Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets

Chris Forman
College of Management, Georgia Institute of Technology, Atlanta, Georgia 30332,
chris.forman@mgt.gatech.edu

Anindya Ghose, Batia Wiesenfeld
Stern School of Business, New York University, New York, New York 10012
{aghose@stern.nyu.edu, bwiesenf@stern.nyu.edu}

Consumer-generated product reviews have proliferated online, driven by the notion that consumers' decision to purchase or not purchase a product is based on the positive or negative information about that product they obtain from fellow consumers. Using research on information processing as a foundation, we suggest that in the context of an online community, reviewer disclosure of identity-descriptive information is used by consumers to supplement or replace product information when making purchase decisions and evaluating the helpfulness of online reviews. Using a unique data set based on both chronologically compiled ratings as well as reviewer characteristics for a given set of products and geographical location-based purchasing behavior from Amazon, we provide evidence that community norms are an antecedent to reviewer disclosure of identity-descriptive information. Online community members rate reviews containing identity-descriptive information more positively, and the prevalence of reviewer disclosure of identity information is associated with increases in subsequent online product sales. In addition, we show that shared geographical location increases the relationship between disclosure and product sales, thus highlighting the important role of geography in electronic commerce. Taken together, our results suggest that identity-relevant information about reviewers shapes community members' judgment of products and reviews. Implications for research on the relationship between online word-of-mouth (WOM) and sales, peer recognition and reputation systems, and conformity to online community norms are discussed.

Key words: digital markets; information processing; social identity; online reviews; Internet retailing; virtual communities; identity disclosure; user-generated content
History: Anil Gupta, Senior Editor. This paper was received on July 15, 2006, and was with the authors 6 months for 4 revisions.

1. Introduction

User-generated online product reviews have proliferated rapidly on the Internet, and such user-generated content has had a profound impact on electronic commerce. Scholars and practitioners alike are concerned with the relationship between online consumer product reviews and sales, but we have yet to understand why, how, and what aspects of online consumer-generated product reviews influence sales. Moreover, research in this arena is fragmented and little attention has been paid to how identity processes implicated in members' posting of reviews may influence how consumers respond to these reviews.

Prior work on the relationship between online product reviews (also called word-of-mouth) and sales has generally assumed that the primary reason that reviews influence sales is that they provide information about the product or the vendor to potential consumers. This literature has thus drawn attention to the product information contained in such reviews (Dellarocas et al. 2005, Reinstein and Snyder 2005, Chevalier and Mayzlin 2006). For example, research has demonstrated an association between how positively a product such as a book or movie is rated by consumers on a site and subsequent sales of the product on that site (Dellarocas et al. 2005, Chevalier and

Mayzlin 2006), or between review volume and sales (Duan et al. 2005, Liu 2006).

Importantly, many on-line product reviews provide information about the reviewer as well as information about the product. While prior research has addressed the link between review valence/volume and sales, little work considers the effect of information that reviewers disclose about themselves. Moreover, with regard to the benefits reviewers derive, work on online peer recognition and reputation systems has primarily focused on the consequences of peer recognition rather than on its antecedents (Resnick et al. 2000). To date, we are not aware of any prior research evaluating the influence of reviewers' disclosure of information about themselves on either peer recognition of reviewers or economic behavior in the form of online product sales.

It is critical to examine the effects of reviewer disclosure of identity-descriptive information for both theoretical and practical reasons. On a theoretical level, the information processing literature has accumulated an extensive body of research suggesting that attributes of an information source have powerful effects on the way people respond to messages (Kelman 1961; Chaiken 1980, 1987; Hass 1981; Mackie et al. 1990). Indeed, the information processing literature has repeatedly demonstrated that attributes of a message source often exert *direct* effects on message recipients' attitudes and behaviors, independent of the message content (e.g., Chaiken and Maheshwaran 1994, Petty et al. 1998, Simpson et al. 2000, Cohen 2003, Menon and Blount 2003, Chang 2004, Pornpitakpan 2004, Kang and Herr 2006). In a different literature, research on computer-mediated communication and virtual social contexts demonstrates that work outcomes are improved when virtual communicators exchange information about themselves as well as task-based information (Walther and Burgoon 1992; Spears and Lea 1992; Walther 1995, 1996; Jarvenpaa and Leidner 1999).

There are also practical reasons to expect that identity-descriptive information about the message source has influence in the online context in particular. On many sites, identity-descriptive information about the reviewer is at least as prominent as product information. For example, on sites such as Amazon information about product reviewers is graphically depicted, highly salient, and sometimes more detailed and voluminous than information on the products they review. Visitors to the site can see reviewers' badges (e.g., "real name" or "top reviewer") as well as personal information about reviewers ranging from where they live to the names of their pets, their nick names, hobbies, professional interests, pictures, and other posted links. Given the extent and salience of social information on product reviewers, it seems worthwhile to inquire whether such information influences the online consumers who are responsible for product sales.

Our work is designed to extend past research in at least five important ways. First, we contribute to work on the antecedents of online word-of-mouth by exploring how identity processes might shape the content and pattern of reviewer disclosure of identity-descriptive information in product reviews. Second, we contribute to research concerning the consequences of online word-of-mouth (Duan et al. 2005, Dellarocas et al. 2005, Chevalier and Mayzlin 2006, Liu 2006) by evaluating whether identity-descriptive information about reviewers predicts online product sales over and above the effects of product information in the reviews. Third, we highlight how disclosure of reviewer location can shape the geographic distribution of sales, thus contributing to emerging work concerning the impact of geography on electronic commerce (Forman et al. 2007). In the process, we also make a methodological contribution in better identifying the relationship between reviews and sales. We use cross-sectional variation in local sales and local reviews to identify the relationship between reviews and sales by differencing out time-invariant local preferences for books and using national reviews as a proxy variable for changes in perceived product quality over time. Fourth, while the growing literature on reputation systems and peer recognition (Resnick et al. 2000, Ghose et al. 2005) has addressed the *consequences* of peer ratings for those rated and consumers who use the ratings, the research has devoted little attention to the *antecedents* of peer recognition, which is the focus of our paper. Fifth, we provide insight into *how* and *why* reviewer disclosure of identity-descriptive information influences the consumers who evaluate the helpfulness of reviews and who purchase the products reviewed.

1.1. Research Context

In this paper, we use the term *online community* to refer to voluntary collectivities whose members share a common interest or experience and who interact with one another primarily over the Internet (see Sproull 2003 for a review of the online community research). The online community we explore is the community of book reviewers and consumers on Amazon—the largest electronic book retailer. Community members have common interests in particular books and are interdependent because members who consume rely on reviewers to post reviews that can inform their purchase decisions while members who review rely on other members for peer recognition in the form of helpful votes. Such a common interest or purpose, combined with interdependence, suggests that book reviewers and consumers on Amazon may function as a common identity community (Prentice et al. 1994, Sassenberg 2002, Ren et al. 2007).

Amazon has definable boundaries for membership because members are required to register to post reviews, rate reviews, or purchase products on the site (Grohol 2006). Many features of the site are designed to increase the salience of users' membership in, and identification with, Amazon. For example, members are greeted by name and they are reminded of past activity on the site (for example, the products they viewed during past visits to the site). Furthermore, they are reminded of their relationship to other consumers on the site with information about activities of people with similar interests (e.g., the purchase behavior of members who viewed the book they clicked on). A variety of site features provide opportunities for members to increase their engagement with the community such as by posting a personal profile, images, and identifying "friends" among other members of the site.

More intimate communities may develop within Amazon. Specifically, reviewers generally review particular types of books: The reviews they post are often for a particular genre of book, such as diet books, children's books, books on current political issues, or romantic fiction. The same focus is likely to be true of the reviews they read (and perhaps cast helpful votes on). Thus, the relevant community on Amazon may be much smaller than Amazon as a whole. Instead, it is likely to be described by a core group of people with a common set of interests in a particular book category or set of categories, who are likely to have repeated encounters with one another around specific books within that book category. While not necessary for social identification, evidence of such community attributes supports our assumption that identification concerns are likely to be salient to many Amazon members (Tajfel and Turner 1979), and makes it reasonable to assume that Amazon members could develop perceptions of familiarity and social connection with other members.

In the context of consumer book reviews on Amazon, message content is in the text of book reviews, but is succinctly summarized in the form of a star rating indicating the valence of the review (positive or negative) that appears prominently at the top of the review, and can range from 1 to 5 stars. Reviews with 5 stars or only 1 star are highly unequivocal, and are probably interpreted as strong information counseling in favor or against purchasing the book (respectively), while reviews with three stars are highly equivocal, probably interpreted as neither a strong positive nor a strong negative recommendation.

The information source in this context is the user who posts a review. Source characteristics may be unavailable, as when reviews are posted anonymously, or may be made available by the reviewer with some attributes appearing right above the text of the review. A variety of different types of personal information can be made available, such as one's real name (which requires providing credit card information to certify one's name and thus receiving a "real name" badge in the form of a sign next to one's name) and geographic location (both of which appear above the review text). Other information such as nickname, interests, birthday, or images require a single mouse click to the reviewer's profile page.

Research has repeatedly demonstrated that social identity concerns and beliefs (which refer to individuals' definition of the self in terms of group-defining attributes) play a vital role in shaping how recipients respond to information about the message source when processing messages (Abrams et al. 1990, David and Turner 1996, Mackie et al. 1990, Wood 2000). Social identity theory is thus an ideal bridge between the two literatures we integrate: the literature concerning individual motivations driving online contribution (which emphasizes identity motives) and

the literature associating the content of these contributions with community-level outcomes (i.e., peer recognition and product sales).

The rest of the paper is as follows. In §2, we describe the theory and the hypotheses. Section 3 describes the data. The empirical methodologies and results are described in §4. In §5 we discuss the theoretical implications of our results, followed by the managerial implications in §6. Section 7 concludes.

2. Theory and Hypotheses

Our analysis of reviewer identity-descriptive information in online product reviews arises at two different levels of analysis: the review level and the product level. At the review level of analysis, we consider whether community norms affect reviewers' willingness to disclose identity-descriptive information about themselves. We then evaluate whether such disclosure is associated with members' rating of the helpfulness of the reviews. At the product level of analysis, we consider whether reviewer identity disclosure affects product sales and if so, how? Figure 1 provides a roadmap of our analyses, showing the major relationships we test. We first turn to the individual review level of analysis.

