Online Shopping and Social Media: Friends or Foes?

July 3, 2017

Yuchi Zhang*

Michael Trusov

Andrew T. Stephen

Zainab Jamal

* Yuchi Zhang is Assistant Professor of Marketing at the Leavey School of Business, Santa Clara University (yzhang6@scu.edu). Michael Trusov is Associate Professor of Marketing at the Robert H. Smith School of Business, University of Maryland (mtrusov@rhsmith.umd.edu). Andrew T. Stephen is LOreal Professor of Marketing at Said Business School, University of Oxford (andrew.stephen@sbs.ox.ac.uk). Zainab Jamal is a researcher in the Business Optimization Lab, Hewlett-Packard (zainab.jamal@hp.com).

Online Shopping and Social Media: Friends or Foes?

Abstract

As social network use continues to increase, an important question for marketers is whether consumers' online shopping activities are related to their use of social networks and, if so, what the nature of this relationship is. On the one hand, spending time on social networks could facilitate social discovery, meaning that consumers "discover" or "stumble upon" products through connections they have to others. Moreover, cumulative social network use could expose consumers to new shopping-related information, possibly with greater marginal value than the incremental time spent on a shopping website. This process may therefore be associated with increased shopping activity. On the other hand, social network use could be a substitute for other online activities, including shopping. To test the relationship between social network use and online shopping, the authors leverage a unique consumer panel data set that tracks individuals' browsing of shopping and social network websites and their online purchasing activities over one year. The authors find that greater cumulative usage of social networking sites is positively associated with shopping activity. However, they also find a short-term negative relationship, such that immediately after a period of increased usage of social networking sites, online shopping activity appears to be lower.

Keywords: social networks, online shopping, electronic commerce, retailing.

In recent years, social media has become an important part of people's daily lives and, therefore, firms' marketing programs. Although no longer new, social media still presents marketing professionals with challenges as they struggle to understand how it is changing the business environment and grapple with identifying opportunities for effectively leveraging online social media channels (e.g., Facebook, Twitter) in ways that drive value. In parallel with the practitioners, academic researchers have been exploring a variety of topics related to social media, including word-of-mouth (WOM) propagation, online social influence, the role of user-generated content and product reviews in stimulating sales, and factors related to the "virality" or diffusion of online content (e.g., Berger and Schwartz 2011; Chevalier and Mayzlin 2006; Kumar et al. 2013; Kumar and Rajan 2012; Moe and Trusov 2011; Stephen and Galak 2012; Toubia and Stephen 2013; Trusov, Bucklin, and Pauwels 2009; for a comprehensive review, see Lamberton and Stephen 2016).

The foremost players in the social media landscape are social networking sites. In essence, social networks are online communities that allow users and firms to share content with other individuals. Social network use among consumers is high and continues to increase. As of September 2016, approximately 1.79 billion people were active Facebook users (Facebook 2016). In the United States, 30% of time spent online is on social networking and social media sites (GWI 2016). Given the widespread use of social networks, it is necessary to understand how this is related to another popular, and economically important, online activity: e-commerce. While firms are increasingly using social networks in their marketing strategies, research by IBM suggests that social media has little impact on e-commerce, with only .34% of online sales referred by social media websites (Del Rey 2013; Gara 2012). However, other industry studies indicate a positive link between sales and consumers' engagement with social media sites such as Twitter, Pinterest, and Facebook (Bercovici 2013).

Despite the growing body of marketing literature on social networks, researchers have paid little attention to the interplay between social network usage and e-commerce activity. In particular, extant research has not examined how consumer engagement in social networks coexists with and is related to consumers' e-commerce activities. In this paper, we aim to shed some light on this question.

We argue that social networks can play a dual role in consumers' e-commerce buying behaviors**.** On the one hand, using social networks could be positively associated with purchasing because consumers on social networks are frequently exposed to information about products and consumption-related activities, ranging from product ads by brands to friends' conversations and opinions about recent shopping experiences (e.g., Chevalier and Mayzlin 2006; Moe and Trusov 2011; Stephen and Galak 2012). Both individuals and firms use social networks to share information that can be broadly described as "consumption related." For example, on their social network accounts, consumers often post photos of recent purchases, share stories about shopping experiences, and describe products that they want to purchase in the future. Firms also frequently use their social network channels to post information about their various offerings.

Intuitively, a person is more likely to buy while browsing an e-commerce site such as Amazon.com (e.g., while actively searching for product information) than when interacting with friends on Facebook. However, over time, a person who is more active in social networks might also be exposed to a greater variety of shopping-related content posted by firms and friends than a person whose interactions with shopping content are limited to goal-driven product search across the e-commerce sites he or she is familiar with and visits. Thus, for the former shopper (with higher social network exposure), "buying" might be less costly than it is for the latter shopper (without much social network exposure) in terms of reduced costs of information search. We argue that repeated exposure to consumption-related content on social networks (1) informs consumers about consumption opportunities (purchase options), which reduces search costs and, as such, makes buying more appealing (Hauser, Urban, and Weinberg 1993), and (2) may be associated with consumption through social and peer referral mechanisms (e.g., Katona, Zubcsek, and Sarvary 2011; Kumar, Petersen, and Leone. 2010; Trusov, Bodapati, and Bucklin 2010). We therefore expect a positive correlation between consumers' social network engagement and their online shopping behaviors.

On the other hand, use of social networks may substitute for time spent on e-commerce sites, thus potentially having a negative relationship with online shopping activities in a more immediate, short-term sense. On average, consumers spend approximately 30% of their Internet time on social media and 8% on online shopping (GWI 2016). According to theories of time allocation (e.g., Becker 1965), how people allocate their time tends to be fixed, and there are costs associated with time allocated to activities. Logically, because there are a fixed number of hours in a day, allocating time to activities, including online shopping versus social network use, may take on a "zero-sum" nature. In other word, if a consumer spends more time using social networks on a given day, he or she may spend commensurately less time doing other online activities such as online shopping.[1]

This duality in how social network usage could be correlated with online shopping activity might explain an interesting and seemingly paradoxical empirical pattern that we observe

[1] Consistent with this logic, model-free evidence in our data reveals a negative correlation between time spent on social networks and online e-commerce websites of –.54 (conditional on overall Internet usage activity).

in our data: during periods of elevated activity on social networking sites, people appear to be less likely to buy online—a result that is alarming to online retailers. However, on a more positive note, we also observe that people who have been engaged in social networking sites for an extended period of time (i.e., cumulative use) tend to buy more and, importantly, tend to buy from a larger number of online retailers. Thus, it appears that cumulative social network use may be associated with an increase in overall shopping behavior in the long run but a decrease in overall shopping behavior in the short run.

With these competing forces in mind, we develop a conceptual framework that builds on classic time allocation literature (Becker 1965; Hauser, Urban, and Weinberg 1993; Ratchford, Lee, and Talukdar 2003) to explain these trends. We then provide empirical evidence in support of our conceptualization using a unique real-world data set of complete online browsing and purchase activity over one year for more than 10,000 individuals. Our data allow us to investigate the associations between social network usage and buying behavior over time.

To preview our empirical results, we find evidence of both positive and negative relationships between social network usage and online shopping activities. Notably, the nature of these relationships depends on the time horizon of past social network usage (i.e., immediate vs. cumulative) and, to a lesser extent, product category. Specifically, our main findings are as follows:

1. Social network usage is both positively and negatively associated with shopping activity: social network engagement over time (i.e., cumulative usage) is positively correlated with shopping activity, whereas immediate usage is negatively correlated with shopping activity.
2. Social network referrals are predictive of shopping activity, but only in the immediate term.
3. The positive relationship between cumulative social network usage and online shopping activity is stronger for product categories that tend to be shared on social networks and/or are often bought as unplanned purchases (e.g., chocolates) and weaker for products that

are less typically socially shared and/or are usually planned purchases (e.g., automotive parts, gift cards).

This research makes several important contributions. First, to the best of our knowledge, this is the first paper to empirically associate individual-level e-commerce activities with the social network usage behavior of a large number of individuals. While extant studies have suggested that electronic WOM (eWOM) can significantly affect sales (e.g., Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Godes and Mayzlin 2004; Kumar et al. 2013; Stephen and Galak 2012), have assessed the impact of social web links and recommendations on online behavior (Goldenberg, Oestreicher-Singer, and Reichman 2012; Mayzlin and Yoganarasimhan 2012; Oestreicher-Singer and Sundararajan 2012), and have linked exposure to another form of media (television advertising) with online shopping (Liaukonyte, Teixeira, and Wilbur 2015), our research fills an important gap in the literature by offering insights into how both immediate and cumulative over-time usage of online social networking sites are related to individuals' e-commerce buying behaviors. Second, this study answers recent calls for more research to consider how using social networks is related to consumers' behaviors *outside* social media channels (Lamberton and Stephen 2016; Stephen 2016). Finally, we propose an original theoretical framework (building on classic time allocation literature) that can help managers resolve some ambiguity in the relationship between social network usage and sales. In doing so, we also contribute to the time allocation literature by exploring more modern uses of time.

Conceptual Framework

As we have noted, our conceptual framework is based on theories of time allocation. Becker (1965) proposes that individuals allocate their time to various activities based on the utility impact of the commodities they derive from those activities.[2] For example, a consumer can choose to allocate his or her time to playing golf or to exercising on a treadmill. Both activities create commodities such as entertainment value and health value, although golf may offer greater entertainment value and exercise may yield more health benefits. The consumer then allocates a portion of his or her time to golf, the treadmill, or a combination of both to produce both entertainment and health-related value. Inherent in this decision is a trade-off between engaging in different activities, due to the fixed amount of available time and the marginal values gained.

The trade-offs that people must make when allocating time to different activities has been studied extensively in the literature. Gronau (1973, 1977) examines the decision to spend leisure time at home and spouses' allocation of household time. Consumers can engage in different leisure and non-leisure activities and tend to make their time allocation decisions as a function of the value they potentially derive from the time invested in each. Hauser, Urban, and Weinberg (1993) provide evidence that consumers allocate their time across different sources when searching for information to be used for future purchase decisions. These time allocation decisions are based on the marginal value they derive from the time dedicated to each source. In these previous examples, allocating time to one activity reduces the time available to other activities. Therefore, certain commodities are gained (e.g., entertainment) at the expense of others (e.g., fitness).

We apply the classic theory of time allocation to our specific context. The Internet offers consumers a variety of sources that provide a wealth of shopping- and entertainment-related content. For the sake of simplicity, we focus on only two types of sites, e-commerce and social

---

[2] We thank the Area Editor for this suggestion.

networks, and two types of value "commodities" that these sites provide, informational value and entertainment value. Note that entertainment value can be construed broadly and can be considered "all other value" (i.e., non-informational) to some extent to reflect the diversity of experiences and types of value that Internet users might derive.[3] For convenience and brevity, however, in developing our theory, we distinguish informational and entertainment value.

