A Multi-Level Examination of the Impact of Social Identities on Economic Transactions in Electronic Markets[1]

Chris Forman
Tepper School of Business
Carnegie Mellon University
cforman@andrew.cmu.edu

Anindya Ghose
Stern School of Business
New York University
aghose@stern.nyu.edu

Batia Wiesenfeld
Stern School of Business
New York University
bweisenf@stern.nyu.edu

Abstract

Three of the most important uses of the Internet today are as an economic marketplace, as a forum for social interaction, and as a source of information. In this paper, we explore how these three activities come together, in the form of emergent social communities built around information exchanges within IT-enabled electronic marketplaces. Drawing on social identity theory, we suggest that the relationship between online consumer reviews and internet product sales is partially explained by social identity processes. Using a unique dataset 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 at the community level linking the prevalence of identity claiming behavior in an online community with subsequent product sales. In addition, we show that when reviewers claim to be from a particular geographic location, subsequent product sales are higher in that region. At the review level of analysis, we show that subsequent reviews conform to identity-claiming norms set in previous reviews, and that identity claiming that conforms to community norms elicits identity granting. Furthermore, our results suggest that the prevalence of identity granting has implications for economic exchange in the form of product sales. Implications for research on word-of-mouth and electronic communities are discussed.

Keywords: *Digital Markets, Social Identity, Online Reviews, Internet Retailing, Virtual Communities, Social Exchange.*

---

1. Introduction

The Internet has had a profound impact on at least three areas of life – the way people shop, the way they socially interact, and the way they exchange information. All three are relevant to consumer product reviews posted in IT-enabled electronic markets. Online consumer product reviews provide information that can facilitate economic exchange, which is the main purpose of electronic marketplaces. However, reviews are also sometimes used as a forum for social exchange, and the social exchange, in part, serves to draw people to such websites, promote purchases, and regulate user behavior. In these ways, geographically-dispersed electronic marketplaces have something in common with corner markets in small villages – social exchanges and economic exchanges are interwoven, and facilitate one another.

This paper provides a multi-level exploration of the important role of social communities in geographically-dispersed electronic marketplaces. We suggest that in addition to providing informational content, online consumer product reviews may be conceptualized as a form of social exchange. Social exchange in the context of consumer product reviews posted on electronic forums may be most apparent in reviewers' disclosure of information about themselves unrelated to the products they review. In this paper, we use social identity theory (Tajfel and Turner, 1979) as a conceptual lens to explore how the online community associated with electronic markets influences the way that reviewers present personal information about themselves, and to investigate the influence that their self-presentations have on social and economic exchanges in the community.

By exploring virtual communities that emerge based on individuals' relationships to specific economic products, the present research provides a powerful means of examining the dynamic interplay between social and economic exchange on the Internet. We contribute to the existing literature by bridging two disparate streams of research – work that explores the informational influence of online reviews on economic behavior, such as how more positive online reviews

increase sales (Dellarocas et al., 2005; Reinstein and Snyder, 2005; Chevalier and Mayzlin, 2006), and

work exploring the social dynamics (a) motivating people to participate in online communities by

posting "electronic word-of-mouth" (e.g., Balasubramanian and Mahajan, 2001; Hennig-Thurau,

Gwinner, Walsh and Gremler, 2004) and (b) influencing other consumers' social response to

postings (Arguello et. al., 2006; Xia and Bechwati, 2006).[2]

Our paper examines the interrelationships between macro-level virtual community attributes

and outcomes (Cummings, Butler and Kraut, 2002; Pavlou and Gefen, 2004) and micro-level

dynamic and reciprocal patterns of social exchange among members (Bartel, 2006). In so doing, our

study leverages findings regarding community dynamics to demonstrate the economic importance of

social exchange, and utilizes findings concerning reviewer dynamics to understand the theoretical

mechanisms underlying community-level effects, answering the question of <u>why</u> reviews lead to

sales. We suggest that social identification and self-categorization processes characterize online

consumer reviews of books posted on electronic marketplaces, arguing that book sales reflect, in

part, participation in an online consumption community. Our results suggest, first, that community

norms regarding reviewer disclosure of personal information are positively related to economic

outcomes in the form of online book sales. This influence is over and above that of the

informational content of consumer reviews (i.e., the positivity, entropy or valence of the ratings) that

have been examined in prior research (Godes and Mayzlin 2004; Dellarocas et al., 2005; Chevalier

and Mayzlin, 2006).

Second, we provide evidence consistent with the notion that online book sales reflect

participation in the consumption community by demonstrating that self-disclosure by individual

---

[2] More broadly the latter stream of research is related to why people contribute to electronic communities of practice. Work in this context includes Wasko and Faraj, (2000), Gu and Jarvenpaa (2003), Subramani and Peddibhotla (2002) and Wasko and Faraj (2005). By considering knowledge as a public good, Wasko and Faraj, (2000) point out that knowledge exchange is motivated by moral obligation and community interest rather than by narrow self-interest. Gu and Jarvenpaa (2003) find that self-interest, reciprocity and identity are three concepts associated with voluntary contributions associated with shared databases. Wasko and Faraj (2005) discuss why individuals help strangers in electronic networks.
.

reviewers is patterned after group norms. Third, our results show that the community legitimates members' norm conforming behaviors by responding with support in the form of helpful votes. This contributes to the identity literature by providing quantitative empirical evidence demonstrating social confirmation of individual attempts to communicate their identity online. Finally, we show that consumption communities characterized by higher levels of social support for member contributions are associated with greater book sale.

Defining terms.

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). In the context of online consumer product reviews in electronic markets, at least two levels of community may exist. At the broadest level, community members may be all of those who post or consume product reviews on a particular site. Membership in this type of community may be defined by actions such as registering on the site (Grohol, 2006). The broad community may play host to a multitude of smaller communities whose members share an experience with a particular economic product such as a specific book, movie or music CD – a group that fulfills the definition of a *consumption community* (Fischer, Bristor and Gainer, 1996). At a minimum, to be a member of the consumption community means that people consume the product that unites the consumption community (e.g., read the book that is rated on Amazon). The consumption community may be further divided into smaller subunits of people who share additional defining attributes (e.g., those Amazon members reading and reviewing a particular book who are all from the same geographic area).

Member identification with these online consumption communities occurs when community participants think of themselves as members of the group; i.e., when the community becomes self-referential (Tajfel and Turner, 1979). Prior research suggests that shared consumption of a particular

product or attachment to a brand may elicit social identification with that product or brand (e.g., Muniz and O'Guinn, 2001; Bhattacharya and Sen, 2003).

While identification is a cognition about the self, it is manifested in social behavior. In particular, people may behave in ways that signal to others that they are members of the group. Social assertions that signal an individuals' identity to others are called *identity claims* (Bartel and Dutton, 2001; Bartel, 2006). In this paper, we focus on the most observable form of identity claims in online consumer reviews – the disclosure of personal information about the reviewer. Members' identity claims play a valuable role in communities because they are public and observable both to members of the community and outsiders. Identity claims not only signal that a member fits into and thus deserves to be viewed as a part of the community, but they also implicitly communicate *community norms*, defined as behaviors that are expected and appropriate in the community, and the community *prototype*, defined as cognitive representations of the group, usually in the form of attributes or actions (Turner, 1987). Thus, in aggregate members' identity claims help to define the community as a social entity by serving as both the basis for and response to community norms and prototypes (Bartel, 2006).

Successful identity claims are those that yield *identity grants*, defined as interaction partners' social verification of an individuals' identity (Bartel and Dutton, 2001; Bartel, 2006). Identity claiming and granting is a means by which individuals can convey or confirm the perception that they are a member in good standing within the community. In digital markets identity grants are institutionalized in the form of peer ratings, or helpful votes. Thus, in our paper identity claiming and granting are micro-level processes that occur between people, where the level of analysis is the product review. However, these micro-level processes are situated within online consumption communities, and thus provide insight into the dynamics linking community attributes to product sales (Pavlou and Gefen, 2004).

