The Economic Impact of User-Generated and Firm-Published Online Content: Directions for Advancing the Frontiers in Electronic Commerce Research

ANINDYA GHOSE

New York University

1. Introduction

An important use of the Internet today is in providing a platform for consumers to disseminate information about products they transact with as well as about themselves. Indeed, through the use of Web 2.0 tools like blogs and opinion forums, a large amount of content is being generated by users in the online world. Consequently, we have seen online markets develop into social shopping channels, and facilitate the creation of online communities and social networks. Firms are also using technology mediated spaces to reveal information about their buyers. The increasing avenues for online content creation have changed the fundamental nature of information in terms of volume, availability and importance. A lot of that publicly available content has concrete economic value that is often embedded in it. To date, businesses, government organizations and customers have not fully incorporated such information in their decision making and policy formulation processes, either because the value of the intellectual capital, or appropriate methods for measuring that value have not been identified. This book chapter is a call for research that aims to measure the economic value of various kinds of user-generated and firm-published content on the web.

As an example, a vital piece of socio-cultural information that is increasingly being published on the web is the *geographical location* of market participants and members of virtual communities. The availability of users' location information, either disclosed by users themselves or published by firms, opens up a plethora of opportunities for research examining how geographical location shapes consumer search and purchase behavior on the Internet. As another example of increasingly ubiquitous user-generated social information, members who use electronic markets as a forum for social interaction, reveal a lot of personal information describing themselves. The availability of this self-descriptive information generated by users in an online community can enable researchers to examine how economic exchanges in the online world are being influenced by social exchanges between various entities.

While the above instances are about the availability of users' personal information, there are more detailed forums that provide information about users' actual experiences with sellers or with products in rich textual format. For example, when buyers cannot deterministically assess the quality of a seller's fulfillment characteristics ex-ante in an electronic market, the textual feedback posted by buyers describing their transaction experience can influence other buyers' purchase decisions, and thus affect sellers' future performances. Similarly, based on the rich theoretical literature that suggests that consumer generated word of mouth influences product sales, we can hypothesize that textual content of user-generated reviews are also likely to influence sales. Most studies of reputation systems or online reviews so far have used only numeric information about sellers or products to examine their economic impact. The understanding that "text matters" has not been fully realized in electronic markets or in online communities. Insights derived from text mining of user-generated feedback can thus provide substantial benefits to businesses looking for competitive advantages.

At the same time, excessive content in the online world can cause information overload amongst individuals resulting in various cognitive costs incurred by users. These human-computer interaction costs include for example, search costs incurred by consumers for locating the right information, cognitive costs of processing textual information prior to making purchase decisions, and decision-making or menu costs incurred by managers for adjusting price information. These costs arise due to delays in information diffusion and are brought about by bounded rationality (limited ability to process complex information) of humans. To date, businesses have generally formulated strategies in the online world without factoring such costs. Hence, such policies can be sub-optimal. This calls attention for the need to identify and measure these costs in order to formulate optimal pricing policies.

The overarching theme across the above phenomena is that much of this user-generated and firm-published online information has an *economic value* that can be measured, monetized and intelligently utilized in formulating business strategies. Extracting this economic value from publicly available online content and leveraging it has become increasingly important for all participants in a competitive market. To identify the economic value of online content, we need to examine three related questions: (i) how does the Internet influence consumers' information-seeking and purchase behavior by providing newer distribution channels, newer forms of online advertising, and unique community forums for social exchanges; (ii) what is the economic value of user-generated content in Internet-mediated spaces such as reputation systems, review forums, and social networking sites, and (iii) how do users' information search and processing costs affect firms' pricing strategies in offline and online markets. Answering these questions requires an inter-disciplinary approach that builds upon theories and tools from multiple fields such as computer science, economics, information systems, machine learning, marketing, social psychology, and statistics to measure how various categories of content on the Internet influence exchanges between participants in digital markets and online communities.

Sections 2-5 constitute the main body of the paper. Section 2 discusses the opportunities in measuring the economic value information on consumers' information-seeking and purchase behavior in electronic markets. This kind of information is embedded in both user-generated and firm-published content. Section 3 discusses the economic value of user-generated textual feedback that is ubiquitous on the Internet such as in reputation systems in electronic markets, product reviews in online communities, product descriptions in used-good markets, and social networking sites. We also discuss some methodologies that could be used to estimate that value. Section 4 analyzes the economic cost of information consumption such as search costs of finding information and the costs of processing textual information incurred by consumers, as well as costs of adjusting product information on electronic markets incurred by firms. It also discusses the impact of search costs and menu costs on the emerging Long Tail phenomenon. In each of these sections, we describe some research opportunities that can build on current work. Section 5 concludes.

2. Consumers' Information-Seeking and Purchase Behavior

The Internet has been thought of as a technological advancement that removes the disparities between underserved communities and the rest of the country. However, we have little understanding of whether the benefits of the online channel (due to increased convenience, wider selection, and lower prices) are influenced by the concentration of offline retailers that vary across geographical locations. Similarly, knowledge about how different kinds of online advertising affect consumers' search and purchase behavior is still in its infancy. The emergence of natural and sponsored search keyword advertising is intrinsically related to user-generated queries on search engines. By examining how keyword attributes and user-level demographics affect user search and purchases, one can estimate the business value of search engine advertising. Finally, by exploring the behavior of members in online virtual communities, future research can potentially examine the dynamic interplay between social and economic exchanges on the Internet. The research opportunities described in this section are based on the notion that an analysis of content that specifies social information of users can increase our understanding of the factors that drive consumer usage of online channels.

2.1 Geographical Location and Online Purchase Behavior

Despite a wealth of research on electronic commerce, very little work has measured how geographical location shapes consumer buying behavior in electronic markets. Do consumers in different locations derive different benefits from using the Internet in terms of selection, convenience and price? Prior studies in this domain have examined substitution between online and offline channels analytically (for example, Balasubramanian 1998, Ghose, Mukhopadhyay and Rajan 2007), and empirically (for e.g., Goolsbee and Brown 2003, Ellison and Ellison 2006, Prince 2006). Since prior work has focused on price differences across channels, future research can examine how changes in offline shopping convenience and product assortments can influence online purchasing behavior. By combining online purchase data with offline data on demographics and availability of local retail channels, one can address an important problem that has been inadequately addressed in statistical e-commerce research: how and why consumers substitute between online and offline channels.

Recent research that analyze this question include Forman, Ghose and Goldfarb (2007) who use data from the web pages on "Purchase Circles" on Amazon where Amazon publishes the geographical locations of its buyers. Purchase Circles are specialized best-seller lists that denote the top-selling books, music, and DVDs across large and small towns throughout the US. The Purchase Circles are organized in multiple layers - first, by state and then within a state, by town and by county. Forman, Ghose and Goldfarb (2007) match this online data on local demand with that on store openings and closings of major offline competitors of Amazon that include discount stores such as Walmart, Target, and large specialty stores such as Barnes & Noble and Borders in order to study how geographic variations in offline selection, convenience and price influence online product purchases. Their findings confirm that consumers do derive considerable benefits from convenience and price. Evidence for selection is demonstrated only for University towns and larger cities. Their results provide empirical support for the assumptions of a widely used framework in models of spatial differentiation that include a direct channel (Balasubramanian 1998). They find that variables and parameters in these models such as offline transportation cost, online shopping disutility cost, market coverage, and the prices of online and offline retailers interact to determine consumers channel choice in a way that is consistent with these models. Moreover, their results are suggestive about the relative magnitudes of some of these parameters, showing that online disutility costs can be significant, even for products such as books for which non-digital attributes are relatively unimportant. By looking at how consumers use online channels to compensate for offline retail supply deficiencies, their research contributes to the marketing literature that uses spatial data to capture variations in supply-side and demand-side factors—such as local consumer preferences (Jank and Kannan 2006).