Consistent with Becker (1965), these commodities enter a consumer's utility function, which the consumer then maximizes by allocating a fixed amount of Internet time available to social networking sites, shopping sites, or some combination of the two. Thus, the benefits of obtaining informational value and entertainment value from these websites must be balanced with the cost (in terms of time) of using those sites. More formally, in line with Becker (1965), Hauser, Urban, and Weinberg (1993), and Gronau (1977), consumers maximize their utility subject to their time constraint:

(1) $$\max \text{Utility} = \text{entertainment value} + \text{informational value}$$

$$\text{entertainment value} = f(E_{sh}, t_{sh}, E_{sn}, t_{sn})$$

$$\text{informational value} = g(I_{sh}, t_{sh}, I_{sn}, t_{sn})$$

$$\text{s.t. } T = t_{sh} + t_{sn},$$

where E and I represent the specific entertainment and informational content available on shopping and social networking websites (subscripted with sh and sn, respectively), t is the time input needed to realize the content on the two types of websites, and T is the total time dedicated to these activities. We assume that both types of websites provide some levels of entertainment and informational content, with social networking sites providing relatively more entertainment

[3] We thank an anonymous reviewer for this suggestion. We note that our model can be expanded to accommodate multiple types of value (i.e., other than entertainment and information) that consumer derive from visiting websites. For example, we also explored a three-dimensional extension to our model (the results are available upon request) and found that it does not substantially change our key takeaways.

content and e-commerce sites providing relatively more informational content (I_{sn} < I_{sh} and E_{sh} < E_{sn}).

We also assume a diminishing marginal return functional form for $f(\cdot)$ and $g(\cdot)$. This means that marginal returns in either shopping or social networking sites diminish with more time spent on the site. This reasoning follows previous research that proposes diminishing returns to the value of information when a consumer spends more time with a given information source (Ratchford, Lee, and Talukdar 2003). Diminishing marginal returns of resource allocation are also well established in the literature not only with respect to time allocation (Gronau 1977; Hauser, Lee, and Talukdar 1993) but also in other contexts, including marketing resource allocation (Dorfman and Steiner 1954; Fischer et al. 2011; Morey and McCann 1983) and customer relationship management (Venkatesan and Kumar 2004). For example, researchers find that marketing budgets should be allocated to different products as a function of incremental sales (Fischer et al. 2011) or proportional to demand elasticities (Dorfman and Steiner 1954). Diminishing returns also occur when firms decide how to allocate their marketing resources to maximize customer lifetime value using, at least in part, consumers' contribution margin (Venkatesan and Kumar 2004).

Following Gronau (1977) and Hauser, Lee, and Talukdar (1993), we specify a log functional form to reflect the customary diminishing marginal returns of time allocated to either shopping sites or social networks sites. We then solve for the time spent on shopping sites and social network sites to show the trade-off effect:[4]

[4] For the sake of brevity, we provide a summary of our analytical results here. Full details are in Web Appendix A. Our analytical model assumes diminishing marginal returns with respect to the information and entertainment value gained from time spent on social network and shopping sites. Our choice of a log functional form is driven mainly by analytical convenience and popularity across previous studies. Our hypotheses are also robust under a more

(2) $$\text{entertainment value} = E_{sh} \times \ln(t_{sh}) + E_{sn} \times \ln(t_{sn})$$

$$\text{informational value} = I_{sh} \times \ln(t_{sh}) + I_{sn} \times \ln(t_{sn})$$

(3) $$t_{sn}^* = \frac{(E_{sn}+I_{sn})t_{sh}^*}{E_{sh}+I_{sh}}.$$

The result implies that the optimal time spent on social networking sites is proportional to the ratio of contributions to the entertainment and information values the two sources produce.

Buying Behavior

Next, we use our time allocation model to help explain consumers' buying behavior, which is central to our empirical study. To demonstrate the association between informational value accumulated from using the Internet and online purchasing behavior, we draw on a well-established premise that choice decisions are a function of the amount of product-related information to which a person is exposed. For example, Meyer (1982) finds that choice decisions may stem from a sequential elimination process that consumers work through by using the information they gather through search. This search process helps people form preferences and leads to purchase activity (Meyer and Sathi 1985). In addition, consumers commonly need to search for product-related information before they make purchase decisions (Branco, Sun, and Villas-Boas 2012). Information search may also increase the utility people gain from their brand choice decisions (Hagerty and Aaker 1984), and it helps consumers reduce their time cost during shopping sessions and enables them to select better-matched products (Ratchford, Lee, and Talukdar 2003). Researchers also find that the availability of product information online (e.g.,

---

general functional specification: $(x) = x^r$, $(0 < r < 1)$, where r is the rate of diminishing marginal returns in entertainment or informational value acquired. We thank the Area Editor for helpful guidance here.

price, attributes, recommendations) contributes to purchase activity (De, Hu, and Rahman 2010; Klein and Ford 2003; Ratchford, Lee, and Talukdar 2003).

In summary, informational value accumulated through the use of social networks (e.g., exposure to other consumers' experiences, recommendations, and stated preferences) and e-commerce sites (e.g., exposure to brands and products) can lower consumers' search and information costs. Accordingly, this may make it easier to buy products, thus increasing consumers' purchasing probability. We model the amount of informational value as follows:

(4) $$\text{informational value} = \mathrm{I_{sh}} \times \ln(\mathrm{t}^*_{\mathrm{sh}}) + \mathrm{I_{sn}} \times \ln(\mathrm{t}^*_{\mathrm{sn}}).$$

The Negative Correlation Between Immediate Social Networking Site Usage and E-Commerce

We first consider the negative relationship between social network usage and e-commerce. In the absence of time allocation considerations, the usage of both social networking and shopping sites is likely to be associated with an increase in online purchasing (through the accumulation of information), with a smaller buying lift for social network usage and a greater buying lift for shopping site usage. However, when this relationship is subject to time allocation constraints, a greater immediate usage of social networking sites will reduce the usage of shopping sites, implying less purchase activity. A consumer may choose to spend more time on social networking sites because the entertainment value is higher on these sites. According to the model we propose, when a consumer spends ε more time on social networking websites, she or he reduces the corresponding portion of Internet time spent on shopping sites. This may arise because the higher entertainment value found on social networking sites may decrease the attractiveness of subsequent shopping sessions, especially for product categories associated with greater entertainment value (e.g., clothing, children's products).[5] Because the informational value obtained from spending ε more time on social networking sites is less than what could have been obtained from spending time on e-commerce websites ($I_{sn} < I_{sh}$), this substitution with respect to time allocation leads to lower informational value—even with diminishing marginal returns —accumulated in the given period and thus correlated with less purchase activity.[6]

We illustrate this mechanism in Figure 1. All else equal, the information extracted per unit of time spent on shopping sites yields higher returns than that per unit of time spent on social networks: $\tilde{I}_{sh} > \tilde{I}_{sn}$ (left side of graph). Thus, purchase probability increases less rapidly when the user is on social networking sites. This leads to our first hypothesis:

H_1: In the short run, usage of social networking sites is negatively correlated with buying behavior (short-term substitution).

[Insert Figure 1 about here]

The Positive Correlation Between Cumulative Social Networking Site Usage and E-Commerce

While we argue that the immediate use of social network sites is correlated with less e-commerce activity, we also propose that cumulative usage of social networks—defined as spending more (vs fewer) days within any given time period on social networking sites—is positively correlated with buying behavior. This follows from Becker's (1965) theory and is in

[5] We also explore these category-specific effects empirically later in the paper.
[6] At the optimal time allocation for social networks, the incremental benefit of spending more time on social network sites with respect to informational value is lower than that of spending more time shopping, even when considering diminishing marginal returns. This result occurs because people allocate time to maximize both informational and entertainment value. Thus, the optimal time allocation for social networks is greater than that if people allocated time based purely on informational value. This means that a reduction in immediate social network usage will increase informational value (for additional details, see Web Appendix A).

line with the common assumption of diminishing marginal returns from the informational value obtained from browsing websites.

Drawing on research on the theory of time allocation, Gronau (1977) and Hauser, Urban, and Weinberg (1993) posit that incremental time spent on a given resource or effort can result in diminished marginal returns. Similarly, others find evidence that diminishing marginal returns occur when allocating marketing resources (Dorfman and Steiner 1954; Fischer et al. 2011; Morey and McCann 1983; Venkatesan and Kumar 2004). Based on this theory, we reason that although social networks may provide less informational value than shopping sites, the initial incremental daily time t spent on social networks (see Figure 1) may result in greater informational value ($\tilde{I}_{sn}$ in Figure 1) than if the incremental time was spent on the shopping sites that have already received a substantial amount of consumer's attention ($\hat{I}_{sh}$ in Figure 1). This implies that spending more days on social network sites can result in greater total informational value extracted (which is correlated with greater purchase activity) than spending less days.

Using the model described in Equation (1), we can assess the impact of cumulative social network usage (e.g., daily usage) by comparing the accumulation of informational value under N and n days of social network usage, where N > n. In the latter scenario, there are N-n days with no social network usage. The expected informational value on each of these no-social-network days is always $\mathrm{I_{sh}} \times \ln(\mathrm{T})$. On days with social network usage, we calculate (in Web Appendix A) the expected informational value as follows:

(5) $$\mathrm{E[informational\ value]} = \mathrm{I_{sn}}[\ln(\mathrm{T}) - 1] + \mathrm{I_{sh}}[\ln(\mathrm{T}) - 1].$$

If we compare these two cases, using social networking sites over N days (compared with n days) generates more informational value under the following condition:

(6) $$\frac{\mathrm{I_{sn}}}{\mathrm{I_{sh}}} > \frac{1}{[\ln(\mathrm{T}) - 1]}.$$

In other words, as long as there is informational content available on social networks, allocating a small ε amount of time to social networking sites increases the amount of informational value, which is correlated with increased purchase activity. It follows that if ε is measured in smaller units (i.e., seconds), T becomes large (total available seconds), and 1/[ln(T) – 1] becomes small.

We note that this is different from the substitution effect proposed previously. In the prior example, we assume that consumers already spend time on social networking sites and that the variation in immediate time spent (greater or less) is correlated with differences with respect to e-commerce activity. Thus, conditional on already allocating time to social networking activities on a particular day, more time spent on these sites results in less informational value due to less time spent on shopping sites. Here, with respect to cumulative social network usage, we are comparing variation in day-to-day social network usage incidence across an extended period of time (days or even months). Our assumption is that consumers can obtain new information on social networking sites on a day-to-day basis, providing them with greater marginal value relative to their initial daily time allocation to social networking activities. Thus, a consumer who frequently engages in social networking activities each day of the week, compared with a consumer who visits these sites only once per week, will potentially generate greater marginal informational value. Overall, this suggests that, in the long run, there is a positive correlation between social networking usage and shopping activities rather than a negative one.

Of course, this outcome relies on the premise of new informational value on social networking sites across different days. We argue that this is indeed the case. Consumers are often exposed to product and consumption-related content as well as opinions and eWOM posted by other consumers (e.g., Chen, Wang, and Xie 2011; Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Godes and Mayzlin 2004; Kumar et al. 2013; Liu 2006; Moe and Trusov 2011)[7] As such, consumers can discover new products to purchase through the content that others (e.g., friends) share. Importantly, consumers are exposed to a *wide variety* of products and consumption-related activities on social networks. Thus, cumulative exposure, even in low amounts, could be correlated with greater e-commerce activity. We provide the formal derivation of this outcome in Web Appendix A.

