recognize that virtual communities are, foremost, a venue for individual interactions. The interactions generate information that benefits all visitors to virtual communities even if they do not participate in the interactions. Theoretical principles from economic and homophily perspectives can explain interaction behavior, but may offer contradictory views of how individuals choose to interact with each other. Extant economic theory suggests that rational decision makers will seek out others with information they do not possess to improve decision making, thereby more information aggregation and potentially higher social welfare are created (Feltham and Demski 1970). That is, as often claimed in popular press, virtual communities serve as a melting pot of ideas and opinions and individuals come together to share and discuss various viewpoints to arrive at meaningful inferences.

Homophily studies, on the contrary, indicates that individuals prefer to associate

[1] ComScore Media Metrix reports that there were 15 million opinions posted on Tripadvisor.com in 2007.

with like-minded people, leading to *community fragmentation* and potentially lower social welfare (Hall et al. 2007; McPherson et al. 2001). Community fragmentation is a situation in which individuals sharing common opinions come together while shunning individuals with different opinions. That is, community is partitioned into virtual knowledge spaces. Van Alstyne and Brynjolfsson (2005) provide analytical results to suggest community fragmentation or balkanization due to bounded rationality (Simon 1957). Since excessive information leads to higher information processing costs, community members may be more selective in interacting with others (e.g. Butler 2001; Gu et al. 2007; Whittaker et al. 1998). Further, individuals may only interact with people of similar opinions to reinforce their own beliefs rather than assimilating different viewpoints (Gilovich 1991). In addition, such selective interactions are amplified on virtual communities through automatic filters, ignoring lists, collaboration recommendations, and other web technologies. Therefore, virtual communities do not necessarily assimilate information. Instead, they could foster fragmented and biased interactions that polarize the communities.

This study addresses the gap in our understanding between the two competing theoretical lenses with contradicting outcomes. In particular, in this research I develop models of behavior under different theoretical anchoring and provide empirical validation to better understand the behaviors.

I study observed individual interactions in virtual financial investment communities. Financial investment communities such as Yahoo! Finance, MSN Money,

and AOL Money & Finance attracted over 62 million unique visitors in February 2008 (Media Metrix 2008). Investors interact with each other to exchange stock tips, investment strategies and rumors. The large number of unique participants and postings raises a number of interesting questions on how investors choose to interact with others. It is also relevant to policy makers and market efficiency since the interaction determines the speed of information (e.g. rumors) to spread and influence trading decisions.

Virtual investment communities provide several advantages to study behaviors. For most virtual community contexts, individuals' beliefs and opinions are usually high dimensional and difficult to observe. However, in virtual investment communities, the beliefs are of one dimension, providing an effective avenue towards measuring online investors' beliefs. Second, unlike in many virtual communities, an individual's belief on stock performance, i.e. the information signal and strength to buy or sell a certain stock, is verifiable from his/her self-declared sentiment about a stock. That is, investors rate whether they are bullish (i.e. buy or strong buy), bearish (i.e. sell or strong sell), or neutral. This avoids mining of noisy texts to understand investor opinions. Third, investing activity has rich literature in using both economic and psychological theories to understand investor behavior.

I collected 17,329 discussion threads within which a total of 72,019 individual postings are considered, from 29 stock message boards for a three-year period from January 2004 to December 2006. Virtual investment communities use discussion threads to facilitate interaction among members. My results find that, despite the

prediction of information economics theory that suggests that rational investors seek others with different information to increase expected utility, online investors are more inclined to interact with like-minded people, reflecting cognitive biases and homophily behavior. I argue that investors seek others to reinforce their beliefs and enhance psychological biases such as illusion of control and illusion of knowledge (Burger 1989; Deci and Ryan 1987; Langer 1975). This behavior creates fragmented communities instead of a unified, informed village.

This research makes three unique contributions. First, this study provides empirical evidence to verify that online investors are mainly motivated to interact with others from cognitive biases and psychological needs rather than economic rationales, in support of homophily studies. To the best of our knowledge, this has not been studied empirically in the literature. Second, this study addresses the important issue of the informational value of virtual communities for practice and businesses. The demonstration of homophily behavior in virtual communities provides insights into the power of word of mouth and managerial implications such as community member segments based on their opinion profiles. Third, I develop a methodology to construct an online investor's information set for stock investments, and to assess similarity of information sets between investors that interact with each other. The methodology also suggests an information economic view of how individuals seek online postings to increase their expected utilities. This has important implications for research in virtual investment communities.

The rest of the chapter is organized as follows. In Section 2.2, I present literature review. Theories and hypotheses for community fragmentation or aggregation are discussed in Section 2.3. In Section 2.4, I present data followed by methodology in Section 2.5. Empirical analysis and discussions are given in Section 2.6.

2.2 Literature Review

There are four major bodies of literature relevant to this study: bounded rationality (Simon 1957), economic rationale for information seeking (e.g. Stigler 1961), homophily in social networks (e.g. Lazarsfeld and Merton 1954; McPherson et al. 2001), and psychological biases such as illusion of control, illusion of knowledge, and overconfidence (e.g. Langer 1975; Barber and Odean 2001a; Kahneman and Tversky 1996; Konana and Balasubramanian 2005).

Bounded rationality (Simon 1957), arising from critical psychological limits to process all information, is one of the major theoretical arguments to explain how individuals seek information and participate in virtual communities. Individuals with bounded rationality are likely to filter out information for more focused attention (Broadbent 1958), and become selective in their participation and information seeking behavior (DeMarzo et al. 2003). Further, a large amount of information creates information overload, a situation in which an individual fails to process and utilize all information (Rogers and Agarwala-Rogers 1975). These arguments have been used to describe virtual community participation. Bounded rationality impacts online network

formation by requiring community members to make trade-offs between information quantity and quality (Gu et al. 2007). Information overload drives users to have a tendency to only attend to simpler messages, end active participation, and generate simpler responses (Jones et al. 2004). Further, users may self-select to online groups based on knowledge and interests because of bounded rationality, leading to cyber-Balkans in electronic communities (Van Alstyne and Brynjolfsson 2005).

Economic theories view information seeking from a cost-benefit perspective (Demski 1967); the benefit of acquiring an additional piece of information depends on its value of increasing one's utility (Birchler and Bütler 2007). Thus, economically rational individuals will seek out different information to improve decision making (Feltham and Demski 1970). This body of literature suggests that individuals will search for dissimilar opinions from virtual communities to lower uncertainty and maximize utility.

The human interaction and information seeking behavior, however, can also be examined through another lens from social science and social psychology literature. Homophily, first coined by Lazarsfeld and Merton (1954), suggests that individuals tend to associate with others who share similar backgrounds or beliefs, often referred to as "similarity breeds connection" (McPherson et al. 2001). Literature suggests that homophilous behavior occurs frequently based on status including socio-demographic factors, or value such as sentiments, beliefs and values (Rogers and Bhowmik 1971; McPherson et al. 2001). Homophily studies are mainly reasoned by social psychological theories, including self-categorization (Yuan and Gay 2006), reinforcement

theory or law of attraction (Macy et al. 2003), and social comparison theory (McPherson et al. 2001). Specifically, social comparison theory (Festinger 1950) and self-categorization (Turner et al. 1987) support that people may use a reference group of those who are similar to themselves to self-identify and form homophily. Further, people will seek out confirmatory information to be rewarded more based on reinforcement theory (Baumeister and Bushman 2007). This body of literature argues that homophily behavior may transfer even in virtual space without any prior ties.

This research also relies on the psychological biases of individuals particularly in the context of online investors. Online investors demonstrate multiple psychological biases that distort decision making and economic outcomes (Barber and Odean 2001a & 2001b). In particular, online investors present illusion of control (Langer 1975), which results when people inappropriately estimate their ability to control events, when, in fact, some events are not controllable. People with illusion of control alter their process of seeking and obtaining realms of information, leading to "illusion of knowledge," and believe that they are more knowledgeable than they really are (Burger 1989; Deci and Ryan 1987). Investors are also often biased by overconfidence (Barber and Odean 2001a), especially when they are less experienced yet successful, and thus will hold stronger beliefs in their ability and evaluation in beating markets. These psychological biases impact how individuals seek information in virtual communities (Konana and Balasubramanian 2005).

The recent studies indicate that information from virtual communities is likely to

influence individuals in many different contexts. Van Alstyne and Brynjolfsson (2005) specifically provide analytical support for Balkanization in virtual communities. However, these studies do not empirically suggest if individuals show homophily behavior. My research provides further insights into virtual community dynamics and adds to the literature.

2.3 Theories and Hypotheses

Whether community fragmentation or information aggregation prevails in virtual investment communities is resulted from individual information seeking and interactions. Two competing theories argue how individuals seek information and interact with each other. Information economics suggests that people will search for information they do not possess to lower uncertainty and increase utility. Psychological perspective, on the contrary, indicates that individuals are driven by cognitive biases to seek confirmatory information to reinforce their original beliefs.

2.3.1 Information Economics Perspective

Information economics considers that the main objective of interactions with others is to obtain information, and information seeking is viewed from a cost-benefit perspective (Demski 1967). The benefit of acquiring an additional piece of information depends on its value of increasing one's utility (Birchler and Bütler 2007). Thus, economically rational individuals will seek out information that lowers uncertainty,

maximizes utility, and improves decision making (Feltham and Demski 1970). In the context of virtual investment communities, if an online investor is economically rational, then when he/she observes a posted message with a sentiment, information economics suggests the following investment decisions:

a. *Invest based on his/her own sentiment if it is the same as the message sentiment.*

b. *Invest based on his/her own sentiment if it is different from the message sentiment, but the latter is weaker.*

c. *Invest against his/her own sentiment if it is different from the message sentiment, and the latter is stronger.*

The above decisions indicate that when the sentiments of the investor and the message are the same, there is no value for the investor to further identify whether the message sentiment is weaker or stronger. However, when the sentiments are different, the investor will explore the strength of the message sentiment by, for example, clicking on the posted message and reading it through. Therefore, economic motivation suggests that individuals have more incentives in exploring messages which sentiments contradict their beliefs to improve the quality of decision making. (For a more detailed illustration, please see Appendix 2.1 for Chapter 2.) This behavior potentially makes virtual communities melting pots of beliefs.

2.3.2 Homophily and Psychological Perspective

Psychological perspective argues that information seeking behavior is heavily influenced by psychological factors and cognitive biases. One of the most well-known cognitive biases of information seeking is anchoring and judgment, i.e. decision makers are likely to rely too heavily on a small set of information when making decisions (Tversky and Kahneman 1974). Specifically, anchoring states that during decision making, individuals tend to anchor on specific information and then adjust their decisions accordingly. Once the anchor is set, individuals demonstrate a bias towards that value.

A particular anchoring bias arises in virtual investment communities when an online investor participates in online discussions with *a priori* self belief in stock investment. This self belief can serves as an anchor from which later sentiments in stock investment are made. Online investors then anchor on the self belief as a starting point and make adjustments towards that sentiment. For instance, if the self sentiment leans towards "buying a certain stock," anchoring bias drives the online investor to have higher evaluation of postings with similar sentiments and lower evaluation of those with opposite sentiments.

The tendency to anchor on self beliefs is due to the psychological biases that online investors are commonly shown to bear (Barber and Odean 2001a & 2001b)—illusion of control (Langer 1975) and illusion of knowledge (Burger 1989; Deci and Ryan 1987). Illusion of control results when people have the bias in estimating that the events are controllable by their ability where in fact they are not. This illusion is

further related to the process of seeking and obtaining realms of information of online investors and thus leads to illusion of knowledge. People with illusion of knowledge believe that they are more knowledgeable than they really are and become overconfident in beating markets (Barber and Odean 2001a). Illusion of control, illusion of knowledge and overconfidence drive online investors to not only anchor on specific information, but particularly on self beliefs to reinforce their original sentiments.

The psychological biases of online investors in associating with similar others lead to the homophily behavior (Lazarsfeld and Merton 1954). Instead of status homophily based on gender, race, etc. widely discussed in physical social networks (e.g. McPherson et al. 2001), online investors form value homophily based on opinions, beliefs, etc. in virtual investment communities. This tendency of interacting with like-minded people leads to a community fragmentation.

Note that the information seeking and interaction behavior in virtual investment communities is not completely accessible. That is, while we can easily figure out who have posted what messages online, people who have "viewed and read" those messages are rather unobservable. Therefore, in this study I limit the information seeking, interaction, and participation behavior only in the more clearly-defined range of posting activities. This behavior can then be assessed by observing individual choices to post messages after reading other messages in online discussion threads.

Based on the arguments above, when individuals are driven by psychological and cognitive biases, they tend to post messages in discussion threads in which like-minded

others express supportive opinions. On the flip side, when individuals are influenced by information economics rationales, they are more likely to participate in discussion threads in which messages with dissimilar sentiments are shown. I used the term "opinion distance" to represent the similarity between the individual belief and the aggregated opinion from a discussion thread. Specifically, opinion distance defines the extent to which an individual's opinion is aligned with the average sentiment across all messages in a certain discussion thread. A large opinion distance indicates that the investor's opinion is dissimilar as the average thread opinion. A small opinion distance, on the other hand, refers to that the two compared opinions are confirmatory. To investigate which of the two competing theories actually holds in virtual investment communities, I propose the following hypotheses:

Hypothesis 1A (homophily and cognitive biases). *For a discussion thread, as opinion distance increases, an investor's probability of posting in this particular thread decreases.*

Hypothesis 1B (information economics rationale). *For a discussion thread, as opinion distance increases, an investor's probability of posting in this particular thread increases.*

The two hypotheses state that individual choice of discussion thread participation leads to either community fragmentation or information aggregation in virtual investment communities. This choice behavior is further moderated by the nature of virtual

investment communities and individual statuses, as suggested by previous research, which will be detailed in the next chapter. In the following, I first present data, methodology and empirical analysis to examine community fragmentation or aggregation in virtual communities.

2.4 Data

I conduct this study in virtual investment communities. These communities, such as Yahoo! Finance, Silicon Investor, etc, provide a venue where online investors voluntarily interact with each other to share investing-related information. Virtual investment communities are presented in the form of stock message boards, or stock forums. Each stock message board covers one stock and individual investors interact with one another by reading and posting online messages in the message boards. Most of the messages in virtual investment communities are the results of a series of interactions among investors on a particular topic. These interactions form discussion threads are shown in Figure 2.1. In a discussion thread, the posted messages present the information that online investors seek, process, and get involved, and by repeatedly interacting with each other, investors find out more details of online postings that help them make investment decisions.

In this study, I collect stock message board messages from Yahoo! Finance, one of

the leading stock investment communities. The data include a sample of 29[2] Dow Jones stocks during a 36-month span from January 1, 2004 to December 31, 2006. The 29 stocks are all large companies, but with various risk levels (beta values), and these stocks are of significant interests of online investors. In each message, investors can not only express their opinion in the message body but also through sentiment labels. In Yahoo! Finance stock message boards, sentiments express investor opinions in five categories: Strong Buy, Buy, Hold, Sell, and Strong Sell. These five categories are coded ranging from -2 (Strong Sell) to 2 (Strong Buy), as shown in Figure 2.2. Not all users provide sentiment labels in their messages though. Two approaches have been used in prior studies to address missing sentiments for these messages. Researchers either use text-mining approach to estimate sentiments (Antweiler and Frank 2004; Das and Chen 2001) or ignore these messages without sentiments all together. It is noted that the accuracy of text-mining approach is quite low. To avoid the influence of noises due to text-mining, I choose the second approach by removing all messages without sentiments for the analysis.

The following attributes are acquired for each message: thread ID, message ID, thread topic, author ID, posting date and time, message content, and sentiment assigned along with every message. Table 2.1 shows the summary of descriptive statistics of the discussion threads of 29 stock message boards in Yahoo! Finance community.

[2] We started tracking virtual investment communities for the 30 Dow Jones stocks on January 1, 2004. We kept tracking the same set of 30 stocks even though the components of Dow Jones Index changes. In November 2005, two of the 30 companies (AT&T and SBC) merged, leaving only 29 stocks for this study.

2.5 Methodology

To provide an initial indication of community fragmentation, I compare inter-thread variation against intra-thread variation in investor sentiments within a certain stock message board on a given day. For a stock message board *s* on a given day *d*, suppose there are N_{sd} unique threads, and in each thread *i*, there are n_i messages in this thread. For each thread *i*, let $\overline{X}_i$ be the average of the sentiments of all messages in thread *i*, X_{ij} be the sentiment value of message *j* in thread *i*, and $\overline{X}$ be the average of the sentiments of all messages on this given day. The inter-thread variation of stock message board s on day d is computed as

$$Var(InterThread)_{sd} = \sum_{i=1}^{N_{sd}} n_i (\overline{X}_i - \overline{X})^2 \Big/ (N_{sd} - 1). \tag{2.1}$$

The intra-thread variation of stock message board *s* on day *d* is then is computed as

$$Var(IntraThread)_{sd} = \sum_{i=1}^{N_{sd}} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2 \Big/ \left(\sum_i n_i - N_{sd} \right). \tag{2.2}$$

I thus define community fragmentation as the ratio of inter-thread variation to intra-thread variation of stock message board *s* on a given day *d*, that is,

$$CommunityFragmentationIndex_{sd} = \frac{Var(InterThread)_{sd}}{Var(IntraThread)_{sd}}. \tag{2.3}$$

The fragmentation analysis is conducted on a daily basis due to the nature of stock market—the stock market operates every weekday, and related events and news are

released accordingly, which affect investor sentiments in stock investment. If online investors interact randomly, we should expect intra-thread variation to be the same as inter-thread variation, and community fragmentation in Equation (2.3) is 1. If online investors prefer interacting with others of different views as predicted by information economics, we expect intra-thread variation to be larger than inter-thread variation, and community fragmentation is less than 1. On the other hand, if online investors prefer interacting like-minded people based on cognitive biases and psychological needs, we expect intra-thread variation to be smaller than inter-thread variation, and community fragmentation is greater than 1.

