Social Dollars: The Economic Impact of Customer Participation in a Firm-sponsored Online Customer Community*

Puneet Manchanda†

Grant Packard‡

Adithya Pattabhiramaiah§

April 2010

This version: Feb 2015

*The authors would like to thank the Editor-in-Chief, the AE, two anonymous reviewers, Utpal Dholakia, Pete Fader, Ron Goettler, Anja Lambrecht, S. Sriram, Melanie Zaglia and seminar participants at the University of Maryland, London Business School, University of Connecticut, CKGSB, Tsinghua University, LMU Munich, HEC Paris, University of California Davis, University of North Carolina, Universidad Carlos III, University of Colorado, University of Rochester and participants at the 2011 ART Forum for valuable feedback and an anonymous firm for providing the data. The standard disclaimer applies. All correspondence by regular mail should be sent to the first author at Ross School of Business, University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109, USA.

†Isadore and Leon Winkelman Professor and Professor of Marketing, University of Michigan, pmanchan@umich.edu

‡Assistant Professor of Marketing, Wilfrid Laurier University, gpackard@wlu.ca

§Assistant Professor of Marketing, Georgia Institute of Technology, adithyap@gatech.edu

Abstract

Many firms operate customer communities online. This is motivated by the belief that customers who join the community become more engaged with the firm and/or its products, and as a result, increase their economic activity with the firm. We describe this potential economic benefit as "social dollars." This paper contributes evidence for the existence and source of social dollars using data from a multi-channel entertainment products retailer that launched a customer community online. We find a significant increase in customer expenditures attributable to customers joining the firm's community. While self-selection is a concern with field data, we are able to rule out multiple alternative explanations. Social dollars persist over the time period observed and arose primarily in the online channel. To assess the source of the social dollar, we hypothesize and test whether it is moderated by participation behaviors conceptually linked to common attributes of customer communities. Our results reveal that posters (vs. lurkers) of community content and those with more (vs. fewer) social ties in the community generated more (fewer) social dollars. We found a null effect for our measure of the informational advantage expected to accrue to products that differentially benefit from content posted by like-minded community members. This overall pattern of results suggests a stronger social than informational source of economic benefits for firm operators of customer communities. Several implications for firms considering investments in and/or managing online customer communities are discussed.

Keywords: Online Customer Communities, Online Customer Behavior, Social Networks, User-Generated Content, Retailing, Field Data.

1 Introduction

People have long communed as consumers. From Apple acolytes to Java junkies, Tupperware partiers to European car clubbers, like-minded consumers have engaged with one another in customer communities – networks of individuals who engage in social interactions regarding their shared enthusiasm for and/or use of specific brandmarks, products or consumption activities (Algesheimer et al., 2005; Bagozzi and Dholakia, 2002; Porter and Donthu, 2008; Rheingold, 1993).

The last decade has witnessed an explosion of marketer interest in these consumer-powered social engines of brand engagement, trust, and loyalty (Porter and Donthu, 2008; Williams and Cothrel, 2000; Wirtz et al., 2013). Consumer adoption of the Internet, social media and mobile technologies has been a central driver of this enthusiasm. Over 70% of Americans (Forrester, 2012) and an estimated one billion people across the planet (Eddy, 2012) are using social media, spending over one third of their waking hours in online environments that allow them to present a public or private profile, establish social ties with like-minded others, and exchange information of mutual interest socially (Boyd and Ellison, 2007).

The movement of like-minded consumers into online communities represents a major business opportunity for firms, whether these communities are embedded in independent websites, firm-operated websites, or third-party social media platforms (Forrester, 2012). A recent IBM survey of 1,709 CEOs from over 60 countries reveals that they believe online customer communities are the second most important means by which they will engage customers in the future (after face-to-face interactions and well ahead of traditional media). Nearly 60% of these executives plan to invest (or invest more) in online communities over the next few years, leading to an expected total annual marketing expenditure in online customer communities of nearly US $5 billion by 2016 (Paul, 2012; Schniederjans et al., 2012; Forrester, 2012).

While firms and their top managers express high levels of confidence in the marketing efficacy of these communities, there is surprisingly little evidence documenting the economic benefits of online customer communities (whether firm-sponsored or hosted by third parties). In fact, doubts have been expressed in industry regarding the positive return on investment of customer communities hosted on third-party social media platforms such as Facebook (Vranica and Raice, 2012). Firms have also articulated other concerns such as loss of control and the potential of consumer backlash when customer communities are hosted by a third-party (Thompson et al., 2011). This has led many companies to invest in building their own online customer communities. We estimate that, depending on the definition of the attributes of a firm-hosted online customer community,[1] between 25 and 50 of the top 100 global brands (Interbrand, 2011) host their own such community. In additon, Forrester (2012) reports that 18% of all businesses around the globe are making online customer community investments independent of third-party platforms.

It is not clear which attributes of customer communities firms should prioritize to maximize their economic outcome. Is providing customers with a rich source of user-generated information from like-minded individuals the key to community success? Alternatively, is it the ability for customers to voice their opinions by posting this content that is more critical? Providing a mechanism for customers to establish formal "friend" ties may also be an essential community attribute. While each of these have been described as key identifiers of online customer communities (Wirtz et al., 2013) and their importance theorized in the literature (e.g., Algesheimer et al., 2010; Bagozzi and Dholakia, 2002; Dholakia et al., 2004; McWilliam, 2000; Porter and Donthu, 2008; Schau et al., 2009), we are unaware of research that directly examines the relationship between social and informational attributes of customer communities and their economic consequences for the firm.

In this paper, we focus on quantifying the incremental economic benefit of such communites and the source(s) of this benefit in relation to specific community participation behaviors using actual transaction and participation data. We label this incremental expenditure "social dollars." We do this with the help of a novel dataset from a multi-channel (online and offline) retailer that decided to launch an online customer community.