We also recognize that in social network sites, consumers are also exposed to negative content that might dissuade them from purchasing particular products (e.g., negative eWOM).[8] However, we argue that two key factors might contribute to the long-term positive impact of consumption-related information. First, it is generally acknowledged that the majority of user-generated content tends to be positive (Chevalier and Mayzlin 2006; Moe and Schweidel 2012). For example, more than half of the reviews on Amazon are 5 stars, with the average being 3.9 (Woolf 2014). In addition, there is a tendency among social network users to present themselves and their experiences in a positive light (Toubia and Stephen 2013). In other words, the "share of voice" for positive consumption-related content is larger than that of negative content. Second, and arguably more important, negative consumption-related information is typically associated with a particular product or service experience (i.e., negative reviews). Clearly, such information might be influential in driving other consumers away from making certain purchases (East, Hammond, and Lomax 2008, Ho-Dac, Carson, and Moore 2013). However, because such reviews are product focused, it is unlikely to suppress consumers' general shopping interest for an entire product category. On the contrary, negative information may actually help consumers

---

[7] We tested the assumption that the use of social network websites correlates with purchase activity, due to exposure to informational value, with a simple experiment (details of the experiment appear in Web Appendix C). The results of this experiment suggest that consumers browsing a social network website (vs. a control website) are more likely to be exposed to consumption-related content, which is associated with greater shopping intentions. Thus, our assumption is supported through experimental data.

[8] We thank an anonymous reviewer for bringing this point to our attention.

find a better match for their preferences, thus increasing overall satisfaction from their shopping experience. Accordingly, we hypothesize that consumption-related content is positively associated with purchase decisions in the long run.

H_2: In the long run, cumulative social network usage is positively correlated with buying behavior.

In addition, we propose that product category moderates the relationships we have described. If social network engagement is positively correlated with online shopping activity because social networking sites continually expose people to new information about products and brands, then that correlation should be stronger for product categories that are more commonly mentioned in people's posts on social networking sites. This is due, for example, to the tendency of social network users to project highly positive self-images to others through their posts on these sites and to feature themselves using certain products, particularly higher-status items or products that signal something favorable about oneself or one's family, as a way to signal a positive self-image (Gonzales and Hancock 2011; Wilcox and Stephen 2013). For example, we would expect a stronger positive correlation in categories such as chocolate (an indulgence) and children's products (indicating one's status as a parent) but not in categories such as automotive parts (which are, for most consumers at least, unglamorous). We also expect a stronger correlation in categories that have a higher incidence of unplanned purchases, as this is associated with casual retail browsing (Moe and Fader 2004; Stilley, Inman, and Wakefield 2010). Finally, we expect the proposed negative correlation between immediate social network usage and online shopping activity to also be moderated by product category. If consumers are getting more entertainment value from social networks, they may be less inclined to also shop for products that provide them with entertainment value in the shopping process. This may be correlated with a reduction in sales for product categories such as clothing, children's goods, and jewelry. Thus, we advance our final hypothesis:

H_{3a}: The correlation between cumulative social networking usage and online purchase incidence is moderated by the product category of the purchased item.

H_{3b}: The correlation between immediate social network usage and online purchase incidence is moderated by the product category of the purchased item.

Empirical Study Using Internet Usage Data

To explore the associations between social network usage and online shopping activity, we take advantage of a rich, individual-level data set provided by a major global market information and measurement company. The data set features a panel of 10,192 individuals whose online activities were tracked during the full year of 2007 and has three components: individual online purchase records, full web-browsing history, and demographic information. The data provider selected these panelists to be a representative sample of the U.S. population.[9]

Because the web-browsing data include details on each site visitation, we investigate the data at the Internet session level.[10] Each Internet session, as defined by our data provider, begins when a user first opens a website, continues when the user opens other websites, and ends when the user closes all opened website pages or is inactive for more than 30 minutes (see Figure 2). This is slightly different from a website session (which begins when a user opens a website and

[9] To provide a representative panel of online users, our data provider recruits an online panel using random digital dialing and online technologies. The exact method is proprietary. However, as a check, we compare statistics of our data sample with that from the general U.S. population (retrieved from the 2010 U.S. Census [US Census Bureau 2010]). In our data (U.S. Census data in parentheses), the average age is 47 (U.S. Census: 47.5 when only counting adults), average income is about $52,000 (U.S .Census: $53,400), male to female ratio is .48 (U.S. Census: .484), and percentage of families with children under the age of 18 is 37% (U.S. Census: 40%). This suggests that our sample is representative of the U.S. population on those characteristics.
[10] We also analyze our data at the daily level. The results are available from upon request and are consistent with respect to our findings.

ends when the user leaves the website by closing the page, progressing to a different domain on the same browser page, or remaining inactive for more than 30 minutes) because it accounts for visits to multiple websites in one sitting. We choose to analyze our data at the Internet session level rather than at the website session level because the latter raises simultaneity concerns. Specifically, once a user opens multiple websites, we are unable to identify the order of use within these concurrent sessions. In contrast, we are able to identify distinct Internet sessions without overlap in usage activity because, by definition, Internet sessions are separated by periods with no Internet usage.

[Insert Figure 2 about here]

Online Purchase Records

The data set includes *all* online purchases made by the panelists between January 1 and December 31, 2007. For each purchase, we know who made the purchase, the purchase date, the retailer (i.e., website domain) where the purchase was made, and the category of purchase. We observe 140,291 distinct purchase transactions made during this period by 7,402 unique users. Figure 3 shows a distribution of purchase activity by user. Of our panelists, 73% purchased at least once during the observation period, and conditional on having made a purchase, each individual made an average of 19 purchases over the course of the year.[11]

[Insert Figures 3 & 4 about here]

We show the aggregate purchasing activity for our panel in Figure 4. As expected, we observe seasonal fluctuations (e.g., spikes in shopping activities around the holiday season). This is especially pronounced in the months of November and December, with approximately 25%

[11] The top 10 retailers and their distribution of purchase and shopping activities can be found in Web Appendix E.

and 31% greater shopping activities compared with the other months. Therefore, in our subsequent empirical analysis, we control for these seasonal variations using time-specific random effects.

From the purchase records, we create three variables to measure panelists' online shopping activities: (1) purchase observation, (2) the number of purchases, and (3) the number of retailers purchased from. The first variable, $PURCHASE_{it}$, records, on a session basis, whether individual i made a purchase in session t. The second variable, $NUMPURCHASE_{it}$, records the number of purchases made in session t, which represents a panelist's shopping activity intensity for that session. The third variable, $NUMRETAILERS_{it}$, records the number of retailers from which a panelist made purchases in each session. We measure each of these variables at the individual session level. Table 1 provides the summary statistics of these variables.

[Insert Table 1 about here]

Web-Browsing History

The second component of our data set is a detailed history of website visitation for all members of the panel during the observation period. This includes records of the domain (e.g., Facebook.com) and domain pages, a time stamp of each website visit, the website category, and duration of the visit. The majority of Internet usage by our panelists in 2007 was related to entertainment, e-commerce, news, search, social communities, and Internet services (see Web Table E3 in Web Appendix E for a full list of categories and subcategories). In our analysis, we focus on two specific subcategories of website visit (i.e., browsing) observations: mass merchandiser and member communities. Mass merchandisers include browsing visits to online retailers. Thus, we have detailed information on each user's shopping browsing activities, captured through shopping duration and shopping sessions. The member communities subcategory includes visits to domains for social networking websites, such as Facebook.com. Given our focus on social networks and online shopping activities, we do not differentiate among other website domain categories that cover the rest of our sample's Internet usage, but we do use this information as a baseline for all other non-social/non-shopping-related online activity.

The first set of measures derived from web browsing covers activities related to social networks. In our data, the social networking sites are Facebook and Myspace, which together account for nearly half of all social network browsing sessions.[12] We combine the two and refer to them as our "social networking" sites because they place a heavy emphasis on the creation of and interaction with social ties (i.e., "friends").[13] The two social networking sites facilitate a rich variety of content exposure through social posts. As a result, we expect that consumers are exposed to a significant variety of new consumption- and shopping-related experiences in their social networks.

We created variables to measure social network usage and engagement as follows. First, for each individual, we record the incidence and total duration of social network sessions during each Internet session. The variables SN-SESSION_{it} and SN-DURATION_{it} record the incidence of social network usage and the usage duration for user i during session t.

Second, we created SN-ACTIVE_{it} to denote whether an individual was participating in social networks before session t. Because our data were left-censored, we do not have

---

[12] We note that Facebook launched its mobile app in July 2008 (after our observation period). Therefore, we expect our data to capture the majority of usage because individuals are limited to web browser–based interactions.
[13] We note that social networks in 2007 (our data time frame) are different from the dominant platforms in the present. In 2007, Facebook contained a news feed on individual profiles with photos and information about groups, interests, updates, and other user-generated content. Myspace was similar in that included a profile and also allowed users to share posted information with friends. At present, Facebook contains an aggregated news feed that combines postings of all friends. We expect that the current format exposes users to a greater variety of consumption-related content (e.g., sponsored product placement) than the format in our data set.

information on when each individual first visited a social networking site. Instead, we infer this action from the data. Figure 5 plots the first day of observed usage of social networking websites for our panel in January 2007. Many users visited social networking sites from the beginning of our data set, which suggests that they may have been active users of social network before January 1, 2007. We also observe a kink in Figure 5 at about 13 days, suggesting that the initial group of individuals may have been previously active and different from the users who we observe arriving later.[14] Thus, we assume that users who we first observe on social networking sites after two weeks are new users and were not active in social networking websites previously. For these "new" users, we set SN-ACTIVE$_{it}$ to 0 before they first visited a social networking website and to 1 after they made their first visit. For all other users, we set SN-ACTIVE$_{it}$ to 1 for all sessions in our data set.[15] To further control for users who were active in the first month, we include a dummy variable, SN-FIRSTMONTH$_{i}$, which equals 1 if we observe an individual on a social networking website during the first month.

[Insert Figure 5 about here]

Third, we created SN-ENGAGEMENT$_{it}$ to measure the cumulative engagement in social networking sites. This variable records the total number of days that user i has been active on social networks before session t.[16] For example, we set this variable to 1 during the first day a user actively visits social networking websites. For the second day with an observed visit, we set

[14] One possible way to deal with left censoring is to drop the early observations in our data. We run a robustness check by dropping the left-censored users from the analysis. The results are consistent with our main results and are available in Table E.5 in Web Appendix E.

[15] To explore the validity of this classification of new versus existing users, we measure the average interarrival time to social networking websites. We find that, on average, consumers visit these social netwrk websites about once a week. For social networking websites, the mean number of days between usage, across individuals, is 6.8 (SD = 5.67). Therefore, we expect that the average user is a new user if he or she has not visited a social networking website during the first two weeks. We also calculate new users as those for whom we observe a first usage on social networking websites four weeks after the beginning of our observational period. The results of this robustness check (available upon request) are equivalent in terms of sign and significance.

[16] To test robustness, we also calculate this variable as the number of sessions and duration on social networking sites. The results are consistent with our main results and are available upon request.

this variable to 2, and so on. We use this variable to capture cumulative usage of social networking sites.

Fourth, we created measures related to users' shopping browsing behavior. We created two variables to account for this: shopping session count and shopping duration. SHOP-SESSION$_{it}$ records the number of unique visits user i made to e-commerce websites in session t, and SHOP-DURATION$_{it}$ records the total time (minutes) user i spent on e-commerce websites in session t. We also record the total Internet session count and duration for user i in session t, exclusive of shopping or social networks, as covariates (NET-SESSION$_{it}$ and NET-DURATION$_{it}$). These variables are important because they allow us to controls for individual i's Internet activity in session t. In other words, they capture the variation in a user's Internet activity over time. In addition, we include cumulative measures of shopping and Internet usage (days that user i has visited the respective site before session t), labeled SHOP-CUMULATIVE$_{it}$ and NET-CUMULATIVE$_{it}$.