2.1. Norms and Reviewer Identity Expression

What leads some community members to furnish identity-descriptive information along with their product reviews? Prior research draws on social identity theory to explain these motivations. In particular, research in online contexts suggests that identity motives play an important role in shaping online behavior (Turkle 1996, Donath 1999, McKenna and

Figure 1 Roadmap of Analyses

Bargh 1999, Ma and Agarwal 2007). While people wish to affirm a clear, consistent, and distinctive sense of self, they also wish to feel connected to others in social groups and to receive identity-affirming (termed self-verifying) feedback from others in the group (Jones and Pittman 1982, Swann 1983, Brewer 1991, Baumeister and Leary 1995, Tyler et al. 1996, Hornsey and Jetten 2004). Self-verification and social identification are desirable because they fulfill self-enhancement needs, reduce uncertainty, and thus enable people to predict and control their world (Lecky 1945, Swann 1983). The notion that identity shapes behavior has been affirmed in fields as diverse as psychology, economics, organizational behavior, marketing, and information systems (e.g., Tajfel and Turner 1979, Dutton et al. 1994, Akerlof and Kranton 2000, Ma and Agarwal 2007).

The need to communicate and verify their identity leads people to provide identity-descriptive information to others that conveys the way they view themselves, which may include the way they dress, act, and what possessions they purchase (e.g., car, house) (Swann 1983, Akerlof and Kranton 2000). Among the behaviors specifically implicated in prior research on self-verification in online contexts are the use of persistent labeling, which refers to using a single, consistent way of identifying oneself (such as "real name" in the Amazon context); and self-presentation, which refers to presenting oneself online in ways that help others identify one (such as posting geographic location or a personal profile in the Amazon context; Ma and Agarwal 2007). Research suggests that online disclosure of identity-descriptive information facilitates the formation of relationships, common bonds, and social attraction that electronic community members value (Ren et al. 2007).

If online self-disclosure is driven in part by (a) the desire for identification with a community and (b) the need for self-verifying feedback from other community members affirming that one is a member in good standing, then reviewer identity expressions should be patterned to follow community norms. Members conform to norms when they have a sense that a definable social group exists with behavior that is similar across group members and when members identify with that social group (Postmes et al. 2000). Norm conformity involves motivated efforts to seek
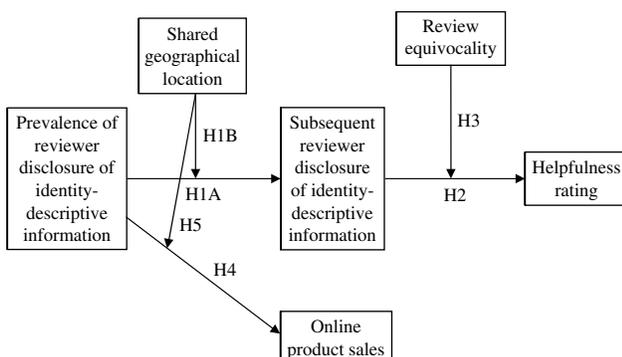
information on the behavior of other community members, to detect patterns in that behavior, and then to behave in ways that are consistent with those patterns. For this reason, prior research suggests that conformity to community norms is strong evidence of both individual social identity motives and group referencing behavior (Postmes et al. 2000, 2005; Sassenberg 2002).

Self-disclosure patterned after community norms thus stands in contrast to user-generated product reviews that are either entirely anonymous or that disclose personal information inconsistent with the patterns or norms in the community. Anonymous reviews require the least amount of effort and motivation, but have no clear identity-related benefits. Reviews that contain self-disclosure but do not follow the normative patterns in the group satisfy the need for self-expression (Jones and Pittman 1982), but provide no social affirmation benefits because they are less interpretable by interaction partners. In contrast, reviewers' self-disclosure that is patterned after community norms communicates reviewers' community identification to others, helps to establish the reviewers' reputation (Resnick et al. 2000), and provides an opportunity to obtain identity-verifying responses from fellow members. Hence, norm conformity is evidence of an investment on the part of an individual contributor signaling that the contributor would like to be viewed as a member of the community (Bartel and Dutton 2001, Bartel 2006).

In Amazon book reviews, norm conformity may be evident in the pattern of disclosure of identity-descriptive information that reviewers exhibit. 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, identification processes are likely to be an important antecedent to reviewer disclosure.

Norms can be inferred from archives of previous reviews that indicate whether prior reviewers tended to provide information such as real name, location, hobby, or birthday. Consistency between the type of information disclosed in previous reviews and the type of information disclosed in a subsequent review is evidence consistent with norm conformity.

HYPOTHESIS 1A. *Disclosure of identity-descriptive information in previous reviews of a particular product will* *be positively associated with disclosure of similar identity-descriptive information in subsequent reviews of that product.*

One way to assess whether identity motives drive self-disclosure is to evaluate whether norm-conforming self-disclosure is greater when community identification is more salient to members. If norm conformity is stronger among people who are known to share an identity group, then it is more likely that identity motives drive such norm conformity. Identification is stronger when the group is more relevant, immediate, and socially present (Dutton et al. 1994). Group relevance, immediacy, and presence is greater when individuals feel that fellow group members are closer in space and time (Latane 1981), and when people feel that they are similar to other group members (Byrne 1971, Turner 1987).

In the context of online consumer reviews of books, community identity may be more salient when members of the community claim to be from the same geographic region because geography is a natural basis for social community (Festinger et al. 1950). Common geography lowers perceived differences in space and time, and serves as a salient basis for feelings of similarity with other members of the group (Wiesenfeld et al. 1999, Ren et al. 2007). Amazon members may review the identity disclosure behavior of prior reviewers, especially attending to and imitating the behavior of those prior reviewers from their geographic location. In sum, sharing an off-line identity (geography) should increase the similarity of members' identity disclosure if such disclosure is motivated by identification.

HYPOTHESIS 1B. *Shared geographical location will increase the positive relationship between the pattern of disclosure of identity-descriptive information from prior reviewers and the pattern of disclosure of identity-descriptive information from subsequent reviewers of a product.*

2.2. Review Valence, Reviewer Identity Disclosure, and Peer Recognition

The previous section suggests that identification processes may be an antecedent shaping reviewer self-disclosure in online product reviews. If self-disclosure is an attempt to signal community identification and thus obtain affirmation and membership status from

the community, it is worthwhile to evaluate whether self-disclosure yields affirmation and status in the form of peer recognition.

As we described earlier, Amazon's reputation or peer rating system allows members to grant helpful votes to reviewers by answering "yes" to whether they found the review helpful. Prior research highlights the importance of reputation systems by clarifying the important consequences of peer recognition for both information contributors and consumers (Resnick et al. 2000, Jeppesen and Fredericksen 2006, Moon and Sproull 2006). This work generally suggests that peer recognition is a positive motivator for information contribution—information contributors report that they value such recognition (Jeppesen and Fredericksen 2006) and the presence of peer recognition systems is positively associated with both the quality and duration of knowledge contribution in electronic communities (Moon and Sproull 2006). Positive reputation and peer recognition also encourages trust, facilitates economic exchange, and even allows providers to command economic premiums (Resnick et al. 2000, Pavlou and Gefen 2004, Ghose et al. 2005). Antecedents of peer recognition—our focus in this paper—are less well understood, and have not been studied in contexts such as Amazon where helpful votes are granted anonymously.

Community members are the peer recognition grantors. Their goal in reading reviews is presumably to make purchase decisions, so peer recognition should be greater for reviews that facilitate purchase decisions. Reviews providing clearly positive evaluations help consumers make a purchase they will value while clearly negative evaluations help consumers avoid a purchase they may otherwise regret. Indeed, prior research suggests that a key factor driving negative word-of-mouth is the desire to warn others and strengthen social bonds (Wetzer et al. 2007). Whether positive or negative, unequivocal reviews should be judged as more helpful because they have clear implications for purchase decisions. In contrast, equivocal reviews are relatively uninformative because they contain ambiguous information (relative to unequivocal reviews) and therefore do not provide a clear guide for members' action.

Reviewers' self-disclosure may also influence perceived helpfulness in at least two respects. First,

if self-disclosure conforms to the norms established in the community, then reviews providing personal information conform to members' expectations and reinforce the community norms, leading members to evaluate the reviewer and the review more positively. Second, self-disclosure provides information about the message source that may increase the perceived usefulness of the message (Hass 1981, Kruglanski et al. 2006). Prior research suggests that message recipients use social information about the source of a message as a heuristic device, drawing on their assessment of the information provider as a simple and convenient decision rule or cognitive shortcut to help them reach judgments and guide action (Chaiken 1980, 1987). Thus, reviews that contain self-disclosure may be judged as more helpful because message source characteristics may be used to reach judgments about the product and guide purchase behavior. Evidence consistent with this logic may be found in the virtual groups and communities literature, which has found that members respond more positively when social information is present than when messages are exclusively task focused (Walther 1992, 1996; Jarvenpaa and Leidner 1999; Xia and Bechwati 2006).

Whether as a signal of norm conformity or as a source of heuristic information to guide purchase decisions, reviewer self-disclosure may be viewed as a behavior warranting reinforcement by other community members. Helpful votes are a way of socially reinforcing fellow community members because such peer recognition is known to be valued by information contributors (Jeppesen and Fredericksen 2006). Thus, we expect:

Hypothesis 2. *Reviews that disclose identity-descriptive information about the reviewer will be rated as more helpful than anonymous reviews.*

If review equivocality (i.e., message content) and reviewer self-disclosure (i.e., message source characteristics) both provide information shaping the perceived helpfulness of reviews, how do these types of information relate to one another? When members consider whether to grant helpful votes to an individual review, a clear and unequivocal review may contain enough information to guide action without members paying attention to source characteristics. However, when the message content is

more ambiguous and equivocal, they may rely more heavily on information about the message source to shape their opinions and actions. This logic suggests that review equivocality may moderate the effect of self-disclosure on perceived helpfulness, such that the positive relationship between self-disclosure and recognition should be greater when reviews are more equivocal.

Hypothesis 3. *The predicted positive association between reviewer disclosure of identity-descriptive information and helpfulness rating of the review will be stronger when the review is more equivocal than when it is unequivocal.*

2.3. Review Valence, Identity Disclosure, and Product Sales

From an economic, strategic, and marketing perspective, consumer product reviews are most important if they influence product sales. Prior research concerning the relationship between online consumer-generated product reviews and sales has primarily focused on the effect of review valence (Dellarocas et al. 2005, Duan et al. 2005, Chevalier and Mayzlin 2006, Liu 2006). One reason the valence of consumer product reviews may influence sales is that it may serve as a proxy for underlying product quality. This is especially relevant for cultural products such as books and movies that are difficult for consumers to evaluate prior to purchase (Senecal and Nantel 2004). Sales may also be positively related to review valence because of the influence of the reviews themselves, even when underlying product quality is controlled (e.g., Chevalier and Mayzlin 2006). For example, reviews may increase the salience of a product in the minds of consumers, they may draw consumers' attention to purchase decision criteria that may facilitate purchases, and they may provide information to consumers about "the social and psychological consequences of the purchase decision" (Brown et al. 2007, p. 4) which may increase sales in the case of positive reviews and diminish sales in the case of negative reviews.