Our data and research approach adds to the literature on online customer communities on multiple dimensions. First, and perhaps most important, we use actual behavioral data to investigate the economic impact to the firm of offering such a community. Second, the availability of consumer panel data before and after the formation of the community allows us to assess the potential implication of self-selection common in analysis of field data (Shriver et al., 2013). While we cannot

[1] We examined three attributes described as identifiers of online or social media-based customer communities (Wirtz et al., 2013) to determine if a firm sponsors such a community. These were: (1) the ability for consumers to create and maintain a personal profile page in the firm's brand or product-centered website, (2) the ability for consumers to create and maintain friend ties in this setting, and (3) the ability to post and consume user-generated content at the website. The most conservative definition is based on the presence of all three attributes while the least conservative definition is based on the presence of at least one attribute.

rule out selection on unobservables with certainty, several falsification tests and robustness checks provide compelling support that self-selection is unlikely to explain our results. Third, the long time series of our data allows us to investigate whether the change in purchase behavior that results from joining the community is a short-term effect driven by the novelty of the event (the formation of the community) or a more persistent phenomenon. Fourth, given the multi-channel nature of our data, we are able to test whether the formation of the community affects behavior differentially across channels. Finally, we use the observed participation behaviors and interactions among community members to isolate mechanisms that may underlie the economic effect of online customer community participation, allowing us to provide theoretically-grounded empirical evidence for how social dollars come about.

Our results find that, in the firm setting we observe, social dollars represent a double-digit percentage increase in customer expenditures. While it is possible that our finding is idiosyncratic, this increase represents an economically significant overall return for the firm as it more than covers the fixed cost of setting up the community as well as the variable cost of operating it. We subject our base estimate of the social dollars to multiple robustness tests and demonstrate that it is indeed robust. For example, we find that the social dollar estimate is robust to selection on observables and unobservables. We find that social dollars persist over time. Furthermore, we do not find evidence of channel cannibalization. We also rule out a series of alternative explanations for the existence of these social dollars.

Finally, the economic effect of the community should be linked not only to joining, but to specific behaviors conceptually linked to the consumer benefits of customer community participation. We examine participation behaviors that prior theoretical and survey-based research suggest are central to the definition and benefits of customer communities (e.g., Algesheimer et al., 2005; Algesheimer et al., 2010; Jang et al., 2008; Dholakia and Vianello, 2009). We predict and demonstrate that consumers who merely lurk (vs. post content) in the community will recognize diminished social benefits from their lack of participation, diminishing their economic engagement with the firm. Similarly, consumers who heighten their social connection to the community by establishing friend ties obtain additional social benefits, leading to heightened economic engagement with the firm. These results demonstrate that customer utilization of specific social attributes of online customer communities are indeed linked to the financial outcomes to which their firm hosts aspire– social dollars.

The rest of the paper is organized as follows. In § 2, we discuss the conceptual and empirical literature in which our research is grounded. We then describe the research setting and data in § 3. Sections § 4 and § 5 describe our modeling and analysis strategy. Finally, we discuss the managerial implications of our findings in § 6 and then conclude in § 7.

2 Conceptual Background

2.1 Benefits of *Mere Membership* in an Online Customer Community

The conceptual definitions of the attributes and consequences of customer communities are rich and varied. The most cited of these is Muniz and O'Guinn's (2001), who described a *brand community* as a, "specialized, non-geographically bound community, based on a structured set of social relationships among admirers of a branded good or service" and offered three specific markers of these communities: shared consciousness, rituals and traditions, and a sense of moral responsibility. Researchers subsequently proposed an expanded conception to describe *marketplace communities* or *consumption communities* as relationships and behaviors not only of customers with brands, but among customers themselves, between the customer and firm, and between the customer and the product-in-use (McAlexander et al., 2002). Research examining these relationships in technology-mediated settings has sometimes described them as *online* or *virtual communities* (Balasubramanian and Mahajan, 2001; Dholakia et al., 2004; Kozinets, 2002; Porter and Donthu, 2008). Overall, *customer community* has come to be the most common term used to describe a group or network of individuals who engage in social interactions regarding their shared enthusiasm for and/or use of specific brandmarks, products, retail environments, or consumption activities whether in online or offline settings (Algesheimer et al., 2010; Bagozzi and Dholakia, 2002; Porter and Donthu, 2008; Rheingold, 1993). Empirical investigations of customer communities include those that operate independently of the brand or firm (e.g., Dholakia et al., 2004) and those that are organized and controlled by a commercial firm i.e., "firm-sponsored" (Kannan et al., 2000; Gruner et al., 2014; Porter and Donthu, 2008).

Firms that operate customer communities are said to have the opportunity to increase customer engagement and loyalty among community members (Fournier and Lee, 2009; Porter and Donthu, 2008; Williams and Cothrel, 2000). These benefits are said to accrue from an increased sense of brand community identification, the perceived belonging that arises through membership in a customer community (Algesheimer et al., 2005; Bagozzi and Dholakia, 2006). The expectation is that the customer's increased sense of belonging, engagement or loyalty will lead to better economic outcomes for the firm, as exemplified by predictions that firm sponsors of customer communities will be "richly rewarded with peerless customer loyalty and impressive economic returns" (Hagel and Armstrong, 1997, p. 2). While Mathwick et al. (2008) were primarily concerned with how these benefits enhance value to customers, they and other researchers also point to increased customer loyalty and commitment, which should lead to incremental economic gains as downstream consequences for the firm (Algesheimer et al., 2005; Balasubramanian and Mahajan, 2001; Porter and Donthu, 2008). Using surveys and self-report data, some academic research has reported an increase in purchase intention among online customer community members (Algesheimer et al., 2005; Porter and Donthu, 2008). Other researchers have shown that enabling consumer membership in a firm-sponsored online customer community is one of seven factors linked to increased purchase intentions and willingness to pay (WTP) a price premium with online retailers (Srinivasan et al., 2002).