Fifth, in addition to social network and shopping metrics, our web-browsing history data allows us to infer referral metrics that identify which domains help lead consumers to online retailers. While we do not have clickstream data to identify referral sources precisely, we use our web history data to infer referrals to e-commerce websites. Specifically, we define a referral site as one that a user visited immediately before going to an e-commerce website if the referral session ended less than five minutes from the start of the shopping session. Therefore, if we do not observe any activity within five minutes before a shopping session, we assume that the shopping session was not initiated by a referral. The referral activity allows us to assess the relationship of the previously visited website with purchase activity. Thus, this helps us identify potential referral effects of social networks, where consumers may discover a new product or referral link and immediately transition to an e-commerce website, from the cumulative effects of social networks. We separately identify three different categories of referral websites: social network, search, and shopping (see Table 2). There are many referrals from both e-commerce websites and search engines (i.e., Google and Yahoo). In contrast, social network visits precede only about 2% of all shopping sessions as referrals. Using this referral information, we created three variables: (1) SN-REFERRAL_{it}, (2) $\text{SEARCH-REFERRAL}_{it}$, and (3) $\text{SHOP-REFERRAL}_{it}$. These are dummy variables equal 1 if we observe a referral to an e-commerce website for individual i in session t.

[Insert Table 2 about here]

Demographics and Control Variables

The final part of the data set comprises demographic information for each panelist and several control variables. Table 3 provides summary statistics on age, gender, household size, income, and number of working members in a household, as well as an indicator for children in the household. Note that we do not have a continuous measure for income, which we instead measure as one of four categories (see Table 3). We use these demographic variables to control for observed heterogeneity in our model. Note that the average age is greater than what might be expected on social networks because our data set captures a representative sample of the U.S. population.

[Insert Table 3 about here]

Empirical Model

We now describe how we model the relationship of individuals' immediate usage of and engagement with social networking sites with their online shopping activities. As we mentioned previously, shopping activity is indicated by three variables: (1) a binary session-level purchase action, (2) the number of purchases made each day, and (3) the number of retailers purchased from each day. Thus, our empirical analysis is composed of these three components.[17]

Purchase Decision

We use a hierarchical binary probit model to assess the relationship between user i's social network activities and his or her decision to purchase from an online retailer in session t. Let $y_{it}^{p} = 1$ if individual i makes a purchase in session t and $y_{it}^{p} = 0$ if we do not observe any purchase activity during that session. The superscript p indicates the parameters and data pertaining to the purchase decision model. We specify individual i's latent purchase utility in session t as follows:

$$\text{(7)} \quad Z_{it}^{*} = \beta_{i1}^{p} \cdot y_{i,t-1}^{p} + \boldsymbol{\beta}_{\mathbf{i2}}^{\mathbf{p}} \times \text{SOCIALNETWORK}_{i,t-1} + \boldsymbol{\beta}_{\mathbf{i3}}^{\mathbf{p}} \times \text{INTERNET}_{i,t-1} + \boldsymbol{\beta}_{\mathbf{i4}}^{\mathbf{p}} \times \text{REFERRAL}_{it} + \boldsymbol{\delta}^{\mathbf{p}} \times \text{DEMOGRAPHICS}_{i} + \epsilon_{it}^{p},$$

where each set of variables is defined as follows: (a) $y_{i,t-1}$ is the purchase decision in the previous session, which accounts for state dependence in purchasing; (b) $\text{SOCIALNETWORK}_{i,t-1}$ is a set of lagged variables describing social network activity; (c) $\text{INTERNET}_{i,t-1}$ contains lagged non-social Internet activity (with separate measures for shopping and non-shopping, non-social sites); (d) REFERRAL_{it} is the three types of referrals before a shopping session (referrals from search, social, and non-social sites); and (e) DEMOGRAPHICS_{i} is the demographic

---

[17] These models are similar in nature to models in the customer relationship management literature (e.g., Venkatesan, Kumar, and Bohling 2007; Verhoef et al. 2010).

information for user i (see Table 4). We note that for each individual, $\boldsymbol{\beta}_{\mathbf{i2}}^{\mathbf{p}}$, $\boldsymbol{\beta}_{\mathbf{i3}}^{\mathbf{p}}$, $\boldsymbol{\beta}_{\mathbf{i4}}^{\mathbf{p}}$, and $\boldsymbol{\delta}^{\mathbf{p}}$ are vectors of parameters that represent the coefficients for each set of variables (e.g., $\mathrm{SOCIALNETWORK}_{\mathrm{i,t-1}}$ is a matrix that contains the eight variables listed in Table 4, and $\boldsymbol{\beta}_{\mathbf{i2}}^{\mathbf{p}}$ are the coefficients for those variables).

Our goal in Equation 7 is to examine the relationship of social network usage and engagement with purchase activity. Of course, we need to control for many factors that may confound our results, such as unobserved individual-level heterogeneity, correlated unobservables, and simultaneity. First, unobserved differences across consumers could potentially bias our results. For example, some consumers may be particularly savvy Internet users and thus be more active on both social networking and e-commerce websites. Our data may also reflect different types of consumers, such that some prefer to participate in social networks while others do not. This is an issue if different types of consumers also exhibit differences in online purchase patterns.

[Insert Table 4 about here]

Second, correlated unobservables are factors related to both social network engagement and online purchase activities but are unobservable to the researcher. One example in our setting is activity bias, where consumers who are active online in a given session might be more likely to visit social networking websites *and* shop online.[18] This becomes a concern if there is high variation in day-to-day Internet use because the relationship between social network usage and online shopping could be driven by general Internet activity (e.g., “lumpy” Internet usage; see Lewis, Rao, and Reiley 2011). Another example is time-varying marketing-mix and webpage

[18] Activity bias occurs when, on some days, users spend a lot of time online (engaging in various activities) and, on other days, spend very little time online. We thank an anonymous reviewer for highlighting this important factor.

content. Online retailers may run time-sensitive promotions on social networking websites, which may result in increased social network usage and online purchase for certain days.

To control for unobserved individual-level heterogeneity and correlated unobservables, we follow Hartmann et al. (2008) and Nair, Manchanda, and Bhatia (2010) by including individual-level random effects. Specifically, we decompose the error term in the Equation 7 into the following:

(8) $$\epsilon_{it}^{p} = \alpha_i + \zeta_t + \nu_{it},$$

where, following the probit model specification, $\nu_{it} \sim N(0,1)$. In addition, α_i is a random effect specific to individual i and controls for time-invariant unobservable customer heterogeneity (e.g., Kumar et al. 2013) and correlated unobservables. For example, potential customer-level heterogeneity could include a person's general preference for online shopping or whether he or she tends to browse websites with particular purchase decisions in mind. For activity bias, Equation 8 allows us to control for unobservable individual-specific factors such as a person's tendency to go online and, as a result, shop and visit social networking websites more (we also capture observable Internet activity using $\text{NET-SESSION}_{i,t-1}$ and $\text{NET-DURATION}_{i,t-1}$, which are time-varying and individual-specific to control for activity bias by capturing any variation in individual i's Internet activity across time). Furthermore, ζ_t is a random effect specific to session t and controls for time-varying unobservable heterogeneity and time-varying correlated unobservables. These include seasonal effects, such as those found in Figure 4 during the holiday shopping period. In addition, given the early stage of social network adoption in 2007, ζ_t also controls for time-varying changes in consumer behavior with regard to social network usage and online shopping. For identification purposes, we assume that ζ_t is independent of the other error terms in the model and is normally distributed with a mean fixed at 0: $\zeta_t \sim \mathrm{N}(0, s)$, where s is the variance hyperparameter for the prior distribution of ζ_t.

Finally, simultaneity may be a concern. Simultaneity occurs if consumers concurrently use social networking websites and shop online. We control for this issue by including lagged social network activity in the model (i.e., lagged duration and sessions) and examine the relationship of these lagged variables with a consumer's current online purchase activity.[19] Thus, we assume that a person's social network usage at time t – 1 is not affected by that person's future shopping activity at time t.

Number of Purchases and Number of Retailers Purchased From

To assess the effects of social network usage on other shopping-related activities, we also examine the relationship between social network activity and both the number of purchases made and the number of retailers purchased from. The number of purchases and the number of different retailers purchased from are count-dependent variables (i.e., nonnegative integers) and are modeled taking this into account using conditional Poisson distributions. Let $y_{it}^{r} = k$ if individual i makes k purchases (r = 1) or makes a purchase at k different retailers (r = 2) during session t and $y_{it}^{r} = 0$ if we do not observe any purchase activity. Here, the r superscript refers to the parameters and data used in the models assessing the number of purchases (r = 1) and the number of retailers purchased from (r = 2). Because we observe many instances in which $y_{it,}^{r} = 0$ (i.e., purchase activity is relatively uncommon at the daily level for any given individual), we use a zero-inflated model. Specifically, we use a finite mixture between a Poisson distribution (with probability P_{it}^{r}) and a degenerate distribution concentrated at zero (with probability $1 - P_{it}^{r}$) to

[19] As a check, we run a robustness test with same-day social network variables, and the findings are consistent.

model the number of purchases or the number of retailers purchased from (Lambert 1992). We specify the likelihood of individual i's purchase activity as follows:

$$\text{(9)} \qquad L(y_{it}^{r}, \lambda_{it}^{r}) = \begin{cases} \frac{P_{it}^{r} \exp(-\lambda_{it}^{r})(\lambda_{it}^{r})^{k}}{k!}, y_{it}^{r} = k \\ (1 - P_{it}^{r}) + P_{it}^{r} \exp(-\lambda_{it}^{r}), y_{it}^{r} = 0 \end{cases},$$

where P_{it}^{r} is the probability that individual i's behavior follows a Poisson distribution with parameter λ_{it}^{r}. We model λ_{it}^{r} as follows:

$$\text{(10)} \qquad \log(\lambda_{it}^{r}) = A_{it}^{r} + \beta^{r}{}_{i1} \times y_{i,t-1}^{r} + \boldsymbol{\beta}_{i2}^{r} \times \mathrm{SOCIALNETWORK}_{i,t-1}$$

$$+\boldsymbol{\beta}_{i3}^{r} \times \mathrm{INTERNET}_{i,t-1} + \boldsymbol{\delta}^{r} \times \mathrm{DEMOGRAPHICS}_{i}.$$

The variables entered into Equation 10 for the conditional mean of the Poisson distribution are the same as those used in the previous hierarchical binary Probit model, with one exception: we do not include the referral variables because we are modeling the number of purchases and shopping activities during a given session rather than a decision to purchase.[20] Again, we used a random-coefficients specification to capture unobserved heterogeneity, correlated unobservables, and activity bias: $A_{it}^{r} \sim N(\overline{A}^{r}, q^{r})$. Thus, the models for number of purchases and number of retailers purchased from are hierarchical zero-inflated Poisson models. Note that the hierarchical nature of the model allows for overdispersion, thus permitting distributional flexibility without using a more complex count-data distribution (e.g., negative binomial, double Poisson).

[20] Given that users can purchase multiple products across various websites, it is less obvious how to attribute the number of purchases or the number of sites with purchase activity to one specific referral site.