In addition to investigating the extent of community fragmentation at the aggregate level, I take one step further to identify fragmentation resulted from the individual level. To do so, I adopt a discrete choice model to analyze individual selection of discussion threads. I consider an investor arriving at a stock message board and observing a list of discussion threads, and analyze how the investor chooses among these discussion threads.

To apply the discrete choice model, we need to define the choice set for each investor at each choice decision. Since virtual investment communities present threads in reverse chronicle order, I define the choice set as the 10 most recent unique discussion threads that investor *i* will see as soon as he/she arrives at the stock message board. I use 10 as the number of the elements in the choice set due to the fact that every time an investor visits a stock message board in Yahoo! Finance, 20 discussion threads will be

presented on the web page, and with just one scrolling-down, about 10 discussion threads are listed. Given the choice set, online investors evaluate each discussion thread in the set and then choose the one with the highest evaluation. I use the multinomial logit (MNL) framework to model the choice process. Specifically, the probability that discussion thread *k* is chosen by investor *i*, given choice set *C* of 10 most recent unique discussion threads, is

$$P(k \mid C) = \frac{\exp(X'_{ik}\,\beta)}{\sum_{m \in C} \exp(X'_{im}\,\beta)}, \tag{2.4}$$

if thread *k* is in choice set *C* and zero otherwise. X_{ik} indicate the discussion thread-related mix variables, which I will explain below.

The main effect of individual choice of discussion threads on community fragmentation or aggregation under consideration is opinion distance. The opinion distance between investor *i* and each choice of discussion thread *k* is computed as the absolute difference of investor *i*'s sentiment value and the average sentiment value of all messages in thread *k*. Notice here we use the average sentiment value of all messages in thread *k* due to the assumption that an investor will not post in a thread until he/she reads through all messages posted in this certain thread.

To examine the effect of opinion distance on individual choice, the following variables are controlled: the number of threads, the number of participants in a thread, stock volatility, thread position, and thread life span. First, for a thread choice, the number of thread is the number of concurrent unique threads in a certain stock message

board that have occurred within one-hour frame prior to the first message time of this thread choice. Second, the number of participants within a thread choice is calculated as the total number of unique users in this thread. Stock volatility is retrieved from the monthly historical stock volatility data downloaded from Chicago Board Options Exchange (CBOE). Two more thread-related variables should also be considered. Thread position indicates the positions of the threads in reverse chronicle order when an investor arrives at the message board. Life span of a thread choice is the time (in minute) elapsed since the first message posting time until an investor makes the thread choice. More details of the variables can be found in Section 3.4 of Chapter 3, in which the driving factors of online investor interactions and homophily behavior are examined at the individual level in virtual investment communities.

2.6 Empirical Analysis and Discussions

Table 2.2 first presents the average community fragmentation measure for each of 29 stock message boards at the aggregate level. The result shows that intra-thread variation is significantly smaller than inter-thread variation. This implies that online investors prefer interacting with like-minded others, indicating investors' behavior is more driven by cognitive biases and psychological needs. The finding aligns with the behavioral investigation of online investors in finance literature.

The Yahoo! Finance stock message data contains a total of 720,190 (=72,019×10) unique decision points across the 29 stock message boards. These decision points refer

to individuals making choices in participating in one of 10 discussion threads. I present the effect of opinion distance on individual choice of discussion participation in Table 2.3. The opinion distance estimate shows that Hypothesis 1A is supported (β=-0.4453). Based on average partial effects (APEs) calculation (Wooldridge 2001, p. 471), for a discussion thread, each unit increase in the opinion distance, on average, partially reduces the probability that this particular thread being chosen by 1.71%. That is, the greater the opinion distance between the investor and the discussion thread, the less likely he/she is to select this thread. This result echoes the community fragmentation analysis, in which investors are motivated by cognitive biases and psychological needs when seeking and exchanging information. The results also indicate that the controlled number of concurrent threads, thread position, and thread life span have significantly negative effects on individual choice of discussion threads, while the number of participants in a thread shows significantly positive effect. The stock volatility does not have any direct effect on individual choice of thread.

The results of this study provide initial empirical validation to previous literature on homophily and on community fragmentation (McPherson et al. 2001; Van Alstyne and Brynjolfsson 2005). An understanding of this research issue is important for both theory and practice. Realizing community fragmentation serves as the first avenue towards understanding how online users obtain information from virtual communities. In the next chapter, I take one step further into exploring the underlying interaction motives and mechanisms that drive this community fragmentation and homophily

behavior. Specifically, unraveling the factors that drive individual choice of discussion threads sheds lights on how the interactions in virtual investment communities impact user opinions and alter their economic decisions.

Chapter 3: Causes of Homophily: Individual Choice and User Interactions in Virtual Investment Communities

3.1 Introduction

In Chapter 2, it is shown that community fragmentation and homophily behavior, instead of a melting pot of opinions, prevail in virtual investment communities. Community fragmentation is resulted from individual interactions and choice of discussion threads. Therefore, it is necessary to unravel the mechanisms of individual choice of participation in virtual discussions to fully understand the cause of homophily behavior and community fragmentation. We note that a deal of prior work has studied user behaviors in various contexts of virtual communities, such as online investors (e.g. Barber and Odean 2001b & 2002; Konana and Balasubramanian 2005) and online consumers (e.g. Li and Hitt 2007). However, the link between how online investors behave and thus self-select to participate in online discussions that drive homophily is still missing. This study, therefore, is one of the first attempts to bridge this gap.

An extensive collection of literature has argued that information overload (e.g. Jones et al. 2004), the existence of uncertainty (e.g. Zhang 2006), and membership size (e.g. Butler 2001) will significantly moderate the information seeking and interaction behavior. Based on the rationales and data in Chapter 2, in this research I further explore individuals' incentives in exchanging information with like-minded people that

are moderated by the nature of virtual investment communities and individual statuses. Recall the data used for the study in Chapter 2, in virtual investment communities, all postings are organized in threads and the threads are organized in the reverse chronicle order based on the most recent posting in each thread. An investor who decides to interact with others in a virtual investment community needs to decide which discussion thread(s) to join. This structure of virtual investment communities particularly provides an ideal environment to study individuals' choice and motives of interaction. First, all interactions among investors are observed by researchers. Second, we can reconstruct the choices of discussion threads by each investor when he/she decides to participate in the community. The observed choice of discussion threads allows us to identify the underlying drivers for individual interactions in virtual investment communities.

With the 72,019 individual participants in 29 stock message boards, my results show that increase in the availability of thread choices, increase in stock volatility, and holding minority opinion are associated with homophily behavior. Membership size, on the other hand, determines whether homophily or heterophily behavior occurs. The first three results are shown to be consistent with psychological needs that drive homophily behavior, and the last result indicates interactions between economic rationales and psychological homophily. This study investigates that the underlying interactive mechanisms of online investors is principally motivated by homophily behavior due to cognitive biases and psychological needs instead of economic concerns.

The rest of this chapter is organized as follows. In Section 3.2, in addition to the

recap of some relevant literature detailed in Chapter 2 including bounded rationality, homophily behavior, information economic rationale of information seeking, and psychological biases of online investors, I will particularly present literature review on individual participation in virtual communities. Theories and hypotheses of individual interactions and choices of discussion threads are given in Section 3.3. I discuss methodology and data in Section 3.4. I present the same set of data from Yahoo! Finance for analysis as in Chapter 2, with some detailed description specifically for individual choice of discussion threads in this section. Empirical analysis and results are given in Section 3.5. Finally I present complete discussions and concluding remarks of the homophily and individual interaction studies in virtual investment communities in Section 3.6.

3.2 Literature Review

This study is related to two main streams of literature: studies on individual participation in virtual communities (Butler 2001; Wasko and Faraj 2005; Kuk 2006; Ma and Agarawal 2007) and studies on homophily in social interactions (e.g. Lazarsfeld and Merton 1954; McPherson et al. 2001).

Individual incentives of participation in virtual communities play an antecedent role for this research. Studies have shown that a variety of social and economic factors influence people participating in online networks. Butler (2001) provides a resourced-based framework that suggests resource availability and benefit provision

jointly influence individual participation in virtual communities. Specifically, while an increasing number of members can make available more resources, it could become more difficult for community members to obtain benefits from the virtual communities, leading to less participation (Asvanund et al. 2003; Butler 2001). Wasko and Faraj (2005) identify that one of the key resources provided by virtual communities is social capital. Their analysis of professional virtual networks finds that professional reputation and network position drive individual participation in virtual communities. Ma and Agarwal (2007) extend the theory further, finding that perceived identity verification promotes participations in virtual communities. As a result, IT artifacts for identity communications offered by virtual communities have a significant influence on participation. These studies provide an in-depth look at the relationship between individuals and virtual communities. I complement these studies by taking a step further to consider how individuals interact within a virtual community. Like physical societies, each virtual community consists of many small groups and each individual interact with only a limited number of groups. How individuals choose to interact has profound influence on information generation and distribution in virtual communities.

Understanding individuals' choice to interact with others is particularly important in virtual investment communities. These communities reach tens of thousands of active members and receive hundreds or even thousands of postings per day. Studies on information overload and bounded rationality (Jones et al. 2004; Simon 1957) highlights the necessary in such cases for individuals to be selective in interacting, as presented in

Section 2.2 of Chapter 2 (e.g. Broadbent 1958; DeMarzo et al. 2003). I extend these studies to provide a systematic analysis of the influence of information overload on choice of interaction in such an information-rich environment.

My research on interactions in virtual communities is also closely related to social science and social psychology literature that examines human interactions in offline communities. Prior homophily studies have largely focused on status homophily rooted in social psychological theories, as discussed in Section 2.2 of Chapter 2. A key difference between virtual communities and physical communities is the lack of commonly observed status artifacts in these communities. Few studies exist to date to examine individual interactions in such communities and what motivate(s) the interactions. My research, therefore, adds to homophily studies to virtual communities and shows that psychological biases of individuals particularly in the context of investors are major drivers of homophily behavior in such settings.

3.3 Theories and Hypotheses

Whether community fragmentation or information aggregation prevails in virtual investment communities is resulted from individual information seeking and interactions. Two competing theories, information economics and psychological perspective, explain how individuals seek information and interact with each other through opposite lenses, as detailed in Chapter 2. Further, extant literature has argued that information overload (e.g. Jones et al. 2004), the existence of uncertainty (e.g. Zhang 2006), and membership

size (e.g. Butler 2001) will significantly moderate the information seeking behavior. In the following I will discuss these three factors separately.

3.3.1 Moderating Role of Information Overload

There is unprecedented amount of information from heterogeneous sources in virtual investment communities. The information may result in information overload, the condition in which the amount of information far exceeds an individual's processing capacity such that he/she cannot process and utilize all input communications (Rogers and Agarwala-Rogers 1975). Individuals demonstrate "bounded rationality" in such cases, as they experience psychological limits in processing information and solving problems at one point of time (Simon 1957).

Information overload plays a significant role in individual interactions within virtual communities. When the amount of information is within an individual's processing ability, there is little need for selective and filtered processing. Such need, however, increases sharply with the presence of information overload (Broadbent 1958; DeMarzo et al. 2003). Jones et al. (2004) show that information overload forces community members to be selective of information—in particular to ignore information that requires significant processing effort and put more weights on information with less processing cost.

In virtual communities, a discussion thread is centered on a particular discussion topic. A large number of discussion thread choices present the possibility for

information overload problem as individuals may only have capability to process a limited number of topics. This is especially a key issue for online investors in virtual investment communities. These communities routinely receive hundreds of threads daily that can overwhelm the most dedicated online investors. The excess number of discussion topics involved in member interactions then leads to information overload and the selective information seeking behavior results. Prior homophily studies, in addition, consistently show that individuals become more selective in a large community (Mollica et al. 2003).

Online investors, as discussed earlier, present various psychological biases that affect their information seeking and interaction behavior. Specifically, anchoring and illusions of control and knowledge drive individuals to seek like-minded opinions to reinforce their original beliefs. Searching for confirmatory information requires less cognitive effort and processing cost with reduced coordination needed and limited amount of disagreement (Rabin and Schrag 1999). Therefore, when online investors become selective in processing information due to information overload, they are more likely to interact with like-minded others, i.e. choosing to post in discussion threads with smaller opinion distance. (Recall that opinion distance is used to represent the similarity between the individual belief and the aggregated opinion from a discussion thread, as presented in Section 2.3 of Chapter 2.) In addition, the number of threads available in virtual investment communities also influences individual selection by providing more opportunities for the investor to identify threads with similar sentiments. I thus propose

the following hypothesis:

Hypothesis 1. *An investor is more likely to post in a discussion thread with smaller opinion distance when there exist more concurrent discussion threads.*

3.3.2 Moderating Role of Uncertainty

Investors are known to overweight their priors relative to new information due to anchoring bias (Tversky and Kahneman 1974) and overconfidence (Barber and Odean 2001a). Recent studies further show that investors overweight prior beliefs even more when there is greater information uncertainty (e.g. Jiang et al. 2005; Zhang 2006). For example, Zhang (2006) studies whether analysts and investors exhibit more behavioral biases when greater information uncertainty exists. It is shown that if analysts under-react to new information when revising their forecasts due to behavioral biases, they under-react even more in the case of greater information uncertainty. Information uncertainty refers to the ambiguity stemmed either from the firm's stock volatility or poor information (Zhang 2006). As more frequent and more speculative investment decisions take place, overconfidence in addition will operationally manifest itself (Konana and Balasubramanian 2005).

Prior research suggests that when there is greater information uncertainty, behavioral biases on stock investment decisions will be further strengthened. Information uncertainty comes from two sources for virtual investment community users.

First, stocks differ in their inherent volatility. Some stocks such as utility stocks present much lower uncertainty than others (e.g. Internet stocks). Stock characteristics thus have a significant influence on performance uncertainty. Second, uncertainty can be found from an individual opinion status as compared with community consensus. Specifically, investors that hold minority opinion compared with the mass perceive greater uncertainty and are more likely to form homophily (e.g. Doyle and Kao 2007). These behavioral biases of online investors result in the tendency to seek out supportive opinions and reinforce their own prior beliefs. Thus I state that:

Hypothesis 2A. *When faced with higher stock volatility, an investor is more likely to post in a discussion thread with smaller opinion distance.*

Hypothesis 2B. *When an investor holds minority opinion, he/she is more likely to post in a discussion thread with smaller opinion distance.*

3.3.3 Role of Membership Size

The effects of membership size on individual information seeking and interactions are two-fold from both economic and psychological perspectives. From economic resource-based viewpoint, members are an important source of "resources" (Butler 2001). These resources mainly include greater information sources available from diverse knowledge sharing among a large community of online members. In addition, for those seeking to increase visibility and audience in virtual communities, a larger group of

community members is preferred over a smaller one with like-minded members (Butler 2001), increasing the audience resources in virtual communities (e.g. Fulk et al. 1996). I thus propose the following hypothesis:

Hypothesis 3A. *An investor is more likely to post in a discussion thread with more participants for greater information sources.*

In the context of virtual investment communities, online investors seek information and audience resources by joining discussion threads. In particular, a larger membership size provides greater possibility and access to economic resources (McPherson 1983). Therefore, to increase the availability of economic resources, online investors are motivated by resource-based rationale to seek out interactions in discussion threads with a larger number of dissimilar participants. Discussion threads with greater opinion distance are thus more likely to be chosen by community members for resources from a large membership size.

On the other hand, a discussion thread with small membership size does not provide the economic resource accessibility as expected by community members. Community members joining smaller discussion threads are thus no longer driven by economic resource-based concerns. Instead, community members may be more motivated by psychological biases in seeking social support such as self-identification, social enhancement, etc. from a smaller group in virtual communities (Dholakia et al.

2004). Particularly, smaller groups of discussions may induce lower level of knowledge diversity, thereby decreasing information sources yet enhancing self-identification and confirmatory bias (Harrison et al. 2002; Mojzisch et al. 2008). Therefore, online investors tend to participate in smaller thread discussions in which there are smaller opinion distances to seek out supportive sentiments based on psychological needs. Based on the above arguments, I propose the following hypothesis:

Hypothesis 3B. *For a discussion thread, if the membership size is small, an investor is more likely to post in this particular thread with the presence of small opinion distance.*

The considered factors and the proposed relationships constitute individual interactions and thread choices in virtual investment communities, as depicted in Figure 3.1.