Two recent studies have examined the consequences of customer community membership using behavioral data. Zhu et al. (2012) found that firm-sponsored online customer community membership was linked to greater financial risk-taking as observed in lending (Prosper.com) and bidding (eBay Germany) behaviors. Algesheimer et al. (2010) examined the behavioral consequences of customer community membership and participation at eBay Germany. They found that bidders and sellers at the auction site became more selective and conservative in their behavior as a result of online community participation, leading to null or negative effects of community membership on individual-level bidding volume, product listings, average amount spent by buyers, and revenue earned by sellers. A unique aspect of the customer communities investigated in the above two studies is that they both exist to "make" markets. Thus, most of the important marketing mix elements (such as product and price) in both of these settings are a function of the actions of independent agents rather than of the firm. The setting and findings from these studies – especially the Algesheimer et al. (2010) study – can be seen as complementary to the setting and findings from this paper. Overall therefore, with the two exceptions mentioned above, there has been little empirical assessment of the impact of mere membership in online customer communities to the firm.

2.2 Benefits of *Behavioral Participation* in an Online Customer Community

If the sense of belonging that arises from mere membership in an online customer community is the foundation on which economic benefits accrue to the firm, the nature and extent of the customer's participation in the community should be the mechanism through which these benefits are moderated. As discussed in the prior section, customer communities are commonly defined by the nature of the attributes and behaviors within them; that is: (a) social interactions that occur among (b) a structured set of social relationships. These behaviors include the social transmission of product experiences, recommendations or advice and developing social relationships among individuals who share common interests (Algesheimer et al., 2010; Brown et al., 2007; Dholakia et al., 2004; Muniz and O'Guinn, 2001; Schau et al., 2009). In this section, we provide additional support for our predictions that the extent to which a community member participates in the community by (a) interacting socially through content contribution of (posting versus lurking), (b) consuming more information posted for products that are more likely to benefit from the "like-mindedness" of community members (preference heterogeneity) and (c), establishing structured social relationships (friend ties) will moderate the economic impact of community membership.

Posting vs. Lurking: Schlosser (2005) describes two types of consumer participants in Internet-based transmissions of product information: posters, who actively share their product experiences online, and lurkers, who read others' postings without communicating themselves. Lurkers are generally believed to represent the majority of people in online customer communities, while a small minority of members generate the content (McWilliam, 2000). A survey-based investigation of online customer communities found that the act of content contribution by community participants was positively related to member brand commitment (Jang et al., 2008). Posters commit more conspicuous public behaviors in a community, leading to an increase in their sense of belonging or engagement with the focal brand or product whether the community is "real" or virtual (Kozinets, 1999; Laroche et al., 2012). An empirical link between community-driven customer engagement of this kind and self-reported purchase behaviors was first shown in a survey of car club community members (Algesheimer et al., 2005). This finding is also in keeping with theorizing by Balasubramanian and colleagues (Balasubramanian and Mahajan, 2001), who suggest that posters may gain approval utility from the benefit of others consuming their contributions to the community, with an expected consequence of increased expenditure with the firm. Notably, while several researchers have reported behavioral or economic impacts of posting on the purchase behavior of the consumers of this information (e.g. Chevalier and Mayzlin, 2006; Godes and Mayzlin, 2004), we are unaware of prior work reporting the impact of posting versus lurking on the poster (or lurker) her self.

Preference Heterogeneity and the Role of Information: Participation in online customer communities should offer consumers unique informational benefits due to the nature of product information available within (versus outside) the community. Most of the product opinions shared by consumers online are offered by anonymous or socially distant sources, giving the consumer little means by which they can assess the personal relevance of the recommendation (Dellarocas, 2003; Ma and Agarwal, 2007). In contrast, both offline and online customer communities are by definition organized around the shared interests of like-minded individuals (Algesheimer et al., 2010; Bagozzi and Dholakia, 2002; Porter and Donthu, 2008) who are more demographically and psychogically similar to one another (Dholakia and Vianello, 2009). Customer communities should thus tend to exhibit heightened homogeneity in individual attitudes (McPherson et al., 2001; Watts et al., 2002). Consumers expect that because homogeneous sources of product information are more likely to share the consumer's own attitudes and preferences, they are more diagnostic to the information task and are more likely to influence purchase intentions (Brock, 1965; Eagly et al., 1978; Gershoff et al., 2001).

In our setting, while we do not observe the extent to which customers share attitudes towards products, we do observe the nature (i.e., category or genre) of some of the products they purchase. To link product attributes to the shared interests or like-mindedness (attitudinal homogeneity) of community participants, we rely on Feick and Higie (1992), who empirically demonstrate that homogeneous sources of product information are particularly persuasive for products with more heterogeneous preference structures (e.g., restaurants) than for products that can be more objectively assessed (e.g., personal computers). In short, "like-minded" people are better sources of information for products for which subjective tastes or preferences vary. If there is an informational benefit to gaining product information from like-minded participants in a customer community, we should observe a disproportionate benefit (i.e., increased expenditure) for products that are higher in preference heterogeneity. Therefore, we predict that the effect of customer community participation on purchases (the social dollar) will be moderated by the extent to which the products purchased are perceived to hold greater preference heterogeneity.

Structured Social Relationships (friend ties): In online customer communities, the list of explicit social connections made among members for the purposes of social interactions are commonly described as "friend ties" (Zaglia, 2013). These ties can be said to represent one form of the structured social relationships among members that are central to brand and customer communities (Muniz and O'Guinn, 2001). When a firm sponsor facilitates the creation of friend ties, it is said to promote social interactions among members, reinforcing their social engagement with the community, and creating "collective value" for both the customer and firm (Dholakia et al., 2004; Schau et al., 2009). A common belief among marketers and researchers interested in brand communities is that the firm component of this value is in part economic– "Forming relationships among like-minded consumers who share one's interest in the brand will be credible and impactful in persuading and bonding customers to the brand, leading them to make more purchase behaviors" (Bagozzi and Dholakia, 2006, p. 46). We are unaware of research that has directly tested the assumed link between community participation (such as posting or friend tie formation) and the expected economic consequence of hosting an online customer community (i.e., increased purchase behavior). For example, while Algesheimer et al. (2005) find relationships between (a) community engagement and participation behavior, and (b) community engagement and brand-related purchase behavior in a survey of car club community members, they do not predict or report a test for the potential link between community participation activities and brand-related purchase behavior. Zhu and her colleagues demonstrate that the degree to which a consumer participates (e.g., posts) in auction and lending communities is linked to their risk-seeking tendencies (Zhu et al., 2012). They are not, however, primarily concerned with the consequences of risk-seeking consumer behavior for the firm. In contrast, the present research examines participatory behaviors linked to the fundamental theorized benefits to the firm of customer community participation.