Methodology

Although we control for many confounding factors, such as activity bias in the previous models, there may still be other potential endogeneity concerns, such as time-varying unobservables. To address other potential sources of endogeneity, we estimate Equations 7–10 using the latent instrumental variables (LIV) approach (Ebbes et al. 2005; Rutz, Bucklin, and Sonnier 2012; Rutz and Trusov 2011; Zhang, Wedel, and Pieters 2009).[21] We also allow for the possibility of correlation of purchase, number of purchases, and number of retailers purchased from using a multivariate normal copula (e.g., Danaher and Smith 2011; Kumar, Zhang, and Luo 2014; Stephen and Galak 2012). The details are provided in Web Appendix B.

Results

Results for Focal Social Network Variables

We now discuss the specific parameter estimates pertaining to the relationship between social network usage and purchase incidence. We report our main results in Table 5, which shows how social network activity is associated with consumers' purchase decisions. We organize this discussion to first focus on the key social network variables. Then, we report the results for the other Internet usage variables and control variables.

The parameter estimates for the social network variables are consistent with our theory and support H_1 and H_2, suggesting that social network usage and engagement are correlated with online shopping activities. Two specific sets of results are worth noting. First, we find that

[21] Before using the LIV approach, we first ran a test to check for potential endogeneity for all "suspect" predictors individually using the Hausman-LIV (HLIV) test statistic (Ebbes 2004)..The HLIV statistics for SN-ENGAGEMENT and SN- SESSION are 6.18 and 5.89, respectively (follows asymptotically a chi-square distribution). This suggests the need to correct for endogeneity. The validity of the LIV application is also subject to some assumptions (i.e. normality of the error term and non-normality of the endogenous regressor). We test for these in Web Appendix G.

immediate (or recent) time spent on social networking websites is associated with a decrease in purchase activity (Table 5, Column 1: SN-SESSION = –.3850, SN-DURATION = –.0516). We also report the results from the two other shopping activity measures that corroborate this finding: the number of purchases and the number of retailers purchased from. Columns 2 and 3 in Table 5 the report results for these measures. Higher immediate usage of social networking websites is associated with making fewer purchases (Column 2: SN-SESSION = –1.2704, SN-DURATION = –.1893). This is in line with H_1. Similarly, more immediate usage of social networking websites is associated with making purchases from fewer websites, which is also consistent with our theory (Column 3: SN-SESSION = –1.8318, SN-DURATION = –.2108).

Second, we find a positive association between cumulative social network usage and shopping activities, which is consistent with H_2. Greater cumulative usage of social networking sites is correlated with a higher probability of purchasing (Column 1: SN-ENGAGEMENT = .0846). This is corroborated by the other two shopping measures available in our data. More social network engagement (i.e., cumulative use) is also positively associated with the number of purchases made (Column 2: SN-ENGAGEMENT = 1.2768) and the number of retailers purchased from (Column 3: SN-ENGAGEMENT = 1.3623).

These findings demonstrate that the immediate and cumulative usage of social networking websites may play two distinct, opposing roles in their relationship with shopping activities. First, these results point to an immediate, short-term negative relationship, such that time spent on social networking is negatively correlated with online shopping, which suggests a substitution effect. Second, our results indicate that there is a cumulative, longer-term positive relationship, such that cumulative usage of or engagement with social networking websites over time is associated with increased purchasing activity. In line with our theory, this may be because people who have been engaged in social networks for an extended period of time are better informed about new shopping options that they learned from their networks. In other words, continuous exposure to new consumption-related content over time may provide greater marginal benefit than if that time was spent on shopping websites.[22]

Results for Other Variables

We also consider the relationship of other Internet-related variables with purchase activity in Table 5. These variables allow us to control for short-term effects (e.g., eWOM) that are correlated with immediate purchase. We find that search referrals (i.e., Google and Yahoo web search) have a positive relationship with the probability of purchase (e.g., SEARCH-REFERRAL = .0075). This means that search activity is associated with subsequent purchase activity, which intuitively makes sense (e.g., Hauser and Wernerfelt 1990; Ratchford and Srinivasan 1993).

We also control for several other important factors. For example, NET-DURATION captures an individual's session-level Internet activity and, along with the individual-level random effect, controls for possible activity bias. Our results suggest a positive relationship between the duration of a person's Internet session usage and both the number of purchases made (NET-DURATION = .0492) and the number of retailers purchased from (NET-DURATION = .0510). SHOP-SESSION and SHOP-DURATION control for the propensity to engage in shopping-related activities. As we expected, these shopping-related activities are positively associated with shopping activities. In the purchase incidence model shown in Table 5,

[22] We also find that number of shopping sessions mediates the relationship between cumulative social network usage and purchase incidence. Thus, our findings suggest that the correlation between increased engagement on social networking websites and the higher likelihood of making online purchases is due in part to increased shopping browsing activity. The results of this analysis are available upon request.

greater shopping duration is related to greater purchase incidence, which means that consumers who visit shopping websites or spend more time shopping are more likely to make a purchase (Column 1: SHOP-SESSION = .0118, SHOP-DURATION = .0078). We also find (Table 5, Column 2) that, consistent with our expectation, more visits and more time spent on e-commerce websites are correlated with the number of purchases (Column 2, SHOP-SESSION = .4887, SHOP-DURATION = .1692). Finally, in Column 3 of Table 5, we find that more time spent on e-commerce websites is associated with purchasing from a greater variety of retailers (Column 3: SHOP-SESSION = .7140, SHOP-DURATION = .1387).

As for the other demographic control variables, we also find that older consumers and those with more income are associated with greater online purchase activity.[23] Higher-income individuals may have more disposable income, which is correlated with greater online spend. We also find that women are more likely than men to make more purchases and buy from a wider variety of retailers. Our results also suggest an interesting relationship of household size and number of children with purchase activity. Having a larger household and/or more children are positively associated with greater purchase incidence but negatively associated with the number of retailers purchased from. We speculate this might be a result of larger households combining their purchases within one retailer. We also find that households with more working members are more likely to purchase online but buy from fewer retailers. Finally, as a check, the time-specific estimates are consistent with our expectations, with higher baseline purchase activity occurring during the December holiday season.

[Insert Table 5 about here]

[23] We also include interactions between the demographic variables and the other variables in the model to test whether demographic characteristics moderate any of the effects. The one interaction was that significant was income x $y_{i,t-1}$ (lag purchase), and it was positive. This suggests that higher-income consumers tend to have a higher purchase probability if they purchased during the previous day. The other interactions were not significant.

Category-Level Analysis

We also empirically examine product category data to provide further evidence that the positive correlation between social network engagement and purchase activities may arise due to greater consumption-related exposure.[24] In accordance with H_{3a}, we expect that product category moderates this positive correlation because consumers are more likely to be exposed to certain categories of products on social networks. In addition, product category might moderate the negative correlation between immediate social network usage and purchase incidence (H_{3b}). To test for this, we identify the product category for each purchase observation in our data. There are 21 categories under which the purchases fall (see Table 6). We use this information to create a new set of purchase incidence dependent variables, which we denote as $\mathrm{y_{itc}} = 1$ if consumer i purchased a product in category c during session t. For each category, we separately estimate the binary probit model using $\mathrm{y_{itc}}$ as the dependent variable.

In Table 6, we report the average partial effects for the key social network variables—both immediate social network usage (SN-SESSION) and cumulative social network usage (SN-ENGAGEMENT).[25] In Table 6, each row shows the partial effects from one purchase incidence model with the respective category purchase incidence described in Column 1 as the dependent variable. Consistent with H_{3a}, our results (Table 6, Column 1) suggest that product category moderates the positive correlation between social network engagement and purchase incidence. Specifically, categories such as clothing, children products, jewelry, and chocolates have significant, positive coefficients (Column 1: 5.079, .700, .428, .370, respectively). In contrast,

[24] We thank an anonymous reviewer for this suggestion.
[25] Partial effects show the effect on the probability of purchase for a unit change in the respective social network variable. We report partial effects for comparison across models. The estimated coefficients and other variables are consistent with our main results and are available upon request.

categories such as tools/hardware, auto parts, or gift cards have nonsignificant effects (Column 1: .132, .915, .018, respectively). This suggests that the positive correlation between social network engagement and purchase likelihood is more pronounced for products that are more likely to be shared on social networks.

Similarly, consistent with H_{3b}, we find that product category moderates the negative correlation between immediate social network usage and purchase incidence. Consumers with higher previous period social network usage may have satisfied their need for entertainment value, thus exhibiting lower purchase activity for entertainment-related products. While all the partial effects are significantly negative, consistent with H_1, we observe variation across the categories. The negative relationships in Column 2 of Table 6 appear stronger for categories that provide greater entertainment value (e.g., –11.046 for clothing, –1.891 for jewelry,–1.681 for children's products). While not perfect, to large extent, these results are in line with our theoretical framework.

[Insert Table 6 about here]

Managerial Implications and Conclusions

Our findings have two important managerial implications. First, our study offers strong evidence of the correlation between social network usage and online sales, which is a hotly debated topic among both marketing academics and practitioners. While some preliminary evidence—for example, that provided by IBM—suggests that social networks are not effective in driving e-commerce sales (Del Rey 2013; Gara 2012), our results suggest that that firms should not be discouraged by weak (immediate) sales performance of their social network campaigns. Rather, the positive association between cumulative social network usage and sales suggests that the expected payoff might be more of a longer-term phenomenon. This finding also suggests that popular performance metrics such as clickthrough rates on social networking sites (i.e., direct referrals) are likely to underestimate the total effects of firms' social network marketing campaigns. Thus, managers should also consider cumulative consumer interactions with their brands on social networking sites and not simply referrals.

Second, managers can benefit from learning that social network engagement is related to people's propensity to shop. Importantly, this in not necessarily tied to exposure to the firm's brand. Rather, continuous social network usage is associated with an increase in shopping activities in general. Building on this insight, managers could target specific groups of individuals on social networks who display more positive associations between cumulative social network usage and shopping activities. They could also advertise specific product categories (e.g., chocolate) on social networks to these individuals.[26]

Using our model, managers can identify this group of consumers by assessing heterogeneity across consumers in their individual-level estimates to determine any correlation between social network engagement and purchase activity.[27] Figure 6 plots the β_i estimates for SN-ENGAGEMENT from Model 1 (i.e., the model with only social networking sites) and suggests that, for our set of consumers, there is significant variation in the correlation between social network engagement and shopping activities (in this instance, the decision to purchase). Thus, firms (e.g., Facebook) could apply our model to their set of users to identify consumers

[26] This is akin to product placement at checkout, when consumers are already in a shopping mindset (Rook 1987). That is, firms could increase sales by strategically marketing products in categories that tend to be shared on social networks to consumers who have this positive relationship.

[27] An alternative way is to identify characteristics of individuals whose shopping sessions are more likely to mediate the relationship between social network engagement and sales. We ran an individual-level mediation analysis for each user in our sample. For 52% of the individuals, shopping sessions partially mediated the relationship between social network engagement and purchase incidence. We then regressed the demographic variables for each individual on a binary variable indicating whether mediation was found for each individual using a logistic model. The results (available upon request) suggest that younger consumers, consumers from larger households, and lower-income consumers are more likely to have shopping session mediation, meaning that social network usage is more likely to be associated with shopping activities. Thus, a firm could also target social network users who are younger, part of a larger household, or have lower income.

who have larger β_i estimates on the SN-ENGAGEMENT variable and target these individuals on social networks.

[Insert Figure 6 about here]

Despite the growing interest of both academics and managers in online social networks, little is known about the individual-level relationships of immediate and cumulative usage in social networks with online shopping activities. In this paper, we study the relationship between consumers' immediate usage of and cumulative engagement with social networks and their shopping activity. Our individual-level social network and shopping data allow us to directly investigate how different aspects of social networks may have varying correlation with the decision to purchase, the number of purchases made, and the number of websites purchased from.