3.4 Methodology and Data

This study uses the same set of data from Yahoo! Finance as described in Chapter 2. The multinomial logit (MNL) framework is adopted to model the choice process. Recall that the probability that discussion thread *k* is chosen by investor *i*, given choice set *C* of 10 most recent unique discussion threads is

$$P(k \mid C) = \frac{\exp(X'_{ik}\,\beta)}{\sum_{m \in C} \exp(X'_{im}\,\beta)}, \tag{3.1}$$

if thread k is in choice set C and zero otherwise. X_{ik} indicate the discussion thread-related mix variables. I explain the choice modeling and the variables in the following and summarize them in Table 3.1.

Independent Variable: Individual Choice of Discussion Threads

As discussed earlier, for each individual arriving at a stock message board and observing a list of discussion threads, I define the choice set as the 10 most recent unique discussion threads for each investor at each choice phase. Thus, 1 indicates that the individual chooses to post in a certain thread, and 0 otherwise.

Dependent Variables (Main Effects)

Three variables constitute the main effects of psychological needs and economic rationales on individual choice of discussion threads.

Opinion distance. The opinion distance between investor *i* and each choice of discussion thread *k* is computed as the absolute difference of investor *i*'s sentiment value and the average sentiment value of all messages in thread *k*. Notice here I use the average sentiment value of all messages in thread *k* due to the assumption that an investor will not post in a thread until he/she reads through all messages posted in this certain discussion thread.

Number of concurrent threads. The number of concurrent threads is calculated as that for a discussion thread, the total number of unique threads in a certain stock message board that have occurred within one-hour frame prior to the first message time of this particular thread. This is to capture the effect of information overload on individual choice of threads resulted from the increasing number of available thread choices.

Number of participants. The number of participants is calculated as that for a discussion thread, the number of unique participants within this particular thread. This accounts for the effect of membership size on individual choice of threads based on economic resource-based viewpoint.

Dependent Variables (Moderating Effects)

To understand the underlying driving forces that affect homophily and community fragmentation of opinions, the interaction terms accounting for the moderating effects on opinion distance according to the hypotheses are considered. In addition to the moderating effects of the numbers of concurrent threads and participants, as detailed above, two more variables for the effects of uncertainty are also included.

Stock volatility. Stock volatility is calculated based on monthly historical stock volatility data downloaded from Chicago Board Options Exchange (CBOE). Each thread choice will then be associated with a stock volatility value at the time period it

occurred.

Opinion status. An individual's opinion status is represented by a dummy variable with 1 indicating that the investor holds a minority opinion and 0 for holding a majority opinion on a stock. Minority opinion is defined as if the sentiment of an investor differs from 90% of his/her peers.

Controls

Our MNL model includes the following control variables. First, at the individual level, information overload can exert multiple influences on investor's selection of discussion threads. In particular, as the posted messages in virtual investment communities refresh frequently, information overload will limit online investors to focus on more recent discussions. The term "recent" here can be explained by the following two concepts.

Thread position. It refers to the position of each discussion thread, explained by cognitive processing costs—the cost of browsing and the cost of clicking to the next page (Brynjolfsson et al. 2006a). For a discussion thread in the choice set, the thread position is relative to other threads in reverse chronicle order when an investor arrives at the message board, that is, 0 means the most recent and 9 means the furthest thread. For an investor, the recent discussion threads are often more visible and accessible to the

investors and thus are more likely to be selected. However, the probability that the earlier discussion threads are viewed and selected drops dramatically, given the nature of virtual investment communities that those earlier discussion threads can be already pushed several web pages away. This "diminishing" probability that a discussion thread is selected can be captured by including a quadratic form of discussion thread position.

Thread life span. The second concept of being "recent" refers to the age of a discussion thread. It is differentiated from the concept of thread position: a discussion thread that is most easily viewed by online investors but with longer life span is considered less interesting to the investors. In contrast to cognitive processing costs incurred by viewing threads from further positions, the concept of life span follows economic rationales that a thread occurred a longer time ago has less informational value. Life span is calculated as the time (in minute) passed since the first message posting time in the discussion thread until the time an investor makes the choice decision.

Stock volatility and opinion status. Finally, the direct effects of stock volatility and opinion status on individual choice of discussions are also controlled for the modeling.

Based on the above variable definitions, to facilitate the MNL modeling, I reorganize the posting data from the 29 stock message boards to the form in which for each of the 72,019 participants, 10 choices of most recent discussion threads are being

considered. For each thread choice, the following information is then acquired or computed: stock ticker, whether or not the investor posted in this thread (1 for yes and 0 for no), opinion distance, the number of concurrent threads, the number of participants, stock volatility, opinions status, thread position, and life span. The descriptive statistics and correlations among the key variables are given in Table 3.2 and Table 3.3, respectively.

3.5 Empirical Analysis and Results

Given the above model and variable description, I present the full MNL modeling results in Table 3.4. The estimates account for the moderating effects of the number of concurrent threads, the number of participants, stock volatility, and minority opinion status on the influence of opinion distance for investors' thread selection. The result shows that Hypothesis 1 is supported (β=-0.1635). It suggests that, for a discussion thread, each unit increase in the number of concurrent threads, on average, partially reduces the probability that this particular thread being chosen by 1.37% with maximum possible opinion distance, yet by 0.92% with minimum possible opinion distance. That is, as the availability of thread choices increases, individuals are more motivated to choose the thread with smaller opinion distance. The result demonstrates that the homophily exacerbates when individuals have more choices and when information overload limits individual capacity to process all available information. Thus, investors are more likely to seek out information from like-minded opinions (smaller opinion

distance).

The results show that Hypotheses 2A and 2B are supported. The estimate β=-0.0622 indicates that, for a discussion thread, each unit increase in stock volatility, on average, partially reduces the probability that this particular thread being chosen by 0.67% with maximum possible opinion distance, yet by 0.06% with minimum possible opinion distance. This finding shows that when the stock volatility is higher, individuals are more motivated to choose the thread with smaller opinion distance. The result also reveals that individuals who hold minority opinions are more likely to interact with like-minded people. By having dummy 1 for minority opinion and 0 for majority opinion, β=-0.0834 indicates that the effect of individuals holding minority opinions to form homophily (by choosing threads with smaller opinion distance) is significantly stronger than that of individuals holding majority opinions by 0.0834 unit. These results present supportive findings for finance literature that when the market is more unstable or when investors hold less popular sentiments, in contrast to being economically rational, they will rely more on cognitive biases and psychological needs when participating in discussions and exchanging information.

The results also support Hypotheses 3A and 3B. With β=1.5206, the main effect of the number of participants on individual choice is shown. For a discussion thread, each unit increase in the number of participants, on average, partially increases the probability that this particular thread being chosen by 3.54%. That is, the higher the number of participants in a discussion thread, the greater the possible information sources,

and the more likely an investor to select this thread, supporting Hypothesis 3A. Estimate β=0.3055 supports Hypothesis 3B for the moderating effect of the number of participants on opinion distance for individual choice. For a discussion thread, each unit decrease in the number of participants, on average, partially reduces the probability that this particular thread being chosen by 1.32% with maximum possible opinion distance, yet increases the probability that this particular thread being chosen by 7.87% with minimum possible opinion distance. That is, individuals who post in smaller discussion threads are more likely to be motivated by psychological biases, while those posting in larger discussion threads are driven more by economic needs. This finding aligns with the fact that, both information economics and psychological factors influence individual interactions in virtual investment communities. Smaller discussion threads have less informational value and therefore people who interact within them are more likely to be driven to seek confirmatory opinions. Larger threads, on the contrary, are considered to have greater informational value. Investors who choose larger threads are more likely to be driven by information economic concerns to seek economic resources.

3.6 Discussions and Concluding Remarks

In the studies of Chapter 2 and Chapter 3, I empirically examine interaction behavior of online investors based on both information economics and psychological theories. I show that investors' behaviors are influenced mainly by psychological considerations. From psychological perspective, investors demonstrate clear biases

towards associating with like-minded people in virtual investment communities and their biases are strengthened when more thread choices are concurrent, when fewer members are participating, when the market is more volatile and when individuals hold minority sentiments. The behavior is consistent with the hypotheses derived from cognitive and psychological theories. However, life span and membership size of discussion threads reduce the influence of cognitive and psychological drivers and show that threads with greater informational sources and value are more likely to be chosen. The results indicate that individual choices are influenced by both economic and psychological considerations, yet psychological needs constitute most drivers of individual choice of discussions.

This study provides empirical validation to previous literature on homophily and on community fragmentation (McPherson et al. 2001; Van Alstyne and Brynjolfsson 2005). In addition, by investigating the underlying individual selection of discussion groups, my research represents one of the first attempts to study individual interaction behavior in virtual communities and identify the informational value of virtual investment communities.

This study has several limitations. First, current analysis does not account for the impact of the interaction activity on individual opinion. Individual opinions can be mediated during the repeated interactions with other investors. A typical example is polarization, the phenomenon that individuals tend to take more extreme opinions after group interactions (Isenberg 1986). It renders us a more complete understanding of the

driving forces of individual selection of discussion threads and community fragmentation if how individual opinions alter during group interactions is unraveled. Second, in addition to the posted discussion threads considered in virtual investment communities and the stock volatility, there are external factors such as financial news that can also impact individual opinions pertaining to certain stocks (Tetlock 2007). Third, the underlying mechanisms of user interactions in stock message boards can be further explored. From the data used in my study, different stock message boards pertain to different levels of online investor participation and interactions. The relationship among the stock characteristics, forum participation activity, and the extent that investor opinions are influenced is not yet explored in this study. Finally, the impact of online investors' interaction behavior on community structure and virtual group formation could be of great interest.

An understanding of these research issues is important for both theory and practice. This research suggests evidence explaining the community fragmentation and individual interaction in virtual investment communities. In addition, realizing community fragmentation serves as the first avenue towards understanding how online users obtain information from virtual communities and how the interactions in virtual investment communities impact user opinions and alter their economic decisions. The business value and market performance resulted from the fragmentation within virtual investment communities are of particular importance. My exploration shows online investors look for like-minded interactions, contrary to the prediction of economic

rationales. Such behavior could have significant implications for market efficiency and raises important public policy questions.

From a practical perspective, the result of the study can be applicable to a broad category of virtual communities, and provide guidelines to practitioners such as virtual community moderators and businesses who are interested in understanding the distribution of information and advertising to virtual community members. For instance, knowing whether and how online individual opinions fragment helps segment virtual community members, enabling businesses to design customized marketing campaign and pricing scheme for different types of individuals. This extra piece of profile information can also be incorporated into virtual community platform designs, serving as an index for virtual community members to judge and value others' opinions.

These studies show how virtual community members interact with each other and how they value opinions in virtual communities, suggesting community fragmentation resulted from private individual choices. The level of community fragmentation can be influenced by virtual community operators when the factors of virtual community member interactions influencing this fragmentation are known. Exploring the issues detailed above will advance the understanding of the value of virtual communities as informational media and how they shape user opinions and preferences in today's networked economy.

In the next chapter, I present a specific study to demonstrate the informational value of virtual communities for businesses in the context of online retailing. In

particular, the research studies how consumer consideration and choice of products are impacted by online recommendations, and thus imply a significant change in the landscape of product competition in online retailers.

Chapter 4: Implications of Consumer Consideration and Choice with Online Recommendations: Product Competition in Online Retailers

4.1 Introduction

Digital commerce has significantly altered the landscape of retailing. One of the most well-known phenomena is the Long Tail. The Long Tail of electronic commerce indicates the longer-lasting availability of slow-moving and rare products (Anderson 2006). Without physical constraints faced by traditional stores, online retailers provide unprecedented product varieties to their customers (Brynjolfsson et al. 2006b; Clemons et al. 2006). This is particularly facilitated by the ability of online retailers to catalog, recommend, and sell a large number of products (Brynjolfsson et al. 2003). As a result, with the help of online retailers, consumers are able to discover, evaluate and shop a greater variety of products than they previously could at brick-and-mortar stores (Brynjolfsson et al. 2003 & 2006b; Ghose et al. 2006). This opportunity to access a wider selection of products may lead to a transformation of the way in which online consumers consider product alternatives to identify the most promising purchases. Specifically, when the number of products vying for a consumer's attention increases, it is likely that more alternatives are considered by the consumer before making purchase decisions (Anderson 2006). Thus, greater product variety will expand a consumer consideration set of product alternatives.

The explosive application of online product recommendations, however, has made the expansion of the consumer consideration set in digital commerce debatable. In fact, prior research has shown with experimentation that the size of consumer consideration set could become smaller due to product recommendations in online retailers (Häubl and Trifts 2000). This consideration set shrinkage could be explained by two main factors. First, whether or not a consumer considers an additional product depends on the trade-off between the marginal benefits and costs of considering it (e.g. Hauser and Wernerfelt 1990; Robert and Lattin 1991). However, the marginal benefit of considering an additional product rapidly decreases with online recommendations because the relative product utilities are now accessible from screened and ranked product alternatives based on prior consumer transactions. As a result, online recommendations lead to a reduction in the number of products needed to be considered (Häubl and Trifts 2000). Second, product recommendations might "reinforce" the position of products on online consumers such that "a rich-get-richer effect for popular products is created, and vice-versa for unpopular products" (e.g. Fleder and Hosanagar 2008; Mooney and Roy 2000). This will result in a reduction in product variety, contradicting to what the Long Tail phenomenon indicates.

The possible shrinkage of consumer consideration of products in online retailers has significant implications for product competition. Product competition arises when consumers make purchase decisions among multiple products that are imperfect substitutes to each other. For example, a consumer in the market for a Digital SLR

camera often considers products offered by multiple manufacturers ranging from Canon to Sony. Each of the manufacturers can also offer more than one product model to target different market segments. However, as the number of products possibly considered by consumes decreases, it suggests that manufacturers now need to face two levels of competition. Products need to compete for not only being ultimately purchased but also being at least first taken into consideration by consumers. This highlights the importance of investigating product competition for both consideration and choice to understand the impact of consideration set shrinkage on product competition in online commerce. Therefore, I address two research questions in this study: a) To what degree do online consumers consider product alternatives in making purchase decisions? and b) How do products compete for consideration and for choice, given the reduced size of consideration set in online retailers?

Marketing literature has developed a number of models that account for the two-level decision making process of consumers, consideration and choice (e.g. Andrews and Srinivasan 1995; Nedungadi 1990; Roberts and Lattin 1991). However, these models require micro level data about individual choice among multiple products—information that is difficult to obtain by manufacturers. In this study, I develop a new model that requires only aggregate demand data to identify the size of consideration set and the resulting product competition in both consideration and choice stages. The model leverages a new form of aggregate data provided by online retailers that identify conditional product demand given consumers considering an alternative

product. A well-known example of such aggregate data is provided by Amazon in the form of "What Do Customers Ultimately Buy After Viewing This Item?". The data provide percentage of customers purchasing an alternative product given they have considered a certain product. The percentage represents statistics on the conditional probability of consumer purchase propensity given their consideration of a particular product. The model shows that we can recover from the statistics of both the probability of each product being considered and the probability that a product being chosen given it is considered by a consumer.

I apply the model to 38,400 unique products collected from Amazon's Electronics category at two different time periods. Besides the conditional purchase data, the data also include sales rank, sale price, and online customer review rating for each product. The results show that the consideration set size of online consumers is fairly small in real-world digital commerce. Specifically, more than half (53.294%) of the purchase decisions are made by consumers considering two or fewer alternative products. This finding leads to the differentiated product competition for consideration and for choice. That is, a product does well for consideration may not do so well for purchase choice, and vice versa. In particular, I show that because of the shrinkage of consideration sets, being considered by consumers plays a dominant role for product competition in online retailers. The findings further indicate that pricing strategies only influence the product competition in choice stage. It suggests that despite overwhelming focus on pricing competition in earlier studies (e.g. Brown and Goolsbee 2002), with the presence of

product recommendations in online retailers, pricing plays a rather limited role in product competition for consideration.

This study sheds lights on the distinctive nature of product competition resulted from the small-sized consideration sets in online retailers. The results provide significant implications for manufacturers. Getting into a consumer's consideration requires establishing product awareness, while ensuring the product being chosen by a consumer requires convincing consumers of the product value (e.g. Lilien et al. 1995). Therefore, different marketing and advertising strategies (e.g. awareness advertising and persuasive advertising) need to be developed to address these challenges in online consideration and choice of products.

The remainder of this chapter proceeds as follows. In Section 4.2, I review the prior literature as it relates to the research topic. In Section 4.3, I present the theories and methodologies to measure consumer consideration and choice and the resulting product competition in both stages. I present the data and modeling results along with discussions in Sections 4.4 and 4.5, respectively. I conclude this chapter in Section 4.6.

4.2 Literature Review

The prevalence of digital commerce in recent years has inspired growing interests in product variety and consumer purchase decisions in online retailers. The first stream of related literature is concerned about the increased product variety provided by online retailers (e.g. Anderson 2006; Brynjolfsson et al. 2003 & 2006b; Oestreicher-Singer and

Sundararajan 2006). In particular, the long tails of electronic commerce introduce the longer-lasting availability of slow-moving and rare products compared with traditional brick-and-mortar stores, which creates a greater variety of products (Brynjolfsson et al. 2003 & 2006b). The authors suggest that the higher level of product variety in online retailers may change the sets of products that are considered profitable and expand the number of products considered by a consumer. Wider selection and lower search costs facilitated by recommendations in digital commerce help consumers discover niche products and thus increase product variety (Anderson 2006).