To sum, we expect that the social dollar will be moderated by the extent to which the community member leverages attributes that should be fundamental to community success: engaging in social interactions (posting versus lurking), leveraging user-generated content more for product categories that should benefit the most from a customer community (category preference heterogeneity), and creating structured social relationships (friend ties). We expect a negative (positive) effect of lurking (posting), and a positive effect of the presence and volume of friend ties on the size of the social dollar at the individual member level. Finally, because these communities should foster social interactions among like-minded individuals, we predict a bigger effect as the preference heterogeneity for a product category increases.

3 Research Setting and Data

Our data come from a large North American retailer of entertainment and informational media products (e.g. books, movies, music).[2] The firm is the largest retailer in its market by sales volume in its core product category, and operates in both retail and online channels, with approximately 10% of total revenues occurring online for the firm's fiscal year 2009.

The firm launched its own online customer community in September 2007. The formation and existence of this community was communicated via mass marketing to consumers and current customers. Marketing communications were comprised of signage in stores, banner advertising on the firm's website, print advertising in national newspapers, and the firm's house opt-in email list. Advertising announcing the launch of the online customer community was untargeted– different customer segments were not given differential exposure to this announcement. Participation in the community was purely voluntary on an "opt-in" basis, and no financial incentive was given to customers to join the community. In addition, after the launch of the community, the firm did not engage in marketing activity that was targeted at community members. In other words, customers who joined the community and customers who did not received the same marketing activity.

Our empirical setting is consistent with this literature's conceptual description of a *customer community*. Specifically, we observe a firm-organized and operated online environment that the firm explicitly describes internally and to the public as a customer community. Individuals who join the customer community share textual and graphical information about themselves and their product preferences and/or recommendations with other customers, graphically display a variety of personal and product-related content on a personal profile page, engage in discussions on community chat boards and establish formal friend ties. Customer community participants contribute a variety of

[2]Due to the proprietary nature of the data, the firm has requested that its identity not be divulged.

user-generated content for the consumption of others who are either within their own network of friend ties (i.e., “private” content) or for the customer community at large (i.e., “public” content). While the content of the customer community interactions we observe is most commonly over products-in-use (McAlexander et al., 2002), we also observe conversations regarding the firm brand (the retailer). Thus, our setting is highly consistent with prior descriptions of a firm-sponsored online customer community (Algesheimer et al., 2010; Brown et al., 2007; Dholakia et al., 2004; Kannan et al., 2000; Muniz and O’Guinn, 2001; Schau et al., 2009).

The data used in our analysis were extracted in January 2009. Using an “*n*th-select” random sampling procedure, the firm generated a random selection of 26,624 community members (from a population of about 266,000 such members) for analysis. The firm provided us with two kinds of data – transactional data and community activity data – for these members. The transactional data represent actual purchases made by these members in the firm’s online and retail (offline) channels. We are able to observe offline purchases for some community members via the use of a firm sponsored loyalty card. Customers could sign up for this card by paying a modest annual fee ($ 20). All customers in our primary analysis sample (across both treatment and control groups) had signed up for the loyalty card, hence there are no differences on this dimension between the two groups. Across the firm’s entire customer database, approximately 16% of customers had a loyalty card and they accounted for approximately 40% of the firm’s total sales revenue. Each record in the transactional data includes the date of the customer’s first purchase, his/her first name, his/her geographic location, and details on each purchase event. Each purchase event indicates the channel and date of purchase, the specific product(s) purchased, customer expenditure net of any standing or promotional discounts received for each product within the transaction, and each product’s category classification. While we focus on the loyalty card holders in our primary analysis because it provides the most conservative estimate of the social dollar and supports a multi-channel view of the effect, to enhance generalizability we also report our analysis restricted to online customers who are not loyalty card holders as part of a robustness check.

The community activity data we observe includes the date members joined the community and the social behaviors in which they have participated within the community. Specifically, we observe the volume of several different types of user-generated content such as peer-to-peer product recommendations, product reviews written, Top 10 lists published, and the number of products (e.g., book cover graphics) displayed on their personal profile page.

There was a difference of fifteen months between the data extract (January 2009) and the formation of the customer community (September 2007). We therefore also asked the firm to provide fifteen months of data before the launch of the community for the random sample extracted. This allowed us to create a “pre” period for comparison. The firm provided transactional data going back to June of 2006 (i.e., fifteen months before the launch of the community), for the full analysis sample. In addition to the sample drawn from the community members described above, we asked the company to provide transactional data on customers who did not participate in the community to create a control group. The firm drew a random sample from customers (the total population was just under one million customers) who had not become members of the customer community during our observation period and who transacted at least once with the firm (online or offline) in the thirty months from June 2006 to January 2009, inclusive. They were able to provide us data for 6,091 online transactional accounts for our control group. Of these accounts, 2,352 were also loyalty card holders, which provides full visibility of their purchase behavior with the firm (online and offline (retail) channels).

In the subsequent discussion, we designate the 15-month period before the launch of the community as T1 (“pre-community,” June 2006 to September 2007, representing five quarters denoted Q1 through Q5) and the period after the launch of the community as T2 (“post-community,” October 2007 to January 2009; quarters denoted as Q6 through Q10). Note that while the exogenous change (the launch of the customer community) occurs at a specific point in time (September 2007), a customer can decide to join the community at any time after the launch. We address this issue in detail and exploit this attribute of the data in our analysis. Taken from the full sample described above, our primary analysis sample includes customers for whom we observe behavior across both sales channels (via the loyalty card) and who transact at least once in T1 and T2. We do this to ensure full visibility of the customer’s expenditure with the firm (across both channels) and to control for differential entry and exit patterns in the treatment and control groups (Blundell et al., 1998). This has the added benefit of making our results as conservative as possible (results without the entry/exit restriction are described in section § 4.4). The application of our primary analysis criterion limits this sample to 7,909 (30% of the full treatment sample) and 1,255 (21% of the full control sample) customers in the treatment and control groups respectively.