Future analytical modeling research can use the findings of this research. In particular, the results from Forman, Ghose and Goldfarb (2007) suggest the usefulness of incorporating the effect of varying offline transportation costs in making optimal product assortment decisions for commodity products in local as well as online stores, and incorporating the effect of product popularity in modeling the impact of product returns on retailers' pricing decisions since the cost of returns to retailers and to consumers are likely to vary by product popularity and distance to stores, respectively.

Further research is needed in this domain to increase our understanding of how geography influences the benefits consumers derive from the Internet. To do so, one needs to look at more disaggregated data such as how individual consumer transactions vary across locations. This can highlight how offline geographical distance between buyers and sellers affects their propensity to transact with each other in online markets. Data on demographic characteristics can be obtained from the Bureau of Labor Statistics (BLS), and matched with the transaction level data. For example, data on population size by county or Metropolitan Statistical Area (MSA) are provided annually by the Census, while demographic data are available at various levels of aggregation in the decennial Census. This can shed further light on how demographic factors contribute to differences in online purchase behavior (Scott-Morton, Zettelmeyer and Risso 2006). Such research will contribute to the literature on the potential of the Internet to reduce the costs associated with distance (Forman et al. 2004, Sinai and Waldfogel 2004) and increase consumer welfare (Brynjolfsson, Smith, and Hu 2003, Bapna, Jank and Shmueli 2006, Ghose, Smith, and Telang 2006).

2.2 Web Search, Online Search Advertising and Social Networks

There exists a vast body of literature that has used click-stream data to study consumer behavior in the online world (Bucklin et al. 2002). Similarly, there has been an emerging body of literature that has studied incentive mechanisms and auction strategies in key word search auctions using game theory and computational methods. However, there exists a lot of potential for research that skates the boundaries of these two streams. Specifically, one can investigate the correlation between Internet search queries, the associated click-through rates and online sales for different time periods, different categories of products and more importantly for consumers in different locations. A more rigorous analysis would involve building predictive econometric models that will take the research one step towards identifying causality between these phenomena.

The data that are needed for these kinds of studies can be obtained from firms that advertise on search engines such as Google, MSN or Yahoo. An ideal dataset would consist of search queries sampled over several weeks, bid prices and per-query search result click-through rates with product and consumer demographic information. In any economic model required for this study, one needs to incorporate the fact that consumers face decisions at two levels – first, when they receive the results of a search engine query, they make a decision whether or not to click on it; two, if they click on a result that is displayed, they can take any one of the following actions–take no action, make a click-through (without making a purchase) or make a purchase. Research in this domain can use a Hierarchical Bayesian modeling framework and estimate the model using Markov Chain Monte Carlo (MCMC) methods (Rossi and Allenby 2003).

Ghose and Yang (2007) provide a first step in this direction. Using a unique panel dataset of several hundred keywords collected from a large nationwide retailer that advertises on Google, they empirically model the relationship between different metrics such as click-through rates, conversion rates, bid prices and keyword ranks. They estimate the impact of keyword attributes on consumer search and purchase behavior as well as on firms' decision-making behavior on bid prices and ranks. They find that the presence of retailer-specific and brand-specific information in the keyword increases click-through rates and conversion rates. Moreover, their analysis provides some evidence that advertisers are not bidding optimally with respect to maximizing the profits. They also demonstrate that as suggested by anecdotal evidence, search engines like Google factor in both the auction bid price as well as prior click-through rates before allotting a final rank to an advertisement. Finally, they conduct a detailed analysis with product level variables to explore the extent of cross-selling opportunities across different categories from a given keyword advertisement.

Additionally, if one were to have access to historical data on online sales for a variety of goods, one would be able to design predictive models by merging pricing and sales data at electronic retailers with the Internet search behavior data. Estimation can be challenging in this area because of data sparsity issues and rarity of clicks. Machine learning and statistical techniques that predict imbalanced or rare responses by sampling the majority class to reduce imbalances can be useful in this context (Chawla et al. 2003, King and Zeng 2001). This research can be extended to investigate how consumers' physical distance from retailers influences their online search and click-through behavior on search engines. The additional data needed for this analysis are the locations of the originating search which is available from advertisers. Future research in this domain can incorporate user's geographical locations to help identify how online advertisements should be customized and targeted across locations. In the business world, this kind of research would be of interest not just to advertisers but also to search engine marketing (SEM) firms.

Another interesting research direction would be to examine the relationship between the textual content in search queries, and economic variables such as bid prices in keyword auctions, bid slots, page numbers and click-through rates. Indeed, one can examine the economic value of various modifiers (basically adjectives or adverbs) used by consumers in online queries and search keywords. This will involve the analysis of the price premiums on keyword bid prices by examining how much the price of a bid change with the addition of a specific modifier. For example, this would involve comparing the difference in bid prices, click-through rates, and conversion rates for a generic keyword such as "*airline tickets*" with a branded keyword like "*Orbitz airline tickets*" or "*discounted airline tickets*" to determine the premium that online advertisers need to pay for a branded advertisement or for a keyword that contains a modifier like "*discounted*". This study will involve a combination of text mining with economics to infer the economic value of text in keywords and search queries. Research in this domain is in its infancy and so there is a lot of potential to address these open questions.

Finally, an emerging area of research that offers many interesting possibilities is that of *Social Search*. Firms are increasingly looking for ways to combine information from social networking sites to improve the quality and accuracy of search with the final aim of providing personalized search. Microsoft has made some headway in this area. The popularity of web-based social networks like Delicious, Technocrati, Flickr, etc. which allow users to tag resources like blog, pictures and webpages, generates a rich data trail that can potentially be exploited to improve and broaden search quality. A natural consequence of improving search quality can be an increase in revenues from sponsored advertisements. From the academic research point of view, a combination of machine learning and statistical techniques can be combined with techniques in keyword generation to determine the most valuable keywords at the individual level to enable increased precision in geo-targeting and contextual advertisements during web search.

2.3 User-Generated Social Information and Economic Exchanges

The Internet has had a profound impact on at least three areas of life – the way people shop, the way they socially interact, and the way they exchange information. All three are relevant to consumer product reviews posted in IT-enabled electronic markets. Online consumer product reviews provide information that can facilitate economic exchange, which is the central function of electronic markets. However, reviews are also sometimes used as a forum for social exchanges, and this, in part, serves to draw people to such websites, promote purchases, and regulate user behavior. Hence, more research is needed that will explore the important role of social communities in geographically-dispersed electronic markets. Some participants in electronic markets also use it as a forum for social interaction, and, in the process, reveal a lot of information about themselves. Drawing on theories from social psychology such as social identity theory and information processing theory, an interesting arena of research is to investigate whether online users view consumer reviews as a forum to form social identities, and if these social exchanges influence economic outcomes in consumer settings.