While some of the reasons that reviews are presumed to lead to sales are associated with the valence of the review, a substantial amount of research on word-of-mouth suggests that perceived attributes of the *reviewer* may shape consumer response to reviews

(e.g., Brown and Reingen 1987). We draw on the information processing literature to suggest that product sales will be affected by reviewer disclosure of identity-related information. Message source characteristics have been found to influence judgment and behavior (Hass 1981; Chaiken 1980, 1987), and at least two possible pathways have been suggested whereby source characteristics might shape product attitudes and purchase propensity.

First, theory and extensive empirical evidence in psychology and marketing suggest that source characteristics have a direct impact on product evaluation regardless of the content of the message transmitted by the source (Hass 1981, Chaiken and Maheshwaran 1994, Petty et al. 1998, Simpson et al. 2000, Cohen 2003, Menon and Blount 2003, Chang 2004, Pornpitakpan 2004, Kang and Kerr 2006). Generally, these findings emerge when recipients process messages in a heuristic manner (whether because they are motivated, disposed, or contextually led to do so). Judgment based on source cues is sometimes referred to as "messenger bias" (Menon and Blount 2003). Indeed, Chaiken and Maheswaran theorized and found that "heuristic processing of the source cue exerted an independent and direct persuasive impact (on attitudes toward a fictional product)" (1994, p. 468). They found that source cues alone, and *not* message content, influenced subjects' judgment of whether a fictional product was superior when subjects heuristically processed the source cues.

Research on source cues and information processing is typically conducted in a laboratory rather than field settings, where subjects are exposed to a single message from a single source. Exposure to a single message and source facilitates careful and systematic processing of message content. However, Amazon members evaluating a single product are likely to see numerous reviews from numerous sources. Indeed, prior research suggests that the online context is more likely to be characterized by information overload than information scarcity which, in turn, increases the likelihood that information will be processed heuristically (Hansen and Haas 2001). If heuristic processing promotes reliance on source cues even to the exclusion of message content, reviewer disclosure of identity-descriptive information should be directly related to product sales.

In sum, if community members identify with and more positively assess reviewers who disclose identity-descriptive information (Walther 1992, 1996; Jarvenpaa and Leidner 1999; Lea et al. 2001) and use their assessment of reviewers as a heuristic shaping their evaluation of the product reviewed (Chaiken 1980, 1987; Chaiken and Maheswaran 1994), then members should be more likely to purchase products reviewed by community members who disclose identity-descriptive information about themselves.

HYPOTHESIS 4. *The prevalence of reviewer disclosure of identity-descriptive information will be positively related to product sales in an online market.*

To the extent that less effortful heuristic processing is responsible for the link between reviewer self-disclosure and sales, this relationship should be stronger when identity is more salient because identity salience increases the likelihood that people will process identity-relevant information (Turner 1987). Shared geography is one type of information triggering identification (Festinger et al. 1950, Byrne 1971, Wiesenfeld et al. 1999, Ren et al. 2007) and therefore increases the influence of identity-relevant information in reviews on product purchases. In sum, if identity processes are responsible for the relationship between disclosure and sales, then the prevalence of identity self-disclosure in reviews from a particular geographic region should differentially predict sales in that geographic region.

HYPOTHESIS 5. *Shared geographical location will increase the positive relationship between the prevalence of reviewer disclosure of identity-descriptive information and subsequent sales of that product.*

3. Methods and Data

3.1. Empirical Context
A major goal of this paper is to explore how social communities and identities influence economic transactions. Therefore, we felt it was essential to obtain a broad measure of the economic demand for the products that we study. To fulfill these requirements, we study reviewer disclosure of identity-descriptive information, reviewer ratings, and economic transactions in the electronic market for books on Amazon.com. Amazon.com is the leading

electronic market for books with over 70% market share (Ehrens and Markus 2000). Moreover, it provides a forum in which members can post and rate reviews of the products sold on the site. In the next section, we detail how we collected information from these forums.

3.2. Data Description
We gathered our data using automated Java scripts to access and parse HTML and XML pages on books available for sale from Amazon. Our sample includes 786 unique books drawn from all major categories. We derived this list from a random sample of books appearing as a best-seller in at least one city between April 2005 and January 2006, based on Amazon's "purchase circles." We use two data sets. The first consists of data on product characteristics, reviews, and reviewers of books in our sample: The unit of analysis in this data set is a product review. The second consists of economic transactions involving these products based on purchases by consumers in different geographical locations in the US: The unit of analysis in this data is either product time or product-location time. We provide more details on each of these data sets below.

3.2.1. Reviewers, Reviews, and Product Characteristics. We collected data on product characteristics, reviews, and reviewers from Amazon. Amazon chronologically archives its product review data. Summary statistics and correlations for each of the variables are included in Tables 1(a) and 1(b).

Table 1(a) Descriptive Statistics for Identity Disclosure and Helpful Votes Regressions

Variable	Observations	Mean	Std. dev.	Min	Max
Number reviews	175,714	1,257.73	1,480.57	1	5,756
Review valence	175,714	4.12	1.31	1	5
Perceived helpfulness	150,114	0.59	0.33	0	1
Real name	169,240	0.36	0.48	0	1
Location	175,714	0.71	0.45	0	1
Identity disclosure	175,714	0.78	0.41	0	1
Equivocal reviews	175,714	0.08	0.27	0	1
Same state	175,714	0.04	0.18	0	1

Notes. Note that not all reviews have been graded as "helpful or not" at the time of data collection. Specifically, reviews very close to the date of data collection were often not yet graded; this is why the number of observations is lower for some of these variables. Similarly some reviewers did not disclose any of the variables.

Table 1(b)   Correlation Matrix for Identity Disclosure and Helpful Votes Regressions

	Number reviews	Review valence	Perceived helpful	Real name	Location	Identity disclosure	Equivocal review	Same state
Number reviews								
Review valence	0.09**							
Perceived helpfulness	−0.1**	0.19**						
Real name	−0.17**	−0.01**	0.05**					
Location	−0.03	−0.32**	−0.04**	0.06**				
Identity disclosure	−0.06**	0.02**	0.045**	0.39**	0.79**			
Equivocal review	−0.03**	−0.26**	−0.07**	0.019**	−0.02	0.01		
Same state	−0.005**	0.001	−0.01**	−0.03**	0.14**	0.11**	0.01**	

Note. ** denotes significance at 1%.

Reviewer Characteristics: Amazon has a procedure by which reviewers can disclose personal information about themselves. We focus our analysis on the categories of information most commonly provided by reviewers: the reviewer's real name, location, nickname, and hobbies. By real name, we refer to a procedure wherein members indicate their legal name and verify it with their credit card data. Members may optionally decide to post information about their geographic location, disclose additional information (e.g., hobbies) or use a nickname (e.g., L. Quido "Quidrock"). We use such data to assess reviewer personal disclosure.

Two forms of identity-descriptive information are available right above the review (real name and location) while additional information appears on a separate profile page. Consistent with the notion that community members may not click past the first page containing the reviews and thus dimensions of disclosure on the first page should be less noisy, we constructed a dummy variable labeled "identity disclosure," which captures each instance where the reviewer has revealed real name, geographic location or both. We aggregate the variable in this way to ease interpretation of our analyses.[1]

Review Characteristics: We collected all reviews of each book in the sample posted chronologically from the time it was released into the market until the time period of our data collection (the week of March 17th, 2006). We modeled our analyses on prior work (e.g., Chevalier and Mayzlin 2006) and included average review valence (on the 1–5 star scale Amazon provides) as well as the number of reviews in our regressions. We also constructed a dummy variable to differentiate between extreme reviews, which are unequivocal and therefore provide a great deal of information to inform purchase decisions, and equivocal reviews that provide less information. Specifically, ratings of 3 were classified as equivocal reviews while ratings nearer the endpoints of the scale $(1, 2, 4, 5)$ were classified as unequivocal.[2]

To assess helpfulness ratings, we used a feature on Amazon at the bottom of each review where readers may rate the review by answering "yes" or "no" to the question, "Was this review helpful to you?" Previous peer ratings appear immediately above the posted review, in the form, "[number of helpful votes] out of [number of members who voted] found the following review helpful:", enabling us to compute the fraction of votes that evaluated the review as helpful.

Product Characteristics: Our data contain product-specific characteristics such as the books' Amazon retail price and the date that the product was released

---

[1] We also ran all our analyses with a variable that captures each instance where the reviewer has engaged in any one of the four kinds of self-disclosure. All our results are very similar to the current results and for brevity we do not include them in the paper. Moreover, regressions using each of our four independent measures of disclosure give qualitatively similar results, though for some types of disclosure the degree of variance explained was reduced. For brevity we have not included the results of these regressions, but they are available on request.

[2] We also ran our analysis with another specification of equivocal reviews where ratings in the middle of the scale (2, 3, and 4) were classified as equivocal while ratings at either of the two endpoints of the scale (1, 5) were classified as unequivocal. The interpretation of our results was virtually identical to our current results, and are hence omitted for brevity.

on the market. We use this latter variable to compute the elapsed time from the date of product release. We also collect the product's Amazon sales rank from our purchase circle data, described below.

3.2.2. Data on Economic Transactions.

Product Sales: For Hypotheses 1A, 1B, 2, and 3 that are tested at the review level, we use the raw data collected from Amazon as described above. For Hypotheses 4 and 5, in which we examine the relationship between disclosure and product sales, we aggregate data to the monthly level. By aggregating data in this way, we smooth potential day-to-day volatility in sales rank. Moreover, it allows us to maintain consistency in our analyses with our purchase circle data (described below) which, because of restrictions on the frequency with which Amazon updates purchase circles, are also organized at the monthly level. For these regressions, sales rank is the average sales rank for the month of analysis, while review valence, number of reviews, and disclosure are equal to the average (or, in the case of number of reviews, sum) of reviews up to the preceding month (we do not include the current month because of concerns that such analysis may reflect simultaneity between reviews and sales, making causality more ambiguous). Control variables—including Amazon price and log of elapsed date—are the averages for that particular month. We exclude products that have fewer than 10 reviews as the relationship between reviews and sales for these products may be different than that for the rest of the sample. Further, we exclude products for which we are missing elapsed date because absence of this information may bias the relationship between reviews and sales. All of our results are robust to including these observations. Descriptive statistics and a correlation matrix for this data are included in Tables 2(a) and 2(b).