A comparison of the two groups on behavioral variables (e.g., total expenditure) is provided in Table 1.[3] The table shows that there is no significant difference in the total expenditure per customer and the number of orders per customer in the fifteen month period before the community was launched. However, average purchase size or the per order dollar expenditure is marginally higher for the control group (and statistically significant). After the launch of the community, the total mean expenditure and the number of orders increases for both the treatment and control groups while the average purchase changes very little for both groups. In terms of demographics, we see some minor differences (see Table 2), with the treatment group having a slightly higher proportion of women, a slightly larger household size and slightly lower access to computers.

4 Estimation Strategy and Results

In this section, we present a roadmap for our analysis approach (as laid out in Table 3) with a brief description of the role played by each of these analyses. First, we describe our modeling approach to estimate the magnitude of social dollars (if they exist) in § 4.1. As noted earlier, we exploit the exogenous formation of the online customer community and the availability of a control group. We replicate our results with demographic and behavioral controls. Given that our data do not originate from a setting characterized by perfect randomization (e.g., a field experiment), we need to ensure that our findings are not driven by self-selection (which would lead to non-random assignment to treatment and control groups). We therefore run a series of analyses controlling for self-selection based on the observables as well as unobservables in § 4.2. Having demonstrated the robustness of social dollars to selection concerns, we then explore the validity of other possible explanations that could account for the social dollar in § 4.3.

4.1 Existence and Magnitude of the Social Dollar

Given the structure of our dataset, we use a panel specification to help us obtain the magnitude of the social dollars. The observation of transactions of the same customers before and after helps us rule out alternative explanations such as selection. The presence of a control group helps us

[3] Overall, the primary analysis sample consists of heavier spenders. The total expenditure before the launch of the brand community is 30% higher for the analysis sample we draw from the treatment group and 73% higher for the analysis sample we draw from the control group.

rule against the influence of exogenous factors that may influence transaction expenditure. We aggregate the detailed purchase data to these two periods (T1 and T2) rather than leveraging a more fragmented time series form to mitigate potential serial correlation and grouped error term effects (Bertrand et al., 2004). The specification we estimate is

$$R_{igt} = \beta_1 I_g + \beta_2 I_t + \beta_3 I_g I_t + \beta_4 X_{ig} + \varepsilon_{igt} \tag{1}$$

where R_{igt} is the outcome of interest (the total dollar expenditure at the individual level for most of our analyses) for consumer i in group $g \in \{Treatment, Control\}$ at time $t \in \{T1, T2\}$. The X_{ig} consists of a vector of county-level demographic variables (e.g, median household income, avg. household size, avg. household educational spending, computer access) and controls for observable differences across customers in our sample.[4] Possible differences between the treatment and control groups are captured by β_1, while β_2 controls for expenditure differences that are common across treatment and control groups between T1 and T2, and ε_{igt} is the error term. The coefficient we describe as the "social dollar" is β_3, which estimates the causal effect of treatment (community membership) on purchase behavior, controlling for biases in permanent group differences and biases within the treatment group due to individual trends across the time periods.

Differential group entry and/or exit represents significant threats to the assumption of no sample composition changes across groups in panel estimation (Blundell et al., 1998). For example, one could expect that a customer community attracts new customers to the firm. From the firm's perspective, it would be natural to include these customers in calculating the returns from launching the community, as these new customers also generate additional purchase revenues for the firm. Note that this makes for a less conservative test than the one we adopt. The criterion for our primary analysis sample that each consumer transact at least once in T1 and T2 ensures that we include only existing, "active" customers in both groups. This resolves uncertainty in the causes of any differential entry or exit by fixing group composition over time and avoids the possible confound of attributing higher post-period revenues to the community, which may have been contributed mainly by new customers. Subsequently, in section § 4.4, we relax the use of both the loyalty card and entry/exit restriction criteria used in our primary analysis to explore their respective impacts on our findings.

[4]These data were collected from the 2006 National Census county-level database.

To conduct our basic analysis, we use the date of the launch of the community as the temporal "break," *even though customers do not all join on that particular date.* Specifically, we classify a customer who joins the community at any point in time (till the end of our data series fifteen months later) into the treatment group. The reason we do this is because otherwise, the same customer will enter both the control and treatment groups, invalidating our identification strategy. It is also important to note that this classification scheme works *against* our finding social dollars. This is because by including customers who do not join right away, we are in effect including "untreated" customers in our treatment group, thus biasing our estimate of social dollars towards zero. In other words, our estimate of social dollars will be conservative. We also exploit this feature of our data (that not everybody in the treatment group joins at the same time) as one of our tests ruling against selection on unobservables.

As presented in Table 4, our results show that the social dollars, as represented by β_3 in equation 1 above, exist (i.e., are statistically different from zero) and are \$ 127.01 in magnitude over the 15 month observation period (approximately \$ 102 on an annualized basis). Using an annual base expenditure of \$ 676.69 (in T2) for the treatment group, social dollars are estimated to be 19% of expenditures post the launch of the customer community. To the best of our knowledge, this is the first empirical result documenting that an online customer community can lead to a direct increase in total customer-level expenditure for its firm sponsor.