In the past, product reviewers had limited opportunities to convey community affiliation because product reviews were nominally focused on the product itself rather than the reviewer. Recent design changes in electronic markets now enable members, who identify with the community, to engage in self-disclosure of social information. For example, on sites such as Amazon information about product reviewers is graphically depicted, highly salient, and sometimes more detailed than information on the products they review. Specifically, it allows reviewers to publicly reveal their *real name, geographical location, professional interests and hobbies*. Such user generated self-descriptive information can be explained in social psychology as members' attempts to convey the "social identity" they wish others to associate with them. If on-line self-disclosure is driven in part by the desire for identification with a community and the need for self-verifying feedback from other community members, then reviewer identity expressions should be patterned to follow community norms. Hence, norm conformity is evidence of an investment on the part of an individual contributor signaling that one would like to be viewed as a member of the community (Bartel and Dutton 2001). In particular, if the types or categories of information that reviewers disclose are consistent with the type of information that is typical or normative in the community, this is suggestive evidence that identification processes can be an important antecedent to reviewer disclosure. Such theories can now be empirically tested with the kind of data available on online review and social networking sites. Furthermore, given the extent and salience of social information regarding product reviewers, it is also possible to inquire whether such information has an influence on the online consumers who are responsible for product sales.

Forman, Ghose and Wiesenfeld (2006) explore some of these phenomena. They estimate econometric models using a panel data set consisting of data on chronologically compiled reviews on these set of products, and the various self-descriptive, social, and personal, information of the reviewers. Using research on information processing (Chaiken 1980), they suggest that in the context of an online community, the social information reviewers provide about themselves is used to supplement or replace product information in shaping community members' purchase decisions and the value they attribute to online reviews. Online community members rate reviews containing identity-descriptive information more positively, but the informative value of reviews attenuates the relationship between reviewer disclosure of identity-descriptive information and community ratings relative to more equivocal reviews. Furthermore, they show that the prevalence of reviewer disclosure of identity information in online reviews is associated with increases in subsequent online product sales after controlling for the valence and volume of reviews.

Future research in this domain can extend an emerging stream of academic work on the relationship between reviews and economic outcomes (for example, Dellarocas et al. 2005, Reinstein and Snyder 2005, Chevalier and Mayzlin 2006), the work exploring motivations for people to post word-of-mouth (for example, Hennig-Thurau et al. 2004), and the work on belonging to a community (for example, Cummins, Butler and Kraut 2002). While Forman, Ghose and Wiesenfeld (2006) do not directly assess the form of information processing Amazon members used, their results are consistent with the notion that people use more heuristic processing of source characteristics when information overload is high. Future research can evaluate whether the number and diversity of messages influence recipients' response to source and message characteristics. While initial research on common bond and common identity suggested that groups could be characterized as *either* common bond or common identity (Prentice et al 1994), subsequent work suggests that the formation of common bonds and common identities may be related to one another (Ren et al. 2007). Thus, future research may consider whether common bonds promote common identity in online communities like Amazon.

This research stream can be combined with *sentiment analysis techniques* from computer science to investigate if sentiment patterns in these reviews can be attributed to the disclosure of self-descriptive information by specific reviewers, and how such reviews are rated by the community. Text mining techniques similar to Kim and Hovy (2004) can be used to automatically find reviewers who hold opinions about a topic. The reviews will be classified as subjective (opinionated) vs. objective (neutral) using techniques such as those in Riloff and Wiebe (2003). In theory, subjective reviews may be rated by the community as less helpful than objective reviews because they provide less useful information to guide purchase decisions. However, this may be less true when the reviews contain more disclosure of personal information because helpfulness ratings of those reviews may be partly driven by the desire to grant membership status. Such propositions can now be empirically tested using the rich trail of data.

3. Economic Value of User-Generated Textual Information

The research described in this section is based on the conjecture that the qualitative information contained in text-based feedback, online reviews of products or descriptions of used goods on the Internet, plays a substantial role in influencing social and economic outcomes in electronic markets. Such studies can combine empirical tools such as state-of-the-art techniques in hedonic modeling, multivariate consumer choice modeling and panel data model, with automated text mining techniques to identify and quantify the impact of this information on economic variables such as revenues, price premiums and resale rates. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others), deriving patterns within the structured data, and finally evaluation of the output. The main goal is to discover patterns and trends in the information contained in textual documents (Hearst 1999, Grobelnik et al. 2001).

3.1 Feedback in Reputation Systems and Economic Impact

When buyers purchase products in an electronic market, they assess and pay not only for the product they wish to purchase, but also for a set of fulfillment characteristics such as packaging, timeliness of delivery, responsiveness of the seller, and reliability of settlement. Most studies of online reputation so far have based a seller's reputation on the numerical rating that characterizes the seller and the level of experience (Dellarocas 2003, Resnick et al. 2006). This implies that there is big potential for future research on reputation systems based on the analysis of text-based feedback posted by buyers. Such feedback describes their transaction experience with the sellers and it seems natural that it will play an important role in establishing reputation in electronic markets, since different sellers in these markets derive their reputation from different characteristics. A comment about "*super-fast delivery*" can enhance a seller's reputation and thus allow it to increase the price of the listed items by a few cents, without losing any sales. On the other hand, the feedback about "*sloppy packaging*" can have the opposite effect on a seller's pricing power. To accomplish this study, one needs information on the transaction price at which the good was sold, the date of sale, all relevant details for competing offers at the time of sale, the number of such used goods listed and corresponding price premiums from an electronic market for used good exchanges. Prior research (Ghose 2006, Ghose, Smith and Telang 2006) has demonstrated a novel way of extracting such information using data from Amazon's used-good marketplaces.

Broadly speaking, this kind of research that analyzes the economics of textual content needs to combine established techniques from econometrics with text mining algorithms from computer science to identify the "value of text" and assign an economic value to each feedback posting, measuring sentiment effectively and without the need for manual labeling of postings. Ghose, Ipeirotis and Sundararajan (2005) is the first known study to demonstrate the value of some of these automated textmining methods. The text mining techniques used in their study are based on research on opinion extraction in the computational linguistics community. The first step in the algorithm consists of parsing the feedback postings to identify the dimensions across which the buyer evaluates a seller. For this task, a part-of-speech (POS) tagger (such as the Stanford Java NLP tagger) is used for each word. The nouns, noun phrases, and verbal phrases are kept as the dimensions of the seller. Then the adjectives and adverbs that refer to the nouns and verbs extracted in the previous step are retrieved, similar to Turney (2002). For instance, Hatzivassiloglou and McKeown (1997) use a supervised learning technique to identify the *semantic orientation* of adjectives. To associate the adjectives and adverbs with the correct dimensions, they use a syntactic parser. The adjective-noun and adverb-verb pairs serve as the basis for further analysis in their paper. Next, using panel data methods, they estimate econometric models that show that textual feedback adds value to the numerical scores, and affects the pricing power of sellers. That is, merchants with negative comments charge lower prices, and merchants with positive comments command higher premiums than their competitors. They show the emergence of a number of unique dimensions of seller reputation (e.g., *shipping, packaging, delivery, responsiveness* and so on), and corroborate that a substantial pricing power is associated with each dimension of seller reputation.