2. Theory and Hypotheses

2.1. Community-Level Processes: Norms and Member Participation

In offline markets, consumers' purchase decisions are frequently influenced by word-of-mouth (Gilly, 1998). Computer-mediated communication facilitated the migration of these conversations to the Internet (Dellarocas, 2003). Recently, research attention has been directed toward understanding the relationship between online consumer product reviews and customer purchase behavior (e.g., Dellarocas et al., 2005; Chevalier and Mayzlin, 2006). To date, however, research concerning the relationship between online word-of-mouth and sales has been limited in taking an exclusively informational perspective – that is, consumer product reviews are viewed as a source of information, and it is the informational value of the product reviews that are expected to influence sales. For example, prior research has found that more positive reviews are associated with greater sales (Godes and Mayzlin, 2004; Dellarocas et al., 2005, Chevalier and Mayzlin, 2006).

While it is clear that the informational value of consumer product reviews does influence sales, social processes may also play a role in the relationship between reviews and product purchase decisions. Specifically, electronic marketplaces provide an opportunity for the emergence of consumption communities – communities within the larger digital market whose members have in common the consumption of a particular product (e.g., Fischer, et al., 1996) and the desire to share, through either posting or reading reviews, their opinions of that product. If consumption of the product serves as the "entry ticket" qualifying people for membership in the community, then sales may reflect community participation. We also suggest that reviews of the product provide information about the community and its membership. One signal of the strength of a community is the prevalence of identity claiming behavior among its members (Bartel, 2006). Stronger consumption communities, in which identity claiming is more prevalent, should motivate people to

join and participate in the consumption community (e.g., through purchasing the product) for several reasons.

Theory suggests that the relationship between sales and online word-of-mouth is embedded in social relationships. In offline contexts, word-of-mouth is socially embedded in identity groupings (Brown and Reingen, 1987; Dellarocas, 2003). Identity concerns and the desire to communicate identity are also prominent motivators influencing the propensity to communicate online in general, and to provide online product reviews in particular (McKenna and Bargh, 1999; Balasubramanian and Mahajan, 2001; Bagozzi and Dholokhia, 2002; Gu and Jarvenpaa, 2003; Dholakia et al., 2004; Hennig-Thurau et al., 2004; Moon and Sproull, 2006). Product reviewers in electronic marketplaces invest substantial energy in projecting identity claims online (Bush and Tiwana, 2005). Identity claims are perhaps most apparent in the personal information that reviewers disclose about themselves that is irrelevant to the products they review. For example, Amazon's top reviewers project not only their opinions of products, but also personal information ranging from the names of their pets (e.g., cats names "Sashi" or "WYSIWYG") to dreams that they have (e.g., "walking in fields of lavender"). They invest in creating and elaborating online profiles indicating their nickname, their hobbies and interests, and their geographic location, and they take steps such as posting credit card information to qualify for a badge that signifies they are using their "Real name".[3] Such self-disclosure publicly communicates the identity they wish others to associate with them.

As the most obvious representatives of the online consumption community, reviewers' identity claims communicate at least two types of information that may motivate others to participate in the community. First, self-disclosure helps potential consumers to feel that members of the community are familiar, increasing feelings of trust (Resnick et al., 2000; Pavlou and Gefen, 2004; Xia and Bechwati, 2006). Familiarity and trust are the basis for psychological bonds motivating

---

[3] "Real name" is the label provided by Amazon to reviewers who provide credit card information when posting information online. It is a signal to demonstrate that reviewer information is truthful and accurate.

potential consumers to become part of the community, which in this instance involves purchasing

the product on which community membership is based.

Second, because identity claiming requires the investment of effort, the prevalence of

identity claiming behavior signals the regard members hold for the community that makes

membership worthy of effort. If the prevalence of identity claiming is evidence of the perceived

value of community membership, it should be positively associated with participation in the

community in the form of product purchase. Thus, in a social identity model, identity claiming may

serve as a proxy for the perceived psychological benefits of community membership (Butler, 2001).

Hypothesis 1a: *Increases in disclosure of self-descriptive information (i.e., revealing Real name, nickname,*
hobbies or geographic location) will be positively associated with subsequent sales of the product.

Certain identity claims should be especially appealing to particular potential community

participants. Specifically, people are likely to examine the content of identity claims to infer how

easily their own attributes would fit the community norms and prototypes that those identity claims

convey. They may compare these attributes to those that are prototypical of the community, with a

closer match increasing the appeal of community membership and thus eliciting participation

(Byrne, 1971; Turner, 1987). Thus, if social identity processes are operative, there should be a match

between the content of reviewer self-disclosures and the attributes of subsequent customers because

self-disclosure functions as identity claims.

In the context of online consumer reviews of books, not all categories of self-disclosure are

equally appealing as the basis of similarity between group members. For example, people are unlikely

to feel pressure to have the same first and last names as other members of the community because

names are highly variable and people are accustomed to thinking of their name as a source of

personal, rather than collective identity. However, they may feel more confident of their fit with the

community if members of the community claim to be from their geographic region because

geography is a more natural basis for social community (Festinger et al., 1950). The relationship between buying behavior in a region and identity claiming from people of that region should exist over and above any spurious similarities in product preferences among people from the same geographic region.

Hypothesis 1b: *An increase in the number of reviews from people who identify that they are from a given geographical location will be positively associated with subsequent sales of that product in that geographic location.*

2.2 Review-Level Processes

While purchasing the product may qualify an individual for membership in the consumption community, research has found that membership is not merely a question of being in or out, but also it is a matter of degree – people desire social affirmation of their membership status, which may lead them to take proactive efforts to assert their identity in the form of identity claims (Bartel and Dutton, 2001; Swann, 1987). As we have suggested above, posting a review is an opportunity to disclose personal information that can serve as the assertion of identity claims.

If the identification processes we hypothesized are responsible for relationships between reviewer self-disclosure and product sales, then patterns of posting and rating of reviews consistent with social identity processes should emerge. In particular, if self-disclosures in product reviews function as identity claims, then we would expect these claims to be patterned in ways that are consistent with community norms because norm conformity communicates that one is a member of the group. Similarly, we would expect to see evidence that the community response to these identity claims is consistent with social processes, such as that identity claims that conform to norms are socially legitimated in the form of identity granting. Finally, the rate of successful identity claiming should make the community more attractive, and therefore should be associated with subsequent community participation in the form of product purchases. Below, we consider the relationship

between dynamics at the product review level of analysis and those at the community level of analysis argued above.

2.2.1 Group Norms and Identity Claiming Self-Disclosure

Research suggests that perhaps the strongest signal of the existence of an identity group is evidence of norms shaping member behavior (Postmes et al., 2000). Adapting oneself to group norms is one way that members can communicate to others in the community as well as those outside the community that one is a full-fledged community member (Bartel, 2006). Thus, norm conformity is one of the clearest indicators of both the existence of an identity-based community and members' desire to claim membership in that community.

Patterns of self-disclosure in online consumer product reviews provide inferences about the existence of norms. Consumer product reviews can be entirely anonymous, thus exhibiting little social content. Alternatively, reviewers can disclose personal information but not follow consistent patterns or norms in the community, thus projecting reviewers' personal identities but no community identification. This satisfies the desire of the individual to express an authentic identity (McKenna and Bargh, 1999) but is less likely to yield social affirmation of that identity because these identity claims are less interpretable by interaction partners. Finally, reviewers' self-disclosure can follow group norms (e.g., consistency in the types or categories of information members disclose about themselves), providing evidence of community identification.[4] Such conformity patterns are easily interpreted by interaction partners, help establish the reviewer's reputation (Resnick et al., 2000), and thus are more likely to yield social affirmation. Moreover, these norms may be easily

---

4 Self-disclosure norms, in particular, are especially valuable in groups because they satisfy two simultaneous needs – the need to belong and the need to feel special and unique (Brewer, 1991). Research suggests that people wish to be part of a group of which they can be proud, but also be affirmed as individuals by other members of that group (Brewer, 1991; Tyler, Degoey and Smith, 1994). Norms emphasize similarity and thus reinforce a sense of belonging. To address the desire to feel special, group members will wish to express their personal characteristics. Both needs are satisfied when people identify with groups that normatively prescribe autonomy and individualism (Hornsey and Jetten, 2004). Put differently, individuals are more motivated to participate in communities having norms institutionalizing the ways that members express their authentic self (McAuliffe, Jetten, Hornsey and Hogg, 2003a, 2003b; Sheldon and Bettencourt, 2002).

inferred from archives of previous reviews (Postmes et al., 2000; Dholakia et al., 2004), which are likely to be available and salient in a particular product consumption community (i.e., in prior reviews posted regarding a particular product). Thus, if self-disclosure is a form of identity claiming, there should be a positive association between the prevalence of particular types of self-descriptive reviewer information in the consumption community and the disclosure of similar types of information among subsequent product reviewers.