Two aspects of this result are noteworthy. First, as noted earlier, the brand name under which this customer community operates is a retail brand selling a variety of individually-branded books, DVDs, CDs, and a selection of ancillary gift items (e.g., bookends, pens, greeting cards). In essence, the retail brand name is an "umbrella" brand associated with a specific assortment of product categories. This can be contrasted with a brand such as Lego that sets up its customer community in the context of Lego's toy building-brick product only. We expect that the social dollars effect in an assorted product community would be weaker than that of a single product community given the potential diffusion of customer interest over multiple brands and/or product categories. Second, our results are quite different from the two other studies that quantified changes in customer behavior as a consequence of joining a customer community (Algesheimer et al., 2010; Zhu et al., 2012). The most direct comparison can be made with Algesheimer et al. (2010), which found that participation in eBay Germany's customer community leads to null or small negative effects on economic outcomes. eBay can be contrasted with most e-commerce websites in that consumers play the role of both seller and buyer. As the authors note in their study, customers who join the community become educated about both the site as well as the behavior of other buyers and sellers. This education leads them to be more efficient and effective in their own marketing behavior, leading to fewer listings. We propose that the customer community we investigate may be more representatitive of firm-sponsored online customer communities in general. In such communites, consumers engage with one another in generally positive social interactions pertaining to shared product or consumption interests rather than in competitive transactions to determine best prices (eBay) or loan interest rates (Prosper.com). Thus, in many ways, our results may be seen as complementary to those found in the above studies with the difference in results likely driven by the distinct nature of the sponsoring firm's business model (and the resulting nature of its online customer community).

Applying our analysis to each of order size and order frequency as dependent variables, we find an increase in order frequency due to membership in the customer community (Table 4). The increase in average purchase size, however, is non-significant. Order frequency appears to drive the majority of the social dollar effect – we observe nearly three additional purchase occasions over the 15 month observation period, representing an 18.4% increase in order frequency. This finding is consistent with the online customer community literature which argues that the array of informational content and opportunities for social engagement available in the community should increase the number of visits the consumer is likely to make to the community website, as well as the conversion rate per visit (Brown et al., 2002; Holland and Baker, 2001; McWilliam, 2000), leading to an increase in order frequency but not necessarily in order value (Schau et al., 2009). Other reasons for the lack of significant change in the order value could be a threshold effect shown in shopping behavior (Wansink et al., 1998) - in this case, the threshold is the monetary value of the order that qualifies for free shipping.

Given that the firm operates in both online and offline channels, a reasonable hypothesis could be that the social dollars arise from differential expenditure across sales channels. Specifically, the online nature of the customer community we observe may cause social dollars to be generated more from the online channel or, in an extreme case, to originate entirely from channel switching. To assess this, we first check to see if there is a significant difference in the share of customer expenditure between the online and offline channels before and after the community is formed. We carry out an analysis for the proportion of total sales for a given customer that comes from the online channel. We find that the online proportion goes up by 13 points (Table 4 column 4), suggesting that the community does shift purchase behavior towards the online channel.

To assess the size of this shift, we replicate our basic analysis by channel and find that the total social dollar magnitude - $ 127.01 - is composed of $ 87.79 from the online channel and $ 39.23 from retail (offline), representing a 37.0% and 8.9% increase in T2 purchases in the respective channels. Two things about this decomposition are noteworthy. First, as predicted by research reporting that the ability to exchange information in customer communities enhances loyalty to e-commerce providers (Srinivasan et al., 2002), 70% of the social dollar arises in the firm's online channel. Second and perhaps more important, community membership seems to increase customer expenditure in the offline channel as well (albeit directionally at p = .12). To our knowledge, academic research has yet to consider the presence and magnitude of cross-channel effects of online social interactions such as those found in customer communities. This finding thus adds to the literature by documenting an economic measure of positive channel spillover for online social interactions.

Finally, we allow for the possibility that observable differences across the treatment and control groups could affect the existence and magnitude of the social dollar. We focus on demographics and tenure with the company before the launch of the community. Recall that we had only first names for each customer in our data. Using a standard "genderizer" list,[5] we were able to infer gender for 82% of the sample (all analyses that follow that include gender are restricted to this 82%.) As shown in Table 2, group comparison t-tests indicate that the proportion of males was slightly lower (albeit significant statistically) in the treatment than control group. There were also minor differences in the average household size and computer access between the two groups, with the average treatment group household being slightly larger. The distribution of the sample across the 12 geographical regions (which correspond approximately to counties) observed in our data was not significantly different across the treatment and control groups ($\chi^2(11) = 13.54$, p=0.26).

We replicate our main analysis with explicit controls for these minor differences across the treatment and control groups. Specifically, we add a vector of demographic controls (including

[5]A list of over 100,000 international first names recommending assigment of records to one of three categories (e.g., "Christopher" = male, "Christina" = female, "Chris" = unclassified).

gender, observed tenure with the firm, average household size, median family income, household educational spend and computer access) to the main specification (equation 1). As can be see from Table 4, we replicate our results after controlling for these demographic differences.[6]

4.2 Dealing with Self-Selection

While an experimental design involving random assignment to conditions is ideal, many field applications of panel estimators observe non-random group assignment (e.g., Autor, 2003; Chevalier and Mayzlin, 2006; Bronnenberg et al., 2010). Given that customers in our field data are not randomly assigned to treatment and control groups, it is possible that the two groups could differ– especially on unobservables– and that such a difference drove treatment group members to join the community. While the model and the use of the "before-and-after" structure of the data reduces these selection concerns by design,[7] we pursue additional robustness checks to more thoroughly assess a selection-based explanation for our results. Given that we observe the behavior of consumers in T1 (before the community), differences on unobservables become an issue only if the unobservables have a differential interaction with the treatment (the formation of the community). For example, it could be that the customers in the treatment group were more engaged with the firm in T1, as previous research has shown that customer communities have a much larger effect on more engaged customers (Algesheimer et al., 2005). Our strategy first examines the possibility of self-selection affecting our results via statistical analyses. We then exploit a feature of our data– the fact that not all consumers join the community at the same time– to create a new control group and check whether self-selection could indeed account for our results.

4.2.1 Selection on Observables

Though customers in the treatment group did not appear to be markedly different from those in the control group based on the demographics we observe (Table 2), it is still possible that observable differences between customers in these two groups may drive differences in their purchases.

[6]We report the estimates of the Social Dollar in Table 4 with the demographics entering as main effects. We also ran analyses where we included interactions (treatment × demographics) and find that our results do not materially change. Detailed results are available from the authors on request.