Their research also contributes to an emerging stream of work that examines sentiments in online communities and auctions (Gu et al. 2007, Pavlou and Dimoka 2006). There exists tremendous potential for future work in this area. For example, using data from eBay one can compare how textual feedback in a reciprocity based reputation system (eBay) differs from that in a market where only buyers rate sellers (Amazon). This is important because a reciprocity-based reputation system has a decreased likelihood of eliciting honest feedback from buyers due to the potential for retaliation by sellers in the case of a negative comment. This opens up the possibility of feedback manipulation. Dellarocas (2005) provides a nice theoretical modeling of this practice but more empirical research is needed to corroborate such practices. This is an area where the combination of econometrics with text mining methods can open up a myriad of research opportunities.

3.2 Sentiments in Product Reviews and Economic Impact

Prior work has shown that the volume and valence of online product reviews influences product sales such as books and movies (Godes and Mayzlin 2004, Dellarocas et. al. 2005, Chevalier and Mayzlin 2006). However, this research did not account for the impact of textual content in those reviews. Similarly, computer scientists have conducted extensive sentiment analysis of reviews and news articles (Wiebe 2000, Pang, Lee and Vaithyanathan 2002, Turney 2002, Dave, Lawrence and Pennock 2003, Hu and Liu 2004, Liu, Hu and Cheng 2005, Kim and Hovy 2004, Popescu and Etzioni 2005) but have not examined their economic impact on sales.

This presents an opportunity for future research to bridge the gap between these two streams of literature. One can examine how sentiment embedded in online reviews performs as a predictor for future demand. To do this, one needs to examine consumer sentiment in online markets. Prior studies have measured the predictive power of word-of-mouth on sales, but have not used a combination of sentiment analysis and econometric experiments to extract more precise information from the text. Ghose and Ipeirotis (2006) conduct an initial study in this domain. After training the classifiers appropriately using technical product descriptions provided by manufacturers, they automatically classify reviews as subjective (highly opinionated) or objective. Then, they measure their impact on revenues at Amazon. This enables them to empirically quantify the effect of textual information in online product reviews on sales in various product categories as well as measure how the extent of subjectivity in the opinions affects consumers' perceptions of review informativeness as measured by peer provided "helpful votes". A limitation of their study is that they only have crude measures for changes in product demand. Future research can use actual data on demand from an online retailer to do a more refined analysis. There is a lot of potential for future researchers to work on ways to score the utility of consumer reviews with a focus on extracting the various features of a review and measuring their impact on sales or informativeness of reviews. Moreover future research can examine the usefulness of data from blogs and social networking sites in predicting sales. Dhar and Chang

(2007) use data from MySpace and Amazon to track the changes in online chatter for music albums before and after their release dates. Such studies can provide insights to marketing managers interested in assessing the relative importance of the burgeoning number of "Web 2.0" information metrics.

3.3 Dimensions of Product Descriptions and Economic Impact

Although e-commerce enables easier search as far as new products are concerned, such standardized search has not yet been implemented in used-good markets because of the diversity in seller or product characteristics. Whereas attributes such as product features can be communicated easily in electronic markets, "non-digital" attributes (product condition and seller integrity) are subject to noise and manipulation. This has the potential to create information asymmetry between buyers and sellers, stemming from the unobservability of quality signals in electronic markets. This informational asymmetry is associated with both an individual seller's reputation and the product's self-reported quality. Hence, in such used-good markets, asymmetric information can lead to market failure such as adverse selection (Akerlof 1970). This failure manifests itself in the fact that sellers with relatively high quality goods need to wait longer than sellers of low quality goods in order to complete a trade (Janssen and Karamychev 2002). Sellers try to minimize this information asymmetry by using textual descriptions of the products such as *pristine condition* or *factory sealed to* signal quality *before purchase*. To what extent can standardized textual descriptions of used goods provided by sellers prior to purchase alleviate these information asymmetries?

To study these questions one can combine automated text mining techniques with econometric methods to infer the dimensions of used-product description that consumers value the most, and quantify the price premiums associated with each dimension. Specifically, using similar text mining methods as in (Ghose, Ipeirotis and Sundararajan 2005), one can identify features of the used good descriptions that mitigate the information asymmetry problem between buyers and sellers, by identifying the *dimensions of product description* that maximize price premiums, resale rates and revenues in used-good markets. Future work on the economics of textual descriptions in used-good markets can build on methods deployed in prior work (Ghose, Telang and Smith 2006) to advance our understanding of the impact of Internet-based exchanges on social welfare. They can also shed light on how such used good descriptions can alleviate the information asymmetry problem in electronic markets as found in Dewan and Hsu (2004) and in Ghose (2006). By quantifying the economic value of sentiment in online reviewer communities, such studies will provide methods that can be extended to quantify the economic impact of textual content on social networks, blogs, and social shopping sites as well as in the emerging practice of "tagging" products.

4. Economic Cost of Information Search, Processing and Modification

To monetize the economic impact of firm-published or user-generated content, a key analysis that needs to be conducted is to estimate the cost of processing and disseminating information during various human-computer interactions. One of the main advantages that Internet channels posit over physical markets is a reduction in "friction" costs. These costs include search costs incurred by consumers for locating product-related and price-related information (Bakos 1997), decision-making costs incurred by managers to adjust prices in response to market conditions (Levy et al. 1997) and cognitive costs of processing textual content in product descriptions incurred by users. There is a lot of potential for research that estimates the cognitive costs incurred by consumers when searching for information or processing textual information online, and the cost of decision-making incurred by managers while adjusting product-level information. This section describes some opportunities in estimating different kinds of cognitive costs incurred by users in the online world.

4.1 Consumer Search Costs

The literature on "rational inattention" argues that observing, processing, and reacting to price change information is not a costless activity. An important implication of rational inattention is that consumers may rationally choose to ignore--and thus not to respond to--small price changes, creating a "range of inattention" along the demand curve (Levy et al. 2006). This applies to online markets too wherein consumers face cognitive search costs of processing the multitude of information published on the Internet (Johnson et al. 2004). This cost includes visiting multiple retailers who differ on various attributes, comparing a diverse set of offerings, and assessing the overall quality of their offerings. Moreover, opportunities for obfuscation (Ellison and Ellison 2006), and the unobserved lack of inventories (Arnold and Saliba 2002) can also create such costs. Indeed, the existence of price dispersion in online markets (Brynjolfsson and Smith 2000, Baye, Morgan and Scholten 2004) has often been attributed to the presence of search costs and retailer menu costs. Unless retailers factor these behavioral aspects into their strategies, market transparency and prices can be suboptimal.

While the presence of search costs and its impact on consumer demand and retailer strategies has been well established theoretically, very few empirical studies exist in this domain. The main focus of existing work has been to quantify these costs at the consumer level. Brynjolfsson, Dick and Smith (2004) found that the cost of an exhaustive search is about $6.45 per consumer on a shopping bot. Hong and Shum (2006) develop a methodology for recovering search cost estimates using only observed price data. Bajari and Hortacsu (2003) quantified the cost of entering into an eBay auction to be $3.20, which includes search costs and other costs related to auction participation. In a related stream of work, Hann and Terweisch (2003) discuss how search costs and other related frictional costs in electronic markets could be substantial and found that the median value of these costs ranging from EUR 3.54 for a portable digital music player (MP3) to EUR 6.08 for a personal digital assistant (PDA).