Hypothesis 2a: *The prevalence of self-descriptive reviewer information (i.e., revealing Real name, nickname, hobbies or geographic location) in recent reviews of a product will be positively associated with identity claiming in the form of the disclosure of similar self-descriptive information by subsequent reviewers.*

Above, we pointed out that aggregate identity claims are even more influential among those who are likely to perceive themselves as more easily fitting the community norms and prototypes – such as those who are from a geographical region well-represented in the community of reviewers. If such social processes are in fact operative, they should also be manifested in whether consumers from regions well-represented by reviewers choose to claim membership status by posting a review. This should be over and above the popularity of the product in the region. Thus, we expect:

Hypothesis 2b: *The prevalence of reviews from a particular geographical location in the consumption community will increase the likelihood that reviews will be posted from the same geographic location in subsequent time periods.*

2.2.2 Identity Granting and Review Ratings

Above, we suggest that online consumer reviews will reflect identity claims in that self-disclosure will be patterned according to group norms, and thus consistent among members of a consumption community. We now consider the processes through which these claims are either affirmed or denied (Swann, 1987; Bartel and Dutton, 2001). Research suggests that a group or community's identity and members' status within that community are the product of social exchange

(Swann, 1987; Bartel, 2006). Specifically, in addition to members *claiming* the social identities that they desire through proactive efforts, the community may affirm those social assertions (labeled identity *granting*) or deny the identity claims (Bartel and Dutton, 2001).

Identity granting is one of the ways that network-based virtual communities self-regulate; that is, 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). Behaviors that conform to community norms may be more successful types of identity claiming because conformity reinforces norms, clarifies the community identity and facilitates coordination (McKenna and Bargh, 1999; Postmes et al., 2004; Xia and Bechwati, 2006). Put differently, if reviewer self-disclosure is a form of identity claiming, and identity claims are more successful in eliciting identity granting when they conform to community norms, then it follows that reviewer self-disclosure that follows group norms will be more likely to yield identity granting.[5]

Perhaps the most important type of response signaling identity granting is positive recognition and approval from others (Dholakia et al., 2004; Jeppesen and Fredricksen, 2006). In user communities associated with online economic marketplaces, membership grants and the social approval they convey may be highly visible, such as in the form of quantitative and qualitative feedback provided to those who assert identity claims (Moon and Sproull, 2006). For example, Amazon has a voting system whereby community members can provide helpful votes to rate the reviews of other community members. The helpfulness ratings then become a visible form of social approval that is presented alongside the reviews themselves, visually juxtaposing identity claims and grants for all to see (Moon and Sproull, 2006).

---

[5] Consistent with this logic, prior research has found an association between the content of postings and the likelihood that they receive a response from the community (Arguello et. al., 2006). Newcomers, who are less likely to be aware of community norms, are less likely to receive a response (Arguello et. al., 2006).

Hypothesis 3*: Identity claiming in the form of disclosure of self-descriptive information by reviewers (i.e., revealing Real name, nickname, hobbies or geographic location) will be positively associated with identity granting by others in the form of helpful votes.*

2.3 Community Ratings and Product Sales

Aggregating from the individual review to the consumption community level of analysis, when individual reviews have a high likelihood of receiving helpful votes in a particular consumption community, people may attribute positive qualities to that consumption community, making them more likely to participate in it Research suggests that the way a group treats its members provides information about the group's identity, and groups that are more responsive to members elicit greater commitment (Tyler et al., 1996; Arguello et. al., 2006). Identity granting in the form of helpful votes is a sign that the community treats members in a responsive manner, which is likely to attract new members to the community. Thus, the prevalence of successful identity claiming assessed in terms of identity granting creates benefits that we expect to lead to membership growth (Butler, 2001). We have argued that in the context of an on-line consumption community, membership in a community is manifested in purchasing the product on which the community is based. Therefore, higher levels of identity granting in the community should be positively associated with product sales because it makes identification with the community more appealing.

Hypothesis 4: *An increase in the percentage of helpful votes will be positively associated with subsequent product sales.*

A summary of our model and hypotheses appears as Figure 1. The circular arrows in the figure signify the operation of reciprocal processes; for example, while rates of identity claiming and

granting may influence product purchase, buying a book after reading helpful reviews may then lead a consumer to post a review in which he/she claims identity.[6]

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 identity claiming, granting, 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 users 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 on product characteristics, identity claims, identity grants, and product sales.

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.[7] We use two datasets; the first consists of data on product characteristics, reviews, and reviewers of books in our sample. The second consists of economic transactions involving these products based on purchases by consumers in different geographical locations in the US. We provide more details on each of these datasets below.

3.2.1 Reviewers, Reviews, and Product Characteristics

---

[6] Note that the construct "Identity Granting Helpful Votes" exists both at the Product Review Level (as a dependent variable to test Hypothesis 3) and at the Consumption Community Level (as an independent variable to test Hypothesis 4).

[7] We derived the list of 786 products from a random sample of books appearing as a best-seller in at least one city over the period April 2005 to January 2006, based on Amazon's "Purchase Circles." More details on Purchase Circles appears below.

We collected data on product characteristics, reviews, and reviewers from Amazon during a 10-day period from March 17 to March 27. Summary statistics for each of the variables are included in Table 1. We describe the construction of each of these variables below.

Reviewer Characteristics: Amazon has a registration procedure in which users provide some information (e.g., email address, password) and choose a user name to register on the site, and users are recognized by their user name when they return to the site. Users may optionally decide to post additional information such as geographical location, disclose additional titles (e.g., "PhD") or use a nickname (e.g., L. Quido "Quidrock"). We use these data to assess identity claims.

For each review that was posted for each book in our dataset, we collected all personal details of the reviewer that we expected to serve as identity claims, such as disclosure of their Real name, their geographical location, their nickname, or their hobbies and professional interests. The presence or absence of Real name, location, nicknames and hobbies was coded as a dummy variable (0 or 1). We also collected recorded whether the reviewer included the title PhD next to their name to serve as a control because this title may be associated with the quality of the review itself.

Review Characteristics: To assess identity granting, we used a feature on Amazon that allows users to rate the helpfulness of reviews. At the bottom of each review, 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 users who voted] found the following review helpful:" We use this variable to compute helpful votes (the fraction of votes that evaluated the review as helpful)

Product Characteristics: The first dataset consists of product-specific characteristics such as the books' list price, its Amazon retail price, and the date that the product was released 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, which is a proxy for sales (e.g., Chevalier and Goolsbee 2003, Ghose and Sundararajan, 2006, Ghose, Smith and Telang, 2006).

3.2.2 Economic Transactions by Geography: The second dataset that we use for this study comes from the web pages on "Purchase Circles" on the Amazon.com web site. Purchase Circles are specialized best-seller lists. The pages denote the top-selling books, music, and DVDs across large and small towns in every state throughout the US. We collected weekly data on purchase circles over a period of ten months between April 2005 and January 2006 for the study. The Purchase Circles are organized in multiple layers- first, by state and then within a state, by town. For larger towns, they are also organized by suburb or county. Thus, the number of geographical locations that are listed for any given state is very large: for example, the state of Pennsylvania alone lists 271 towns. As a result, the data collection procedure produced a large panel data set for econometric analysis.

We use this data to compute two separate measures of economic transactions. Descriptive statistics are listed in Table 2. First, we use the data to compute the top sellers across all towns that Amazon tracks. For each town, Amazon provides a list of top 10 or top 20 best sellers for each product category. We use this information to compute a dummy variable indicating whether the product is in the top 10 in a particular location. This local popularity rank is our primary dependent variable for the analysis in Hypothesis 1b. Amazon also provides on its web site the national sales rank within that product category (books, music, videos, etc.) for each product that it sells. 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 is ranked #5 in Las Vegas, Nevada, and #6 in Great Falls, Montana.