[7]Panel estimators have been used by other researchers in similar contexts to exploit the advantages they provide via the elimination of individual-level differences across analysis groups/conditions (e.g., the elimination of book-website specific fixed-effects in Chevalier and Mayzlin, 2006).

We therefore ran analyses to check if shopping behavior in the pre-period (expenditure, purchase frequency, variability in expenditure - all at the quarterly level) could predict membership in the community. We found that these behaviors did not predict membership in a logit model.

We also carried out a matching analysis using a a "kernel matching estimator" (Heckman et al., 1998b). The matching estimator computes the effect of a treatment on the treated by matching each treated person with an untreated person based on observable characteristics. The results of the analysis (shown in Table 5 column 3), show that the social dollar estimate is significant and 20.99% in magnitude. We also replicated this result using a "nearest neighbor" matching estimator (Heckman et al., 1998a; Abadie and Imbens, 2006) which produced an estimate for the social dollar of 21.86%. The results were also substantively invariant to the number of nearest neighbors chosen for the matching process (using matching based on 1, 5 and 20 nearest neighbors). Finally, the results were also insensitive to the choice of kernel (all additional results are available from the authors on request). Overall therefore, these analyses suggest that observable differences do not predict the propensity to join and the existence of the social dollar.

4.2.2 Selection on Unobservables

In this section, we carry out two types of analyses to see if selection on unobservables can explain our results. In the first set of analyses, we carry out different formal statistical tests to check for this. In the second set of analyses, we exploit a unique feature of our data to set up a control group by using customers from the treatment group before they select into the treatment.

Table 5 column 4 presents results of a selection model based on the Heckman selection framework. We do not find evidence for selection as the selection correction term λ is insignificant with a p-value = 0.85 (for more details on the implications of the significance of λ, see Heckman, 1979, p. 158 and Puhani, 2000, p. 55). An interesting observation is that ρ, the correlation coefficient between the selection equation and outcome equation errors, is negative at -0.0527. This suggests that a positive shock to the probability of selection reduces the treatment effect through a negative shock to the outcome (see Shaver, 1998; Lemke and Reed, 2001). In other words, the negative directional result above indicates that if there was selection, accounting for it is likely to increase the social dollar magnitude relative to our current estimate (which is based on there being no selection). The estimate of social dollars based on the Heckman selection framework is 32.82%, again supporting our main result as conservative. We also employed a semi-parametric estimator (Choi, 1990; Rosenbaum and Rubin, 1983; Heckman et al., 1998a; Heckman et al., 1998b) that takes a weaker stance on the joint distribution of errors in the outcome and selection equations. The estimate of the social dollar using this analysis was 20.29% (details of this analysis are available from the authors on request).

To further test for the effect of unobserved selection on our estimates, we report two other analyses in the Online Appendix - the Rosenbaum bounds approach and the Relative Correlation Restriction (RCR) approach. The results from both of these analyses also suggest that selection based on unobservables is unlikely to have driven our results.

We now turn to a different approach to check for selection on unobservables. Recall that while the community formation appears as an exogenous shock to customers of the firm, customers are not required to join the community at the same time. In the five quarters after the community was launched, the proportion of our treatment group customers who joined was 44%, 17%, 14%, 14% and 11% in Q6, Q7, Q8, Q9 and Q10 respectively. In other words, the majority of the customers (56%) in our (right-truncated) sample joined after the first quarter.[8]

Once the community was available to customers, if there is a differential interaction between the availability of the customer community and the unobservable attribute(s), then a change in transaction behavior should manifest itself even if a (future) member had not yet joined the community. We present a series of analyses in which we show that social dollars are not driven by *whether* some customers join the community but that they occur only once they choose to join the community. In other words, that it is the act of joining and not the mere availability of the treatment that impacts transaction behavior. In fact, the temporal joining data discussed above already suggest that the availability of the customer community was not enough to sort customers on an unobservable attribute. If this had been the case, a large majority of customers would have joined the community immediately (in Q6).

In the first analysis, we group customers who join the customer community within specific quarterly time intervals into cohorts and contrast cohort behavior across time to see if this impacts the size of the social dollar. A challenge here, however, is that the control group definition for these

[8]As noted earlier, the dispersion in the community join date also suggests that our main social dollars estimate ($ 127.01 or 19% of T2 expenditure) is conservative as many customers did not "benefit" from the community until later in T2.

cohorts is not obvious. This is because our current control group by definition consists of people who *do not* join the community during the 15 month period of its operation that we observe. We therefore use a different strategy to test for the possibility that differences in the time period in which customers join the community impacts the social dollar. We first consider all customers who join the community in the first quarter after the formation of the community (Q6) as our treatment group cohort. The control group for this cohort includes all customers in the treatment group who did not join the community until after the first quarter of its operation; that is, they joined the community between Q7 and Q10, inclusive (Table 6 Panel B, column 1). If our prediction (that behavior changes *when* rather than *whether* they join) is correct, until these customers join the community, we should be able to treat them as control group customers. This analysis can be seen in the same spirit as the falsification analysis carried out in Goldfarb and Tucker (2011).

We perform the analysis for pre (T1) and post (T2) periods of equivalent length limited by the duration of the T2 period for which the treatment cohort were community members. Analysis for subsequent "join cohorts" proceeds similarly (Table 6 Panel B, columns 2-4). As shown in the table, the social dollar effect is positive and significant for all four cohorts. This suggests that even when we restrict our analysis to all customers who possessed the (presumed) unobservable attributes that interacted with the treatment, and divided them into treatment and control groups as above, the social dollar effect is present and significant. Thus, it is not whether they join the community but when they do that matters.