However, these studies have largely focused on measuring consumer search costs, with less attention being paid on how search costs affect consumer demand structure or how they can affect retailers' business strategies. Not much work exists that looks at consumer demand structure in electronic markets and whether search costs differ across online retailers. It is generally believed that it is difficult to measure price search costs directly since it requires individual observations of consumer expectations of price distributions. However, by inferring the nature of consumer demand structure and by analyzing consumer price sensitivity to price changes, it is feasible to verify the presence of price search costs.

The presence of price search costs also implies that it takes time for price information to dissipate among consumers and demand to adjust slowly (Radner 2003). Future research can quantify the impact of this information delay on firm's optimal pricing policies. To do so, one needs to estimate the time it takes for pricing information to percolate in the market and show that the distribution of search cost evolves dynamically overtime. Till date, empirical studies on retailing have implicitly assumed that retailers face a demand function with constant price elasticity for any kind of a price change (Chevalier and Goolsbee 2003, Ellison and Ellison 2004, Ghose, Smith and Telang 2006), and used the estimated price elasticity to make inferences about competition and welfare in online marketplaces. This approach, however, does have a limitation. The assumption of constant price elasticity could lead to biased estimates if the difference in price elasticity for price changes in opposite directions is large. For example, if a retailer faces low price elasticity for price increases, but high price elasticity for price reductions, the constant elasticity assumption will average the price elasticity, thereby underestimating the real competitiveness and sensitivity in the market. One exception to the constant elasticity assumption is a recent study by Baye et al. (2005) which shows that price elasticity on shopbots increases dramatically after a retailer reduces its price to become the lowest price firm on the market. Future research can explore the impact of having a constant elasticity assumption in much greater depth.

4.2 Cost of Processing Textual Content

The above described stream of research can be extended to explicitly estimate the cognitive cost of evaluating retail offers in online markets. The Internet has facilitated the creation of markets that feature a dramatically wide selection of retailers. When a consumer places a product search for a unique product, they obtain a list of several offers with varying attributes (price, seller reputation, delivery schedules, etc), and extended product descriptions. All else equal, in order to make a selection amongst the offers displayed on the computer screen, consumers read these description, process the information and evaluate the best offer for purchase. In the presence of such information overload, it is natural to expect individuals to bear some cognitive cost of scrolling down the screens, and processing the textual description in each offer. Even if users do not cognitively process all the information displayed on the computer screen, a “scan and discard” vs. “scan and read more” decision leads to a cognitive cost of thinking (Shugan 1980).

Hence, rather than leaving it up to the consumer to cope with this distracting overload, providers often try to present first the most salient items in their inventory while taking into account the visual real estate available on a given device. Search engines such as Google or Yahoo do not exhibit all their search results on one webpage, but rather prioritize and display them on consecutive pages whose value is assumed to be decreasingly lower to the user (Huberman and Wu 2006). Hence, a natural direction for future research would be to examine the effect of summarizing product descriptions on product sales: short descriptions reduce the cognitive load of consumers but increase their uncertainty about the underlying product characteristics whereas a longer description has the opposite effect.

By drawing methods from the literature on Summarization in Computational Linguistics and NLP, one can suggest efficient ways of presenting product information in electronic markets by determining the optimal number of words and sentences in a given product description that minimize consumer cognitive cost. This requires an empirical analysis of how economic variables like product sales, price premiums and revenues are affected by the textual content in seller-contributed descriptions. Based on readability metrics such as the FOG, SMOG and Flesch-Kinciad metrics that estimate the reading difficulty of webpages, techniques such as Support Vector Machine-based approaches that automatically recognize reading levels from user queries (Liu et al. 2004), and advances in statistical language models that incorporate both content and surface linguistic features (Si and Callan 2001, Collins-Thompson and Callan 2005), the cost of consumer information processing can be quantified. This will involve building text classifiers, incorporating features based on reading level, and non-text features such as average line-length. Using discrete choice models, one can then estimate the likelihood of a consumer selecting a certain offer conditional on price and other seller attributes.

4.3 Managerial Decision-Making Costs

An important requirement towards establishing a frictionless market on the web is the need for retailers to act upon information and adjust prices without having to incur substantial costs in the process. Prior academic work has used a variety of techniques and methods to study pricing processes and the menu costs incurred by firms in order to improve our understanding of how firms set and adjust prices. In this stream of literature, menu costs consist of both the physical as well as the managerial costs of changing prices.

Lessons from physical markets show that price adjustment is a difficult, costly and time-consuming process and that changing prices "is a complex process, requiring dozens of steps and a non-trivial amount of resources" (Levy et al. 1997, p. 792). Thus, retailers need to incur substantial price adjustment cost to implement price changes. These costs include physically printing price labels, affixing it to retail products, updating POS systems for price changes, correcting errors and supervising these price change activities (see e.g. Levy et al. 1997). Besides these factors, there are substantial managerial costs involved as well. To set and adjust prices effectively, firms need to gather and analyze data that may be in different parts of the organization. They need the ability to assess a wide variety of market factors, from costs to customer segments, to estimating market demand, to understanding customer psychology, to anticipating competitors' reactions and so on. On top of this, one needs to add the complexities of firms selling multiple products through multiple distribution channels, and to multiple customers, often internationally (Zbaracki et al. 2005). Consistent with these aspects, recent empirical studies of price adjustment processes by Levy et al. (1997, 1998), Slade (1998), and Dutta et al. (1999, 2002) conclude that the price adjustment cost associated with these processes may be significant across industries in many offline markets.

Online retailers face a different environment. The increased market transparency due to the availability of information (Granados, Gupta and Kauffman 2005a, 2005b) helps firms monitor rivals' pricing activities. With no price labels to print, there is little physical labor cost involved and hence in theory, firms can adjust prices costlessly. This can have significant implications for consumer cognitive search costs. If menu costs are negligible, and retailers can adjust prices costlessly, then we are likely to see more frequent price changes. Such frequent price changes would impose a higher cognitive search cost on consumers. However, online retailers carry SKUs with much higher value than those in a typical offline stores, and often receive much higher sales volume. In this environment, each price experiment requires careful analysis. As a result, online retailers may incur substantial managerial decision-making costs leading to price rigidity. These costs are similar to data engineering costs, misclassification costs and active learning costs in the literature on human computer interactions and machine learning (Turney 2000). Indeed, Bergen, Kauffman and Lee (2005) find that Amazon and BN change prices less than once every 90 days, a surprisingly low frequency for online firms. Understanding the exact cause of price stickiness and identifying the magnitude of managerial costs of online retailers would thus be a natural next step in improving our understanding of the nature of consumer search costs.