Economic Transactions by Geography: The data we use on economic transactions by geography comes from the "purchase circles" on the Amazon.com web site. Amazon's purchase circles are specialized bestseller lists. The pages denote the top-selling books across large and small towns in every state throughout the US. For each month between April 2005 and January 2006, we collected monthly data on sales for

Table 2(a)   Descriptive Statistics for Sales Rank Regressions

Variable	Observations	Mean	Std. dev.	Min	Max
Log of sales rank	3,139	6.15	1.21	0	11.45
Amazon retail price	3,139	14.33	12.22	0	156.95
Missing retail price	3,139	0.012	0.11	0	1
Log of elapsed date	3,139	6.88	1.29	3.04	9.74
Average review valence	3,139	4.26	0.44	1.77	5
Log of total reviews	3,139	4.82	1.27	2.48	8.47
Percent identity disclosure	3,139	0.8	0.08	0.48	1

Table 2(b)   Correlation Matrix for Sales Rank Regressions

	Log sales rank	Amazon price	Missing price	Log of elapsed date	Average review valence	Log of total reviews	Pct identity disclosure
Log of sales rank							
Amazon retail price	0.08**						
Missing retail price	0.15**	−0.13**					
Log of elapsed date	0.15**	−0.19**	0.028				
Average review valence	0.07**	−0.005	−0.04*	0.24**			
Log of total reviews	−0.23**	−0.21**	−0.004	0.32**	−0.18**		
Percent identity disclosure	0.015	0.12**	0.017	−0.38**	−0.01	−0.23**	

Note. ** and * denote significance at 1%, and 5%, respectively.

each location in the purchase circles. The purchase circles are organized in multiple layers—first by state, then within a state, by town.

We use this data to examine the relationship between disclosure and sales across local geographic markets. For each town, Amazon provides a list of top 10 best sellers for each product category. Our dependent variable is a dummy indicating whether the product appears in the top 10 in a particular location in a particular time period. So, for example, in April 2006 the product *Angels and Demons* by Dan Brown was ranked #27 in books nationally (national sales rank) while it was ranked #5 in Las Vegas, Nevada, and #6 in Great Falls, Montana (local sales rank). Our purchase circle data were collected monthly, so our unit of observation is a product-location-month.

For independent variables, we again compute the nationwide average review valence and the total number of reviews over all prior months. To measure how review valence, number of reviews, and disclosure by community members in the same geographic region influence local sales, we compute average review

valence, disclosure, and the (log of the) total number of reviews in the state of the location under analysis. For example, if the location under analysis was Philadelphia, PA, we would compute these quantities for the state of Pennsylvania. Because some state-product-months will have no prior reviews, we include a dummy that indicates whether at least one review has been written in a state as a control. We also collect data on the product's sales rank from our purchase circle data, which is used in our empirical analyses on national sales.[3] Descriptive statistics and correlations for these data may be found in Tables 3(a) and 3(b).

4. Empirical Methodology and Results

In this section, we discuss the models we use to test Hypotheses 1 through 5 and the results of estimating these models. Our use of secondary data from Amazon's electronic market required us to estimate separate models for the hypotheses. In each section, we briefly describe the empirical model used, our identification strategy, and our results. As in §2, we begin with hypotheses at the review level and then move to the analyses of product sales.

4.1. Disclosure of Identity-Descriptive Information

Based on identity-based norm conformity, Hypothesis 1A suggests that prior book reviewers' disclosure of identity-descriptive information will predict the extent to which subsequent reviewers of the book reveal similar information.

4.1.1. Disclosure of Identity-Descriptive Information. In the first set of analyses, our dependent variable is $DISCLOSURE_{jr}^{p}$, a binary variable that indicates whether review r that has been posted for product j has posted personal information of type p (which can be real name, location, or our composite variable

---

[3] We have no sales rank data for product-months in which a product does not appear in any purchase circle. As a robustness check, we reestimated our models using only products that were in the sample for the full 10 months and our results were qualitatively the same.

Table 3(a)   Descriptive Statistics for Purchase Circle Regressions

Variable	Observations	Mean	Std. dev.	Min	Max
Product appears in local top 10	8,300,573	0.015	0.12	0	1
Log of elapsed date	8,300,573	6.45	1.57	−0.4	9.74
Retail price	8,300,573	8.58	12.81	0	156.95
Missing retail price	8,300,573	0.43	0.49	0	1
Log of total reviews	8,300,573	4.78	1.28	2.48	8.47
No state reviews	8,300,573	0.74	0.44	0	1
Log of total state reviews	8,300,573	1.42	1.25	0	6.21
Average review valence nationwide	8,300,573	4.22	0.46	1.77	5
Average review valence statewide	8,300,573	3.12	1.97	0	5
Percent disclosing real name nationwide	8,300,573	0.38	0.16	0.06	0.91
Percent disclosing real name statewide	8,300,573	0.30	0.33	0	1

identity disclosure). For each of the disclosure variables, we estimate the following fixed effects panel data model:

$$DISCLOSURE_{jr}^{p} = \alpha + \beta DISCLOSURE_{jr-1}^{p} + \Omega' X_{jr} + \mu_j + \varepsilon_{jr}, \quad (1)$$

where $DISCLOSURE_{jr-1}^{p}$ is a dummy variable that indicates whether the prior review disclosed personal information of type p.[4] μ_j is a product fixed effect that controls for differences in the average propensity of reviewers to reveal personal information across books. X is a vector of control variables that includes the review valence of the product and the log of the number of reviews. Our primary interest is in measuring parameter β, which captures the relationship between prior and subsequent disclosure of identity-descriptive information.[5]

---

[4] The results from other combinations such as considering the previous 5 or 10 reviews give qualitatively similar results. The results of these alternative models are available from the authors on request.

[5] Use of lagged dependent variables such as $DISCLOSURE_{jr-1}^{p}$ fail the "strict exogeneity" assumption commonly used in panel data models, and can lead to estimates of β that are biased downward in data with short panels (Wooldridge 2002). However, if the number of time periods in the panel is large (as it is in our sample), the bias asymptotically goes to zero (Hsiao 2003). To explore whether this bias influenced our results, we used a difference-in-difference instrumental variables panel data estimator with three-period lags as instruments, as recommended in Hsiao (2003). The results, which are available on request, were qualitatively the same.

Table 3(b)   Correlation Matrix for Purchase Circle Regressions

	Appears in local top 10	Log of elapsed date	Retail price	Missing retail price	Log of total reviews	No state reviews	Log of total state reviews	Average review valence nation	Average review valence state	Percent real name nationwide	Percent real name statewide
Appears in local top 10											
Log of elapsed date	−0.03**										
Retail price	0.06**	−0.04**									
Missing retail price	−0.11**	−0.10**	−0.58**								
Log of total reviews	0.14**	0.27**	−0.06**	−0.15**							
No state reviews	0.05**	0.14**	−0.05**	−0.05**	0.45**						
Log of total state reviews	0.12**	0.19**	−0.05**	−0.09**	0.72**	0.67**					
Average review valence nationwide	−0.05**	0.26**	−0.04**	0.04**	−0.19**	−0.08**	−0.12**				
Average review valence statewide	0.04**	0.19**	−0.06**	−0.04**	0.4**	0.93**	0.61**	0.1**			
Percent real name nationwide	0.01**	−0.57**	0.11**	−0.09**	−0.4**	−0.19**	−0.29**	−0.05**	−0.19**		
Percent real name statewide	0.04**	−0.12**	−0.0004	−0.02**	0.13**	0.53**	0.27**	−0.08**	0.49**	0.24**	

Note. ** denotes significance at 1%.

Table 4 presents the results of estimating this model. Consistent with Hypothesis 1A, there was a positive relationship between the self-descriptive information disclosed by previous and subsequent reviewers. The coefficients can be interpreted as the change in the likelihood of posting personal information when the prior reviewer also posts the same personal information. Thus, if the prior reviewer discloses information, then the likelihood of subsequent disclosure will increase by 0.7 percentage points for location, 0.9 percentage points for identity disclosure, and 15 percentage points for real name. Thus, the quantitative impact of real name disclosure on future real name disclosure is quite large—the average disclosure rate for real name is 36.6%, which translates into a 41% increase when evaluated at the mean—however the effect of prior location disclosure on future location disclosure is relatively small. All relevant parameter estimates are significant at the 1% level.[6]

4.1.2. Disclosure of Identity-Descriptive Information by Geographic Location. If shared geography increases identity salience, it should influence the extent to which subsequent reviewers from the same location engage in similar kinds of disclosure of identity-descriptive information because identity-based norm conformity should be greater when identity is more salient. To test Hypothesis 1B, we use the same panel data set that we used in §4.1.1. In addition, we introduce a new variable $SAMESTATE$ if the previous reviewer was from the same state as the subsequent reviewer. Our dependent variable is $DISCLOSURE_{jr}^p$ as defined before. For each of the disclosure variables, we estimate the following fixed effects panel data model:

$$
\begin{aligned}
DISCLOSURE_{jr}^p \\
= \alpha + \beta DISCLOSURE_{jr-1}^p + \gamma SAMESTATE_{jr} \\
+ \lambda DISCLOSURE_{jr-1}^p \times SAMESTATE_{jr} \\
+ \Omega' X_{jr} + \mu_j + \varepsilon_{jr},
\end{aligned}
$$

where $SAMESTATE_{jr}$ is an indicator variable equal to 1 when the prior review is from the same state i. Again, μ_j is a product fixed effect that controls for average differences in the likelihood that a review will be posted for a product. Our primary interest is in estimating the parameter vector λ that measures how the effects of prior disclosure are moderated when previous and subsequent reviewers are from the same state. We interpret positive coefficients on λ as support for Hypothesis 1B.[7]

---

[6] Note that the R^2 values used in all models include the estimated fixed effects. R^2 values for the "within" estimator for these equations are available on request.

[7] Note that we are only able to perform these analyses for the case where p = "real name;" similar regressions examining the propensity to disclose location or identity disclosure will confound the effects of location disclosure with location in the same state (since $SAMESTATE = 1$ only when location is disclosed). For both Hypotheses 1A and 1B, the results using disclosure of other kinds of self-descriptive information such as Hobbies and Nicknames are very similar to our current results.

Table 4 How Do Prior Reviewers' Decisions to Postidentity Descriptive Information Influence the Likelihood of a Subsequent Reviewer Posting a Review with the Same Information?