4. Empirical Methodology and Results

In this section we discuss the models we use to test hypotheses 1 through 4. Our use of secondary data from Amazon's electronic market required us to estimate separate models for each of the four hypotheses. In each section we briefly describe the empirical model used, our identification strategy, and then our results.

4.1 Effects of Identity Claiming on Purchase Behavior

Our first set of models is used to link social exchanges to economic outcomes. Our broad goal is to examine how the prevalence of identity claims in the form of self-disclosure influences economic purchase decisions. We do this in two ways. First, we examine how the disclosure of personal information in reviews such as Real name, location, nicknames and hobbies influence product purchases (Hypothesis 1a). Second, we examine how the presence of reviews from the same geographic region influence purchase behavior above and beyond the information content from reviews posted nationwide (Hypothesis 1b).

4.1.1 Revelation of personal information and purchase decisions

We first estimate the relationship between sales rank and disclosure of personal attributes such as Real name, nickname, hobbies and location.[8] The unit of observation in our analysis is a product-month, and the dependent variable is $\log(SALESRANK_{is})$, the log of sales rank of product i in month s. We estimate the following regression:

$$\log(SALESRANK_{is}) = \alpha + \beta_1 AMAZONPRICE_{is} + \beta_2 \log(ELAPSEDDATE_{is})$$
$$+ \gamma_1 PCT-REALNAME_{is-1} + \gamma_2 PCT-NICKNAME_{is-1} +$$
$$\gamma_3 PCT-HOBBIES_{is-1} + \gamma_4 PCT-LOCATION_{is-1} + \mu_i + \varepsilon_{is}$$

where

$AMAZONPRICE_{is}$, and $\log(ELAPSEDDATE_{is})$ are product controls, and

---

[8] Note that prior work in this domain has generally transformed the dependent variable (salesrank) into quantities using the specification similar to Chevalier and Goolsbee (2003), Brynjolfsson, Hu and Smith (2003), Ghose and Sundararajan (2006), and Ghose, Smith and Telang (2006). That was usually done because those papers 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.

$PCT - REALNAME_{is-1}$, $PCT - NICKNAME_{is-1}$, $PCT - HOBBIES_{is-1}$, and $PCT - LOCATION_{is-1}$ are variables that capture personal disclosure of information.[9] We estimate product-level fixed effects to control for differences in average sales rank across products.

Table 3 displays the results of the model. Note that increases in sales rank mean lower sales, so a negative coefficient implies that increases in a variable *increases* sales. Columns (1) through (5) show that the coefficients of *PCT-NICKNAME, PCT-HOBBIES,* and *PCT-LOCATION* are each negative and the latter two are statistically significant while *PCT-NICKNAME* is marginally significant, lending support for Hypothesis 1a. The coefficient on *PCT-REALNAME* is small and insignificant. The coefficients in this model can be interpreted as partial elasticities: a 10% increase in *PCT-NICKNAME* is associated with a 0.3% decline in sales rank; a 10% increase in *PCT-HOBBIES* is associated with a 0.1% decline in sales rank; a 10% increase in *PCT-LOCATION* is associated with a 0.5% decline in sales rank. Other control variables would suggest that, as expected, sales of new books decrease over time, and sales decrease as Amazon's price increases, which is also consistent with prior work (Ghose, Smith and Telang 2006).

4.1.2 Revelation of geographic location and purchase decisions

In this section we use data from Amazon purchase circles to identify how prior posts from a geographic sub-community (in this case, a state) will influence the probability that a book will be one of the top sellers in a location. Our dependent variable is $TOP10RANK_{ijs}$, a binary variable that is equal to one when a product *j* is in the top 10 books purchased in location *i* in month *s*. We estimate the regression:

$$TOP10RANK_{ijs} = \alpha + \gamma'X_{js} + \beta_1 TOTREVIEWS_{js-1} + \beta_2 TOTRATINGS_{ijs-1}$$
$$+ \beta_3 STATEREVIEWS_{ijs-1} + \beta_4 STATERATINGS_{ijs-1} + \mu_{ij} + \varepsilon_{ijs}$$

---

[9] *ELAPSEDDATE* is the difference between the date of data collection and the release date of the book.

X_{js} 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. In keeping with prior research (Godes and Mayzlin 2004; Dellarocas et. al. 2005; Chevalier and Mayzlin 2006), we examine how the volume and valence of reviews influence sales. However, in contrast to prior work, our primary interest is in understanding how the volume and valence of reviews in the same geographic community influence sales as represented by the parameters β_3 and β_4.

Table 4 shows that increase in total reviews from the same state have a significant impact on the likelihood that a book will appear in the top 10 in a purchase circle, providing support for Hypothesis 1b. Consistent with social identity models, the influence of identity claiming is greater when there is more similarity between the content of identity claims and the attributes of potential members. Column (1) shows that a one standard deviation increase in the log of the lagged number of reviews in the same state—from 3.265 to 5.95—will increase the likelihood that a product will appear in the top 10 for that location by 2.0 percentage points. This result is robust to a variety of changes in the basic specification.

4.2 Self-Disclosure of Information

Hypothesis 2 suggests that because reviewers' disclosure of personal information when submitting reviews to Amazon may be interpreted as indicative of the norms in the community, it can influence the extent to which subsequent reviewers reveal similar information. In this section, we quantify how the probability that a reviewer for a particular book will reveal their personal information will be influenced by the decisions of earlier reviewers to disclose similar personal information. We conduct two sets of analyses: revelation of Real name, nickname, location, and hobbies and interests; and a more specific analysis concerning similarity in geographic location.

4.2.1 Revelation of personal information

In the first set of analyses, our dependent variable will be $PERSONAL\_INFO_{jr}^{p}$, a binary variable that indicates whether review r that has been posted for product j has posted personal information of type p. (which could be Real name, nickname, location, and hobbies and interests). For each of the four self-disclosure variables, we estimate the following fixed effects panel data model:

$$PERSONAL\_INFO_{jr}^{p} = \alpha + \beta PERSONAL\_INFO_{jr-1}^{p} + \Omega'X_{jr} + \mu_{j} + \varepsilon_{jr}$$

where $PERSONAL\_INFO_{jr-1}^{p}$ is the percentage of the prior 10 reviewers who provide personal information of type p.[10] 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 whether the person had a PhD, the average rating 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 self-descriptive information.

Table 5 presents the results of estimating this model. Consistent with Hypothesis 2a, there was a positive relationship between the self-descriptive information disclosed by previous and subsequent reviewers. All parameter estimates are significant at the 1% level. The coefficients can be interpreted as the change in the likelihood of posting personal information when each of the prior 10 reviewers also posts the same personal information. These marginal effects are quite similar across specifications. ranging from an increase of 71 percentage points for hobbies to 81 percentage points for location.

4.2.2 Revelation of geographic location

---

[10] Amazon displays an average of 10 reviews per page. We consider the prior ten reviews since when a user clicks on the product review page on the Amazon site, (s)he sees the ten reviews on that first page. The results from other combinations such as considering the prior 5 reviews or immediate prior review give similar results.

If geographic information indicates the composition of the community and thus influences interpretations of a reviewers' similarity to that community, it should influence the extent to which subsequent reviewers from the same location post reviews. To test this hypothesis, we construct a panel data set that indicates whether a review of a particular book has been posted in a particular US state on a particular day. Our unit of analysis is a product-state-day rather than product-state-review because of our need to construct a panel that is balanced across states and because reviews have a one-to-one mapping with states. Because multiple reviews for a product are often posted on the same day, we examine reviews that are posted with a lag of three to five days to be consistent with the analysis for Hypothesis 2a.