The wider selection of products and lower search costs offered in digital markets leads to the relevant literature relates to the retailer-level competition on the Internet (e.g. Baye et al. 2004; Clay et al. 2001; Smith and Brynjolfsson 2001). Specifically, price competition in online retailers has been extensively studied. Brown and Goolsbee (2002), for instance, find that online markets reduce search costs of consumers by enabling online price comparisons. As a result, product prices are reduced significantly. Recent studies further show that price competition in online retailers can be mitigated by other factors. Chevalier and Goolsbee (2003) study price sensitivity of online consumers in online retailers, BarnesandNoble.com and Amazon.com. Their research indicates that Barnes & Noble faces much stronger online price competition from Amazon than Amazon does from Barnes & Noble. There is also significant switching cost in online retailers (Chen and Hitt 2002). Shopbots on electronic markets, in addition, provide opportunities for online retailers to leverage brand names and pricing

strategies to differentiate their products (Smith 2002).

The idea that the Long Tail allows for a wider selection of product variety, however, shows mixed results in previous literature. Specifically, collaborative filters might reinforce the recommended products on consumers such that the already popular products become more popular, and vice versa for unpopular products (Mooney and Roy 2000; Fleder and Hosanagar 2008). As a result, a smaller set of products is constantly reinforced and the number of product alternatives consumers may consider will drop accordingly. In addition, relative product utilities can be accessible from online recommendations based on prior consumer preferences and purchases (Häubl and Trifts 2000). Therefore, the presence of online recommendations decreases the marginal benefits of considering more product alternatives. This leads to a reduced *consideration set*, the set of products consumers may consider before making purchase decisions.

The possible shrinkage of consideration sets before purchases highlights the importance of a different level of competition in addition to price competition—the product competition in online retailers. This draws upon the literature that consumers may not consider all available products when making purchase decisions in a variety of marketing studies (Hauser and Wernerfelt 1990; Nedungadi 1990; Roberts and Lattin 1991). These studies have shown that, from the angles of ad hoc, deterministic, or probabilistic models, consumers' consideration of products will significantly impact their ultimate choices of products. In particular, faced with cognitive limitations, complex choice tasks, and evaluation costs, consumers will resort to phased decision strategies

(Gensch 1987). The phased consumer decision making process mainly involves two stages. In the *consideration stage*, consumers consider a smaller set of products and form the choice set. In the *choice stage*, consumers evaluate every product in the choice set and purchase the one with the highest evaluation.

Different effective marketing strategies for products to be "aware of" and to be "thought of high value" to consumers have also been explored. Product awareness has significant effects on consumer choice, requiring different marketing strategies to address the issues (McMahon 1980). In particular, products with higher quality may not be preferable by consumers given poorer product awareness (Hoyer and Brown 1990). Products also compete to remain in consumers' implicit memory rather than for a short-term stimulus to be purchased more likely (Lee 2002). This research vein is relevant to my study because it accounts for the possible impacts on strategies due to product competition resulted from the drastic change in consideration and choice in online retailers. Further, product competition for different purposes—for being considered versus being chosen—may result in different strategic concerns in pricing, advertising, and manufacturing for the manufacturers.

4.3 Theories and Methodologies

4.3.1 Revealed Preferences in Online Retailers

Online retailers provide a variety of product sales statistics that reveal consumer preferences. In this study, I investigate a particular type of such statistics provided by

Amazon.com—"What Do Customers Ultimately Buy After Viewing This Item?," as shown in Figure 4.1. The figure indicates that Amazon not only provides the products ultimately bought by consumers after viewing the item, but also provides percentages of consumers who do so for each alternative product. These percentages are essentially conditional probabilities of consumers purchasing product *Y* after they have considered product *X*[3]. As shown in Figure 4.1, Amazon provides such statistics for the top four alternative products that often cover a large amount of purchases. This allows us to construct a conditional probability matrix for a given set of product alternatives. For example, Table 4.1 shows four digital camera alternatives from Amazon.com. In each cell, Amazon presents the percentages of consumers who purchased the column product after having considered the row product.

Two observations could be drawn from Table 4.1. First, a significant percentage of consumers purchase the product they just viewed. The percentage is 66%, 78%, 81%, and 66% for the four products in Table 4.1. The surprisingly high percentage indicates that there is limited product competition among four products mainly being considered in online commerce. If each consumer considered two products before they make a purchase decision, then the average probability of purchasing product *X* after viewing product *X* should be 50%. The results from Table 4.1 indicate that on average each consumer considers less than two products in making purchase decisions despite the unprecedented level of product variety indicated by the Long Tail. Second, consumers

[3] *X* and *Y* could be the same product.

demonstrate clear brand preference. Consumers who viewed product *X* but ultimately purchased an alternative product is more likely to purchase a product from the same brand than from a rival brand. For example, consumers who viewed Sony S700 but did not purchase Sony S700 are more likely to buy another Sony camera (i.e. Sony DSCW) than buying either of the Canon cameras. I observe the same phenomenon for Canon. Consumers who viewed Canon A570 but did not purchase Canon A570 are more likely to buy Canon A560 than either of the Sony Cameras.

The objective of the study is to first show the number of product alternatives generally being considered by consumers and second, the resulting product competition from the conditional probabilities provided by Amazon. I start with showing a simple illustration on how to identify consideration stage and choice stage from the conditional probability using a case of two products. I then present the formal model in the next subsection.

Consider only two differentiated products are competing in a market. Consumers make purchase decisions in two stages. In the *consideration stage*, they form a consideration set which can include either one of two products or both. In the c*hoice stage*, they make purchase decisions. Each consumer purchases one and only one product.[4] If his/her consideration set consists of one product, the consumer just purchases the product in the consideration set. If his/her consideration set consists of

[4] This assumption is necessarily because Amazon's statistics are reported on all consumers who made a purchase. Consumers who viewed a product but did not make any purchase are excluded from the statistics.

both products, the consumer chooses the one that yields the higher utility. The purchase decision in this case is modeled with a logit model. The challenge we face is that we do not have data on individual consumer consideration sets or purchase decisions. Instead, we only observe the aggregate conditional probability of purchase propensity given a product is considered. That is, we observe four statistics for a market of two products X and Y, namely $P(X|Y \in C)$, $P(X|X \in C)$, $P(Y|X \in C)$, and $P(Y|Y \in C)$, where $P(i|j \in C)$ in this case refers to the probability of consumers purchasing product i given that they have considered product j. The ultimate goal is to identify the two stages of product competition using the conditional probabilities $P(i|j \in C)$.

We note that product competition is determined by four variables related to consideration and choice. Let $D(X)$ be the probability that a consumer will consider product X and $D(Y)$ be the probability that a consumer will consider product Y. Let $B(X)$ be the utility provided by product X and $B(Y)$ be the utility provided by product Y. $D(X)$ and $D(Y)$ determine the probability that a certain size of consideration set occurs, and describes product competition for consumers' consideration. $B(X)$ and $B(Y)$, on the other hand, describe product competition for consumer purchases given both are considered. I show below the four variables can be identified from the four conditional probabilities of purchase propensity. I first note that the conditional probability of purchase propensity $P(i|j \in C)$ equals the joint probability that a consumer considers j but ultimately purchases i divided by his/her marginal probability of considering j, that is, $P(i|j \in C)$ = Prob[(purchase i)$\cap$(consider j)] | Prob[consider j]. Note that an individual's

probability of considering j is simply $D(j)$. In addition, an individual's probability of purchasing i and considering j equals the probability that the individual considers both products i and j *and* that he/she chooses product i. Based on the above notations, we can now identify for the two products X and Y, the probability that a certain size of consideration set occurs and the competition for consideration ($D(X)$ and $D(Y)$) and in choice ($B(X)$ and $B(Y)$), using the conditional probabilities of purchase propensity, $P(i|j \in C)$.

First, in the two-product example, the consideration set C can be represented as set $\{X\}$, $\{Y\}$, or $\{X,Y\}$. In the consideration stage, the probability that product X only is included in consideration set C can then be represented as $P(X \in C) = D(X)(1 - D(Y))$, and the probability that product Y only is included in consideration set C can be represented as $P(Y \in C) = (1 - D(X))D(Y)$. Similarly, the probability that both products X and Y are considered is represented as $P((X \in C) \cap (Y \in C)) = D(X)D(Y)$. The probability that a certain size of consideration set occurs can then be computed as

$$P(\textit{Consideration Set Size} = 1) = D(X)(1 - D(Y)) + (1 - D(X))D(Y). \quad (4.1)$$

$$P(\textit{Considerat ion Set Size} = 2) = D(X)D(Y). \quad (4.2)$$

In the choice stage, the probability that X is purchased given a certain consideration set depends not only on its purchase propensity but also on what is included in the consideration set, that is, for product X,

$P(X \mid \{X\}) = 1$, $P(X \mid \{Y\}) = 0$, and

$$P(X \mid \{X, Y\}) = \frac{\exp(B(X))}{\exp(B(X)) + \exp(B(Y))}, \text{ using logit model.}$$

Similarly, for product *Y*, we have $P(Y \mid \{X\}) = 0$, $P(Y \mid \{Y\}) = 1$, and

$$P(Y \mid \{X, Y\}) = \frac{\exp(B(Y))}{\exp(B(X)) + \exp(B(Y))}, \text{ using logit model.}$$

Using the above calculations, the $D(X)$, $D(Y)$, $B(X)$, and $B(Y)$ can be identified implicitly with the conditional probabilities of purchase propensity, $P(i|j \in C)$ based on the following equations:

$$P(X \mid X) = \frac{D(X)(1 - D(Y)) + D(X)D(Y)(\exp(B(X))/(\exp(B(X) + \exp(B(Y))))}{D(X)(1 - D(Y)) + D(X)D(Y)}. \tag{4.3}$$

$$P(Y \mid X) = \frac{D(X)D(Y)(\exp(B(Y))/(\exp(B(X) + \exp(B(Y))))}{D(X)(1 - D(Y)) + D(X)D(Y)}. \tag{4.4}$$

$$P(Y \mid X) = \frac{D(X)D(Y)(\exp(B(X))/(\exp(B(X) + \exp(B(Y))))}{D(Y)(1 - D(X)) + D(X)D(Y)}. \tag{4.5}$$

$$P(Y \mid Y) = \frac{D(Y)(1 - D(X)) + D(X)D(Y)(\exp(B(Y))/(\exp(BX) + \exp(B(Y))))}{D(Y)(1 - D(X)) + D(X)D(Y)}. \tag{4.6}$$

In the following subsection, I present full extension of the two-product example to a formal model.

4.3.2 Aggregate Two-Stage Consideration Choice Model

It is straight forward to extend the two product analysis to multiple products.

The approach has been used in prior marketing literature that focuses on individual level analysis (e.g. Andrews and Srinivasan 1995). I show this approach can be extended to aggregate level data. First, in the *consideration stage*, assume D_i to denote the probability that any given product i is considered. D_i is crucial in this study: the presence of D_i implies the way in which consumers consider a limited set of products facilitated by online recommendations. This further implies the competition of a certain product being able to be brought into a consumer's considerations. Product competition reaches its highest intensity when D_i reaches 100% for all products. This would indicate that consumers consider all products in making purchase decisions.

In the consideration stage, if the consideration probability D_i of one product is significantly higher among other products in a given set of product alternatives, the probability of this particular product being the only product in a consumer's consideration set becomes much higher. Similarly, if the D values of two products are both significantly higher than others, then the consumer is more likely to consider two products in his/her consideration set before making purchases. Formally, given the product groups of four product alternatives provided by Amazon.com, I calculate the probability of that a certain size of product consideration set occurs for each group as follows.

$$P(\textit{Consideration Set Size} = 1) = \sum_i \left(D_i \prod_{j \neq i} \left(1 - D_j\right) \right). \tag{4.7}$$

$$P(\textit{Consideration Set Size} = 2) = \sum_i \sum_{j \neq i;\, j > i} \left(D_i D_j \prod_{k \neq i \neq j} (1 - D_k) \right). \tag{4.8}$$

$$P(\textit{Consideration Set Size} = 3) = \sum_i \left((1 - D_i) \prod_{j \neq i} D_j \right). \tag{4.9}$$

$$P(\textit{Consideration Set Size} = 4) = \prod_i D_i. \tag{4.10}$$

The consideration probability D_i, therefore, implies the product competition for being considered by consumers. Specifically, the probability that a certain consideration set C_k of product alternatives will occur is calculated in terms of D_i:

$$P(C_k) = \prod_{i \in C_k} D_i \prod_{i \notin C_k} (1 - D_i), \quad k \in \{1,2,...,2^n - 1\} \text{ for } n \text{ products.} \tag{4.11}$$

In the choice stage, the products contained in the choice set are being considered and evaluated by consumers. I thus consider B_i to denote the utility of product i from inherent product-specific features such as price, quality, functionality, etc, which differentiate one product from another for consumer choice. With B_i, I use the multinomial logit model (MNL) to assess the probability that product i is chosen given choice set C_k, that is,

$$P(i \mid C_k) = \frac{\exp(B_i)}{\sum_{j \in C_k} \exp(B_j)}, \tag{4.12}$$

if product i is in C_k and zero otherwise (Roberts and Lattin 1991).

Given Equations (4.11) and (4.12), we are able to derive the probability that a consumer chooses product i given that he/she has considered product j. I denote this

conditional probability as $P_{i|j} = P(i \mid j \in C)$. The conditional probability is determined by two factors, as mentioned in the two-product example: the probability that a consumer's consideration set contains *j* and that he/she purchases product *i*. Formally, we have

$$P_{i|j} = P(i \mid j \in C) = \frac{\sum_k [P(C_k)P(i \mid C_k)I(j \in C_k)]}{\sum_k [P(C_k)I(j \in C_k)]}, \tag{4.13}$$

where $I(j \in C_k)$=1 if true or 0 otherwise.

The above formulation identifies the conditional purchase probability for a randomly selected consumer. That is, $P_{i|j}$ is the probability of choosing product *i* after considering product *j* for a given consumer. Amazon.com aggregates the observed choices and presents the average preference of a large number (*N*) of previous consumers. We note that the individual event of a given consumer choosing product *i* after considering product follows a Bernoulli distribution with $p = P_{i|j}$. Based on the central limit theorem, the distribution of the Amazon's aggregate measure follows a normal distribution $N(\mu, \sigma)$, where

$$\mu = P_{i|j} \text{ and } \sigma^2 = P_{i|j}(1 - P_{i|j})/N, \tag{4.14}$$

where *N* is the total number of previous consumers who purchased product *i* after having considered product *j*.

To calibrate the proposed model, maximum likelihood estimation is adopted.

According to Equations (4.13) and (4.14), let $P^{*}_{ij,t}$ denote the observed probability that product i is purchased after considering product j at period t across n products, $i, j \in \{1,2,...,n\}$. The log likelihood function for the model is then written as

$$L(D_i, B_i) = \log C - \frac{T}{2}\log\left[P_{i|j}\left(1 - P_{i|j}\right)/N\right] - \frac{\sum_{t=1}^{T}\sum_{j=1}^{n}\sum_{i=1}^{n}\left(P^{*}_{ij,t} - P_{i|j}\right)^2}{2\left(P_{i|j}\left(1 - P_{i|j}\right)/N\right)}. \qquad (4.15)$$

Equation (4.15) provides the approach with which the consideration probabilities (D_i) and the product utilities (B_i) can be identified from revealed preferences in conditional probabilities of purchase propensity. Recall that D_i denotes the probability that product i will be considered by a consumer, and B_i denotes the utility for product i, which reflects inherent brand-specific features. Thus, the estimation would illustrate the product competition in consideration stage and in choice stage, depending on the values of D_i's and B_i's, respectively.

4.3.3 Demand Estimation

The derivation of the probability that a certain choice set C_k occurs, $P(C_k)$ in Equation (4.11), and the probability that a certain product is purchased given it is considered in the choice set, $P(i \mid C_k)$ in Equation (4.12), facilitate the estimation of product demands. Specifically, given choice set C_k, the demand of product i can be estimated from the probability that it will be considered and the probability that it will be chosen among each type of consideration set. That is,

$$P(i) = \sum_k P(i \mid C_k) P(C_k). \tag{4.16}$$

Equation (4.16) reveals that product demands on Amazon.com can be estimated from the readily available statistics of conditional purchase propensity. In particular, demand for a product pertains to its capability to compete for consideration and its product utility evaluated by consumers.

4.3.4 Price and Quality Effects in Product Competition

In the above section I show that both stages of product competition can be identified from a one-period observation of conditional purchase propensity. In this section, I further extend the methodology to separate price effect from quality effect in the setting of a panel data. The extended approach is based on the notion that price changes influence the value of a product thus may affect product competition within consideration set for choice. On the other hand, price changes have little influence on product competition for being considered.