Despite the extensive literature on menu costs (Levy et al. 1997, Levy 2006) in offline markets, no prior work has estimated the actual magnitude of such decision making costs in electronic markets. Ghose and Gu (2007) take one step in this direction by econometrically estimating and quantifying the magnitude of managerial costs of price adjustment faced by firms in electronic markets. They use the well known random walk theory model (Dixit 1991) to show that if the optimal price follows a random walk, the optimal solution for the retailer is to change the price if and only if the absolute difference between the actual price and the optimal price is larger than a constant. This enables them to empirically infer menu costs from online retailers' actual price change decisions given their demand fluctuations. A key contribution of their research is the development of statistical methods to separate long-term demand changes that affect retailers' price decisions from transitory demand changes that do not affect retailers' decisions. Future research could extend that stream of work by adopting dynamic programming techniques to explicitly model the online retailer's decision making process for price changes such as in Aguirregabiria (1999) and Kano (2006). This kind of research will contribute to recent research that measures the costs of online transactions (Hann and Terweisch 2003, Hann et al. 2005), measuring switching costs online (Chen and Hitt 2002), the welfare changes that online channels provide to consumers (Brynjolfsson, Hu, and Smith 2003, Ghose, Smith, and Telang 2006) and also to the newly emerging work in Computer Science that measures the readability and display of various categories of online content (Kumaran et al. 2005).

4.4 Impact of Search Costs and Menu Cost on the Long Tail

User level search costs and firm level menu costs are likely to influence pricing and product assortment strategies. The Internet is known to facilitate the discovery of lesser-known and obscure products. It has been argued that collectively these relatively less popular products could make up a significant portion of sales for online retailers (Anderson 2006). This phenomenon has been termed as the "Long Tail", and is often presented as a strategic opportunity to reach consumers in niche markets who were previously too costly to serve. While existing articles on the Long Tail highlight the strategic role played by lower "product search costs" in promoting the emergence of the Long Tail (Bryjolffson, Hu and Smith 2006), more rigorous research is needed to demonstrate the strategic role of "price search costs" in necessitating the emergence of the Long Tail. The impact of search costs on retailers' operations and marketing strategies has been studied both theoretically and empirically. It is now known that a failure to incorporate consumer search costs into the assortment planning process may lead the retailer to erroneously narrow product assortments (Cachon, Terswiesch and Xu 2005). Prior theoretical studies have also shown that the presence of such search costs could affect the shape and dynamics of consumer demand, resulting in significant difference in retailer's optimal product assortments. For example, Cachon, Terswiesch and Xu (2006) demonstrate that while lower search costs can intensify price competition, it can also expand the pool of customers visiting a retailer. The market expansion effect can justify a broader assortment, which in turn, can lead to higher prices and profits.

While prior studies on the Long Tail phenomenon focus on consumer demand for less popular products, they have not formally analyzed retailers' incentives to provide such products. This presents an exciting opportunity for future research. Rather than focusing on promoting popular products that face aggravated price competition, should online retailers shift product assortments towards niche products to take advantage of milder price competition in the emerging Long Tail? It also seems natural that such incentives will be related to the extent of menu costs faced by retailers in making price changes for popular versus unpopular products. This implies that research that explores the differences in magnitude of menu costs across different categories of products and different kinds of price changes will add to our understanding of firm-level incentives in promoting the Long Tail.

A related research question is to look at the composition of product returns for Internet retailers. Is there evidence that popular products are returned more often than obscure products?

Or is it likely to be vice-versa? From a managerial perspective, this is important because it highlights whether growth in variety is driven by demand-side factors or by supply-side factors. Thus, any research that investigates the impact of the Long Tail on the Internet also needs to account for the frequency with which products are returned in online markets.

5. Conclusion

With technological progress crossing newer frontiers, the amount of user-generated and firm-published content in web-based systems is growing exponentially. A significant portion of this content has concrete economic value that is embedded within it. Prior research has partially studied the impact of content, for instance, by analyzing the impact of numeric information. The aim of this article is a call for research that expands our understanding of that economic value by measuring the benefits and costs from analyzing other forms of content. Such value can be extracted, for example, from the content that specifies the social information of participants in online exchanges, the feedback and product description information that is captured in conversations between individuals, and finally from estimating the costs incurred by individuals during Internet usage, such as costs of locating and processing information and decision-making costs of adjusting information. There are several interesting research opportunities in this domain.

Such research can have lots of managerial and research implications. First**,** by studying the economic value of various kinds of textual information on the Internet, one can go well beyond analyzing the impact of numeric information such as the valence and volume of reviews and reputation scores that has been done in prior work. This will enable researchers to recommend actionable changes in the design of feedback systems, electronic market based communities and social networks. By producing novel ways of measuring the value of user-generated online content, the research activities will make actionable recommendations to practitioners to improve the design of feedback systems and display of information in online markets. In particular, firms are constantly trying to come up with incentive systems that prevent various kinds of spamming, gaming and content manipulation. For example, the analysis of the information in qualitative textual feedback can lead to the design of more robust reputation systems. Similarly, by mining the product descriptions in social shopping sites and used-good markets, future research in design science can come up with tools that help businesses display information efficiently by minimizing the information overload on consumers (for example, Ghose, Ipeirotis and Sundararajan 2005), and maximizing profits.

There are several other ways in which user-generated content can affect e-commerce firms. For example, Delicious, a social bookmarking website exploits users' social network to produce more relevant search results and hence provide better monetization of content. Since user-generated content has led to the emergence of communities in many online markets (Forman, Ghose and Wiesenfeld 2007), the availability of such data provides a plethora of opportunities for researchers interested in studying how online network based word-of-mouth marketing affects transactions. A pioneering work in this domain is that by Hill, Provost and Volinsky (2006, 2007) who find that consumers linked to a prior customer adopt a telecommunication service much faster than baseline groups selected by the best practices of the firm's marketing team.

Future research on how offline geographical locations affect online purchase and search behavior can provide important insights into the continuing debate over whether there exists a geographic "digital divide" in the US by providing concrete evidence or lack thereof. This becomes important in light of recent anecdotal evidence that despite the ubiquity of the Internet, many low-income, rural and small-town communities are being left out of this information revolution, and are deprived of the economic opportunities it offers. Moreover, research in this arena has the potential to demonstrate that a user's geographic community continues to play a role in how he or she behaves in electronic communities. Recent research suggests that users in electronic communities may prefer to communicate with users that share similar socio-economic and geographic neighborhoods (Van Alstyne and Brynjolfsson 2005). Further, it will help businesses understand the role of social information in driving consumer behavior in online communities, and provide prescriptive insights to businesses such that advertising and marketing strategies in online channels can be customized by users' geographical locations. It will help create incentives for underserved communities to participate in Internet commerce by understanding the mindsets of users and their needs for certain products in environments that are culturally mediated. By exploring virtual communities that emerge based on individuals' relationships to specific economic products, future research can advance our understanding of the dynamic interplay between social and economic exchanges on the Internet.

Finally, a key component towards measuring value from online content is an understanding of user costs in searching, processing and modifying information from an economic perspective. At present we have little insight into how to measure various kinds of cognitive costs incurred by humans while interacting with computers to process information. An explicit understanding of the magnitude of costs incurred by users during internet usage will contribute to optimal policies that lead to social and economic benefits for all participants. Such search and menu costs are also related to the emerging phenomenon of the Long Tail by influencing the incentives firms have for stocking niche and obscure products. It would be useful to see more academic research sprouting in these areas.

Acknowledgements

I thank Foster Provost for many helpful comments.

References

Aguirregabiria, V. 1999. The Dynamics of Markups and Inventories in Retailing Firms. *Review of Economic Studies*, 66, 275-308.

Akerlof, G. A. 1970. The Market for ‘Lemons’: Quality Uncertainty and The Market Mechanism. *Quarterly Journal of Economics*. vol. 84, 488-500.