Independent variable	(1) Real name	(2) Location	(3) Identity disclosure	(4) Real name
Prior disclosure	0.15 (0.003)**	0.01 (0.002)**	0.01 (0.002)**	0.06 (0.003)**
Prior disclosure * Same state				0.05 (0.01)**
Same state	−0.03 (0.005)**			−0.05 (0.01)**
Review valence	0.01 (0.001)**	0.01 (0.001)**	0.01 (0.001)**	0.01 (0.001)**
Log of total reviews	−0.095 (0.27)	0.78 (0.26)**	0.64 (0.21)**	−0.09 (0.28)
R-squared	0.10	0.02	0.03	0.10
Adj R-squared	0.09	0.02	0.03	0.09
F	$F_{(4,167,964)} = 236.73^{**}$	$F_{(3,174,144)} = 82.11^{**}$	$F_{(3,174,144)} = 95.60^{**}$	$F_{(5,167,921)} = 154.22^{**}$

Notes. This table shows whether disclosure of identity-descriptive information in prior reviews increases the probability that subsequent reviewers will reveal the same information. The dependent variable is a dummy for the relevant identity-descriptive variable. The variable Same State is excluded for location and identity disclosure because Same State = 1 only when location is disclosed. The main independent variable of interest is whether the prior reviewer disclosed the same information. Columns 1, 2, and 3 indicate results for disclosure of real name, location, and identity disclosure, while Column 4 presents results for interactions of real name with location in the same state (Hypothesis 1B). All models use ordinary least squares with product-level fixed effects. Robust standard errors are listed in parenthesis. ** and * denote significance at 1%, and 5%, respectively. The total number of observations is 168,709.

Column 4 of Table 4 provides the results of these regressions. Consistent with Hypothesis 1B, there was a positive relationship between the disclosure of identity-descriptive information by previous and subsequent reviewers if reviewers were from the same geographical region. As before, β can be interpreted as the change in the likelihood of posting personal information when the prior reviewer also posts the same personal information, while λ can be interpreted as the increase in likelihood because of prior disclosure when the prior reviewer is also from the same state. All relevant parameter estimates are significant at the 1% level. Prior disclosure of real name increases the likelihood of subsequent real name disclosure by 6.1 percentage points, however this jumps to 11.1 percentage points when the prior disclosure came from a reviewer in the same state. These increases are significant compared to the average likelihood of disclosing real name of 36.6%, and suggest that prior disclosure of real name by a reviewer from the same state increases the likelihood of real name disclosure by 30.3%.

Finally, note that the coefficient on same state is negative, which suggests that disclosure of real name is less likely when previous reviewers have disclosed their state (unless previous reviewers from their state have disclosed their real name as well, which increases the likelihood that subsequent reviewers from their state disclose their real name). Disclosing

real name is more effortful because of the credit card registration process, and this might lead reviewers to use location as a means of self-disclosure unless those in their state disclose real name as well.

4.2. Peer Ratings of Reviews
Hypothesis 2 suggested that community members would rate reviews containing self-descriptive information as more helpful than anonymous reviews. The dependent variable, $HELPFUL_{jr}$, is operationalized as the ratio of helpful votes to total votes received for a review r issued for product j. Our baseline specification takes the following form:

$$HELPFUL_{jr} = a_0 + a_1(EQUIVOCAL_{jr}) + a_2(DISCLOSURE_{jr}) + \Omega'X_{jr} + \mu_j + \varepsilon_{jr}. \tag{4}$$

$EQUIVOCAL$ is an indicator variable that is equal to 1 if the review was an equivocal review, and 0 otherwise. It assesses the information content of the review. $DISCLOSURE_{jr}$ is a dummy variable that is equal to 1 if the review indicated real name or location (i.e., if "identity disclosure" is equal to 1). X is a vector of control variables that includes the log of the number of reviews for that product. μ_j is a product fixed effect that controls for differences in the average helpfulness of reviews across books. We interpret positive coefficients on a_2 as evidence of support for Hypothesis 2.

The above equation can be estimated using a simple panel data fixed effects model. However, one concern with this strategy is that the posting of personal identity information such as real name or location may be correlated with some unobservable reviewer-specific characteristics that may influence review quality. If true, such correlation would lead to inconsistent estimates of a_1. To control for this potential problem, we use a Two Stage Least Squares (2SLS) regression with instrumental variables (Wooldridge 2002). Specifically, we instrument for $DISCLOSURE_{jr}$ in the above equation using lagged values of real name and location. The intuition behind the use of these instrument variables is that they are likely to be correlated with the relevant independent variables such as $DISCLOSURE_{jr}$ (as demonstrated by our tests of Hypothesis 1), but uncorrelated with unobservable characteristics that may influence the dependent variable. For example, the use of a real name in prior reviews is likely correlated with the use of real name in the subsequent reviews, but uncorrelated with unobservables that determine perceived helpfulness for a given review. The result of this regression is included in Column 1 of Table 5. Reviewers who disclose real name or location have 12.2 percentage points more helpful votes than an otherwise identical reviewer. These are significant increases when compared to a mean value of perceived helpfulness of 59.6%. In sum, our results provide support for Hypothesis 2. Column 1 also shows that equivocal reviews receive a significantly lower fraction of helpful votes (coefficient -0.062, significant at the 1% level).

Figure 2

Percentage Helpful Votes as a Function of Identity Disclosure and Review Equivocality

To test Hypothesis 3, we added the interaction term $EQUIVOCAL_{jr} \times DISCLOSURE_{jr}$ to the regression model shown in Equation (4). The results in Column 2 of Table 5 show that equivocal reviews are rated as less helpful than clearly positive and negative reviews. Of greater importance, however, the coefficient of the interaction of review equivocality and identity disclosure is positive, statistically significant, and similar in magnitude to that of equivocality. To illustrate the nature of the significant interaction effect, we followed the procedure recommended by Aiken and West (1991), charting predicted values of the dependent variable at 1 standard deviation above and 1 standard deviation below the mean for the independent variables. As can be seen in Figure 2, the positive relationship between reviewer disclosure and perceived helpfulness is significantly attenuated when reviews are unequivocal relative to when they are more equivocal. These results lend support to Hypothesis 3, suggesting that community members are less responsive to reviewer disclosure of identity-descriptive information when reviews are unequivocal (and therefore more informative) than when reviews are equivocal.

Table 5   Identity Disclosure and Perceived Helpfulness

Independent variable	(1)	(2)
Identity disclosure	0.12 (0.03)**	0.10 (0.035)**
Equivocal review	−0.06 (0.003)**	−0.26 (0.06)**
Log of total reviews	−0.15 (0.19)	−0.15 (0.19)
Equivocal review * Identity disclosure		0.26 (0.08)**
R-squared	0.03	0.03
F	$F_{(780,148,657)} = 18.78^{**}$	$F_{(780,148,657)} = 18.57^{**}$

Notes. The dependent variable is equal to percentage of helpful votes. Standard errors are listed in parenthesis. ** and * denote significance at 1%, and 5% respectively. All models use 2SLS to instrument for disclosure variables using lagged values of the same variables. The fixed effects are at the product level. The total number of observations is 149,442.

4.3.   Effects of Identity Disclosure on Sales

Our next set of models is used to link reviewer disclosure of identity-descriptive information to economic outcomes (i.e., product sales). To do this, we first examine how the persuasive content of reviews—as measured by review valence—influences

sales. Second, consistent with a heuristic processing model, we examine how the disclosure of personal information in reviews such as real name and location directly influences product sales (Hypothesis 4). Last, we examine how the prevalence of disclosure of identity-descriptive information in reviews from a particular geographic region influences sales in that region above and beyond the effect of disclosure in reviews posted nationwide (Hypothesis 5).

4.3.1. Disclosure of Identity-Descriptive Information and Sales. We first estimate the relationship between sales rank and (1) the average valence of reviews as well as (2) the percentage of reviews disclosing real name or location (identity disclosure).[8] We estimate the following regression:

$$\log(SALESRANK_{jt})$$
$$= \alpha + \beta_1 \log(ELAPSEDDATE_{jt}) + \beta_2 AMAZONPRICE_{jt}$$
$$+ \beta_3 MISSPRICE_{jt} + \beta_4 \log(TOT - REVIEWS_{jt-1})$$
$$+ \beta_5 AVG - VALENCE_{jt-1} + \gamma_1 PCT - IDENTITY$$
$$- DISCLOSURE_{jt-1} + \mu_j + \varepsilon_{jt}. \tag{2}$$

Our specification is adapted from prior work by Chevalier and Mayzlin (2006), who examine how price, number of reviews, and review valence influence product sales on Amazon and Barnes and Noble. The unit of observation in our analysis is a product-month, and the dependent variable is $\log(SALESRANK_{jt})$, the log of sales rank of product j in month t. $AMAZONPRICE_{jt}$, and $\log(ELAPSEDDATE_{jt})$ are product controls from the same month.[9] We estimate product-level fixed effects to control for differences in average sales rank across products. These fixed effects are algebraically equivalent to including a dummy for every product in our sample, and so they will enable us to control for differences in the average quality of products.

[8] Note that prior work in this domain has generally transformed the dependent variable (sales rank) into quantities using the specification similar to Ghose et al. (2006). That was usually done because those authors were interested in demand estimation. However, in this case we are not interested in estimating demand, and hence we do not need to make the transformation.

[9] *MISSPRICE* is a dummy variable equal to 1 when we are missing price data. *ELAPSEDDATE* is the difference between the date of data collection and the release date of the book.

Thus, any relationship between sales rank and review valence will not reflect differences in average quality across books, but rather will be identified from changes over time in sales rank and review valence within books, diminishing the possibility that our results reflect differences in average unobserved book quality rather than aspects of the reviews themselves.

Our primary interest is in examining the association between $PCT - IDENTITY - DISCLOSURE_{jt-1}$ and sales (Hypothesis 4). To maintain consistency with prior work, we also examine the association between $AVG - VALENCE_{jt-1}$ and sales. However, prior work has shown that review valence may be correlated with product-level unobservables that may be correlated with sales. In our setting, although we control for differences in the average quality of books through our fixed effects, it is possible that changes in the popularity of the book over time may be correlated with changes in review valence. Thus, this parameter reflects not only the information content of reviews, but also may reflect exogenous shocks that may influence product popularity. Similarly, the variable $TOT - REVIEWS_{jt-1}$ will also capture changes in book popularity or perceived book quality over time; thus, β_4 may reflect the combined effects of a causal relationship between number of reviews and sales (Duan et al. 2005) and changes in unobserved book popularity over time.

Columns 1 through 3 in Table 6 display the results of the model. Note that increases in sales rank mean lower sales, so a negative coefficient implies that increases in a variable *increase* sales. Average review valence was not a significant predictor of sales in our analyses. This is true both with (Column 1) or without (Column 2) controlling for the effects of reviewer disclosure.