Let p_{it} denote the price of product i at time t. The utility of a given product at time t can then be represented as

$$B_{it} = a_i - bp_{it}. \tag{4.17}$$

Note that we could estimate product utility B_i from Equation (4.15), but the availability of panel data allows us to further identify a_i and b. Substitute B_{it} in Equation (4.17) for (4.12), the probability that product i is purchased given it is

considered at time t can be rewritten as a function of a_i, b, and p_{it}:

$$P_t(i \mid C_k) = \frac{\exp(a_i - bp_{it})}{\sum_{j \in C_k} \exp(a_j - bp_{jt})}, \tag{4.18}$$

and thus the conditional probability of purchase propensity in Equation (4.13) can be rewritten as, at time t,

$$P_{i|j,t} = P_t(i \mid j \in C) = \frac{\sum_k [P_t(C_k) P_t(i \mid C_k) I(j \in C_k)]}{\sum_k [P_t(C_k) I(j \in C_k)]}. \tag{4.19}$$

The demand for product i at time t, denoted as Q_{it}, can therefore be derived from the estimation in Equation (4.16) and from Equation (4.18):

$$Q_{it} \equiv P_t(i) = \sum_k P_t(i \mid C_k) P_t(C_k) = \sum_k \left[P_t(C_k) \frac{\exp(a_i - bp_{it})}{\sum_{j \in C_k} \exp(a_j - bp_{jt})} \right]. \tag{4.20}$$

4.4 Data

The data I collect are from publicly available statistics of purchase propensity at Amazon's "What Do Customers Ultimately Buy After Viewing This Item?" as shown in Figure 4.1. The data are extracted using automated scripts to access and parse HTML pages from the retailer. I start with the collection of 38,400 unique products under Electronics category in 9,600 product competition groups, as illustrated earlier in Table 4.1. Each product is accompanied with the conditional percentages of consumers buying one product after viewing another for the top three alternative products plus the

product itself. The data collection results in a total number of 153,600 (= 38,400×4) data points per collect. For every electronic product, the following information is included: conditional purchase propensity, sales rank and sale price.

The data were collected in two separate time periods. The first set was collected as one-time data points on May 15th, 2008 (Period 1), and the second set was collected one week later on May 22nd, 2008 (Period 2). The data collection results in a short panel data with significant changes in price and sales rank for the products. Table 4.2 lists summary statistics for the data. Note that in Table 4.2, C1 refers to the conditional purchase percentage for the most purchased product after viewing the product page, e.g. in Figure 4.1 it refers to 53%. In most cases, the most purchased product is the product being viewed. Similarly, C2 refers to the conditional purchase percentage of the second most purchased product after viewing the product page, i.e. in Figure 4.1 it refers to 18%. C3 and C4 follow the same rationales for the third and the fourth most purchased product, respectively.

4.5 Empirical Results and Discussions

The estimation is conducted for each group of substitute product alternatives. Figure 4.2 shows the average distribution of consideration set sizes across all product groups, according to Equations (4.7) to (4.10). The *y*-axis represents the percentage of consumers that have the consideration set size corresponding to the *x*-axis value. It shows that almost all consumers have a consideration set of two or fewer product

alternatives. More than 40% of consumers consider only one product when making purchase decisions. The results indicate that, although facing unprecedented product varieties provided by online retailers, in reality online consumers consider only a fairly small set of products when making purchase decisions due to the presence and reinforcement of online recommendations.

As discussed above, this drastic shrinkage of consideration set has important impacts on product competition in online retailers. For illustration, I first discuss product competition among the four camera alternatives listed in Table 4.1. Table 4.3 shows the estimation results of the consideration probability (*D*'s) and product utility (*B*'s) for Sony S700, Sony DSCW, Canon A570, and Canon A560. Sony DSCW has the highest probability (*D* = 58%) of being considered by consumers. Sony S700 is second most likely to be considered with *D* = 12.6%. Canon A570 and A560, on the other hand, have relatively low consideration probabilities (*D*=6% and 2%, respectively). However, the estimation shows that the most considered product, Sony DSCW, does not provide the best value to consumers. The relative utility of the camera is 0.055, presenting a relatively low value compared with Sony S700 (relative utility is 0.206) and only a slightly better value compared with Canon A570 (relative utility is 0.013). The relatively lower utility of Sony DSCW also indicates that it is less likely to be purchased after being considered along with other products in the choice set. The results suggest that the two stages of product competition are particularly distinctive when the consideration set is fairly small. While Sony DSCW does well for competition in

consideration, it does poorly for competition in choice.

I conduct the above analysis for each group of substitute product alternatives. The summary of the estimations of consideration probabilities (*D*'s) and relative product utilities (*B*'s) are presented in Table 4.4. I first depict the maximum likelihood estimation results of consideration probabilities for the 38,400 products from Equation (4.15) in Figure 4.3. Figure 4.3 suggests that only around 1% of products have more than 50% probability of being considered by consumers in the market for related products. The statistics indicate a fierce product competition for shrunk consideration in online commerce, as a large number of products are rarely considered by most consumers. On the other hand, the results also suggest relatively low level of competition based on product value as consumers consider few products in making purchase decisions. Figure 4.3 also addresses concerns about the influence of the Long Tail phenomenon. While online retailers have substantially increased product variety, the analysis indicates that that niche products continue to face serious obstacles as they are rarely considered by consumers.

The scatter plot in Figure 4.4 illustrates the competition for consideration and for choice, represented by the values of consideration probabilities and product utilities, respectively, for 38,400 products. The results also imply that, while many products present comparable capability in competing for consideration and for choice, there is a great amount of products that do well for consideration but not so well for purchase choice, and vice versa. This demonstrates that there is a clear differentiation at the two

levels of product competition—competition for consideration and for choice.

In the following discussions, I further uncover the factors that attribute to product competition in consideration stage and in choice stage respectively, given that the consideration set is fairly small.

4.5.1 Effects of Consideration Probability and Product Utility on Purchase

I first examine the role of product consideration probability in product competition. Specifically, to observe the impact of consideration probability on ultimate product choice, in Figure 4.5 I re-plot Figure 4.3 to a histogram in which in *x*-axis, "Rank 1" refers to the most-likely-purchased products in each product group, "Rank 2" refers to the second-most-likely-purchased products in each product group, etc. *Y*-axis refers to the average consideration probability of the products in each rank. Figure 4.5 shows that the most-likely- and the second-most-likely-purchased products present comparable average consideration probabilities, while the average consideration probabilities for products in Rank 3 and Rank 4 decrease significantly. This reveals that there is a greater competition for being considered by consumers and that the consideration probability plays a critical role in resulting in greater variation in product purchases.

Product utilities (*B*'s), however, present a different effect on purchase probability. To demonstrate the effect of product utility on product purchase and market share, I remove the influence of consideration probability by assuming that all products in a competing product group are considered by consumers. Therefore, the probability of a

given product being purchased, according to Equation (4.16), is simply

$$P(i) = \sum_{k} P(i \mid C_k) = \frac{\exp(B_i)}{\sum_{j \in C_k} \exp(B_j)}. \tag{4.21}$$

Figure 4.6 presents the histogram of average purchase probability for each product rank, according to Equation (4.21). From the figure, we can see that each product rank reveals very comparable average purchase probabilities. This indicates that the competition for being chosen within the consideration set is little: product utility does not contribute much to the variation in purchases. Compare Figure 4.6 with Figure 4.5, we observe that the impact of consideration probability on product purchase is greater than that of product utility. Specifically, the results indicate that consideration probability, rather than product utility, plays a more dominant role in leading to the variation in ultimate product choice and purchase. On the other hand, product competition within the consideration set is rather limited. In the following, I further separate the price effect from the quality effect within the consideration set on product purchases.

4.5.2 Results of Price and Quality Effects on Product Competition

As discussed earlier, the proposed methodology is further extended to separate the product quality effect (*a*'s) from price effect (*b*'s) for product utility estimation. In particular, the consideration probabilities (*D*'s), quality effects (*a*'s) and price effect (*b*) are directly estimated from Equation (4.19). Table 4.5 shows a small example of the

four digital cameras, with *a*'s and *b* estimated, representing product quality and price parameters respectively.

The separation of quality effect from price effect allows for a further investigation of the effect of product utility on purchase. To examine the quality effect, I control for the price effect by assuming that the prices remain the same across all products in a competing product group. Therefore, the effect of product utility on purchase will solely depend on the product quality. I show in Figure 4.7 the histogram of average purchase probability (based on product quality) for each product rank. The figure shows that, controlling for price effect, each product rank has very comparable average purchase probabilities. This indicates that product quality, within the consideration set, plays a rather limited role in affecting product purchases.

The analysis decomposes the factors that would result in the variation of product purchase probabilities, including consideration probability, product utility, and quality effect. The empirical results show that in online retailers in which the consideration set may shrink with the presence of online recommendations, there becomes more significant competition for being considered by consumers than being chosen from the consideration set.

4.5.3 Demand Estimation

The parameters, including consideration probability and product utility as a function of quality effect (*a*'s) and price effect (*b*'s), are further used for demand

estimation, according to Equations (4.17) to (4.20). Figure 4.8 presents the distribution of demands for products estimated from consideration probability and product utility. The results show that less than 10% of the products have the values for demand above 0.1, while over 90% of the products show the values for demand below 0.1. Specifically, in this figure to the right is the “Long Tail,” and to the left are the very few products that dominate the demands.

4.5.4 Decomposition of Product Competition

To summarize, to understand the structure of product competition in online commerce due to the shrinkage of consideration set, I decompose the competition on three dimensions. First, products differ in their propensity of being considered by a consumer. Second, products differ in their inherent quality. And third, products differ in their prices. The three dimensions show important implications for companies by corresponding to a firm’s advertising, quality and price strategy respectively. To assess the influence of each dimension, I isolate the influence of the other two dimensions. First, I assess the influence of consideration on product competition by estimating the probability of each product being considered by a consumer. Figure 4.5 presents the result. Second, I assess the influence of product utility by estimating the probability of each product’s being purchased under the assumption that every consumer considers all products in a production competition group. This approach allows us to remove the influence of variations in consideration probability and focus on the influence of product

utility. Figure 4.6 presents the result. Finally, I assess the influence of product quality by estimating the probability of each product's being purchased under the assumption that every consumer considers all products in a product competition group and that all products are offered at the same price. This approach allows us to remove the influence of consideration effect and the influence of price effect. Figure 4.7 presents the results. In Figure 4.9, I put the three histograms of probabilities side-by-side. The results indicate that when consideration set is small, consumer consideration plays a significant role in driving up heterogeneity in product demands. The average purchase probabilities for each product rank estimated from product utility and quality, on the other hand, do not vary significantly.

4.6 Conclusions

In this study, I show that the sizes of consideration sets of online consumers are fairly small, evidence from real-world business data of Amazon.com. The possible occurrences of shrunk consideration sets in online retailers have significant implications for online product competition.

Specifically, I analyze product competition for limited consideration by consumers and for choice within consideration sets in online retailers. I find that despite the large number of product varieties provided by online retailers, most consumers consider fewer than two products in making purchase decisions due to the presence and reinforcement of online recommendations. As a result, there is very limited competition

within a consumer's consideration set. Changes to product quality and price therefore have little influence on consumer demand. On the other hand, the analysis shows that majority of variations in product demand is due to differences in products' propensity of being considered. This suggests that competition for consumer consideration plays a dominant role in online product competition.

The findings have significant implications for firms' product, price and marketing strategies for online retailing. Contrary to the convention wisdom that the Internet significantly broadens consumers' product knowledge and access to product specific information, the findings indicate that increasing consumer product awareness remains to be the most important task for firms, while pricing and product positioning strategies have more limited value to product competition.

By developing a model to measure product competition using only publicly available aggregate purchase statistics of consumers provided by online retailers, I also contribute to the literature from a methodological perspective. While prior two-stage consideration choice model requires micro-level data that is usually not available to firms, I show that the model can be identified using only aggregated information in online markets. This is particularly important for electronic commerce research given that micro level transactions are not generally accessible. I also extend the model to panel data setting, thus enabling price effect to be separated from quality effect in the analysis.

This research also provides unique contributions to practice. The findings in this study imply the importance of effective and differentiated marketing strategies for online

retailing. In particular, the results show that product competition for consideration in online retailers may be fiercer than that for choice. Therefore, products competing for consideration and for choice will require different marketing efforts made by product manufacturers. For example, while improving product-specific quality and functionality has long been acknowledged as the core for manufactures, it is suggested that increasing brand awareness to consumers with more sophisticated advertising skills may be even more critical for online commerce.

This study has a number of limitations. Since the consideration process of consumers is not directly observable in online retailers, in the proposed model I assume that a product's probability of being considered is independent from other products. Second, my methodology can only be applied to substitute products but not to complementary products. Complementary consumption requires different concerns about consumers' consideration and choice (Gentzkow 2007), and the issues are not covered in this study.

Appendices

Appendix for Chapter 2

Appendix 2.1 Information Economic Rationale of Information Seeking

Consider an example of stock message boards in which online investors view and share investment information. Let *s* denote the true state of an investment in a certain stock. For explanation purpose, I assume there are two states, $s \in \{s_1, s_2\}$, representing *s* is either a positive or a negative state. Notice that the model can be easily extended to the situation in which there is a finite set with states more than two. Each online investor *i* is endowed with a noisy signal s_i about the true state. Let p_i be the probability of observing the true state by individual *i*, i.e. the accuracy of his/her signal. We then have

$$\begin{aligned} &\Pr[(s_i = s) \mid s] = p_i \\ &\Pr[(s_i \neq s) \mid s] = 1 - p_i \end{aligned},$$

where $s \in \{s_1, s_2\}$ and $p_i > \frac{1}{2}$ (i.e. all investors are at least weakly informed.)

Each online investor's endowed information set can be specified as $\{s_i, p_i\}$, where s_i is the signal he/she observes and p_i is the probability that he/she observes the true state. Assume the correct decision yields a payoff of α while the incorrect decision leads to a loss of α. The objective of the investor is to make an investment that

maximizes his/her expected payoff EU :

$$EU = \sum_{s\in\{s_1,s_2\}} (\alpha \times \Pr[(s_i = s) \mid s] \times \Pr(s)) - \sum_{s\in\{s_1,s_2\}} (\alpha \times \Pr[(s_i \neq s) \mid s] \times \Pr(s))$$
$$= \alpha \times p_i - \alpha(1 - p_i) = (2p_i - 1)\alpha. \quad (2.1.1)$$

Equation (2.1.1) accounts for four scenarios in which the investor's observation is the same as the true state (s_1 or s_2) or the investor's observation is different from the true state (s_1 or s_2). It suggests that, without additional information, an investor invests according to his/her signal as long as his/her accuracy level is above 50 percent. Virtual investment communities allow the investor to improve his/her decision by interacting with others and obtain their information.

I consider an individual arriving at a virtual investment community and observing messages from others. Messages arrive sequentially and the individual considers one message at a time. Each message is posted by another member with information set $\{s_j, p_j\}$. In online communities, while the signal of a message is easily observable, its accuracy is more difficult to assess. It requires the individual to click through the message and often to interact with the person who posted this message. To model this feature, I assume that it is free for an individual to observe the signal of a message but he/she needs to incur a cost *c* to obtain its accuracy through interactions with the other investors. I show below that information economics predicts that the individual has more incentive to interact with investors with opposite opinions than those sharing the same opinion of a stock.

To see the result, I start with a simpler case where the individual observes both the

signal s_j and the accuracy p_j of a message. A rational investor will make inference of the true state using Bayes rule based on the message as well as his/her own signal. If the observed signal from virtual investment communities is the same as the investor's own signal, then the probability that the signal observed by investor *i* represents the true state is:

$$P\left[s \mid (s_i = s \cap s_j = s)\right] = \frac{\Pr[(s_i = s \cap s_j = s) \mid s]\Pr(s)}{\Pr[(s_i = s) \cap (s_j = s)]} = \frac{p_i p_j}{p_i p_j + (1 - p_i)(1 - p_j)}. \quad (2.1.2)$$

On the other hand, if the observed signal from virtual investment communities is different from the investor's own signal, then the probability that the signal represents the true state is:

$$P\left[s \mid (s_i = s \cap s_j \neq s)\right] = \frac{\Pr[(s_i = s \cap s_j \neq s) \mid s]\Pr(s)}{\Pr[(s_i = s) \cap (s_j \neq s)]} = \frac{p_i(1 - p_j)}{p_i(1 - p_j) + (1 - p_i)p_j}. \quad (2.1.3)$$

Given the inference, it is straightforward to show the following lemma:

Lemma 1. An investor endowed with information set $\{s_i, p_i\}$ *makes the following investment decisions upon observing a message with signal* s_j *and accuracy* p_j *:*

(1) Invest based on his/her own signal if the signal of the message is the same as the signal of his/her own.

(2) Invest based on his/her own signal if the signal of the message differs from his/her own signal and the strength of the message is weaker.

(3) Invest against his/her own signal if the signal of the message differs from his/her own signal and the strength of message is stronger.

Proof of Lemma 1

(1) Invest based on his/her own signal if the signal of the message is the same as the signal of his/her own signal.

Proof. If the signal of message *j* is the same as investor *i*'s own signal, then the probability that the signal observed by investor i represents the true state is:

$$P_1 = P\left[s \mid (s_i = s \cap s_j = s)\right] = \frac{p_i p_j}{p_i p_j + (1-p_i)(1-p_j)}. \tag{2.1.4}$$

Since I assume that investors are at least weakly informed, that is, $p_i > 1/2$, $p_j > 1/2$, we have

$$P_1 = \frac{p_i p_j}{p_i p_j + (1-p_i)(1-p_j)} > \frac{p_i p_j}{p_i p_j + p_i p_j} = \frac{1}{2}. \tag{2.1.5}$$

The expected utility for investor *i* is $(2P_1 - 1)\alpha > 0$. Therefore, investor *i* will invest based on his/her own signal.