Our dependent variable will be $STATE - REVIEW_{ijt}$, a binary variable that indicates whether a review has been posted in state i for product j on day t. We estimate the following fixed effects panel data model:

$$STATE - REVIEW_{ijt} = \alpha + \sum_s \beta_{t-s} REVIEW_{jt-s} + \sum_s \gamma_{t-s} REVIEW_{jt-s} \times STATE_{ijt-s}$$
$$+ \sum_s \delta_{t-s} REVIEW_{jt-s} \times REGION_{ijt-s} + \mu_{ij} + \varepsilon_{ijt}$$

where $REVIEW_{jt-s}$ indicates that a review has been posted for the book in some prior day, $STATE_{ijt-s}$ is an indicator variable that is equal to one when at least one of these reviews is from state i, and $REGION_{ijt-s}$ is an indicator variable that is one when at least one of the reviews is from the same geographic region. μ_{ij} is a product-state fixed effect that controls for unobservable preferences that a state may have for certain books that is constant over time (e.g., the book *California: A History* may be more popular in California than in other states) as well as differences in the average propensity for states to post reviews (e.g., California may post more reviews than other

states due to its larger size). The variable $REVIEW_{jt-s}$ controls for changes in the popularity of a book over time.

Our primary interest is in estimating the parameter vectors γ and δ that measure the impact of prior reviews in the same state and region on the likelihood that a review will be posted on a particular state-product-day. We interpret positive coefficients on γ and δ as support of hypothesis 2b.

Table 6 provides the results of these regressions. Our analyses indicate that a review is more likely to be posted in a particular state-day when reviews have been posted in that state or region recently, supporting Hypothesis 2b. This is true even controlling for the average popularity of the book in the state and the average popularity of the book nationwide at a point in time. All parameter estimates are significant at the 1% level. Column (3) shows that a prior posting in the same state on the previous day increases the likelihood that a review will be posted by 6.7%; reviews posted on earlier days also influence the likelihood of reviews but are declining in importance as one moves further back in time (from a marginal effect of 4.6% for reviews from two days ago to a marginal effect of 2.7% for reviews from five days ago). In order to identify regional effects, we also form 4 regional clusters based on socio-cultural similarities.[11] The details of these clusters are given in the appendix. We construct regional dummies based on these clusters and also include these regional dummies in our analyses. We find that prior reviews from the same region also influence the likelihood of posting a review, however the results are weaker. This is not surprising, since despite the clustering regions contain substantial diversity. This shows that, consistent with social identity theory, within the online community the influence of similarity to other members of broad geographic regions is weaker than the influence of similarity to members of smaller regions.

---

[11] These regions are the Midwest, Northeast, South, and West. Further details are available from the authors upon request.

4.3 Identity granting

4.3.1 Self Disclosure and Identity Granting

Hypothesis 3 suggested that identity granting in the form of helpful votes would be positively associated with identity claiming behavior in the form of disclosure of self-descriptive information. The dependent variable $HELPFUL_{jt}$ 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: $HELPFUL_{jr} = \alpha_0 + \alpha_1(PERSONAL\_INFO_{jr}) + \alpha_2(PHD_{jr}) + \Omega'X_{jr} + \mu_j + \varepsilon_{jr}$

$PERSONAL\_INFO_{jr}$ is a vector of dummy variables for the self-disclosure variables such as reviewers' Real name, nickname, location, and hobbies and interests. μ_j is a product fixed effect that controls for differences in the average helpfulness of reviews across books. X is a vector of control variables that includes the log of the number of reviews and the Amazon retail price of that book. We also include a variable that controls for whether the reviewer has a Ph.D., since reviewers with higher education may be perceived as providing more helpful reviews. We interpret positive coefficients on α_1 as support of Hypothesis 3.

The above equation could be estimated using OLS. However, one concern with OLS estimation is that the posting of identity claims such as Real name or nickname may be correlated with unobservables that influence review quality. Some reviewers may put more effort into their reviews than others, and the amount of effort that one puts into their review may be correlated with an individual's propensity to engage in identity claiming behavior. If true, such correlation would lead to inconsistent estimates of α_1. To control for this potential problem, we instrument for $PERSONAL\_INFO_{jr}$ in the above equation using Real name, location, nickname, and hobbies using lagged values of the same variable. For example, our instrument for Real name is whether the prior reviewer posted Real name. The intuition behind the use of these IVs is that they are likely to

be correlated with the relevant independent variables but uncorrelated with unobservables that may influence the dependent variable. For example, the use of a Real name in prior reviews is correlated with the use of Real name in the subsequent reviews but uncorrelated with unobservables that determine helpful votes for a given review.

The results of our regressions are included in Table 7. Our analysis reveals that reviewers who choose to disclose their Real name receive more helpful votes, as do reviewers who reveal other information about themselves, such as their location, their nickname, and their hobbies and interests. The fraction of helpful votes that reviewers who post their Real name receive is 2.7 percentage points higher than an otherwise similar reviewer who does not post Real name, similar results are found for nickname (3.4 percentage points), hobbies and interests (8.6 percentage points), and location (3.7 percentage points). In sum, our results provide strong support for Hypothesis 3. [12]

4.3.2 Social Approval and Sales

Lastly, to evaluate Hypothesis 4, in column (6) and (7) of Table 3 we examine the hypothesis that increases in the percentage of helpful votes will lead to increases in sales rank. We ran a variant of the regression in section 4.1.1 with *PCT-HELPFUL* replacing our disclosure variables. Because we are concerned that changes in helpful votes may be correlated with unobserved changes in product popularity, we instrument for *PCT-HELPFUL* using *PCT-REAL NAME, PCT-NICKNAME, PCT-HOBBIES,* and *PCT-LOCATION* plus the percentage of reviewers who report having a PhD. In column (7) we further control for changes in product popularity by adding average reviewer rating: note that in this specification we do not interpret average rating as a causal variable reflecting the valence of reviews, rather it is included to control for changes in product popularity. As in the analyses reported for Hypothesis 1a, a negative coefficient implies that an increase in a variable increases sales. In both cases the coefficient on *PCT-HEPLFUL* is negative and significant

---

[12] We also estimated the above equation via OLS and the results were qualitatively similar to our baseline estimates in Table 7.

at the 5% level, supporting Hypothesis 4; using the results in column (6), a 10% increase in *PCT-HELPFUL* will lead to a 0.7% improvement in sales rank.

5. Discussion

Our results suggest that patterns of interaction that bear the hallmarks of identity- and community-based social exchange permeate the posting and rating of product reviews on electronic marketplaces. Specifically, the prevalence of self-disclosure in reviews posted in the community was associated with increased book sales. At the review level of analysis, apparent community norms concerning self-disclosure were associated with subsequent identity claiming behavior wherein members' self-disclosure behavior matched the norms set by previously-posted reviews. Identity claiming on the part of members elicited identity granting on the part of the community, in the form of identity-affirming social approval in response to conformity behavior. Finally, the rate of successful identity claiming assessed by prevalence of helpful votes in the community was associated with improved sales.

There are meaningful barriers to identification and community emergence in the context of product reviews posted on digital marketplaces. For example, economic rather than social exchanges are the presumed objective of sites like Amazon, and reviews are presumed to focus on the product being reviewed rather than the person providing the review. Despite this, social exchange clearly plays a role in structuring individual and community behavior in these contexts. These findings suggest that identity confirmation should be considered among the benefits created by electronic communication that have been the focus of prior research (Butler, 2001; Arguello et. al., 2006).

Much of the existing research on social identification and online communities focuses either on the individual community member (e.g., their motivations, their behaviors), or on the characteristics of the community. Only a handful of studies simultaneously explore both sides of the social exchange – the individual and the community (e.g., Arguello et. al., 2006; Moon and Sproull,

2006). The present paper thus fills an important gap in this literature by offering a quantitative analysis of the reciprocal patterns of influence shaping the community and member behaviors. Our multi-level analysis offers insight into the economic value of social behavior as well as providing evidence that the social processes that we hypothesize are in fact operative in the online community we study. The present study also helps to fill a gap in the social identification literature concerning the process by which identities emerge socially. To the best of our knowledge, this paper is the first to empirically examine notions of identity claiming and granting. Furthermore, it is the first paper to address how proactive social behaviors in the form of identity claiming and granting emerge in the context of online consumption communities, thus contributing to the growing body of literature using the lens of social identity theory to understand online communities (e.g., Postmes et al. 2000, Bagozzi and Dholakia, 2002; Gu and Jarvenpaa 2003). Our results suggest that consumption communities united primarily by their connection to an economic product may be like offline communities; economic exchange in both contexts may be embedded in social relationships.