However, Column 2 shows that the coefficient of identity disclosure is negative and statistically significant, implying that a 1 standard deviation increase in identity disclosure leads to a 14.9% decline in sales rank. These results support Hypothesis 4 that the prevalence of reviewer disclosure of identity-descriptive information would be associated with higher subsequent sales, and are consistent with prior research in the information processing literature supporting a direct effect for source characteristics on product evaluations and purchase intentions when

10.99

Table 6   How Does Identity Disclosure Influence Sales Nationally and Locally Within States?

	National sales			Local sales	
	(1)	(2)	(3)	(4)	(5)
Log of elapsed date	1.11 (0.13)**	1.10 (0.13)**	1.10 (0.13)**	−0.01 (0.001)**	−0.004 (0.001)**
Amazon retail price	0.07 (0.01)**	0.07 (0.01)**	0.07 (0.01)**	0.0002 (0.0001)**	0.0002 (0.0001)**
Missing retail price	2.76 (0.25)**	2.76 (0.25)**	2.77 (0.25)**	−0.02 (0.004)**	−0.02 (0.003)**
Log of total reviews	0.27 (0.12)*	0.31 (0.12)*	0.30 (0.12)*	0.01 (0.002)**	0.01 (0.003)**
Dummy for any state reviews				−0.01 (0.004)**	−0.015 (0.005)**
Log of total state reviews				0.01 (0.002)**	0.01 (0.002)**
Average review valence nationwide	0.04 (0.23)	0.15 (0.24)	0.13 (0.24)	0.01 (0.005)*	0.01 (0.005)*
Average review valence statewide				−0.0003 (0.001)	−0.0004 (0.001)
Percent identity disclosure		−1.95 (0.92)*			
Percent real name disclosure nationwide					0.04 (0.02)*
Percent real name disclosure statewide					0.005 (0.002)**
Percent identity disclosure (products with low number of reviews)			−1.36 (1.012)		
Percent identity disclosure (products with intermediate number of reviews)			−1.82 (2.17)		
Percent identity disclosure (products with large number of reviews)			−10.84 (5.16)*		
R-squared	0.82	0.82	0.82	0.32	0.32
Adj *R*-squared	0.77	0.77	0.77	0.32	0.32
F	$F_{(5,2,473)} = 72.48$**	$F_{(6,2,472)} = 64.82$**	$F_{(8,2,470)} = 49.28$**	$F_{(8,532)} = 8.27$**	$F_{(10,532)} = 7.81$**

Notes. Columns 1, 2, and 3 in this table show whether increases in disclosure of identity-descriptive information is associated with an increases in national sales. The dependent variable in these regressions is log (Sales Rank), so a negative coefficient is interpreted as a decrease in Sales Rank or an increase in sales. Product level fixed effects control for average differences in product quality. Columns 4 and 5 show whether disclosure of identity-descriptive information is associated with an increase in regional sales, and whether that relationship is amplified when disclosure comes from reviewers in the same state. The dependent variable is a variable that is equal to 1 when a product is in the local top 10 in sales, so a positive coefficient is interpreted as an increase in the likelihood that a product will appear in the local top 10, or an increase in local sales. Product-location fixed effects control for differences in regional preferences for products. All models include robust standard errors in parentheses. Sample excludes products with fewer than 10 reviews and products missing release date information.

information is processed heuristically (Chaiken and Maheshwaran 1994, Petty et al. 1998, Simpson et al. 2000, Cohen 2003, Menon and Blount 2003, Chang 2004, Pornpitakpan 2004, Kang and Herr 2006). As expected, our control variables suggest that sales decrease as Amazon's price increases (Chevalier and Mayzlin 2006, Ghose et al. 2006).[10]

4.3.2. Disclosure of Geographic Location and Sales. In this section we use data from Amazon purchase circles to identify how prior posts from a geographic subcommunity (in this case, a state) will influence the probability that a book will be one of the

[10] When we run the same analysis with reviews in prior month only all of the results on the identity disclosure variables and their interactions with average ratings remain qualitatively the same.

top sellers in that location. We estimate the regression

$$
\begin{aligned}
TOP10RANK_{ijt} \\
= \alpha + \delta X_{jt} + \beta_1 \log(TOT - REVIEWS_{jt-1}) \\
+ \beta_2 AVG - VALENCE_{jt-1} \\
+ \beta_3 \log(TOT - STATE - REVIEWS_{ijt-1}) \\
+ \beta_4 AVG - STATE - VALENCE_{ijt-1} + \beta_5 NO \\
- STATE - REVIEWS_{ijt-1} + \gamma_1 PCT - REALNAME \\
- DISCLOSURE_{jt-1} + \gamma_2 PCT - STATE \\
- REALNAME - DISCLOSURE_{ijt-1} + \mu_{ij} + \varepsilon_{ijt}, \quad (3)
\end{aligned}
$$

where $TOP10RANK_{ijt}$ is a dummy variable equal to 1 when product j appears in the local top 10 list for location i during time t. As before, $AVG - VALENCE_{jt-1}$ is the average review valence while $AVG - STATE - VALENCE_{ijt-1}$ is the average valence of

reviews in the state. X_{jt} is a vector of product-specific attributes that is changing over time, while μ_{ij} is a product-location fixed effect that controls for average preferences for books across locations.[11] Our focus is on measuring the effects of local and nationwide disclosure on local sales.

Table 3 shows the descriptive statistics for our purchase circle data. The number of data points is very large (over 8.3 million observations) because of the way our data were constructed, with hundreds of products distributed across hundreds of locations over 10 months. A couple of things are worthwhile pointing out about these data. First, these are limited dependent variable models. For many product-locations, we never observe the product appearing in the top 10. Thus, while we have a very large number of observations, we have less variation in our data than, say, traditional regression models of sales. In these types of limited dependent variable models, it is not uncommon to have very low *R*-squared values (Athey and Stern 2002). Still, because we have a very large number of observations, we adopt a conservative approach and report results as statistically significant only when they are significant at the 5% level or better.

Second, we use a unique econometric identification strategy to uncover the relationship between reviews and sales. A common concern in measuring the relationship between reviews and sales is establishing causality: Better reviews may be correlated with sales because such products may have higher average quality. Similarly, products with more reviews may have more sales because they are more popular. We offer a new approach to identify the relationship between reviews and sales. Like Chevalier and Mayzlin (2006), we pursue a "difference in differences" approach; however, unlike this prior work we do not rely on the similarity of review systems in different electronic markets such as those sponsored by Amazon and Barnes and Noble. Instead, we use product-location fixed effects to control for average product quality and differences in product preferences

across geographic locations. Our results do not reflect differences in average preferences for books across locations that will be correlated with reviews and sales; such differences in preferences will be absorbed by our fixed effects. Further, we control for changes in product popularity over time through the variables $\log(TOT-REVIEWS_{jt})$ and $AVG-VALENCE_{jt}$. Identification in our model comes through changes in sales and changes in reviewer behavior within a state over time.

Third, in contrast to the sales rank regressions described in the prior section, we examine the effects of disclosing real name on local sales, rather than examining disclosure of real name or location. This is for identification purposes; we observe the state a reviewer is in only when location is disclosed, so we are unable to separately identify the effects of location disclosure from reviewers in the same location on local sales. As before, we also examine how the valence of reviews influences sales. However, our primary interest is in understanding how increases in the prevalence of reviewer disclosure of identity-descriptive information in reviews from the same geographic community influence sales as represented by the parameter γ_2 (Hypothesis 5).

Columns (4) and (5) in Table 6 show that increases in identity disclosure in the same state will have a significant impact on the likelihood that a book will appear in the top 10 in a purchase circle, providing support for Hypothesis 5: Column 5 shows that a 1 standard deviation increase in the percentage of real name disclosure from the same state increases the likelihood of a local top 10 appearance by 0.5 percentage points. In our sample the average likelihood of a top 10 appearance is 1.5%; thus this 1 standard deviation increase translates into a quantitatively significant 33.3% increase in likelihood of top 10 appearance when evaluated at the mean.

Columns (4) and (5) in Table 6 also show that increases in nationwide valence increase sales, however increases in statewide valence do not. This is true both when we exclude (Column 4) and include (Column 5) real name disclosure. Thus, in contrast to our results in §4.3.1, we have evidence that review valence will be positively related to sales. However, as noted above we must interpret these results with some caution because they reflect changes in the popularity of the book over time.

---

[11] We also cluster our standard errors within products because of repeated values of observations within products. We experimented with clustering our standard errors within product-states (and have run these results without clustering standard errors) and the results are qualitatively similar.

Overall, our results are consistent with the notion that community members are influenced by the social information reviewers disclose about themselves—in particular, product sales are greater when disclosure of identity-descriptive information is more prevalent. Furthermore, identity disclosure from reviewers in a particular geographic location is associated with product sales in that same geographic location over and above the effect of general reviewer identity disclosure.[12,13]

5. Discussion

Our results show the relevance of identity-based processes in shaping the information consumers provide in their product reviews and in shaping members' response to those reviews both with respect to peer recognition (i.e., helpful votes) and purchase decisions. Identity-based norm conformity patterns are evident in our data in the type of identity-relevant information that reviewers disclose in their product reviews, and identity-based heuristic information processing is evident in consumers' response to reviews. In sum, identity-related antecedents of reviewer self-disclosure appear to shape the community response to that disclosure.

[12] To test for any concerns about multicollinearity, we ran the purchase circle regressions in Equation (3) using models that only have the statewide or nationwide ratings, but not both, simultaneously, and the results suggest that this is not a concern. Furthermore, results using total disclosure rather than percent disclosure give qualitatively very similar results. These are available from the authors on request.

[13] We also find that an increase in the volume of national and state reviews is positively associated with sales. Moreover, note that while the sign on Amazon price in the PC regression is positive, interpreting the sign should be done with caution because it simply indicates that increases in price lead to increases in the probability that the book appears on the list of most popular books. Amazon does not disclose how it ranks the most popular books in its purchase circle (based on units sold or revenues), thus these results may capture how price increases lead to higher revenues. We also ran the same regressions using relative price (Amazon price minus list price normalized by list price) and found that—consistent with prior work by Forman et al. (2007) who use similar data—increases in relative price lead to a decline in local sales. Moreover, our main results on disclosure of identity-descriptive information are similar in this alternative regression.