(2) Invest based on his/her own signal if the signal of the message differs from his/her own signal and the strength of the message is weaker.

Proof. If the signal of message *j* is different from investor *i*'s own signal, then the probability that the signal observed by investor *i* represents the true state is:

$$P_2 = P\left[s \mid (s_i = s \cap s_j \neq s)\right] = \frac{p_i(1-p_j)}{p_i(1-p_j)+(1-p_i)p_j}. \tag{2.1.6}$$

Since I assume $p_i > 1/2$, $p_j > 1/2$, and if the strength of the message is weaker, that is, $p_i > p_j$, we have

$$\frac{p_i}{p_j} > \frac{1-p_i}{1-p_j} \Rightarrow p_i(1-p_j) > (1-p_i)p_j. \tag{2.1.7}$$

$$P_2 = \frac{p_i(1-p_j)}{p_i(1-p_j)+(1-p_i)p_j} > \frac{p_i(1-p_j)}{p_i(1-p_j)+p_i(1-p_j)} = \frac{1}{2}. \tag{2.1.8}$$

The expected utility for investor i is $(2P_2 - 1)\alpha < 0$. Therefore, investor i will invest based on his/her own signal.

(3) Invest against his/her own signal if the signal of the message differs from his/her own signal and the strength of message is stronger.

Proof. Similarly as the proof in (2), since I assume that $p_i > 1/2$, $p_j > 1/2$, and if the strength of the message is stronger, that is, $p_i < p_j$, we have

$$\frac{p_i}{p_j} < \frac{1-p_i}{1-p_j} \Rightarrow p_i(1-p_j) < (1-p_i)p_j. \tag{2.1.9}$$

$$P_2 = \frac{p_i(1-p_j)}{p_i(1-p_j)+(1-p_i)p_j} < \frac{p_i(1-p_j)}{p_i(1-p_j)+p_i(1-p_j)} = \frac{1}{2}. \tag{2.1.10}$$

The expected utility for investor i is $(2P_2 - 1)\alpha < 0$. Therefore, investor i will invest against his/her own signal.

Lemma 1 indicates that when the signal from the message is the same as the investor's own signal, an individual's investment decision is the same regardless of the accuracy of the message. However, when the two signals are different, the investment decision depends upon the accuracy of the message.

I now consider the situation when online investors can readily observe a message's signal but require extra cost *c* to identify its accuracy. Lemma 1 indicates that when the signals are the same, there is no value for the investor to identify the accuracy of a message, since the additional information will not influence his/her decision. However, when the signals are different, the investor has more incentive to identify the accuracy of the message:

Proposition 1. An investor endowed with information set $\{s_i, p_i\}$ *makes the following information seeking decisions upon observing a message with signal* s_j *and unknown* p_j *with distribution f(p).*

(1) Seek no further information if the signal of the message is the same as the investor's own signal.

(2) Seek information on accuracy of the message if the signal of the message differs from the investor's own signal and if the information seeking cost is not too high.

Proof of Proposition 1

To prove the result formally, I assume an individual has a prior belief of distribution p_j as follows

$$p_j = \begin{cases} p_0 & \text{with probability } q, \\ p_1 & \text{with probability } (1-q). \end{cases}$$

If both p_0 and p_1 are smaller than p_i, then Lemma 1 indicates that investor *i* always make the investment decision based on his/her own signal regardless of the realization of p_j. Therefore, there is no incentive for the investor to explore the accuracy of the message. Similarly, if both p_0 and p_1 are greater than p_i, investor *i* always makes the investment decision against his/her own signal regardless of the realization of p_j, and there is no incentive for the investor to explore the accuracy of the message. The only situation in which identifying the message accuracy matters to an investor is when the strength of his/her signal is between p_0 and p_1.

Without loss of generality, I assume $p_0 > p_i > p_1$. Let $p^* = E(p_j) = p_0 q + p_1(1-q)$, the expected value of p_j, then there are two cases. The first case is when $p^* > p_i$. If the investor does *not* click on the posting, then the probability of the investor making the correct decision is p^*. The value of information is therefore $(p^* - p_i)\alpha$. On the other hand, if the investor clicks on the posting, the accuracy could be either p_0 or p_1. If p_0 is found, the investor will invest based on what

the posting signal suggests, and the value of information is $(p_0 - p_i)\alpha$. On the contrary, if p_1 is found, the investor will make the decision based on her own signal, and the value of information is 0. The expected value of knowing the realization of p_j is therefore $(p_0 - p_i)\alpha q$. The value of the click can be calculated as the difference between the value of knowing the realization of p_j and the value of not knowing the realization, that is,

$$(p_0 - p_i)\alpha q - (p^* - p_i)\alpha = (p_i - p_1)(1-q)\alpha > 0.$$

Given the cost of clicking, the net utility is $(p_i - p_1)(1-q)\alpha - c.$

Likewise, for the second case in which $p^* < p_i$, I can show that the value of the click is calculated as

$$(p_0 - p_i)\alpha q - 0 = (p_0 - p_i)\alpha q > 0.$$

Given the cost of clicking, the net utility is $(p_0 - p_i)\alpha q - c.$

In sum, the proposition above suggests that economically rational individuals have more interests in exploring messages that contradict their belief to improve the quality of decision making.

Figure 2.1: A Discussion Thread Example (© 2007 Yahoo! Finance Message Board: MSFT)

Messages in Topic

Minimum rating: 2 stars + unrated What's this?

Subject	Author	Rating	Time of Post (ET)
ABI Says Linux Will Dominate Mobile Market	wottowwottow	★★★★★	8-Apr-07 01:05 pm
Re: ABI Says Linux Will Dominate Mobile Market	creepy_word_game	Rate it	8-Apr-07 02:18 pm
Re: ABI Says Linux Will Dominate Mobile Market	intentoserjusto	Rate it	8-Apr-07 02:28 pm
Re: ABI Says Linux Will Dominate Mobile Market	creepy_word_game	Rate it	8-Apr-07 02:33 pm
Re: ABI Says Linux Will Dominate Mobile Market	intentoserjusto	Rate it	8-Apr-07 02:36 pm

Figure 2.2: Coded Sentiment Values in Yahoo! Finance Stock Message Boards

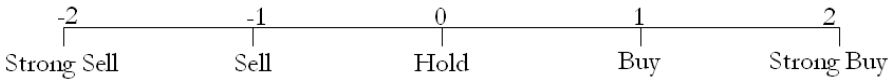

Table 2.1: Descriptive Statistics: Yahoo! Finance Stock Discussion Threads

Stock Ticker	Avg Daily Messages	Avg Daily Tagged Threads	Sentiment Average	Sentiment Standard Deviation
AA	54.43	6.70	0.90	1.27
AXP	11.96	1.10	(0.83)	1.46
BA	173.27	6.35	1.00	1.21
C	36.95	2.99	0.72	1.41
CAT	20.47	5.27	1.08	1.56
DD	31.14	1.27	0.70	1.63
DIS	53.08	5.54	0.42	1.61
EK	9.53	1.81	(0.14)	1.46
GE	351.14	19.27	0.89	1.47
GM	264.60	54.29	(0.24)	1.80
HD	37.44	7.83	0.74	1.66
HON	13.44	0.63	0.82	1.40
HPQ	263.72	7.31	0.05	1.47
IBM	88.03	9.54	0.46	1.75
INTC	234.92	89.71	0.24	1.81
IP	12.29	1.38	0.49	1.36
JNJ	23.75	3.84	1.00	1.31
JPM	96.94	2.29	(0.42)	1.51
KO	64.06	0.91	0.57	1.35
MCD	23.17	1.28	0.42	1.56
MMM	36.26	0.97	0.67	1.45
MO	154.63	10.20	0.91	1.07
MRK	98.60	4.99	0.35	1.62
MSFT	552.60	34.44	(0.35)	1.59
PG	20.49	3.00	1.00	1.17
T	47.58	3.55	(0.07)	1.57
UTX	12.77	2.69	0.83	1.22
WMT	488.41	52.01	0.34	1.95
XOM	102.77	12.67	0.81	1.82

Table 2.2: Community Fragmentation: Intra-thread Variation vs. Inter-thread Variation

T-Tests		
Difference	**DF**	**t Value**
Var(IntraThread) – Var(InterThread)	18000	-152.38***

*** p < .01

Table 2.3: Effect of Opinion Distance on Individual Choice of Discussion Threads

Variables	**MNL Estimate (Std. Err.)**	**Hypotheses**	**Supported**
Opinion Distance	-0.4453 (0.0036)***	H1A H1B	Yes No
# of Concurrent Threads	-1.4601 (0.0143)***		
# of Participants	1.6952 (0.0138)***		
Thread Position (linear term)	-0.3030 (0.0068)***		
Thread Position (quadratic term)	-0.0049 (0.0008)***		
Life Span	-0.0004 ($8.6E^{-6}$)***		
Stock Volatility	0.0380 (0.2271)		
Pseudo R-square	60.13%		
Observations	72,019		

*** p < .01

Appendix for Chapter 3

Figure 3.1: Effects of Psychological and Economic Concerns on Opinion Distance and Resulting Thread Choices

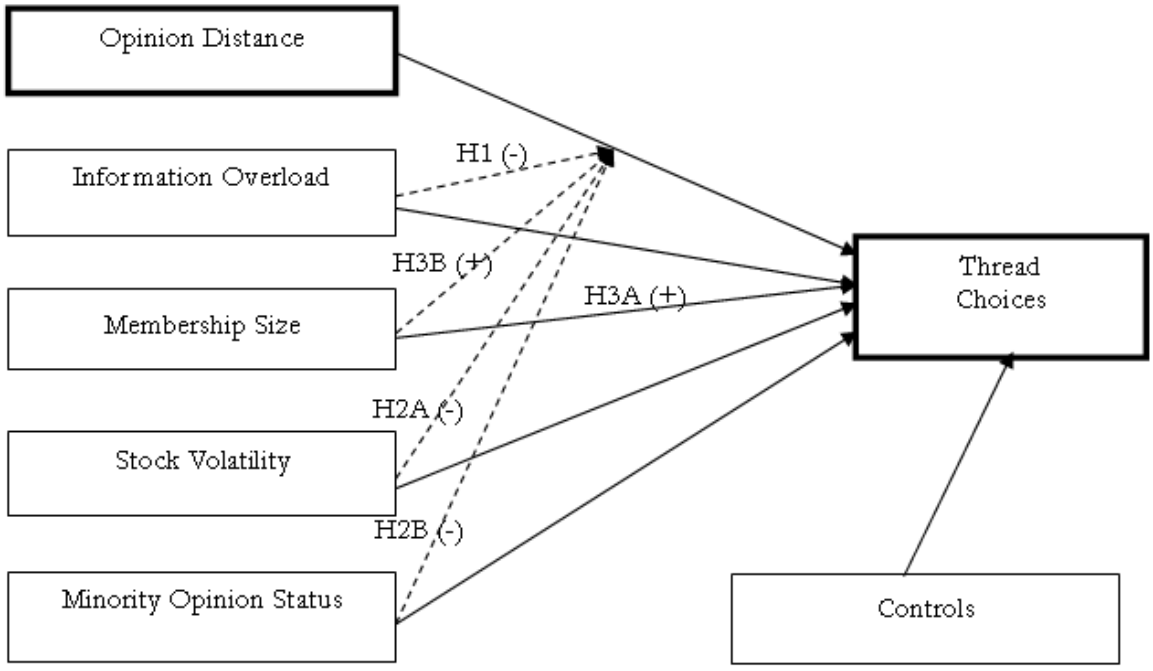

Table 3.1: Variable Definitions for Individual Choice Model

Variables	Definitions
Opinion Distance	The absolute difference of the sentiment value of the investor and the average sentiment of all messages in a certain discussion thread.
# of Concurrent Threads	For a discussion thread, the total number of unique threads in a certain stock message board that have occurred within one-hour frame prior to the first message time of this thread.
# of Participants	For a discussion thread, the number of unique participants within this thread.
Stock Volatility	The monthly-based historical stock volatilities retrieved from CBOE.
Opinion Status	A dummy variable 1 indicates that the investor holds a minority opinion and 0 for a majority opinion on a stock. Minority opinion is defined as if the sentiment of an investor differs from 90% of his/her peers.
Thread Position	For a discussion thread in the choice set, the position relatively to other threads in reverse chronicle order when an investor arrives at the message board: 0 means the most recent and 9 means the furthest thread.
(linear term)	Discussion threads with more recent positions are often more visible and accessible to the investors and thus are more likely to be selected.
(quadratic term)	To capture the fact that the probability that the earlier discussion threads are viewed and selected drops dramatically.
Life Span	The time in minute passed since the first message posting time in the discussion thread until the time an investor makes the choice of thread.

Table 3.2: Descriptive Statistics: Full Thread Choice Data in Yahoo! Finance

Variable	Mean (Std dev)	Max	Min
Opinion Distance	1.359 (1.753)	4	0
# of Concurrent Threads	5.586 (9.324)	201	1
# of Participants	6.131 (7.427)	40	1
Stock Volatility	22.014 (12.433)	111.469	6.247
Opinion Status	0.265 (0.441)	1	0
Life Span	890.606 (3789.990)	43179	1
Total Observations: 72,019			

Table 3.3: Correlation Matrix of the Key Variables

Variable	1	2	3	4	5
1 Opinion Distance	1.00				
2 # of Current Threads	0.042	1.00			
3 # of Participants	0.044	0.509	1.00		
4 Stock Volatility	0.035	0.301	0.331	1.00	
5 Opinion Status	0.404	0.087	0.092	0.042	1.00

Table 3.4: Individual Choice of Discussion Threads: Full MNL Estimation Results

	Estimates (Std. Err.)	Hypotheses	Supported
Opinion Distance	-0.5360 (0.0242)***		
# of Concurrent Threads	-1.3671 (0.0151)***		
# of Participants	1.5206 (0.0146)***	H3A	Yes
# of Concurrent Threads × Opinion Distance	-0.1635 (0.0066)***	H1	Yes
# of Participants × Opinion Distance	0.3055 (0.0080)***	H3B	Yes
Stock Volatility × Opinion Distance	-0.0622 (0.0079)***	H2A	Yes
Minority Status × Opinion Distance	-0.0834 (0.0074)***	H2B	Yes
Control Variables			
Thread Position (linear term)	-0.3002 (0.0068)***		
Thread Position (quadratic term)	-0.0043 (0.0007)***		
Life Span	-0.0004 ($8.6E^{-6}$)***		
Stock Volatility	0.1087 (0.2296)		
Opinion Status	0.0787 (0.000)		
Pseudo R-square	60.53%		
Observations	72,019		

*** $p < .01$

Appendix for Chapter 4

Figure 4.1: An Example of "What Do Customers Ultimately Buy After Viewing This Item?" on Amazon.com