Arnold, M. and Saliba, C. 2002. Price Dispersion in Online Markets: The Case of College Textbooks. Working Paper.

Bakos, Y. 1997. Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science* 43(12), 1676-1692.

Balasubramanian, S. 1998. Mail versus mall: a strategic analysis of competition between direct marketers and conventional retailers. *Marketing Science* 17(3), 181-195.

Bapna, R., Jank, W. and Shmueli, G. 2006. Consumer Surplus in Online Auctions. Working Paper, Indian School of Business.

Bartel, C.A, and Dutton, J.E. 2001. Ambiguous organizational memberships: Constructing social identities in interactions with others. in M.A. Hogg and D. Terry (Eds.), *Social identity processes in organizational contexts* (pp. 115-130). Philadelphia, PA: Psychology Press.

Baye, M. R., J. Morgan, and Scholten, P. 2004. Price Dispersion in The Small and in The Large: Evidence from an Internet Price Comparison Site” *J. of Industrial Economics* 52(4), 463-496.

Bergen, M. E., Kauffman R. and Lee D. 2005. Beyond The Hype of Frictionless Markets: Evidence of Heterogeneity in Price Rigidity on The Internet. *Journal of Management Information Systems* 22 (2): 57 - 89.

Brynjolfsson, E. and Smith, M. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science*, 46(4), 563-585.

Brynjolfsson, E., Hu, J., and Smith, M. 2003. Consumer Surplus in The Digital Economy: Estimating The Value of Increased Product Variety. *Management Science* 49(11): 1580-1596.

Brynjolfsson, E., Dick, A. and Smith, M. 2004. Search and Product Differentiation at an Internet Shopbot. Working Paper, MIT.

Bucklin, R., J. Lattin, A. Ansari, D. Bell, E. Coupey, S. Gupta, J. Little, C. Mela, A. Montgomery, and J. Steckel. 2002. Choice and the Internet: From Clickstream to Research Stream. *Marketing Letters*, 13(3), 245-258.

Cachon, G., C. Terwiesch, and Y. Xu. 2005. Retail assortment planning in the presence of consumer search. *Manufacturing and Service Operations Management*. **7**(4). 330-346.

Cachon, G., C. Terwiesch, and Y. Xu. 2006. On the effects of consumer search and firm entry on multiproduct competition, forthcoming, *Marketing Science.*

Chaiken, S. 1980. Heuristic Versus Systematic Information Processing And The Use of Source Versus Message Cues In Persuasion. *Journal of Personality And Social Psychology*, 39(5): 752-766.

Chawla, N., N. Japkowicz, and A. Kolcz. 2003. Proceedings of the icml2003 workshop on learning from imbalanced data sets.

Chen, P., and Hitt, L. 2002. Measuring Switching Costs and The Determinants of Customer Retention in Internet-Enabled Businesses: A Study of The Online Brokerage Industry. *Information Systems Research* 13(3): 255-274.

Chevalier, J. and Goolsbee, A. 2003. Measuring Prices and Price Competition Online: Amazon and Barnes and Noble. *Quantitative Marketing and Economics* 1(2): 203-222.

Chevalier, J., and Mayzlin, D. 2006. The Effect of Word of Mouth Online: Online Book Reviews. *Journal of Marketing Research.*

Collins-Thompson, K. and Callan, J. 2005. Predicting reading difficulty with statistical language models. Journal of The American Society for Information Science and Technology. 56(13), 1448 – 1462.

Cummings, J, Butler, B., and Kraut, R. 2002. The Quality of Online Social Relationships. *Communications of The ACM* 45(7). 103-108.

Dave, K., Lawrence, S., and Pennock, D. 2003. Mining The peanut gallery: Opinion extraction and semantic classification of product reviews. *World Wide Web Conference.*

Dellarocas, C. 2003. The Digitization of Word-of-Mouth: Promise and Challenges of Online Reputation Systems. *Management Science* 49 (10), 1407-1424.

Dellarocas, C. 2005. Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms. *Management Science* 52 (10), October 2006, 1577-1593.

Dellarocas, C., Awad, N., and Zhang, M. 2005. Using Online Ratings as a Proxy of Word-of-Mouth in Motion Picture Revenue Forecasting. Working Paper, University of Maryland.

Dewan S., and Hsu, V. 2004. Adverse Selection in Electronic Markets: Evidence from Online Stamp Auctions. *Journal of Industrial Economics*. 17 (4) 497-516.

Dhar, V. and Chang, E. 2007. Does Chatter Matter? The Impact of User-Generated Content on Music Sales. NYU-CeDER Working Paper.

Dixit, A. 1997. Analytical Approximations in Models of Hysteresis. *Review of Economic Studies*, 58, 141-151.

Ellison, G., and Ellison, S. 2006. Search, Obfuscation, and Price Elasticities on the Internet. Working Paper, Massachusetts Institute of Technology.

Forman, C., Goldfarb, A. and Greenstein, S. 2004. How Did Location Affect Adoption of The Commercial Internet? Global Village, Urban Leadership, and Industry Composition. Working Paper, Tepper School of Business, Carnegie Mellon University.

Forman, C., A. Ghose, and B. Wiesenfeld. 2007. Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. forthcoming, *Information Systems Research*.

Forman, C., A. Ghose, and A. Goldfarb. 2007. Competition Between Local And Electronic Markets: How The Benefit Of Buying Online Depends On Where You Live Working Paper, forthcoming, *Management Science*..

Ghose, A., Ipeirotis, P., and Sundararajan, A. 2005. The Dimensions of Reputation in Electronic Markets. Working Paper, New York University.

Ghose, A. 2006. Internet Exchanges for Used Goods: An Empirical Analysis of Trade Patterns and Adverse Selection. Working Paper, New York University.

Ghose, A., and P. Ipeirotis. 2006. Does Text Matter in Word-of-Mouth? The Economic Impact of Reviews on Sales and Helpfulness. Working Paper, New York University.

Ghose, A., M. Smith, and Telang, R. 2006. Internet Exchanges for Used Books: An Empirical Analysis of Welfare Implications, *Information Systems Research* 17(1), 3-19.

Ghose, A., Mukhopadhyay, T., and Rajan, U. 2007. Impact of Internet Referral Services on Supply Chain, *Information Systems Research*. 18(3), September.

Ghose, A., and Gu, B. 2007. Estimating Menu Costs in Electronic Markets. Working Paper, New York University.

Ghose, A., and Yang, S. 2007. An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets, Working Paper, New York University.

Goldfeld, S., and Quandt, R. 1973. A Markov Model for Switching Regressions, *Journal of Econometrics* (1:3), 1973, pp. 3-16.

Goolsbee, A. 2000. in a World without Borders: The Impact of Taxes on Internet Commerce. *Quarterly Journal of Economics* 115(2): 561-576.

Granados, N., Gupta, A., and Kauffman, R. 2005a. IT-Enabled Transparent Electronic Markets: The Case of The Air Travel Industry. *Forthcoming in Information Systems and e-Business Management.*

Granados, N., Gupta, A., and Kauffman, R. 2005b. Market Transparency Strategy and Multi-Channel Strategy: Modeling and Empirical Analysis. Working Paper, University of Minnesota.