We also contribute to recent literature that seeks to measure the impact of consumer reviews on firms' sales in two ways. First, we employ 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 since both the number and valence of consumer reviews may be correlated with unobservable product quality. Prior research on the causal relationship between reviews and sales has either employed a proxy variable strategy (Reinstein and Snyder 2005) or has utilized a difference-in-differences approach (Chevalier and Mayzlin 2006). Our research strategy uses cross-sectional variation in local sales and local reviews to better identify the relationship between reviews and sales by differencing out time-invariant local preferences for books and by using national reviews as a proxy variable for changes in perceived product quality over time. Besides providing a new identification strategy, we also contribute to understanding how reviews may contribute to

sales. Prior work has focused primarily on the informational value of reviews on sales, including the volume, valence, and dispersion of ratings (Dellarocas et. al. 2005). Complementing prior work, we demonstrate that the social value of reviews, or changes in the characteristics of *reviewers* and the responsiveness of the community can have a significant impact on firm sales.

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 permit us to assess member identification. Future research may fruitfully be directed at evaluating the identification 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. In addition, our results represent a series of snapshot views of a dynamic, reciprocal and iterative process by which community members claim and grant identity (see Figure 1). While these snapshots may be appropriate to the context of online consumer product reviews on Amazon because the opportunities for interaction between community members are highly constrained, future research examining identity claiming and granting may benefit from examining contexts that allow richer social exchange. In such contexts, it may be possible to evaluate the effect of identity grants to specific individuals on community norms or individual reviewer identity claiming.

6. Managerial implications

Social exchange is a crucial component of the business model of many electronic markets, and our findings suggest that consumer product reviews may be a form of social as well as informational exchange. It is now being increasingly accepted that the conversations that consumers have with one another online are an important source of word-of-mouth (WoM) data. Our results suggest that social benefits and identity formation are major reasons consumers publish their experiences on opinion platforms. These results have implications for the design of electronic markets. Identifying motives can enable electronic market makers to design their platform in a more

customer-oriented way by addressing the reasons community members post reviews (Dellarocas, 2003). For example, electronic retailers could build discussion forums for frequent users of the platform who may prefer an environment in which they can more flexibly reply to other users' posts. Indeed, Amazon users can now start threaded discussions on topics related to specific products. By facilitating interactive communication, such discussion groups can go some way towards fostering a sense of community.

Our results on the effect of geographical location also point to the importance of physical location in online communities. The Internet was touted to make geographic location irrelevant. If members in online communities identify more strongly with others from the same physical location, that implies that geographic distance and proximity still plays a vital role in influencing consumer behavior online. This can have several implications for viral marketers who use online markets as an additional channel of sales to complement their existing distribution channels.

Our results also show that community members who reveal more information about themselves tend to be rewarded by other members through identity-granting. This shows the emergence of self-governing communities in online markets that need little moderation or administration. Since the disclosure of self descriptive variables tends to affect product purchases in local as well as aggregate consumption communities, electronic markets may benefit from new mechanisms that make it easier for consumers to reveal information about themselves. However, an ongoing concern in such markets is whether users may misrepresent themselves (Resnick et al., 2000). New methods that provide incentives for truthful information revelation may serve to increase the sense of community among members and lead to increased revenues.

Encouraging consumers to invest in identity claiming may benefit vendors in another way. The ability to increase web site stickiness is a major concern for web site operators since competition is "just a click away." One strategy used by such operators to increase stickiness and

engender consumer lock in has been to encourage consumers to make web site-specific investments such as the entry of customer data (Chen and Hitt 2002). In our environment, online identities achieved through identity claiming and identity granting represent substantial site-specific investments that consumers value. Efforts to encourage these investments should provide substantial benefits to operators in the form of lock in.

References

Arguello, J., B.S. Butler, L. Joyce, R. Kraut, K.S. Ling, X. Wang. 2006. Talk to me: Foundations for successful individual-group interactions in online communities. *Proceedings of the ACM Conference on Human Factors in Computing Systems*. New York: ACM Press.

Bagozzi, R.P., Dholakia, U.M. 2002. Intentional social action in virtual communities. *Journal of Interactive Marketing*, 16(2): 2-21.

Balasubramanian, S., V. Mahajan. 2001. The economic leverage of the virtual community, *International Journal of Electronic Commerce* **5** (3): 103–138.

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. In MA Hogg and D. Terry (Eds.), *Social identity processes in organizational contexts* (pp. 115-130). Philadelphia, PA: Psychology Press.

Bhattacharya, C.B., S. Sen. 2003. Consumer-company identification: a framework for understanding consumers' relationships with companies. *Journal of Marketing* 67: 76-88.

Brewer, M. 1991. The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*. 17(5): 475-482.

Brown, J., P. H. Reingen, 1987. Social Ties and Word-of-Mouth Referral Behavior. *The Journal of Consumer Research* 14(3): 350-362.

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

Bush, A.A., A. Tiwana. 2005. Designing sticky knowledge networks. *Communications of the ACM*. 48(5) 66-71.

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

Byrne, D. 1971. The attraction paradigm. New York: Academic Press.

Campbell, D.T. 1958. Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science* 3: 14-24.

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

Chen, P., S. Dhanasobhon, M. Smith 2006. All Reviews are Not Created Equal: The Disaggregate Impact of Reviews and Reviewers at Amazon.com. Working Paper, Carnegie Mellon University.

Chevalier, J., A. Goolsbee. 2003. Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics* I(2): 203-222.

Chevalier, J., D. Mayzlin. 2006. The effect of word of mouth online: Online book reviews. *Journal of Marketing Research*, forthcoming.

Cummings, J, B. Butler, R. Kraut. 2002. The quality of online social relationships. *Communications of the ACM* 45(7): 103-108.

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

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

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

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.* Palo Alto, CA: Stanford University Press.

Fischer, E. J. Bristor, B. Gainer 1996. Creating or Escaping Community? An Exploratory Study of Internet Consumers' Behaviors. In *Advances in Consumer Research*, Vol. 23, (eds). Kim P. Corfman and John Lynch, p. 178–182, Provo, UT: Association for Consumer Research.

Ghose, A., A. Sundararajan. 2006. Evaluating pricing strategy using ecommerce data: Evidence and Estimation Challenges. Forthcoming, *Statistical Science.*

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

Gilly, M. C. 1998. A dyadic study of interpersonal information search. *Academy of Marketing Science* 26(2): 83-100.

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

Godes, D., J. Silva (2006). The Dynamics of Online Opinion. Working Paper.

Grohol. J. 2006. Anonymity and online community identity matters. http://www.alistapart.com/articles/identitymatters

Gu, B., S. Jarvenpaa. 2003. Are contributions to P2P technical forums private or public Goods? –An empirical investigation. *Proceedings of the 2003 Workshop on Economics of P2P.*

Hennig-Thurau T., K.P. Gwinner, G. Walsh, D.D Gremler. 2004. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *J. Interactive Marketing* 18 (1): 38-52.

Hornsey, M.J., Jetten, J. 2004. The individual within the group: Balancing the need to belong with the need to be different. *Personality and Social Psychology Review* 8(3): 248-264.

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

McAuliffe, B. J., J. Jetten, M. Hornsey, M.A. Hogg. 2003a. Differentiation between and within groups: The benefits and costs of individualist and collectivist group norms. Working Paper.

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

Moon, J.Y., L. Sproull. 2006. The role of feedback in managing an internet-based volunteer workforce. Manuscript under review.

Muniz, A.M., T. O'Guinn. 2001. Brand community. *Journal of Consumer Research* 27(4): 412-432.

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

Postmes, T., R. Spears, M. Lea. 2000. The formation of group norms in computer-mediated communication. *Human Communication Research* 26, no. 3 (July 1): 341-371.

Reinstein, D., C. Snyder 2005. The influence of expert reviews on consumer demand for experience goods: A case study of movie critics *Journal of Industrial Economics* 53(1): 27-51.

Resnick, P., R. Zeckhauser, E. Friedman, K. Kuwabara. 2000. Reputation systems. *Communications of the ACM 43*(12): 45-48.