Overall, our results are consistent with the notion that user-generated product reviews influence community members through a heuristic process in which community members focus on source characteristics (reviewer self-disclosure of identity-descriptive information) in addition to, and sometimes instead of, the content of the reviews themselves. Reviewer disclosure of identity-descriptive information was significantly and positively associated with both perceived helpfulness at the review level of analysis and sales at the product level of analysis. When community members could be thought of as considering the information in a single review from a single reviewer (as when members assign helpful votes to a particular review) the effect of disclosure is attenuated by the informative value of the review. In particular, reviewer disclosure of identity-descriptive information is a stronger predictor of perceived helpfulness when reviews were more equivocal than when they were unequivocal. However, when community members scan multiple reviews from multiple reviewers (as is likely when members make product purchase decisions), source characteristics predicted sales, but message content had less impact. Specifically, in the sales rank regressions when review valence and reviewer disclosure of identity-descriptive information were included in the model, reviewer disclosure was positively related to sales while review valence was not a significant predictor. These results are consistent with prior research, such as Duan et al. (2005) and Liu (2006), who did not find a significant relationship between review valence and sales, but our results differ somewhat from those obtained by Chevalier and Mayzlin (2006) who found a small but significant relationship. In the purchase circle regressions, our finding that nationwide review valence was positively associated with regional sales is more consistent with the results Chevalier and Mayzlin (2006) report.

Our findings support the notion that when faced with an overload of information in the form of numerous reviews from numerous reviewers, community members process information heuristically, using source characteristics as a convenient and efficient heuristic device on which to base their product purchase decisions. Prior research in the information processing literature is consistent with our findings (e.g., Chaiken and Maheshwaran 1994), but

some prior experimental work in which subjects are given a single message from a single source has found that message responses are predicted by an interactive combination of message source and message content (Pompitakpan 2004). In particular, subjects in those studies appear to weight message content by the credibility or expertise of the message source (Hass 1981, Gilly 1998)—a more complex and cognitively taxing form of information processing than attending to source cues and/or message content independently. Posthoc analyses of our data provide no evidence of a significant interactive relationship between reviewer identity disclosure and review valence on sales. It is possible that information overload in the Amazon context motivates members to process information heuristically rather than systematically.

To evaluate whether community members' reliance on heuristic processing of source characteristics was a response to information overload, we compared the effects of reviewer disclosure of identity-descriptive information for products that had numerous reviews to those relatively few. Splitting the products in our sample by tertile according to the number of reviews in the first month in our sample, we examined the coefficients for identity disclosure by tertile.[14] As shown in Column 3 of Table 6, the effect of information disclosure is montonically increasing by tertile. Reviewer disclosure of identity-descriptive information is significantly associated with greater sales when review volume is highest, but not when it is lower. While these results do not directly assess the form of information processing Amazon members used, they are consistent with the notion that people use more heuristic processing of source characteristics when information overload is high (Hansen and Haas 2001), contrasting with laboratory results where subjects view a single message from a single source.

Our findings are the first to explain the antecedents of reviewer disclosure of identity-descriptive information in online consumer product reviews. These results suggest that community members may be willing to

engage in an effortful process to search for, identify, and then adapt to existing community norms at least with respect to the information they provide about themselves. This has implications for recent research on common identity and common bond groups online (Postmes et al. 2001, Ren et al. 2007). 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 bonds and identities may be related to one another (Ren et al. 2007). Amazon reviewers and members are cooperatively interdependent, with reviewers providing information that other members value, and members granting peer recognition that reviewers value. Interdependence and a common purpose facilitate common identity rather than common bond (Postmes et al. 2000). Reviewer identity-descriptive information may be used to learn about a reviewer as a unique individual, but it may also be used to create social categorizations such as those based on geographic location, which enhance common identity rather than common bond. Future research may consider whether in online communities like Amazon, common bonds and identities are synergistic.

The present study fills several gaps in the literature relating product reviews to sales. First, our study highlights the complementary role of message source (i.e., reviewer) information over and above the role of message content (i.e., review) as a predictor of sales. Second, our study examines the economic value of user-generated content by analyzing why and how online reviews influence sales; namely, product reviews appear to be heuristically processed, with community members more likely to purchase products when they can identify and perhaps identify *with* reviewers. This relationship appears to hold regardless of the evaluative information about the product that the reviews contain.

A third contribution to the literature linking product reviews and sales is methodological. Specifically, in our regressions used to test Hypothesis 5 we use a unique strategy to measure the causal link between reviews and sales. As is well known, identification of the causal link between reviews and sales is problematic because both the number and valence of consumer reviews may be correlated with

---

[14] We tried different methods of splitting the sample, according to prior month reviews and using different thresholds (quartiles, quintiles, etc.). The results were qualitatively similar to those we present.

unobservable product quality. Prior research on the causal relationship between reviews and sales has used either a proxy variable strategy (Reinstein and Snyder 2005) or a difference-in-differences approach (Chevalier and Mayzlin 2006). Our research strategy uses cross-sectional variation in local sales and local reviews to identify the relationship between reviews and sales by differencing out time-invariant local preferences for books and using national reviews as a proxy variable for changes in perceived product quality over time.

Our paper also makes a contribution to the literature on the individual motivations and goals driving content creation in online communities. In particular, while prior research has generally focused on the personal value that information providers derive or expect to derive from their online contributions (Jeppesen and Fredricksen 2006, Ma and Agarwal 2007), our study demonstrates that there are meaningful community-level and economic implications of these activities in the form of community norms and product sales. Moreover, there may be a match between the identity-expressive goals of information providers and the identity-search goals of community members because community members grant helpful votes to reviewers who engage in persistent labeling (Ma and Agarwal 2007), and such peer recognition is valued by reviewers (e.g., Jeppesen and Fredricksen 2006).

Our findings extend the growing research on reputation systems (Resnick et al. 2000, Ghose et al. 2005) by exploring some important antecedents of peer recognition and reputation. While prior research has addressed the consequences of reputation systems for both contributors who are rated (Jeppesen and Fredricksen 2006, Moon and Sproull 2006) and consumers who use the ratings for evaluation purposes (Pavlou and Gefen 2004), little research has addressed what factors predict the valence of those ratings. Our research suggests that reviewer disclosure of identity-descriptive information and review equivocality shape peer recognition. Moreover, we build theory and insight by explaining how and why reviewer disclosure and review equivocality combine to predict peer recognition, with the effects of one type of information contingent on the level of the other type of information. These findings are important because

peer recognition and reputation is one of the ways that network-based virtual communities self-regulate, i.e., draw members into the process of monitoring one another and defining what behaviors are valued and encouraged or harmful and discouraged (Pavlou and Gefen 2004, Moon and Sproull 2006).

5.1. Limitations
One limitation of the present paper is that while identification processes are the hypothesized mechanism explaining the patterns that we found, our data does not directly measure the level of Amazon member identification. Future research may fruitfully be directed at evaluating the identity perceptions underlying our findings and may allow the use of alternative analysis techniques, such as structural equations modeling, that were not feasible with our data.

We use product and product-location fixed effects to control for unobservable differences in product quality and consumer preferences across location. Thus, identification of our hypotheses arises from changes in disclosure over time within a product or within a product location. One limitation is that such fixed effects are unable to control for changes in product popularity.

Another limitation of the paper is the nature of the data set itself. Some of the variables in our data are proxies for the actual measure that one would need for more advanced empirical modeling. Future research may fruitfully be devoted to analysis of the text of reviews to more precisely measure review quality and the polarity of review content, such as by combining automated textmining and sentiment analysis with econometrics (Ghose et al. 2005, Das and Chen 2007) or using decision trees to map nonlinear interaction effects between product sales and user-generated content (Dhar and Chang 2007).

Our sample is also restricted. Our analysis focuses on the sales of books (i.e., experience goods) at one e-commerce retailer. The influence of reviewer information on sales may be different for a different sample of products (e.g., products other than books like search goods) for a different retailer. Additional work in other online contexts will be needed to evaluate whether reviewer self-disclosure has antecedents and consequences that are similar to those we obtained. Also, while our results are qualitatively strong in general, there are exceptions such as the impact of prior

disclosure of location information on the likelihood of subsequent location disclosure.

6. Managerial Implications

One of the most consistent findings of our study is the value of reviewer disclosure of identity-descriptive information in encouraging online product sales. This finding suggests that online retailers may be able to increase sales on their site by taking actions to encourage reviewers to reveal more identity-descriptive content about themselves. Furthermore, we find that community norms observable in prior reviews shape reviewer disclosure. Identity disclosure norms are likely to be more influential when members can easily identify the norms, suggesting that identity disclosure and sales may be enhanced when, for example, reviewer identity-descriptive information is visually aligned so community members can more easily detect patterns in reviewer disclosure.

The results of our purchase circle regressions suggest that consumers are responsive to the level of disclosure in reviews identified as coming from their geographic location. Our results suggesting that offline geographical location can influence consumers' online behavior are consistent with prior work showing that geography influences consumers' purchases in electronic markets (Forman et al. 2007). These findings suggest that online retailers may benefit from customizing the order of reviews that individual members see in the forum. For example, members who log in from a particular location may derive greater value from easier access to reviews from their same geographic location. Also, consumers may gravitate to sites that allow them to sort reviews based on particular reviewer characteristics, a tool that is currently unavailable on many sites.

Our study suggests that when making product purchase decisions, community members use source characteristics as a heuristic cue, saving them the effort and time of interpreting message content in a more fine grained manner. Recognizing the important role of source characteristics in the purchase and peer rating process, web designers may benefit from carefully considering the tradeoffs between the prominence of product information and source characteristics. For example, given space constraints, it may be important to consider whether the text of

reviews or the identity of reviewers is featured more prominently.

7. Conclusion

Our findings suggest that in the world of online consumer reviews and sales, we may need a more nuanced understanding of the old adage that "there is no such thing as bad publicity." Our findings suggest that negatively valenced "publicity" may not necessarily be bad for sales. Bad publicity may exist, however, in the form of anonymous reviews that deprive community members of the type of social information they can use to create a means of identification, reinforce community norms, and provide a heuristic enabling them to more easily evaluate a reviewer and a product, and to make purchases that reflect this evaluation.

References

Abrams, D., M. Wetherall, S. Cochrane, M. A. A. Hogg, J. C. Turner. 1990. Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization. *British J. Psych.* **29** 97–119.

Aiken, L. S., S. G. West. 1991. *Multiple Regression: Testing and Interpreting Interactions.* Sage, Newbury Park, CA.

Akerlof, G. A., R. E. Kranton. 2000. Economics and identity. *Quart. J. Econom.* **115**(3) 715–754.

Athey, S., S. Stern. 2002. The impact of information technology on emergency health care outcomes. *RAND J. Econom.* **33**(3) 399–432.

Bartel, C. A. 2006. Negotiating membership status in organizational groups. Working paper.

Bartel, C. A., J. E. Dutton. 2001. Ambiguous organizational memberships: Constructing social identities in interactions with others. M. A. Hogg, D. Terry, eds. *Social Identity Processes In Organizational Contexts.* Psychology Press, Philadelphia, 115–130.

Baumeister, R. F., M. R. Leary. 1995. The need to belong: Desire for interpersonal attachment as a fundamental human motivation. *Psych. Bull.* **117**(3) 497.