In summary, our results provide support for both positive and negative relationships across three key indicators of online shopping activities. First, we find that more immediate time spent on social networking websites is associated with a lower purchase probability, a lower number of purchases, and a lower number of websites purchased from. Drawing on Becker's (1965) theory of time substitution, we attribute this to immediate time spent on social networks cannibalizing time spent on online shopping. Second, we find that engagement in social networks has a positive relationship with all three shopping activities. We hypothesize that this positive association is driven by exposure to new consumption-related information aiding the shopping search process. Social networks facilitate exposure to a wide variety of content from friends or other network ties, and consumers may encounter new information as they discover the products that their social network buys or discusses.

Our research is not without limitations. First, while we have individual-level data on web-browsing and purchase activity, we can only infer that social network usage is associated with shopping activities. While our experiment in Web Appendix C provides support for causal effects, we do not make any causal claims based on our empirical analysis. In addition, because we do not observe the content that consumers were exposed to on these websites (largely due to privacy reasons), we are unable to determine whether exposure to product-related information definitely occurred. Instead, our data and empirical analysis allow us to identify effects that are consistent with these theorized mechanisms. Future research could investigate whether social discovery or priming is taking place. Second, in a similar sense, we capture the immediate and cumulative usage of social networks with two proxies: previous session/duration spent on the website and cumulative daily website visitation. While we expect that these variables reflect actual usage and engagement, there is the possibility that individuals are not actively engaged with the website but simply have the page open in the background. Third, our purchase data only capture online shopping activity and do not include offline purchases. While we expect to find similar positive and negative relationships between social network usage and offline shopping, we are unable to explicitly test for this. Finally, our results for the estimates of search, social network, and shopping referrals are subject to the “last-click bias.” A richer data set may provide more insight into the various attributions of purchase, whether driven by social network usage or search that occurred over a longer period of time.

References

Becker, Gary (1965), "A Theory of the Allocation of Time," *The Economic Journal*, 75 (299), 493–517.

Bercovici, Jeff (2013), "Study Finds Pins on Pinterest Drive Sales and Have Legs," (accessed February 19, 2015), http://www.forbes.com/sites/jeffbercovici/2013/11/15/study -finds-pins-on-pinterest-drive-sales-and-have-legs/.

Berger, Jonah, and Eric Schwartz (2011), "What Drives Immediate and Ongoing Word of Mouth?" *Journal of Marketing Research*, 48 (September), 869–80.

Branco, Fernando, Monic Sun, and J. Miguel Villas-Boas (2012), "Optimal Search for Product Information," *Management Science*, 58 (11), 2037–56.

Brooks, Stephen P., and Andrew Gelman (1998), "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, 7 (4), 434–55.

Chen, Yubo, Qi Wang, and Jinhong Xie (2011), "Online Social Interactions: A Natural Experiment on Word of Mouth Versus Observational Learning," *Journal of Marketing Research*, 48 (April), 238–54.

Chevalier, Judith, and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (August), 345–54.

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets," *Marketing Science*, 29 (5), 944–57.

Danaher, Peter J., and Michael S. Smith (2011), "Modeling Multivariate Distributions Using Copulas: Applications in Marketing," *Marketing Science*, 30 (1), 4–21.

De, Prabuddha, Yu (Jeffrey) Hu, and Mohammad S. Rahman (2010), "Technology Usage and Online Sales: An Empirical Study," *Management Science*, 56 (11), 1930–45.

Del Rey, Jason (2013), "Social Media's Cold, Hard Reality: It Still Doesn't Drive E-Commerce Sales, IBM Says," (accessed December 2, 2013), http://allthingsd.com/20131129/social-medias-cold-hard-reality-it-still-doesnt-drive-e-commerce-sales-ibm-says.

Denuit, Michel and Philippe Lambert (2005), "Constraints on Concordance Measures in Bivariate Discrete Data," *Journal of Multivariate Analysis*, 93 (1), 40–57.

Dorfman, Robert, and Peter O. Steiner (1954), "Optimal Advertising and Optimal Quality," *The American Economic Review*, 44 (5), 826–36.

East, R., K. Hammond, and W. Lomax (2008), "Measuring the Impact of Positive and Negative Word-of-Mouth on Brand Purchase Probability," *International Journal of Research in Marketing*, 25 (3), 215–24.

Ebbes, Peter, Michel Wedel, Ulf Böckenholt, and Ton Steerneman (2005), "Solving and Testing for Regressor-Error (in)Dependence When No Instrumental Variables Are Available: With New Evidence for the Effect of Education on Income," *Quantitative Marketing and Economics*, 3 (4), 365–92.

Edwards, Yancy D., and Greg M. Allenby (2003), "Multivariate Analysis of Multiple Response Data," *Journal of Marketing Research*, 40 (August), 321–34.

Facebook (2016), "Facebook Reports First Quarter 2016 Results and Announces Proposal for New Class of Stock," (accessed June 1, 2016), https://investor.fb.com/investor-news/press-release-details/2016/Facebook-Reports-First-Quarter-2016-Results-and-Announces-Proposal-for-New-Class-of-Stock/default.aspx.

Fischer, Marc, Sönke Albers, Nils Wagner, and Monika Frie (2011), "Practice Prize Winner—Dynamic Marketing Budget Allocation Across Countries, Products, and Marketing Activities," *Marketing Science*, 30 (4), 568–85.

Gara, Tom (2012), "Social Media Has a Black Friday #Fail," (accessed December 2, 2013), http://blogs.wsj.com/corporate-intelligence/2012/11/26/social-media-has-a-black-friday-fail/.

Godes, David, and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545–60.

Goldenberg, Jacob, Gal Oestreicher-Singer, and Shachar Reichman (2012), "The Quest for Content: How User-Generated Links Can Facilitate Online Exploration," *Journal of Marketing Research*, 49 (August), 452–68.

Gonzales, Angela, and Jeff Hancock (2011), "Mirror, Mirror on My Facebook Wall: Effects of Exposure to Facebook on Self-Esteem," *Cyberpsychology, Behavior and Social Networking*, 14 (1-2), 79–83.

Gronau, Reuben (1973), "The Intrafamily Allocation of Time: The Value of the Housewives' Time," *The American Economic Review*, 63(4), 634–51.

Gronau, Reuben (1977), "Leisure, Home Production, and Work: The Theory of the Allocation of Time Revisited," *Journal of Political Economy*, 85 (6), 1099–1123.

GWI globalwebindex (2016), “Social Media Captures 30% of Online Time” (accessed February 1, 2017), https://www.globalwebindex.net/blog/social-media-captures-30-of-online-time.

Hagerty, Michael R., and David A. Aaker (1984), “A Normative Model of Consumer Information Processing,” *Marketing Science*, 3 (3), 227–46.

Hartmann, Wesley R., Puneet Manchanda, Harikesh Nair, Matthew Bothner, Peter Dodds, David Godes, Kartik Hosanagar, and Catherine Tucker (2008), “Modeling Social Interactions: Identification, Empirical Methods and Policy Implications,” *Marketing Letters*, 19(3-4), 287–304.

Hauser, John, Glen Urban, and Bruce Weinberg (1993), “How Consumers Allocate Their Time When Searching for Information,” *Journal of Marketing Research*, 30 (November), 452–66.

Hauser, John, and Birger Wernerfelt (1990), “An Evaluation Cost Model of Consideration Sets,” *Journal of Consumer Research*, 16 (4), 393–408.

Hayes, Andrew (2013), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York: Guilford Press.

Ho-Dac, Nga N., Stephen J. Carson, and William L. Moore (2013), The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *Journal of Marketing*, 77 (November), 37–53.

Katona, Zsolt, Peter Pal Zubcsek, and Miklos Sarvary (2011), “Network Effects and Personal Influences: The Diffusion of an Online Social Network,” *Journal of Marketing Research*, 48 (June), 425–43.

Klein, Lisa R., and Gary T. Ford (2003), “Consumer Search for Information in the Digital Age: An Empirical Study of Prepurchase Search for Automobiles,” *Journal of Interactive Marketing*, 17 (3), 29–49.

Kumar, V., Vikram Bhaskaran, Rohan Mirchandani, and Milap Shah (2013), “Creating a Measurable Social Media Marketing Strategy: Increasing the Value and ROI of Intangibles and Tangibles for Hokey Pokey,” *Marketing Science*, 32 (2), 194–212.

Kumar, V., J. Andrew Petersen, and Robert P. Leone (2010), “Driving Profitability by Encouraging Customer Referrals: Who, When, and How,” *Journal of Marketing*, 74 (September), 1–17.

Kumar, V., and Bharath Rajan (2012), “Social Coupons as a Marketing Strategy: A Multifaceted Perspective,” *Journal of the Academy of Marketing Science*, 40 (1), 120–36.

Kumar, V., Xi (Alan) Zhang, and Anita Luo (2014), “Modeling Customer Opt-In and Opt-Out in a Permission-Based Marketing Context,” *Journal of Marketing Research*, 51 (October), 403–19.

Lambert, Diane (1992), “Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing,” *Technometrics*, 34 (1), 1–14.

Lamberton, Cait, and Andrew T. Stephen (2016), “A Thematic Exploration of Digital, Social Media, and Mobile Marketing Research’s Evolution from 2000 to 2015 and an Agenda for Future Research,” *Journal of Marketing*, 80 (November), 146–72.

Lewis, Randall A., Justin M. Rao, and David H. Reiley (2011), “Here, There, and Everywhere: Correlated Online Behaviors Can Lead to Overestimates of the Effects of Advertising,” In *Proceedings of the 20th International Conference on World Wide Web*. New York: Association for Computing Machinery, 157–66.

Liaukonyte, Jura, Thales Teixeira, and Kenneth Wilbur (2015), “Television Advertising and Online Shopping,” *Marketing Science*, 34 (3), 311–30.

Liu, Yong (2006), “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” *Journal of Marketing*, 70 (July), 74–89.

Mayzlin, Dina, and Hema Yoganarasimhan (2012), “Link to Success: How Blogs Build an Audience by Promoting Rivals,” *Management Science*, 58 (9), 1651–68.

McCulloch, Robert, and Peter E Rossi (1994), “An Exact Likelihood Analysis of the Multinomial Probit Model,” *Journal of Econometrics*, 64 (1–2), 207–40.

Meyer, Robert (1982), “A Descriptive Model of Consumer Information on Search Behavior,” *Marketing Science*, 1(1), 93–121.

Meyer, Robert, and Arvind Sathi (1985), “A Multiattribute Model of Consumer Choice During Product Learning,” *Marketing Science*, 4 (6), 41–61.

Moe, Wendy W., and Peter S. Fader (2004), “Dynamic Conversion Behavior at E-Commerce Sites,” *Management Science*, 50 (3), 326–35.

Moe, Wendy W., and Michael Trusov (2011), “The Value of Social Dynamics in Online Product Ratings Forums,” *Journal of Marketing Research*, 48 (June), 444–56.

Moe, Wendy W. and David A. Schweidel (2012), “Online Product Opinion: Incidence, Evaluation and Evolution,” *Marketing Science*, 31 (3), 372–386.

Morey, Richard C., and John M. McCann (1983), "Estimating the Confidence Interval for the Optimal Marketing Mix: An Application to Lead Generation," *Marketing Science*, 2 (2), 193–202.