Sheldon K.M., B.A. Bettencourt. 2002. Psychological need satisfaction and subjective well-being within social groups. *British Journal of Social Psychology* 41: 25-38.

Sproull, L. 2003. Online communities. The Internet Encyclopedia, Bidgoli.

Subramani, M., Peddibhotla. 2002. Contributing to document repositories - An examination of prosocial behavior. Working Paper, University of Minnesota.

Swann, W. B., Jr. 1987. Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology 53*: 1038-1051

Tajfel, H., Turner, J. C. 1979. An integrative theory of intergroup conflict. In W. G. Austin and S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA: Brooks/Cole.

Turner, J. 1987. Rediscovering the social group: A self-categorization theory. Oxford: Blackwell.

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. *Journal of Personality and Social Psychology.* 70(5): 913-930

Wasko, M., S. Faraj. 2000. It is what one does: Why people participate and help others in electronic communities of practice. *Journal of Strategic Information Systems* 9 (2-3): 155-173.

Wasko, M., S. Faraj. 2005. Why Should I Share? Examining Knowledge Contribution in Electronic Networks of Practice. *MIS Quarterly* 29(1): 1-23.

Xia, L., N.N. Bechwati. 2006. Positive "word of mouse": The role of personalization. *Proceedings of the American Marketing Association*, 17, 107.

Consumption Community Level
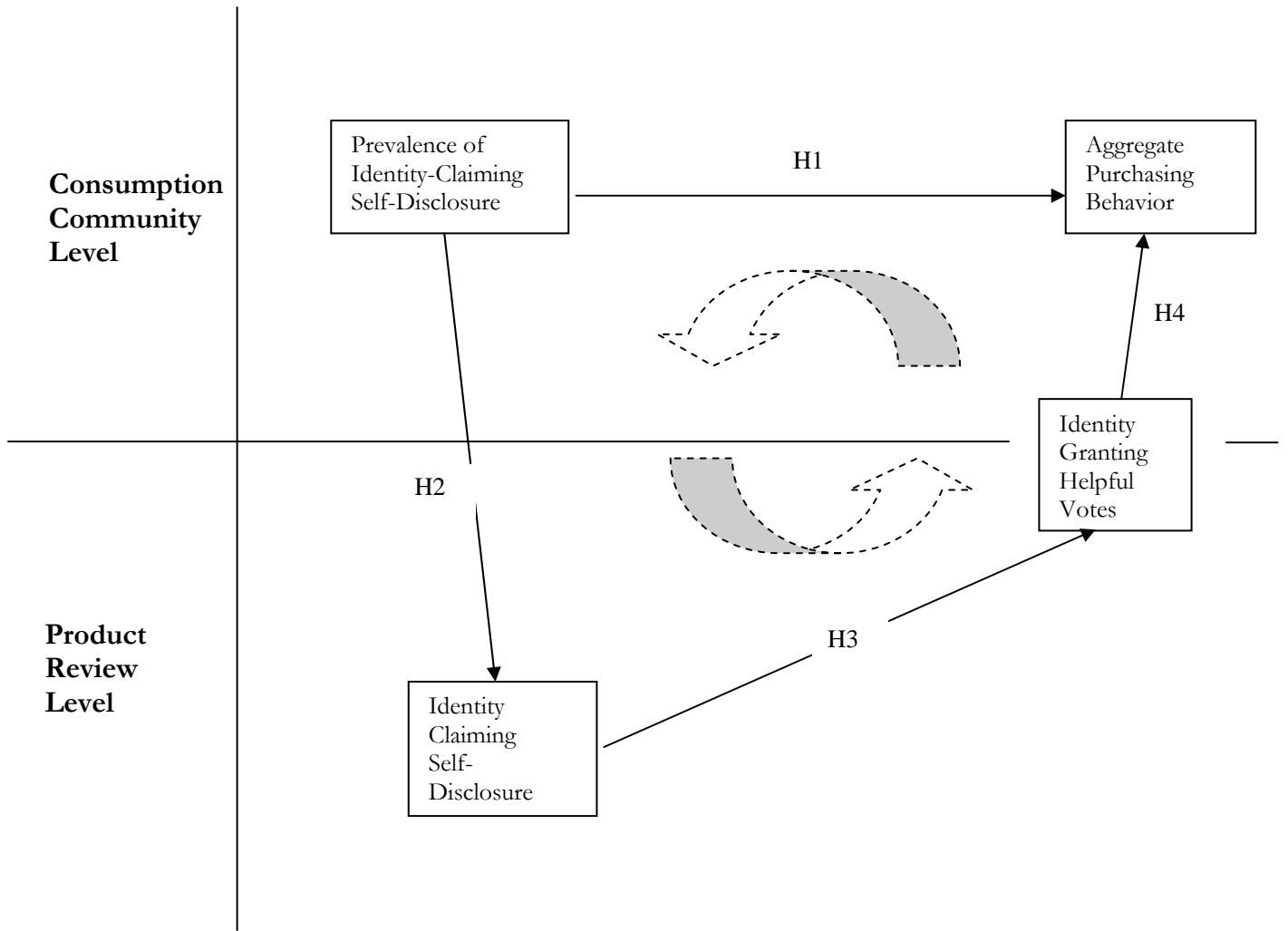
Product Review Level

Prevalence of Identity-Claiming Self-Disclosure

H1

Aggregate Purchasing Behavior

H4

Identity Granting Helpful Votes

H2

Identity Claiming Self-Disclosure

H3

Figure 1: Summary of Model and Hypotheses

Table 1: Descriptive Statistics for Product Data

Variable	Observations	Mean	Std. Dev.	Min	Max
ASIN	237561	328.7474	197.8424	1	786
List Price	233584	18.71348	10.85526	1.5	156.95
Amazon Price	233582	12.81825	7.615922	1.5	156.95
Average Rating	237559	4.16224	.5309869	2	5
Number Reviews	237559	1195.105	1410.138	1	5756
Review Rating	237561	4.136929	1.307897	1	5
Helpful Votes	197708	8.420023	26.03262	0	2457
Number of Votes	197708	15.1071	37.10222	1	2688
Real Name	220630	.3496533	.4768615	0	1
Nick Name	220653	.677897	.4672833	0	1
Hobby	220578	.152853	.3598465	0	1
Review Date	237544	15899.72	866.6416	-21887	16886
Higher Education(PhD)	237561	.0044325	.0664298	0	1
Location	237561	.6438136	.4788722	0	1

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 and this is why the number of observations is lower for some of these variables.

Table 2: Descriptive Statistics for Product Data

Variable	Observations	Mean	Std. Dev.	Min	Max
Dummy for Top 10 Book in Location	12257860	0.0093	0.0958	0	1
Log of Elapsed Date	12257860	4.0567	3.2730	-1.8458	9.7431
Dummy for Missing Elapsed Date	12257860	0.2611	0.4393	0	1
Log of Lagged Total Recommend.	12257860	3.6248	2.3254	0	8.5932
Dummy for No Lagged Recommendations	12257860	0.1909	0.3930	0	1
Log of Lagged Total State Recommend.	12257860	1.0533	1.3066	0	6.2710
Dummy for no State Recommend.	12257860	0.4829	0.4997	0	1
Avg. Lagged Nationwide Ratings	12257860	3.4556	1.7381	0	5
Avg. Lagged Statewide Ratings	12257860	2.1848	2.1915	0	5
Pct Lagged "High Reviews" Nationwide	12257860	0.5255	0.3114	0	1
Pct Lagged "Low Reviews" Nationwide	12257860	0.0553	0.0788	0	1
Pct Lagged "High Reviews" Statewide	12257860	0.3248	0.3910	0	1
Pct Lagged "Low Reviews" Statewide	12257860	0.0379	0.1221	0	1

Table 3: How does self-disclosure influence purchase behavior?