Brewer, M. 1991. The social self: On being the same and different at the same time. *Personality Soc. Psych. Bull.* **17**(5) 475–482.

Brown, J., P. H. Reingen. 1987. Social ties and word-of-mouth referral behavior. *J. Consumer Res.* **14**(12) 350–362.

Brown, J., A. Broderick, N. Lee. 2007. Word of mouth communication within online communities: Conceptualizing the online social network. *J. Interactive Marketing* **21**(3) 2–20.

Byrne, D. 1971. *The Attraction Paradigm*. Academic Press, New York.

Chaiken, S. 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *J. Personality Soc. Psych.* **39**(5) 752–766.

Chaiken, S. 1987. The heuristic model of persuasion. M. Zanna, J. Olson, C. Herman, eds. *Social Influence*: *The Ontario Symposium*, Vol. 5. Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ, 3–39.

Chaiken, S., D. Maheswaran. 1994. Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *J. Personality Soc. Psych.* **66**(3) 460–473.

Chang, C. C. 2004. Country of origin as a heuristic cue: The effects of message ambiguity and product involvement. *Media Psych.* **6**(2) 169–192.

Chevalier, J., D. Mayzlin. 2006. The effect of word of mouth online: Online book reviews. *J. Marketing Res.* **43**(3) 345–354.

Cohen, G. 2003. Party over policy: The dominating impact of group influence on political beliefs. *J. Personality Soc. Psych.* **85**(5) 808–822.

Das, S. R., M. Chen. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Sci.* **53**(9) 1375–1388.

David, B., J. C. Turner. 1996. Studies in self-categorization and minority conversion: Is being a member of the out-group an advantage? *British J. Soc. Psych.* **35**(1) 179–199.

Dellarocas, C., N. Awad, M. Zhang. 2005. Using online ratings as a proxy of word-of-mouth in motion picture revenue forecasting. Working paper.

Dhar, V., E. Chang. 2007. Does chatter matter? The impact of user-generated content on music sales. CeDER Working Paper 07-06, New York University.

Donath, J. 1999. Identity and deception in the virtual community. M. A. Smith, P. Kollock, eds. *Communities in Cyberspace*. Routledge Publishers, London, 29–59.

Duan, W., B. Gu, A. B. Whinston. 2008. The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *J. Retailing.* Forthcoming.

Dutton, J. E., J. M. Dukerich, C. V. Harquail. 1994. Organizational images and member identification. *Admin. Sci. Quart.* **39**(2) 239–264.

Ehrens, S., A. Markus. 2000. Amazon.Com: There's an "R" in E-Tailing. Epoch Partners Internet Company Report, November 13, 4.

Festinger, L., S. Schacter, K. Back. 1950. *Social Pressures in Informal Groups: A Study of Human Factors in Housing*. Stanford University Press, Palo Alto, CA.

Forman, C., A. Ghose, A. Goldfarb. 2008. Competition between local and electronic markets: How the benefit of buying online depends on where you live. *Management Sci.* Forthcoming.

Ghose, A., P. Ipeirotis, A. Sundararajan. 2005. The dimensions of reputation in electronic markets. Working paper, New York University.

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

Gilly, M. C. 1998. A dyadic study of interpersonal information search. *Acad. Marketing Sci.* **26**(2) 83–100.

Grohol, J. 2006. Anonymity and online community identity matters. http://www.Alistapart.Com/Articles/Identitymatters.

Hansen, M., M. Haas. 2001. Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. *Admin. Sci. Quart.* **46**(1) 1–28.

Hass, R. G. 1981. Effects of source characteristics on cognitive responses and persuasion. R. E. Petty, T. M. Ostrom, T. C. Brock, eds. *Cognitive Responses in Persuasion*. Lawrence Erlbaum, Hillsdale, NJ, 141–172.

Hornsey, M. J., J. Jetten. 2004. The individual within the group: Balancing the need to belong with the need to be different. *Personality Soc. Psych. Rev.* **8**(3) 248–264.

Hsiao, C. 2003. *Analysis of Panel Data*. Cambridge University Press, Cambridge, UK.

Jarvenpaa, S. L., D. E. Leidner. 1999. Communication and trust in global virtual teams. *Organ. Sci.* **10**(6) 791.

Jeppesen, L. B., L. Fredricksen. 2006. Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organ. Sci.* **17**(1) 45–63.

Jones, E. E., T. S. Pittman. 1982. Toward a general theory of strategic self-presentation. J. Suls, ed. *Psychological Perspectives on the Self*, Vol. 1. Lawrence Erlbaum, Hillsdale, NJ, 231–262

Kang, Y. S., P. M. Herr. 2006. Beauty and the beholder: Toward an integrative model of communication source effects. *J. Consumer Res.* **33**(1) 123–130.

Kelman, H. C. 1961. Processes of attitude change. *Public Opinion Quart.* **25** 57–78.

Kruglanski, A. W., X. Y. Chen, A. Pierro, L. Mannetti, H. P. Erb, S. Spiegel. 2006. Persuasion according to the unimodel: Implications for cancer communication. *J. Comm.* **56**(1) S105–S122.

Latane, B. 1981. The psychology of social impact. *Amer. Psychologist* **36** 343–356.

Lea, M., R. Spears, D. Degroot. 2001. Knowing me, knowing you: Anonymity effects on social identity processes within groups. *Personality Soc. Psych. Bull.* **27**(5) 526–537.

Lecky, P. 1945. *Self-Consistency*: *A Theory of Personality*. Island Press, New York.

Liu, Y. 2006. Word-of-mouth for movies: Its dynamics and impact on box office receipts? *J. Marketing* **70**(3) 74–89.

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

Mackie, D., L. Worth, A. Asuncion. 1990. Processing of persuasive in-group messages. *J. Personality Soc. Psych.* **58**(5) 812–822.

McKenna, K. Y., J. A. Bargh. 1999. Causes and consequences of social interaction on the internet: A conceptual framework. *Media Psych.* **1** 249–270.

Menon, T., S. Blount. 2003. The messenger bias: A relational model of knowledge valuation. *Res. Organ. Behav.* **25** 137–186.

Moon, J. Y., L. Sproull. 2008. The role of feedback in managing an Internet-based volunteer workforce. *Inform. Systems Res.* **19**(4).

Pavlou, P., D. Gefen. 2004. Building effective online marketplaces with institution-based trust. *Inform. Systems Res.* **15**(1) 37–59.

Petty, R. E., D. T. Wegener, P. H. White. 1998. Flexible correction processes in social judgment: Implications for persuasion. *Soc. Cognition* **16**(1) 93–113.

Pornpitakpan, C. 2004. The persuasiveness of source credibility: A critical review of five decades' evidence. *J. Appl. Soc. Psych.* **34**(2) 243–281.

Postmes, T., R. Spears, M. Lea. 2000. The formation of group norms in computer-mediated communication. *Human Comm. Res.* **26**(3) 341–371.

Postmes, T., R. Spears, A. T. Lee, R. J. Novak. 2005. Individuality and social influence in groups: Inductive and deductive routes to group identity. *J. Personality Soc. Psych.* **89**(5) 747–763.

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

Reinstein, D., C. Snyder. 2005. The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *J. Indust. Econom.* **53**(1) 27–51.

Ren, Y., R. E. Kraut, S. Kiesler. 2007. Applying common identity and bond theory to design of online communities. *Organ. Stud.* **28**(3) 377–408.

Resnick, P., R. Zeckhauser, E. Friedman, K. Kuwabara. 2000. Reputation systems. *Comm. ACM* **43**(12) 45–48.

Sassenberg, K. 2002. Common bond and common identity groups on the Internet: Attachment and normative behavior in on-topic and off-topic chats. *Group Dynam.* **6**(1) 27–37.

Senecal, S., J. Nantel. 2004. The influence of online product recommendations on consumers' online choices. *J. Retailing* **80**(2) 159–169.

Simpson, E., T. Snuggs, T. Christiansen, K. E. Simples. 2000. Race, homophily, and purchase intentions and the black consumer. *Psych. Marketing* **17**(10) 877–889.

Spears, R., M. Lea. 1992. Social influence and the influence of the "social" in computer-mediated communication. M. Lea, ed. *Contexts of Computer-Mediated Communication.* Harvester-Wheatsheaf, London, 30–65.

Sproull, L. 2003. *Online Communities. The Internet Encyclopedia.* H. Bidgoli, ed. John Wiley & Sons, New York.

Swann, W. B. 1983. Self-verification: Bringing social reality into harmony with the self. J. Suls, A. G. Greenwald, eds. *Social Psychological Perspectives on the Self*, Vol. 2. Lawrence Erlbaum, Hillsdale, NJ, 33–66.

Tajfel, H., J. C. Turner. 1979. An integrative theory of intergroup conflict. W. G. Austin, S. Worchel, eds. *The Social Psychology of Intergroup Relations.* Brooks/Cole, Monterey, CA, 33–47.

Turkle, S. 1996. Our virtual selves. *Boston* **88**(11) 78–81.

Turner, J. 1987. *Rediscovering the Social Group: A Self-Categorization Theory.* Blackwell, Oxford.

Tyler, T. T., P. Degoey, H. Smith. 1996. Understanding why the justice of group procedures matters: A test of the psychological dynamics of the group-value model. *J. Personality Soc. Psych.* **70**(5) 913–930.

Walther, J. B. 1992. Interpersonal effects in computer-mediated interaction: A relational perspective. *Comm. Res.* **19**(1) 52–90.

Walther, J. B. 1995. Relational aspects of computer-mediated communication: Experimental observations over time. *Organ. Sci.* **6**(2) 186–203.

Walther, J. B. 1996. Computer-mediated communication: Impersonal, interpersonal and hyperpersonal interaction. *Comm. Res.* **23**(1) 1–43.

Walther, J. B., J. K. Burgoon. 1992. Relational communication in computer-mediated interaction. *Human Comm. Res.* **19**(1) 50–88.

Wetzer, I. M., M. Zeelenberg, R. Pieters. 2007. Never eat in that restaurant, I did!: Exploring why people engage in negative word-of-mouth communication. *Psych. Marketing* **24**(8) 661–680.

Wiesenfeld, B. M., S. Raghuram, R. Garud. 1999. Communication patterns as determinants of organizational identification in a virtual organization. *Organ. Sci.* **10** 777–790.

Wood, W. 2000. Attitude change: Persuasion and social influence. *Annual Rev. Psych.* **51** 539–570.

Wooldridge. 2002. Econometric analysis of cross section and panel data. MIT Press, Cambridge, MA.

Xia, L., N. Bechwati. 2006. Positive word of mouse: The role of personalization. *Proc. Amer. Marketing Assoc.* **17** 107.