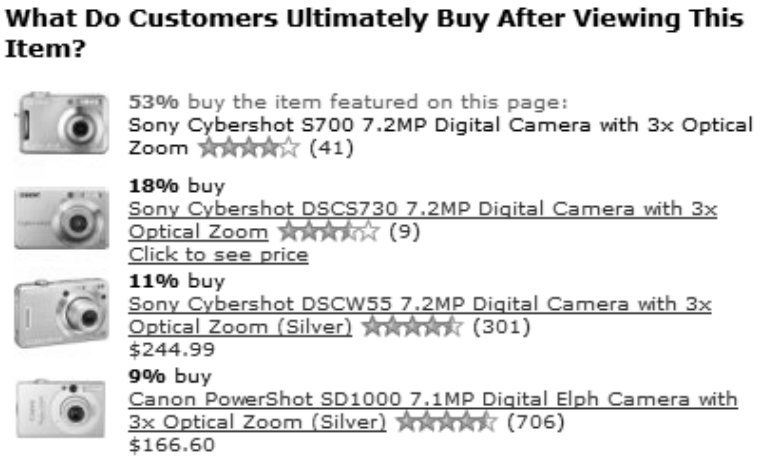

Figure 4.2: Histogram of Probabilities of Different Consideration Set Sizes

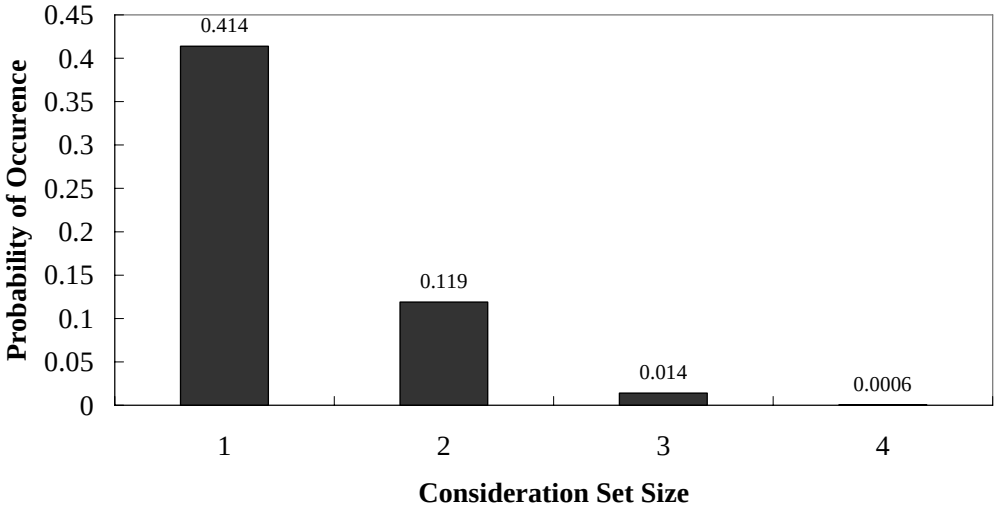

Figure 4.3: Distribution of Consideration Probabilities

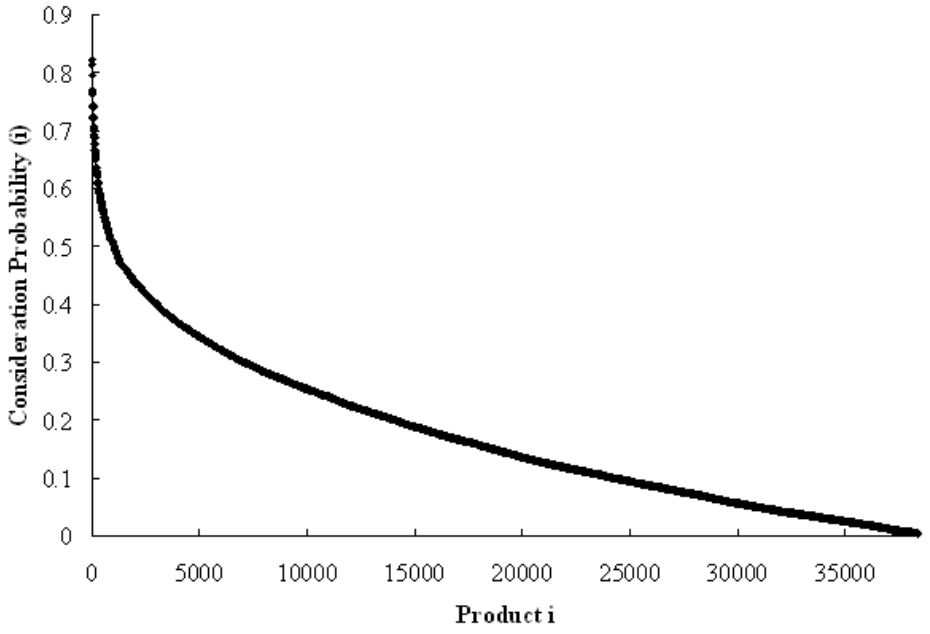

Figure 4.4: Scatter Plot of Consideration Probabilities vs. Product Utilities

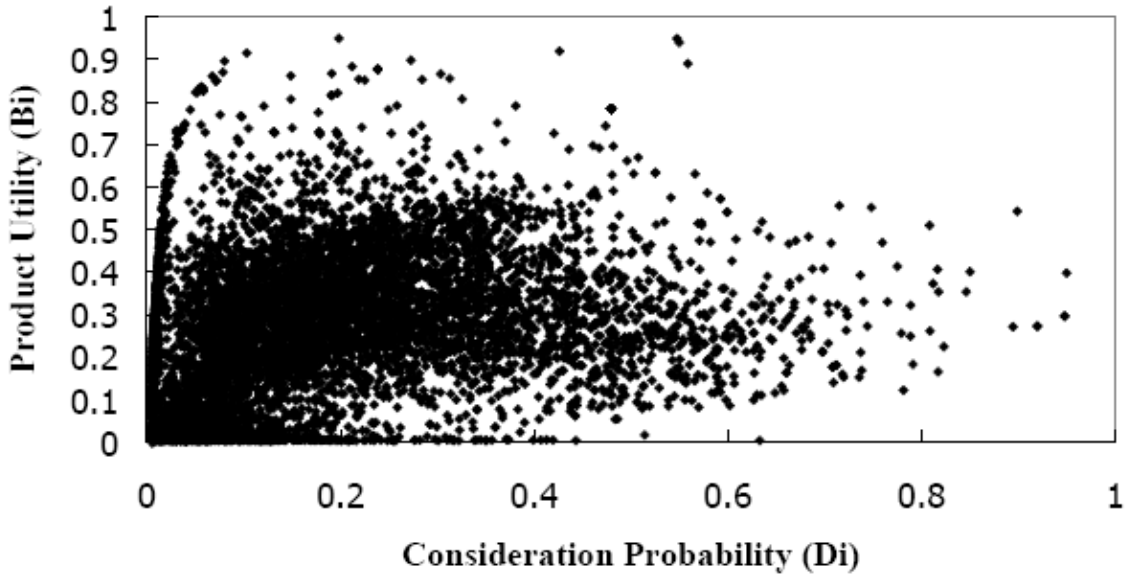

Figure 4.5: Average Consideration Probability for Each Product Rank

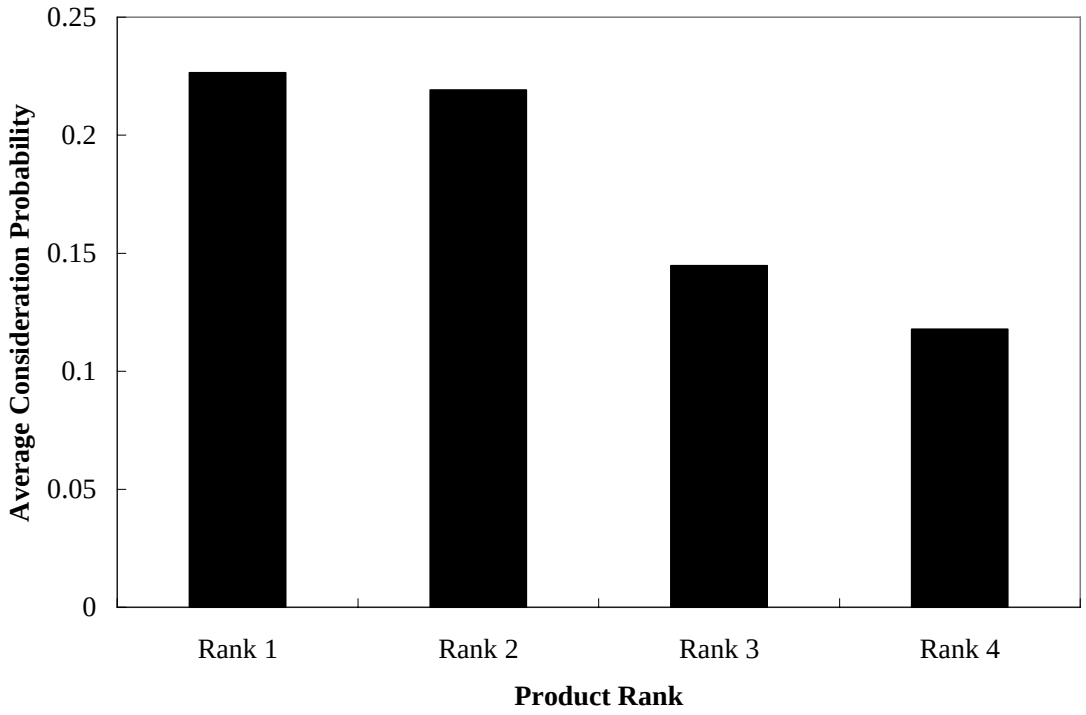

Figure 4.6: Average Purchase Probability (from Product Utility) for Each Product Rank

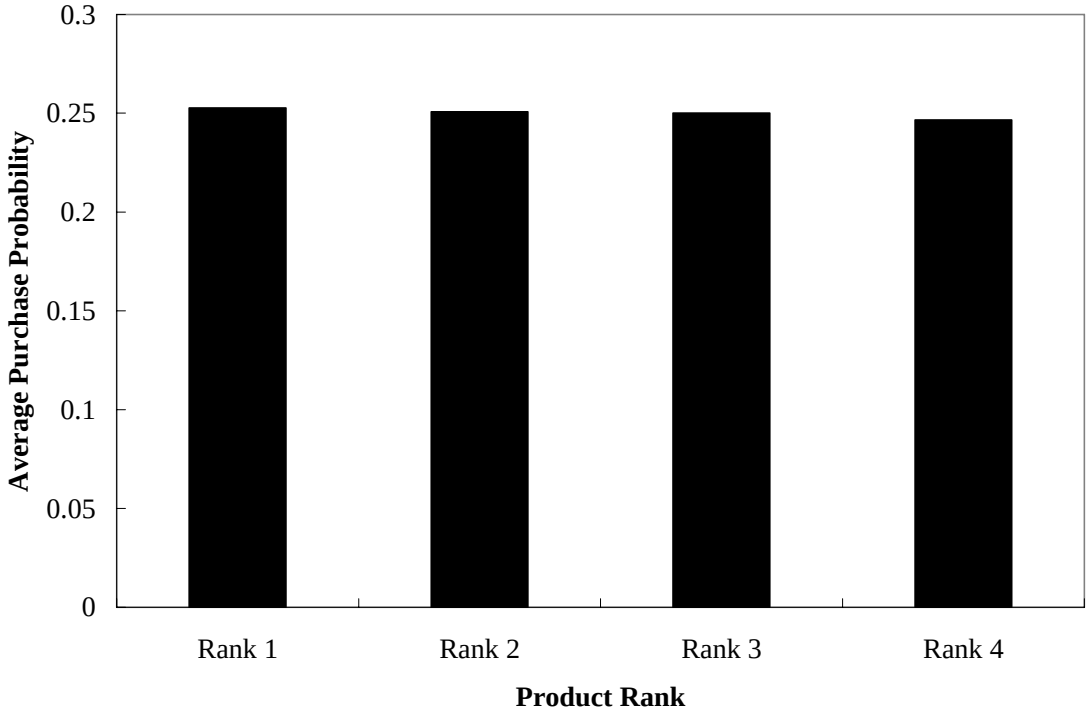

Figure 4.7: Average Purchase Probability (from Product Quality) for Each Product Rank

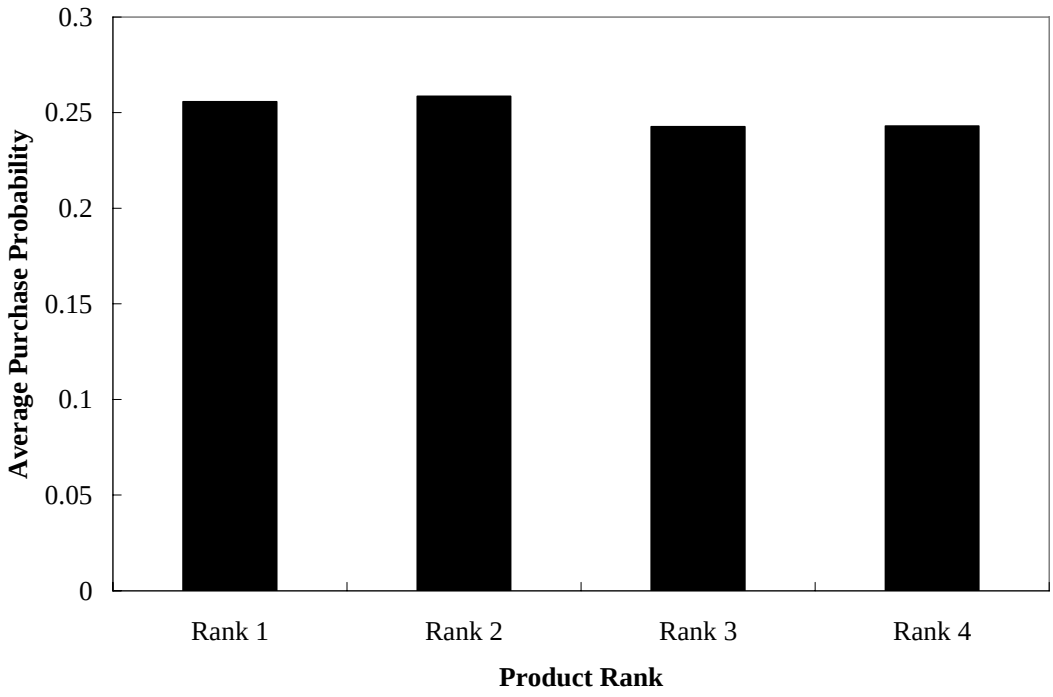

Figure 4.8: Estimated Demand Distribution

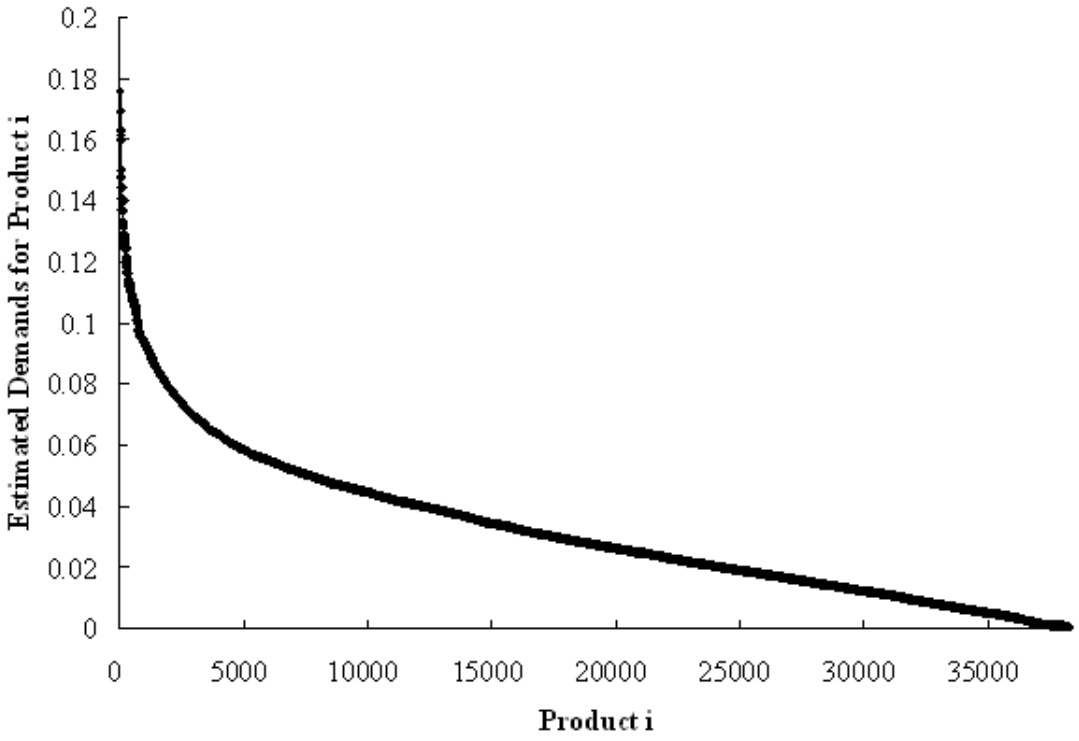

Figure 4.9: Side-By-Side Comparison of Average Probabilities from Product Consideration, Product Utility and Product Quality

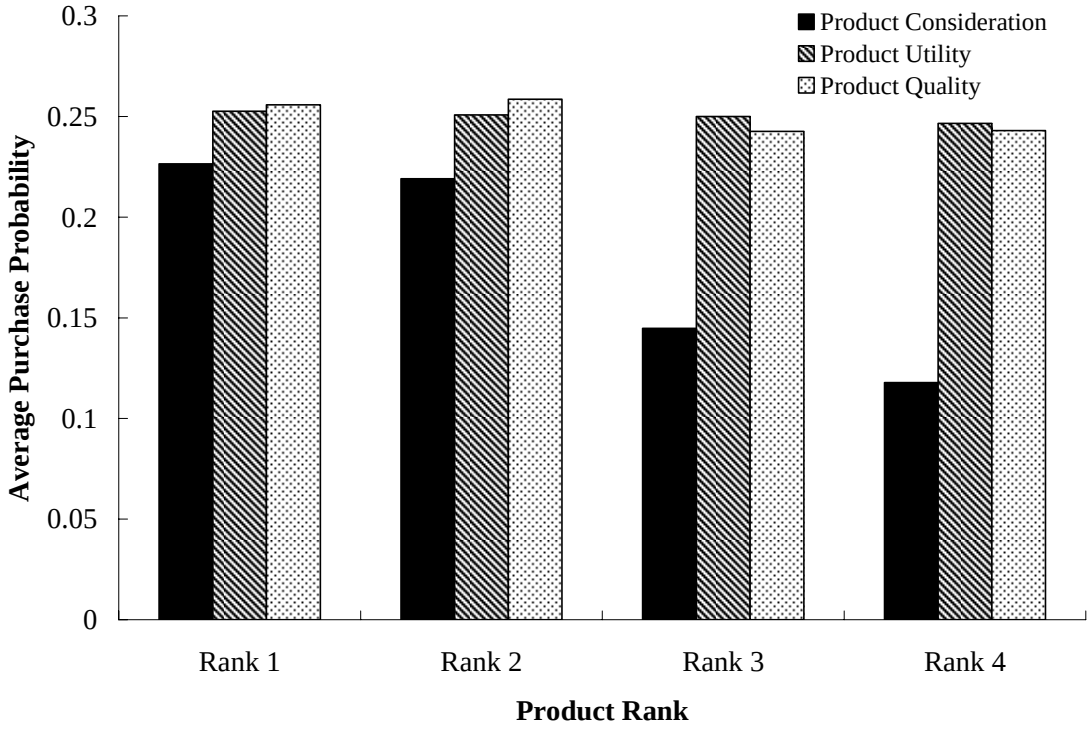

Table 4.1: The Conditional Probabilities of Buying "Row" Products Given Having Considered "Column" Products