Nair, Harikesh S., Puneet Manchanda, and Tulikaa Bhatia (2010), "Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders," *Journal of Marketing Research*, 47 (October), 883–95.

Oestreicher-Singer, Gal, and Arun Sundararajan (2012), "The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets," *Management Science*, 58 (11), 1963–81.

Ratchford, Brian T., and Narasimhan Srinivasan (1993), "An Empirical Investigation of Returns to Search." *Marketing Science*, 12 (1), 73–87.

Ratchford, Brian T, Myung-Soo Lee, and Debabrata Talukdar (2003). "The Impact of the Internet on Information Search for Automobiles," *Journal of Marketing Research*, 40 (May), 193–209.

Rook, Dennis (1987), "The Buying Impulse," *Journal of Consumer Research*, 14 (2), 189–97.

Rutz, Oliver J, Randolph E. Bucklin, and Garrett P. Sonnier (2012), "A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising," *Journal of Marketing Research*, 49 (June), 306–19.

Rutz, Oliver J., and Michael Trusov (2011), "Zooming In on Paid Search Ads: A Consumer-Level Model Calibrated on Aggregated Data," *Marketing Science*, 30 (5), 789–800.

Sklar, A. (1959), "Fonctions de Repartitions à n Dimensions et Leurs Marges," Public Institute of Statistics of the University of Paris 8, 229–31.

Stephen, Andrew T. (2016), "The Role of Digital and Social Media Marketing in Consumer Behavior," *Current Opinion in Psychology*, 10 (August), 17–21.

Stephen, Andrew T., and Jeff Galak (2012), "The Effects of Traditional and Social Earned Media on Sales: A Study of a Microlending Marketplace," *Journal of Marketing Research*, 49 (October), 624–39.

Stilley, Karen, Jeffrey Inman, and Kirk Wakefield (2010), "Planning to Make Unplanned Purchases? The Role of Discretionary Budgets in In-Store Decision Making," *Journal of Consumer Research*, 37 (2), 264–78.

Tanner, Martin A., and Wing Hung Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82 (398), 528–40.

Toubia, Olivier, and Andrew T. Stephen (2013), "Intrinsic Versus Image-Related Motivations in Social Media: Why Do People Contribute Content to Twitter?" *Marketing Science*, 32 (3), 368–92.

Trusov, Michael, Anand V. Bodapati, and Randolph E. Bucklin (2010), "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research*, 47 (August), 643–58.

Trusov, Michael, Randolph E Bucklin, and Koen Pauwels (2009), "Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing*, 73 (September), 90–102.

U.S. Census Bureau (2010), "People and Households: Data By Subject," (accessed February 1, 2016), https://www.census.gov/population/projections/data/national/2008.html.

Venkatesan, Rajkumar, and V. Kumar (2004), "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy," *Journal of Marketing*, 68 (October), 106–25.

Venkatesan, Rajkumar, V. Kumar, and Timothy Bohling (2007), "Optimal Customer Relationship Management Using Bayesian Decision Theory: An Application for Customer Selection," *Journal of Marketing Research*, 44 (November), 579–94.

Verhoef, Peter, Rajkumar Venkatesan, Leigh McAlister, Edward Malthouse, Manfred Krafft, and Shankar Ganesan (2010), "CRM in Data-Rich Multichannel Retailing Environments: A Review and Future Research Directions," *Journal of Interactive Marketing*, 24 (2), 121–37.

Wilcox, Keith, and Andrew T. Stephen (2013), "Are Close Friends the Enemy? Online Social Networks, Self-Esteem, and Self-Control," *Journal of Consumer Research*, 40 (1), 90–103.

Woolf, Max (2014), "A Statistical Analysis of 1.2 Million Amazon Reviews," (accessed March 21, 2016), http://minimaxir.com/2014/06/reviewing-reviews/.

Yang, Sha, Yuxin Chen, and Greg M. Allenby (2003), "Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics*, 1 (3), 251–75.

Zhang, Jie, Michel Wedel, and Rik Pieters (2009), "Sales Effects of Attention to Feature Advertisements: A Bayesian Mediation Analysis," *Journal of Marketing Research*, 46 (October), 669–81.

TABLE 1
Summary Statistics of Shopping Activities

Variable	Mean	Std. Dev.	Min	Max
PURCHASE	0.008	0.090	0	1
NUMPURCHASE	0.02	0.357	0	49
NUMRETAILERS	0.009	0.098	0	4

* Reported at the user-session level

TABLE 2
Statistics on Referrals to Shopping Sites

Referral Type	Number of Sessions	Percent of Total
Referrals from Social Networks	37,834	2%
Referrals from Search	109,276	6%
Referrals from Other Shopping	168,845	9.34%
No/Other Referrals	1,491,060	82.52%
Total Shopping Sessions	1,807,015	100.00%

TABLE 3
Demographic Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Age	47.73	14.00	17	99
Gender*	0.43	0.50	0	1
Houshold size	2.64	1.23	1	5
Income**	1.55	0.90	0	3
Working members	1.46	0.91	0	5
Children***	0.37	0.48	0	1

*Gender variable is 1/0 for male/female

**Income is discrete (0-3) for ($0-25k, $25k-50k, $50k-100k, $100k+)

***Children variable is 1/0 for with/without children

TABLE 4
Description of Variables

Category of Variables	Variables	Description
SOCIALNETWORK	SN-SESSION	Indicator for visit to social network websites during the previous session
	SN-DURATION	Total duration during visits to social network websites during the previous session
	SN-ENGAGEMENT	Variable indicating the number of days the user has visited a social network site
	SN-ACTIVE	Dummy variable indicating whether user has ever visited a social network site
	SN-FIRSTMONTH	Dummy variable indicating whether user has visited the social network site during the first month
INTERNET	SHOP-SESSION	Indicator for visit to shopping websites during the previous session
	SHOP-DURATION	Total duration during visits to shopping websites during the previous session
	SHOP-CUMU:LATIVE	Variable indicating the number of days the user has visited a shopping site
	NET-SESSION	Indicator for visit to any website (exclusive of shopping or social networks) during the previous session
	NET-DURATION	Total duration during visits to all websites (exclusive of shopping or social networks during the previous session
	NET-CUMULATIVE	Variable indicating the number of days the user has visited any non-social media, non-shopping site
REFERRAL	SN-REFERRAL	Dummy variable indicating whether the referral website was from social networks
	SEARCH-REFERRAL	Dummy variable indicating whether the referral website was from search
	SHOP-REFERRAL	Dummy variable indicating whether the referral website was from other shopping sites
DEMOGRAPHICS	GENDER	Gender of user
	HOUSEHOLD-SIZE	Number of people in user's household
	AGE	Age of user
	INCOME	Income of user
	WORKING-MEMBERS	Number of working members in user's household
	CHILDREN	Dummy indicating whether user has children

TABLE 5
Main Results

Model	Y = Purchase (Probit Model)	Y = # of Purchases (ZIP Model)	Y = # of Retailers with purchase (ZIP Model)
Social Network Variables			
SN-SESSION	**-0.385***	**-1.2704***	**-1.8318***
	[-0.2644 , -0.5174]	[-1.1119 , -1.4051]	[-1.5796 , -1.9349]
SN-DURATION	**-0.0516**	**-0.1893***	**-0.2108***
	[0.0791 , -0.0908]	[-0.1632 , -0.2143]	[-0.1933 , -0.2574]
SN-ENGAGEMENT	**0.0846***	**1.2768***	**1.3623***
	[0.1329 , 0.0446]	[1.3108 , 1.2356]	[1.3863 , 1.3362]
SN-REFERRAL	0.007	N/A	N/A
	[0.0162 , -0.0098]	N/A	N/A
SN-ACTIVE	**0.0339***	**0.0463***	**0.0973***
	[0.0443 , 0.0266]	[0.0899 , 0.0025]	[0.158 , 0.0338]
SN-FIRSTMONTH	**0.0303***	0.0116	0.0021
	[0.0391 , 0.0228]	[0.0196 , -0.0068]	[0.0167 , -0.0125]
Other Variables			
LAG-Y	**0.0354***	**0.0641***	**0.1903***
	[0.0467 , 0.0277]	[0.0982 , 0.0316]	[0.2312 , 0.1554]
NET-SESSION	**-0.0132***	**-0.2324***	**-0.3704***
	[-0.0035 , -0.0204]	[-0.1657 , -0.3102]	[-0.2831 , -0.4229]
NET-DURATION	0.0016	**0.0492***	**0.051***
	[0.0111 , -0.0074]	[0.0694 , 0.0257]	[0.0735 , 0.0288]
SHOP-SESSION	0.0118*	0.4887*	0.714*
	[0.0175 , 0.0054]	[0.5535 , 0.4165]	[0.82 , 0.6115]
SHOP-DURATION	**0.0078***	**0.1692***	**0.1387***
	[0.0116 , 0.0039]	[0.1973 , 0.1444]	[0.1629 , 0.1102]
NET CUMULATIVE SESSION	0.0006	**-0.8839***	**-0.8839***
	[0.0378 , -0.0358]	[-0.7629 , -0.9691]	[-0.7789 , -0.9719]
SHOP CUMULATIVE SESSION	0.0006	**0.1284***	**0.1638***
	[0.0161 , -0.014]	[0.1803 , 0.0629]	[0.2122 , 0.0865]
SEARCH-REFERRAL	**0.0075***	N/A	N/A
	[0.0153 , 0.002]	N/A	N/A
Demographic Variables			
AGE	**0.001***	**0.0719***	**0.0157***
	[0.0013 , 0.0007]	[0.0845 , 0.0595]	[0.0258 , 0.009]
GENDER	-0.0059	**-0.0613***	**-0.0431***
	[0.002 , -0.013]	[-0.0553 , -0.0683]	[-0.0306 , -0.0649]
HOUSEHOLD-SIZE	**0.0035***	**0.0262***	**-0.0627***
	[0.0085 , 0.0005]	[0.0314 , 0.0195]	[-0.0527 , -0.0735]
INCOME	0.0261	-0.0047	**0.0762***
	[0.0381 , -0.0124]	[0.0004 , -0.0107]	[0.0837 , 0.071]
WORKING-MEMBERS	**0.0111***	**-0.0082**	**-0.1082***
	[0.0161 , 0.0066]	[0.0004 , -0.014]	[-0.0999 , -0.1167]
CHILDREN	**0.0118***	**0.0893***	**-0.0478***
	[0.021 , 0.0048]	[0.097 , 0.0855]	[-0.0329 , -0.0575]
INCOME^2	**0.0101***	**0.0435***	**0.0238***
	[0.0133 , 0.0054]	[0.0517 , 0.0279]	[0.0291 , 0.0016]
N	2,134,689	2,134,689	2,134,689
LL	-1,242,133	-248,315	-202,977
DIC	1,740,305	-20,363	7,726
Hit Rate (holdout)	0.562	0.929	0.987
MSE (holdout)	0.438	1963.448	161.488
MAD (holdout)	0.430	0.185	0.014