M. Grobelnik, D. Mladenic, and N. Milic-Frayling. 2000. Text Mining as Integration of Several Related Research Areas: *Report on KDD'2000 Workshop on Text Mining.*

Gu, B., Konana, P., Rajagopalan, B., and Chen, H.M. 2007. Competition Among Virtual Communities and User Valuation: The Case of Investor Communities. *Information Systems Research*, 18(1) 68-85.

Hann, I., and Terwiesch, C. 2003. Measuring The Frictional Costs of Online Transactions: The Case of a Name-Your-Own-Price Channel. *Management Science*, 49(11), 1563 - 1579.

Hann, I., Savin, S., and Terwiesch, C. 2005.Online Haggling and Price-Discrimination at a Name-Your-Own Price Channel: Theory and Application. *Management Science*, 51(3), 339-351.

Hatzivassiloglou, V., and McKeown, K.R. 1997. Predicting The Semantic Orientation of Adjectives. *Proceedings of The 35th Annual Meeting of The Association for Computational Linguistics (ACL'97)*. 174-181.

Hearst, M. 1999. Untangling Text Data Mining. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 1999.

Hennig-Thurau T., Gwinner, K.P., Walsh, G., and Gremler, D. 2004. Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on The Internet? *Journal of Interactive Marketing* 18 (1), 38-52.

Hill, S., F. Provost, and C. Volinsky. 2006. Network-based Marketing: Identifying likely adopters via consumer networks. *Statistical Science* 21 (2) 256–276.

Hill, S., F. Provost and C. Volinsky. 2007. Learning and Inference in Massive Social Networks. Proceedings of *The 5th International Workshop on Mining and Learning with Graphs*.

Hong, H., and Shum, M., 2006. Using Price Distributions to Estimate Search Costs. Forthcoming, *Rand Journal of Economics.*

Hu, M., and Liu, B. 2004. Mining and Summarizing Customer Reviews. *Proceedings of The 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04)*. 168-177.

Jank, W., and Kannan, P.K. 2006. Understanding Geographical Markets of Online Firms using Spatial Models of Customer Choice. *Marketing Science*, 24 (4), 2006, p.623--634.

Janssen, M., and Karamychev, V. 2002. Cycles and Multiple Equilibria in The Market for Durable Lemons. *Economic Theory* 20, 579—601.

Kano, K. 2006. Menu Costs, Strategic Interactions, and Retail Price Movements. Working Paper, University of British Columbia.

Kim, S., and Hovy, E. 2004. Determining The Sentiment of Opinions. *Proceedings of COLING-04*. pp. 1367-1373. Geneva, Switzerland.

Kim, S., and Hovy, E. 2006. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. *Proceedings of The Workshop on Sentiment and Subjectivity in Text* at The Joint COLING-ACL Conference. Sydney, Australia.

G. King and L. Zeng. 2001. Logistic regression in rare events data. *Political Analysis*, 9(2):137–163.

Kumaran, G., Jones, R. and Madani, O. 2005. Biasing Web Search Results For Topic Familiarity. *Proceedings of The 14th ACM international conference on Information and knowledge management.*

Levy D., Bergen M., Dutta, S., and Venebirgeo, D. 1997. The Magnitude of Price adjustment cost: Direct Evidence from Large U. S. Supermarket Chains. *Quarterly Journal of Economics*, 112(3), 791-825.

Levy, D., S. Ray, H. Chen, and Bergen, M. 2006. Asymmetric Wholesale Price Adjustment: Theory and Evidence. *Marketing Science*, 25(2), 131-154.

Liu, B., Hu, M., and Cheng, J. 2005. Opinion Observer: Analyzing and Comparing Opinions on The Web. *Proceedings of The 14th international conference on World Wide Web.*

Liu, X., Croft, B., Oh, P., and Hart, D. 2004. Automatic Recognition of Reading Levels from User Queries. *Proceedings of The 27th annual international ACM SIGIR conference on Research and development in information retrieval.*

Pang, B., Lee, L., and Vaithyanathan, S. 2002. Thumbs up? Sentiment classification using machine-learning techniques. *Proceedings of the Empirical Methods in Natural Language Processing*, 79-86.

Pavlou, P., and A. Dimoka. 2006. The Nature And Role Of Feedback Text Comments In Online Marketplaces: Implications For Trust Building, Price Premiums, And Seller Differentiation. *Information Systems Research* 17(4) 392-414.

Prentice, D. A., Miller, D. T., Lightdale, J. R. 1994. Asymmetries in attachments to groups and to their members: Distinguishing between common-identity and common-bond groups. *Personality & Social Psychology Bulletin, 20*(5), 484-493.

Popescu, A., and Etzioni, O. 2005. Extracting Product Features and Opinions from Reviews, *Proceedings of HLT-EMNLP* 2005.

Prince, J. 2006. The Beginning of Online/Retail Competition and its Origins: An Application to Personal Computers. Forthcoming, *International Journal of Industrial Organization.*

Radner, R. 2003. Viscous Demand. *Journal of Economic Theory*. 112 (2), 189-231.

Reinstein, D. and Snyder, C. 2005. The Influence of Expert Reviews on Consumer Demand For Experience Goods: A Case Study of Movie Critics. *Journal of Industrial Economics*, 53(1), 27-51.

Ren, Y., R. E. Kraut, S. Kiesler. 2007. Applying Common Identity And Bond Theory To Design of Online Communities, Forthcoming. *Organization Studies.*

Resnick, P., Zeckhauser, R., Swanson, J., and Lockwood, K. 2006. The Value of Reputation on eBay: A Controlled Experiment. Forthcoming in *Experimental Economics.*

Riloff, E. and Wiebe, J. 2003. Learning Extraction Patterns for Subjective Expressions. *Conference on Empirical Methods in Natural Language Processing (EMNLP-03). ACL SIGDAT. Pages 105-112.*

Rossi, P., G. Allenby. 2003. Bayesian statistics and marketing. *Marketing Science*, 22, 304-329.

Scott-Morton F., F. Zettlemeyer, and J. Risso. 2006. How the Internet Lowers Prices: Evidence from Matched Survey and Auto Transaction Data. forthcoming, *Journal of Marketing Research.*

Shugan, S. 1980. The Cost of Thinking. *Journal of Consumer Research*. 7(2), p. 99-111.

Si, L. and J. Callan. 2006. A Statistical Model For Scientific Readability. *Proceedings of the Tenth International Conference on Information and Knowledge management.* Atlanta, Georgia, USA

Sinai, T., and J. Waldfogel. 2004. Geography and The Internet: Is The Internet a Substitute or Complement for Cities? *Journal of Urban Economics* 56: 1-24.

Turney, P. 2000. Types of Cost in Inductive Concept Learning. *Proceedings of The Workshop on Cost Sensitive Learning at The Seventeenth International Conference on Machine Learning (WCSL) at ICML-2000.*

Turney, P. D. 2002. Thumbs Up Or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of The 40th Annual Meeting of The Association for Computational Linguistics*. 417-424.

Van Alstyne, M., and Brynjolfsson, E. 2005. Global Village Or Cyber-Balkans? Modeling and Measuring The Integration of Electronic Communities. *Management Science* 51(6): 851-868.

Wiebe, J. 2000. Learning Subjective Adjectives from Corpora. *Proceedings of The. 17th National Conference on Artificial Intelligence (AAAI-2000).* Austin, Texas, July 2000.