Independent Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Amazon Price	0.1176 (0.0069)**	0.1172 (0.0069)**	0.1175 (0.0069)**	0.1177 (0.0069)**	0.1174 (0.0069)**	0.1164 (0.0074)**	0.116 (0.0074)**
Log of Elapsed Date	0.078 (0.0273)**	0.0785 (0.0273)**	0.0796 (0.0273)**	0.0789 (0.0273)**	0.0803 (0.0273)**	0.0884 (0.0296)**	0.0912 (0.0298)**
Dummy for No Recommendations	0.0388 (0.0463)	-0.0101 (0.0468)	0.0103 (0.0405)	-0.027 (0.0493)	-0.0581 (0.0579)	-0.9477 (0.4600)*	-0.5643 (0.2778)*
Pct Real name	0.0066 (0.0436)				0.0201 (0.0439)		
Pct Nickname		-0.0784 (0.0443)+			-0.0399 (0.0474)		
Pct Hobbies			-0.1349 (0.0552)*		-0.1182 (0.0585)*		
Pct Location				-0.0966 (0.0463)*	-0.0929 (0.0466)*		
Pct Helpful Votes						-1.3936 (0.6495)*	-1.3498 (0.6734)*
Average Reviewer Rating							0.0803 (0.0535)
Constant	3.7165 (0.2138)**	3.7703 (0.2148)**	3.7352 (0.2129)**	3.7739 (0.2144)**	3.8031 (0.2163)**	4.6041 (0.4711)**	4.2246 (0.3233)**
Observations	3534	3534	3534	3534	3534	3534	3534
Number of pindex	763	763	763	763	763	763	763
R-squared	0.0871	0.087	0.0871	0.0872	0.0872	0.0827	0.0838

Table 3: Product-level fixed effects model with standard errors in parentheses. Columns (7) and (8) use 2SLS to instrument for Pct Helpful Votes using Pct Real name, Pct Nickname, Pct Hobbies, Pct Location, and Pct reporting PhD. ** , * and + denote significance at 1%, 5% and 10%, respectively.

Table 4: How does the presence of reviews in a state influence sales of products in that state?

Independent Variable	(1)	(2)	(3)
Log of Elapsed Date	-0.0029 (0.0007)**	-0.0029 (0.0007)**	-0.0029 (0.0007)**
Dummy for Missing Elapsed Date	-0.0144 (0.0029)**	-0.0146 (0.0032)**	-0.0146 (0.0032)**
Log of Lagged Total Recommendations	0.0062 (0.0013)**	0.0063 (0.0013)**	0.0063 (0.0013)**
Dummy for No Lagged Recommendations	0.0066 (0.0029)**	0.0099 (0.0108)	0.0057 (0.0055)
Log of Lagged Total State Recommendations	0.0090 (0.0021)**	0.0090 (0.0021)**	0.0090 (0.0021)**
Dummy for no Lagged State Recommendations	0.0126 (0.0037)**	0.0110 (0.0046)*	0.0095 (0.0032)**
Lagged Avg. Nationwide Ratings		0.0007 (0.0022)	
Lagged Avg. Statewide Ratings		-0.0004 (0.0007)	
Pct Lagged "High Reviews" Nationwide			-0.0008 (0.0049)
Pct Lagged "Low Reviews" Nationwide			-0.0095 (0.0182)
Pct Lagged "High Reviews" Statewide			-0.0041 (0.0019)*
Pct Lagged "High Reviews" Statewide			-0.0071 (0.0035)*
Number of Observations	12257860	12257860	12257860
R^2	0.2084	0.2084	0.2085

Table 4: OLS product-location fixed effects regression with robust standard errors in parentheses. ** , * and + denote significance at 1%, 5% and 10%, respectively.

Table 5: How do prior postings of the self-descriptive information influence the likelihood of posting a review with the same information?

Independent Variable	(1)	(3)	(3)	(4)
	Real name Dummy	Nickname Dummy	Hobbies Dummy	Location Dummy
Prior Posting of Real name	0.78*** (0.006)			
Prior Posting of Nickname		0.81*** (0.006)		
Prior Posting of Hobbies			0.71*** (0.011)	
Prior Posting of Location				0.75*** (0.007)
Average Rating	0.0029*** (0.0004)	-.002*** (0.0004)	-.0018*** (0.0003)	-.0003 (.00048)
Log of Number of Reviews	.000048 (.000069)	.0001* (.00006)	-1.24e-06 (0.00002)	-.0000537 (.000071)
PhD Dummy	0.022 (0.022)	0.055* (0.019)	0.067*** (0.024)	-0.025 (0.024)
Observations	525598	525555	525545	540768
R^2	0.76	0.61	0.68	0.72

Table 5: This table shows whether the revelation of self-descriptive information in prior reviews increases the probability that subsequent reviewers will reveal the same information. The dependent variable is the dummy for the relevant self-descriptive variable. The main independent variable of interest is the probability of prior posting of the same variable. All models use OLS with product-level fixed effects. Standard errors are listed in parenthesis; ***, ** and * denote significance at 1%, 5% and 10%, respectively. These are robust standard errors.

Table 6: How do prior postings in the same state influence the likelihood of posting a review?

Independent Variable	(1)	(2)	(3)
Prior posting in state [t-1]	0.0692	0.068	0.0671
	(0.0003)**	(0.0003)**	(0.0003)**
Prior posting in state [t-2]	0.0484	0.0468	0.0458
	(0.0003)**	(0.0003)**	(0.0003)**
Prior posting in state [t-3]	0.0377	0.0354	0.0341
	(0.0003)**	(0.0003)**	(0.0003)**
Prior posting in state [t-4]		0.0321	0.0302
		(0.0003)**	(0.0003)**
Prior posting in state [t-5]			0.0269
			(0.0003)**
Prior posting in region [t-1]	0.0025	0.0023	0.0021
	(0.0001)**	(0.0001)**	(0.0001)**
Prior posting in region [t-2]	0.0024	0.0022	0.002
	(0.0001)**	(0.0001)**	(0.0001)**
Prior posting in region [t-3]	0.0023	0.002	0.0018
	(0.0001)**	(0.0001)**	(0.0001)**
Prior posting in region [t-4]		0.0018	0.0016
		(0.0001)**	(0.0001)**
Prior posting in region [t-5]			0.0015
			(0.0001)**
Prior posting—all areas [t-1]	0.0025	0.0021	0.0019
	(0.0000)**	(0.0001)**	(0.0001)**
Prior posting—all areas [t-2]	0.0026	0.0022	0.0019
	(0.0000)**	(0.0001)**	(0.0001)**
Prior posting—all areas [t-3]	0.0025	0.0021	0.0018
	(0.0000)**	(0.0001)**	(0.0001)**
Prior posting—all areas [t-4]		0.0022	0.0018
		(0.0001)**	(0.0001)**
Prior posting—all areas [t-5]			0.0018
			(0.0001)**
N	16,980,750	16,980,750	16,980,750
R^2	0.0188	0.0210	0.0226

Table 6: The dependent variable is equal to one when someone from a state issues a review on a day. Standard errors are listed in parenthesis; ** , * and + denote significance at 1%, 5% and 10%, respectively. All models use OLS with product-state fixed effects. These are "classical" standard errors assuming iid observations within groups. Robust standard errors give similar results.

Table 7: Self-Disclosure and Identity Granting

Independent Variable	(1)	(2)	(3)	(4)	(5)	(6)
PhD Dummy	.06*	.27***	.28***	.21***	.28***	.25***
	(.038)	(.025)	(.025)	(.026)	(.025)	(.025)
Log (Number of Reviews)	-0.0002**	-0.0002**	-0.0002**	-0.0002**	-0.0002**	-0.0002**
	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)
Equivocal Review	-0.1***	-0.1***	-0.1***	-0.1***	-0.1***	-0.1***
	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
Real Name	.027***					.02***
	.006)					(.006)
Nickname			.034***			.011*
			(.004)			(.006)
Hobbies				.086***		.039***
				(.006)		(.008)
Location					.037***	.002
					(.003)	(.007)
Observations	287346	287302	287355	287299	294524	287299
R^2	0.02	0.03	0.02	0.02	0.02	0.03

Table 7: The dependent variable is Log (Helpful). Standard errors are listed in parenthesis; ***, ** and * denote significance at 1%, 5% and 10%, respectively. All models use 2SLS to instrument for Real name, Nickname, Hobbies, and Location using lagged values of the same variables. The fixed effects are at the product level.