Notes. The 95% Bayesian Credible Interval is reported in brackets

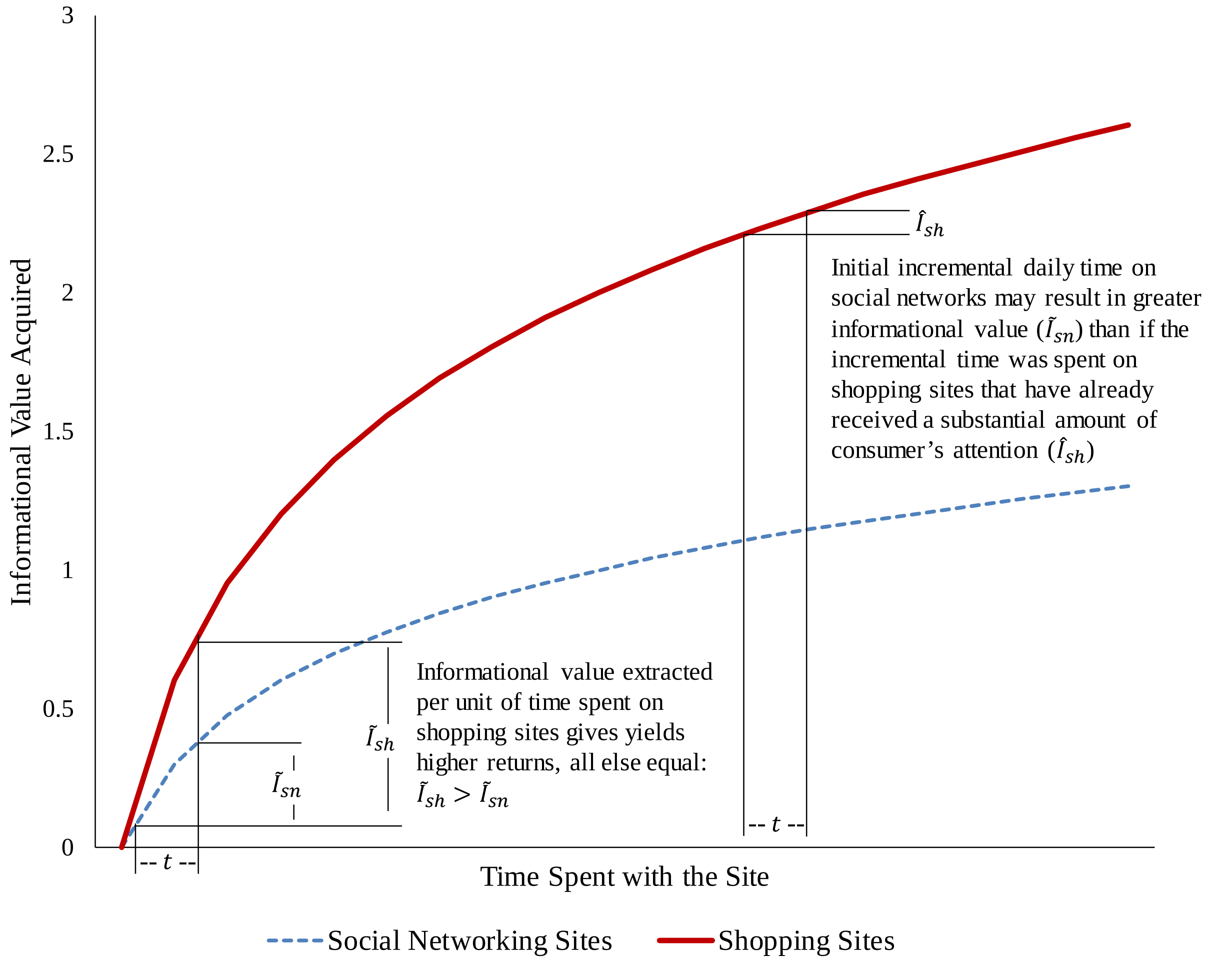
* Indicates that 0 is not contained 95% Bayesian Credible Interval

TABLE 6
Category-Specific Average Partial Effects for Social Network Variables

Category Description	*SN-ENGAGEMENT*	*SN-SESSION*
Audio CD	0.345	**-0.568**
Auto Parts	0.915	**-1.592**
Bath and Beauty	0.051	**-0.198**
Books	**0.126**	**-0.129**
Children	**0.700**	**-1.681**
Chocolate / Candy	**0.370**	**-0.299**
Clothing	**5.079**	**-11.046**
Computer	0.151	**-0.165**
Computer Software	0.021	**-0.297**
Electronics	0.127	**-0.178**
Flowers	**1.812**	**-1.811**
Gift Cards	0.018	**-0.127**
Jewelry	**0.428**	**-1.891**
Magazines	**0.072**	**-0.588**
Photo	**0.181**	-0.543
Shoes	**0.097**	**-0.259**
Sports / Outdoors	**0.355**	**-0.499**
Theater / Entertainment	**0.276**	**-0.951**
Tools / Hardware	0.132	**-0.573**
Toys and Games	**0.081**	**-0.240**
Video	**0.639**	**-0.637**

*Notes: Each row consists of a separate model, with the dependent variable specified as whether a purchase was made (or not) in that particular category. We report the partial effects (scaled by 1000 for presentation purposes) for the social network session and engagement variables. Bolded values indicate that zero is not contained in the 95% bayesian credible interval. We test for the differences between the most affected product categories (i.e. clothing, jewelry, and children's products) and least affected categories (i.e. computers, books, and gift cards) to be statistically significant.

FIGURE 1
Informational Value Acquired from Shopping and Social Network Sites



Notes: All else equal, the information extracted per unit of time spent on shopping sites yields higher returns than time spent on social networks: $\tilde{I}_{sh} > \tilde{I}_{sn}$ (left side of graph). Over the long run, the incremental daily time t spent on social networking sites could result in greater informational value ($\tilde{I}_{sn}$) than if that time t was spent on the shopping sites that have already received a substantial amount of consumer's attention (the right side of the graph - $\hat{I}_{sh}$).

FIGURE 2
Internet Session Versus Website Session

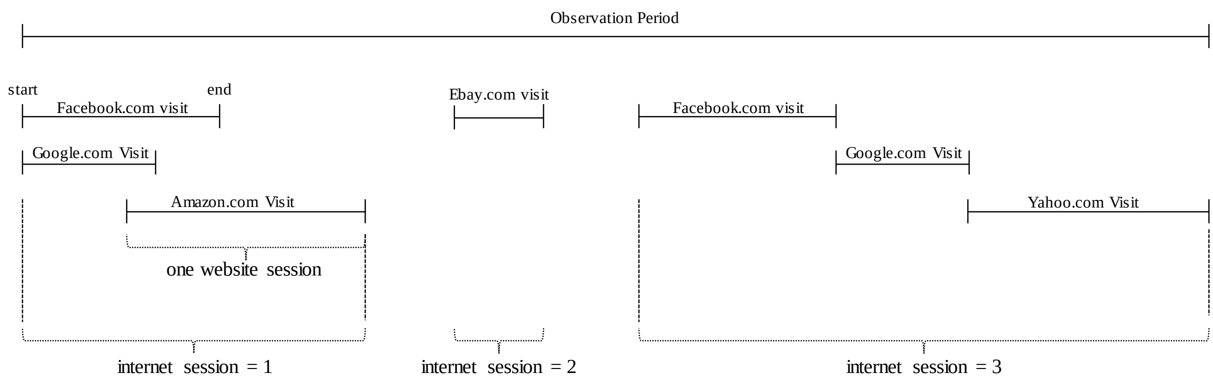


FIGURE 3
Distribution of Purchase Activity Per User

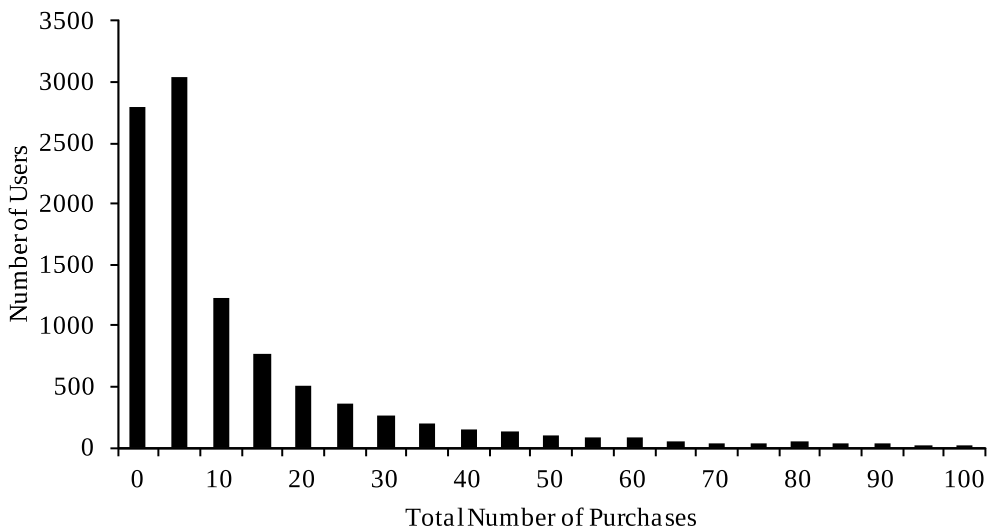


FIGURE 4
Daily Purchase Activity (Aggregate)

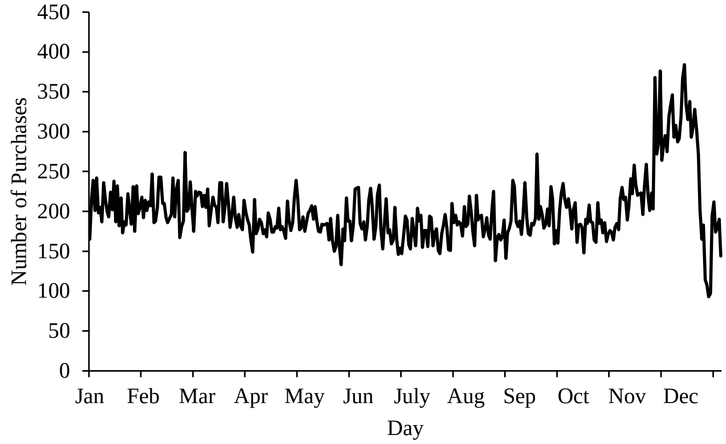


FIGURE 5
Day of First Observed Social Network Usage

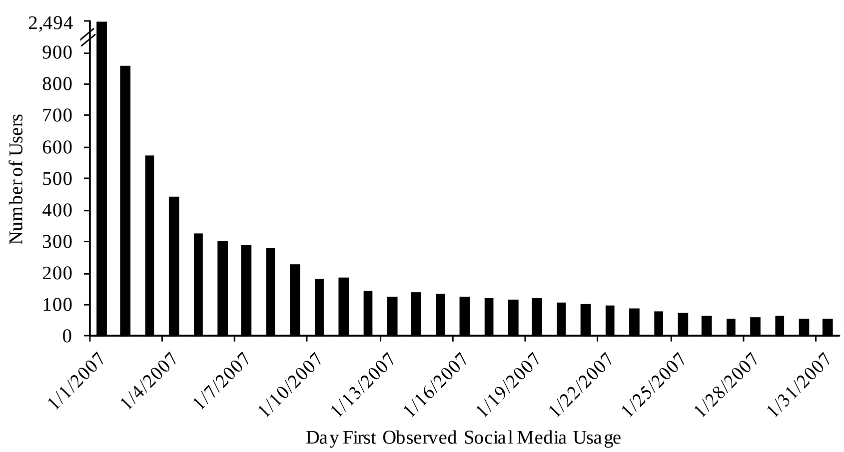


FIGURE 6
Distribution of Individual-Level Estimates for Engagement