Percentage	Sony S700	Sony DSCW	Canon A570	Canon A560
Sony S700	66%	17%	7%	6%
Sony DSCW	7%	78%	5%	0.5%
Canon A570	0.5%	0.5%	81%	7%
Canon A560	0.5%	0.5%	25%	66%

Table 4.2: Summary Statistics of Data

Variable	Observation	Mean	Std dev	Min	Max
		Period 1			
Sales Rank	38,400	345.49	3,337.75	1	249,242
Sale Price	38,400	116.40	209.59	0.01	2,799.77
Rank 1 Purchase Propensity (C1)	38,400	69.90%	0.14	25%	99%
Rank 2 Purchase Propensity (C2)	38,400	12.28%	0.11	0.5%	45%
Rank 3 Purchase Propensity (C3)	38,400	5.42%	0.05	0.5%	32%
Rank 4 Purchase Propensity (C4)	38,400	3.28%	0.03	0.5%	17%
		Period 2			
Sales Rank	38,400	356.11	2817.74	1	152,754
Sale Price	38,400	118.43	223.45	0.01	2,599.00
Rank 1 Purchase Propensity (C1)	38,400	70.17%	0.14	27%	99%
Rank 2 Purchase Propensity (C2)	38,400	12.27%	0.11	0.5%	45%
Rank 3 Purchase Propensity (C3)	38,400	5.43%	0.05	0.5%	33%
Rank 4 Purchase Propensity (C4)	38,400	3.37%	0.04	0.5%	20%

Table 4.3: Estimation Results of A Sample Product Group

Parameters	Estimates (std err)
Consideration Probabilities	
$D_{\text{Sony S700}}$	0.126 (0.051)**
$D_{\text{Sony DSCW}}$	0.580 (0.140)**
$D_{\text{Canon A570}}$	0.060 (0.024)**
$D_{\text{Canon A560}}$	0.020 (0.007)**
Product Utilities	
$B_{\text{Sony S700}}$	0.206 (0.011)**
$B_{\text{Sony DSCW}}$	0.055 (0.002)**
$B_{\text{Canon A570}}$	0.013 (0.008)**
$B_{\text{Canon A560}}$	0

**: $p < .05$

Table 4.4: Summary of Estimations of Consideration Probability and Product Utility

Parameter	Mean	Std dev	Min	Max
Consideration Probability (D_i)	0.184	0.156	0.004	0.950
Relative Product Utility (B_i)	0.279	0.151	0.001	0.942

Table 4.5: Estimation of a's and b For Product Utility: An Example of Digital Cameras

		MLE Estimates (std err)	
Product	**Product Utility (B)**	***A***	***B***
Sony S700	0.265	0.301 (0.0023)**	0.0004 (0.00012)**
Sony DSCW	0.091	0.125 (0.0014)**	
Canon A570	0.094	0.186 (0.0012)**	
Canon A560	0.006	0.089 (0.0036)**	

**: $p < .05$

References

Anderson, C. 2006. *The Long Tail: Why the Future of Business Is Selling Less of More*. New York: Hyperion.

Andrews, R. L., T. C. Srinivasan. 1995. Studying consideration effects in empirical choice models using scanner panel data. *Journal of Marketing Research* 32(1) 30-41.

Antweiler, W., M. Z. Frank. 2004. Is all that talk just noise? The information content of Internet stock discussion Boards. *Journal of Finance* 59(3) 1259-1294.

Asvanund, A., K. Clay, R. Krishnan, M. D. Smith. 2003. An empirical analysis of network externalities in peer-to-peer music sharing networks. *Available at SSRN: http://ssrn.com/abstract=433780*

Barber, B. M., T. Odean. 2001a. Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics* 116(1) 261-292.

Barber, B. M., T. Odean. 2001b. The Internet and the investor. *Journal of Economic Perspectives* 15(1) 41-54.

Barber, B. M., T. Odean. 2002. Online investors: Do the slow die first? *Review of Financial Studies* 15(2) 455-488.

Baumeister, R. F., B. J. Bushman. 2007. *Social Psychology and Human Nature*. Wadsworth Publishing.

Baye, M. R., J. Morgan, P. Scholten. 2004. Price dispersion in the small and in the large: Evidence from an Internet price comparison site. *Journal of Industrial Economics* 52(4) 463-496.

Birchler, U., M. Bütler. 2007. *Information Economics*. Routledge.

Broadbent, D. 1958. *Perception and Communication*. London: Pergamon Press.

Brown, J. R., A. Goolsbee. 2002. Does the Internet make markets more competitive? Evidence from the life insurance industry. *Journal of Political Economy* 110(3) 481-507.

Brynjolfsson, E, Y. Hu, M. D. Smith. 2003. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science* 49(11) 1580-1596.

Brynjolfsson, E, A. A. Dick, M. D. Smith. 2006a. Search and product differentiation at an Internet shopbot. *MIT Sloan Working Paper No. 4441-03*.

Brynjolfsson, E, Y. Hu, M. D. Smith. 2006b. From niches to riches: The anatomy of the Long Tail. *Sloan Management Review* 47(4) 67-71.

Burger, J. M. 1989. Negative reactions to increases in perceived personal control. *Journal of Personality and Social Psychology* 56(2) 246-256.

Butler, B. S. 2001. Membership size, Communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research* 12(4) 346-362.

Chen, P.-Y., L. M. Hitt. 2002. Measuring switching costs and the determinants of consumer retention in Internet-enabled businesses: A study of the online brokerage industry. *Information Systems Research* 13(3) 255-274.

Chevalier, J., A. Goolsbee. 2003. Measuring prices and price competition online: Amazon and Barnes and Noble. *Quantitative Marketing and Economics* 1 203-222.

Chevalier, J. A., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* 43(3) 345-354.

Clay, K., R. Krishnan, E. Wolff. 2001. Price and price dispersion on the web: Evidence from the online book industry. *Journal of Industrial Economics* 49(4) 521-539.

Clemons, E., G. Gao, L.M. Hitt. 2006. When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems* 23(2) 149-171.

Das, S. R., M. Y. Chen. 2001. Yahoo! for Amazon: Sentiment parsing from small talk on the web. EFA 2001 Barcelona Meetings. *Available at SSRN: http://ssrn.com/abstract=276189*

Deci, E. L., R. M. Ryan. 1987. The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology* 53(6) 1024-1037.

Dellarocas, C. 2003. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science* 49(10) 1407-1424.

Dellarocas, C., X. Zhang, N. F. Awad. 2008. Exploring the value of online product ratings in revenue forecasting: The case of motion pictures. Forthcoming at *Journal of Interactive Marketing*.

DeMarzo, P. M., D. Vayanos, J. Zwiebel. 2003. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics* 118(3) 909-968.

Demski, J. 1967. An accounting system structured on a linear programming model. *Accounting Review* 42(4) 701-712.

Dholakia, U. M., R. P. Bagozzi, L. K. Pearo. 2004. A social influence model of consumer participation in network- and small-group-based virtual communities. *International Journal of Research in Marketing* 21(3) 241-263.

Doyle, J., G. Kao. 2007. Friendship choices of multiracial adolescents: Homophily, blending, or amalgamation? *Social Science Research* 36 633-653.

Duan, W., B. Gu, A. B. Whinston. 2008. Do online reviews matter?—An empirical investigation of panel data. Forthcoming at *Decision Support Systems*.

Emerging Media Dynamics Executive Reports. 2006. http://www.ipmediamonitor.com/downloads/tocmicrosoftreport1106.pdf

Feltham, G. A., J. S. Demski. 1970. The use of models in information evaluation. *Accounting Review* 45(4) 623-640.

Festinger, L. 1954. A theory of social comparison processes. *Human Relations* 7(2) 117-140.

Fleder, D. M., K. Hosanagar. 2008. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. NET Institute Working Paper No. #07-10. *Available at SSRN: http://ssrn.com/abstract=955984*

Fortune 2007. Nielsen: Facebook growth outpaces MySpace. November 15, 2007. http://bigtech.blogs.fortune.cnn.com/2007/11/15/nielsen-facebook-growth-outpaces-myspace/

Fulk, J., A. J. Flanagin, M. E. Kalman, P. R. Monge, T. Ryan. 1996. Connective and

communal public goods in interactive communication systems. *Communication Theory* 6(1) 60-87.

Gensch, D. H. 1987. A two-stage disaggregate attribute choice model. *Marketing Science* 6(3) 223-239.

Gentzkow, M. 2007. Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review* 97(3) 713-744.

Ghose, A, M. D. Smith, R. Telang. 2006. Internet exchanges for used books: An empirical analysis of product cannibalization and welfare impact. *Information Systems Research* 17(1) 3-19.

Gilovich, T. 1991. How We Know What Isn't So: The Fallibility of Human Reasons in Everyday Life. Free Press.

Godes, D., D. Mayzlin. 2004. Using online conversations to measure word of mouth communication. *Marketing Science* 23(4) 545-560.

Gu, B., P. C. Konana, B. Rajagopalan, H. M. Chen. 2007. Competition among virtual communities and user valuation: The case of investing-related communities. *Information Systems Research* 18(1) 68-85.

Hall, C. C., L. Ariss, A. Todorov. 2007. The illusion of knowledge: When more information reduces accuracy and increases confidence. *Organizational Behavior and Human Decision Processes* 103(2) 277-290.

Harrison, D. A., Price, K. H., Gavin, J. A., Florey, A. T. 2002. Time, teams, and task performance: Changing effects of surface- and deep-level diversity on group functioning. *Academy of Management Journal* 45 1029-1045.

Häubl, G., V. Trifts. 2000. Consumer decision making in online shopping environment: The effects of interactive decision aids. *Marketing Science* 19(1) 4-21.

Hauser, J. R., B. Wernerfelt. 1990. An evaluation cost model of consideration sets. *Journal of Consumer Research* 16(3) 393-408.

Hof, R. D. 2005. The Power of Us. *BusinessWeek*, June 20, 2005.

Hoyer, W. D., S. P. Brown. 1990. Effects of brand awareness on choice for a common, repeat-purchase product. *Journal of Consumer Research* 17(2) 141-148.

Isenberg, D. J. 1986. Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology* 50(6) 1141-1151.

Jiang, G., C. M. C. Lee, Y. Zhang. 2005. Information uncertainty and expected returns. *Review of Accounting Studies* 10 185-221.

Jones, Q., G. Ravid, S. Rafaeli. 2004. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research* 15(2) 194-210.

Kahneman, D., A. Tversky. 1996. On the reality of cognitive illusions. *Psychological Review* 103(3) 582-591.

Konana, P. C., S. Balasubramanian. 2005. The social-economic-psychological model of technology adoption and usage: An application to online investing. *Decision Support Systems* 39(3) 505-524.

Kuk, G. 2006. Strategic interaction and knowledge sharing in the KDE mailing list. *Management Science* 52(7) 1031-1042.

Langer, E. 1975. The illusion of control. *Journal of Personality and Social Psychology* 32(2) 311-328.

Lazarsfeld, P., R. K. Merton. 1954. *Friendship as a Social Process: A Substantive and Methodological Analysis*. In Freedom and Control in Modern Society, Morroe Berger, Theodore Abel, and Charles H. Page, eds. New York: Van Nostrand, 18-66.

Lee, A. 2002. Effects of implicit memory on memory-based versus stimulus-based brand choice. *Journal of Marketing Research* 39(4) 440-454.

Li, X., L. M. Hitt. 2007. Self-selection and information role of online product reviews. Forthcoming at *Information Systems Research*.

Lilien, G., A. Rangaswamy. 2004. *Marketing Engineering*. Trafford Publishing.

Ma, M., R. Agarwal. 2007. Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research* 18(1) 42-67.

Macy, M. W., J. A. Kitts, A. Flache, S. Benard. 2003. Polarization in dynamic networks:

A Hopfield model of emergent structure. *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers* (2003) 162-173.

McMahon, H. W. 1980. *TV Loses the "Name Game" but Wins Big in Personality*. Advertising Age (December 1) p. 54.

McPherson, M. 1983. The size of voluntary organizations. *Social Forces* 61(4) 1045-1064.

McPherson, M., L. Smith-Lovin, J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27(1) 415-444.

Media Metrix, comScore. 2008. http://ir.comscore.com/releasedetail.cfm?ReleaseID=300388.

Mojzisch, A., S. Schulz-Hardt, R. Kerschreiter, D. Frey. 2008. Combined effects of knowledge about others' opinions and anticipation of group discussion on confirmatory information search. *Small Group Research* 39(2) 203-223.

Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63(2) 81-97.

Mollica, K. A., B. Gray, L. K. Treviño 2003. Racial homophily and its persistence in newcomers' social networks. *Organization Science* 14(2) 123-136.

Mooney, R. J., L. Roy. 2000. Content-based book recommending using learning for text categorization. In *Proceedings of the 5th ACM Conference on Digital Libraries* 195-204.

Nair, H., P. Manchanda, T. Bhatia. 2006. Asymmetric peer effects in physician prescription behavior: The role of opinion leaders. *Stanford GSB Working Paper*.

Nedungadi, P. 1990. Recall and consumer consideration sets: Influencing choice without altering brand evaluations. *Journal of Consumer Research* 17 (12) 263-276.

Ostreicher-Singer, G., A. Sundararajan. 2006. Recommendation networks and the long tail of electronic commerce. *2nd Annual Statistical Challenges in Electronic Commerce Research Symposium*.

P&G report 2007. https://secure3.verticali.net/pg-connection-portal/static/external/files/cd_brochure

WEB.pdf

Portfolio.com 2008. http://www.portfolio.com/news-markets/top-5/2008/04/23/Amazon-Earnings-2008-Q1

Rabin, M., J. L. Schrag. 1999. First impressions matter: A model of confirmatory bias. *Quarterly Journal of Economics* 114(1) 37-82.

Roberts, J. H., J. M. Lattin. 1991. Development and testing of a model of consideration set composition. *Journal of Marketing Research* 28(4) 429-440.

Rogers, E. M., R. Agarwala-Rogers. 1975. *Organizational Communication*. In G. J. Hanneman & W. J. McEwen (Eds.) Communication and Behavior (pp. 218-236). Reading: MA: Addison Wesley.

Rogers, E. M., D. K. Bhowmik. 1971. Homophily-heterophily: Relational concepts for communication research. *Public Opinion Quarterly* 34(4) 523-538.

Simon, H. A. 1957. *Models of Man, Social and Rational*. John Wiley & Sons, New York.

Smith, M. D. 2002. The impact of shopbots on electronic markets. *Journal of the Academy of Marketing Science* 30(4) 442-450.

Smith, M. D., E. Brynjolfsson. 2001. Consumer decision-making at an Internet shopbot: Brand still matters. *Journal of Industrial Economics* 49(4) 541-558.

Stanford GSB News. 2007. http://www.gsb.stanford.edu/news/research/mktg_nair_drugs.shtml

Stigler, G. J. 1961. The economics of information. *Journal of Political Economy* 69(3) 213-225.

Tetlock, P. C. 2007. Giving content to investor sentiment: The role of media in the stock market. Forthcoming at *Journal of Finance*.

Turner, J. C., M. A. Hogg, P. J. Oakes, S. D. Reicher, M. S. Wetherell. 1987. *Rediscovering the Social Group: A Self-Categorization Theory*. Oxford: Blackwell.

Tversky, A., D. Kahneman. 1974. Judgment and uncertainty: Heuristics and biases.

Science. 185(4157) 1124-1131.

USA Today. 2006. *The Guys Behind MySpace.com*. USA Today, February 12, 2006. http://www.usatoday.com/money/companies/management/2006-02-12-myspace-usat_x.htm

Van Alstyne, M., Brynjolfsson, E. 2005. Global village or cyber-Balkans? Modeling and measuring the integration of electronic communities. *Management Science* 51(6) 851-868.

Wasko, M., S. Faraj. 2005. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly* 29(1) 35-57.

Whittaker, S., L. Terveen, W. Hill, L. Cherny. 1998. The dynamics of mass interaction. In *Proceedings of the Conference of Computer Supported Cooperative Work*, ACM Press, New York.

Wooldridge, J. M. 2001. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Yuan, Y., G. Gay. 2006. Homophily of network ties, bonding and bridging social capital in distributed teams. *Journal of Computer-Mediated Communication* 11(4) article 9.

Zhang, X. F. 2006. Information uncertainty and stock returns. *Journal of Finance* 61(1) 105-137.

Vita

Hsuan-Wei Chen was born on April 24, 1980 in Taipei, Taiwan to parents Horng-Shi Chen and Li-Hua Shih. After graduating from Taipei First Girls High School in 1998, she attended National Taiwan University, where she received her B.S. and M.S. degrees in Computer Science and Information Engineering in June 2002 and June 2004, respectively. In August 2004, she entered the doctoral program of Management Science and Information Systems of McCombs School of Business at the University of Texas at Austin.

Permanent address: 12F-2, 82, Sec 2, Fuhsing S Rd, Taipei, Taiwan 106

This dissertation was typed by the author.