We next focus on a second cohort-based analysis of customers who join the customer community. However, here we compare behavior within cohorts by comparing a cohort's quarterly expenditure post joining the community with their own quarterly expenditure pre joining the community. Specifically, we compare the average T1 quarterly revenue for customers who ended up joining the community in a specific quarter with their average revenue in each quarter after becoming members. For example, for the cohort of customers who joined in the first quarter after the community launch (Q6, column 1 in Table 6 Panel A) we observe their revenue in each of four T2 quarters after the join quarter, and compare each of these four quarters after joining to the T1 quarterly mean. The analysis proceeds similarly for those who joined in later quarters until those who joined in Q9 (column 4), for whom we only observe one additional quarter in T2 after this time (Q10). As shown in Table 6 Panel A, nine out of the ten quarterly comparisons in the table are statistically significant. These results further support the argument that there is a significant change in the behavior of these customers based on when (rather than whether) they join the community.

In the above analysis, note that we do not have a control group. One way to reduce the impact of not having a control group would be to shorten the window before and after joining the community in a “regression discontinuity” style analysis. For this analysis, we use the day of joining as the “origin” and contrast mean expenditure before joining with mean expenditure just after joining. The shorter the temporal window on each side of the treatment, the less likely that factors besides the treatment will affect outcomes (Imbens and Lemieux, 2008; Hartmann et al., 2011). We examine behavior in the shortest possible window we observe - a day - as well as two, three, four, five, six and seven days.[9] As expected (Table 9), the mean expenditure increases (for both pre and post launch) as the duration of the window gets longer. We find that the post-launch mean expenditure is higher than the pre-launch expenditure and that the difference is statistically significant for all of the windows we consider.

Overall, the analyses presented above suggest that it is unlikely that a differential interaction between unobservable attributes and the treatment was a material driver of our findings. While it is impossible for us to rule out the effect of all possible unobservables with certainty, it is unlikely, given our robustness analyses, that they played a role in our estimation of social dollars.

4.3 Alternative Explanations for the Social Dollar

In this section, we explore a variety of alternative explanations for our finding of the social dollar, considering both cross-sectional and longitudinal explanations.

4.3.1 The Role of Outliers

We first focus on the possibility that the existence and magnitude of the social dollar is primarily driven by outliers on the expenditure dimension (i.e., a cross-sectional difference). In other words, it is possible that a few heavy spenders are driving the result. To check this, we divide our treatment and control groups into expenditure quartiles using their total purchases in T1 as the baseline.

[9] The sample size changes for each window as we need to drop treatment group customers for whom the end of the “after” window exceeds the end date of our data.

We then carry out a separate analysis for each quartile. As shown in Table 7, the social dollar is statistically significant for all four quartiles and has the strongest statistical significance level for the bottom three quartiles. In terms of magnitude, the largest coefficient estimates are for the middle two quartiles, approximating the "average" customer in our primary analysis sample. The statistically significant, but relatively weaker effect in the top quartile is driven by the variance in this high purchase volume group. It is also likely that the firm already accounts for a greater share of wallet for these heavy spending customers, thus leading to a smaller estimate for their social dollars. Overall, this analysis strongly suggests that our results are not driven by a change in expenditure for a small minority of customers, and especially not only by customers who already spent heavily with the firm prior to the launch of the community.

4.3.2 The Novelty Effect

We next explore whether social dollars may arise due to the novelty of the community. It is possible that customers respond positively to the community as soon as it is launched but then lose interest and revert to their normal (pre-community) purchase behavior ie., a longitudinal explanation. To test this, we estimated the social dollar using our analysis on a rolling quarter basis. We first focused on three months after the launch of the community to create a treatment time period and used three months before the launch of the community as the control time period. Thus, all activity in the first quarter after the launch of the community is contrasted with the first quarter before the launch of the community. We then extend the treatment time period to the second quarter, i.e., months four through six after the launch of the community and add the corresponding control time period prior to the launch of the community. As shown in Table 8, the social dollar persists over time, with the significance of the difference the weakest in the one quarter window but becoming very strong in the two to five quarter windows. The quarterly change in social dollars over each prior quarter appears persistent at \$ 22.63 (\$ 44.25 less \$ 21.62), \$ 20.41, \$ 37.21 and \$ 25.14 from the second through fifth quarter in T2, respectively.

4.3.3 Differential Trends in Customer Loyalty

Another factor that could potentially be driving the social dollars estimate is that there is a differential response to the firm across the treatment and control groups over time. For example, customers who end up in the treatment group start liking the firm more and more before the launch of the community. This could result in a situation where the total expenditure in T1 for the treatment and control groups is not statistically different (as shown in Table 1), but the increased liking results in significant divergence between the treatment and control groups. With the passage of time, this divergent trend could widen the gap between the two groups – a difference that could improperly be ascribed to customers joining the online community. To test for this possibility, we perform across-group trend analysis of total revenue for the treatment and control groups. The statistical analysis we carry out is a mixed-effects model estimated by restricted maximum likelihood (Verbeke and Molenberghs, 2000; Wallace and Green, 2001). This approach is preferred over traditional repeated measures using GLM methods as it allows for a more accurate depiction of serial correlation and correlated error structure, and can accommodate unbalanced group sizes. This model is represented as

$$R_{iq} = X_i\beta + Z_q u + (X_i\beta\, Z_q u) + e_{iq} \tag{2}$$

where R_{iq} is a 5×1 vector representing the total revenue of customer i in quarter q within the five quarters of T1, predicted by the fixed component of analysis group ($X_i\beta$), the random time component ($Z_q u$) and their interaction ($X_i\beta\, Z_q u$). To control for expected serial correlation and correlated error stucture in the within-customer revenue trend we allow an AR(1) process on the error term (e_{iq}). The interaction term ($X_i\beta\, Z_q u$) - that would indicate a difference across comparison groups in the linear slope of the purchase trend across the five quarters of T1 is non-significant (t-test = 0.50, $p = 0.62$). Given the quarterly purchase trend approximates an inverted-U shape (see Table 2), we also specified a model adding a quadratic main effect and interaction for time. The quadratic interaction term was also non-significant (t-test = 1.47, $p = 0.14$), failing to support a difference in curvilinear trends.[10] As an additional test, in Table 2, we present simple group mean comparisons by quarter within the T1 period, which also supports non-significant differences in each of the five quarters prior to the community launch.

These results allow us to rule out the possibility that any differences that we find between the

[10]The results were also identical using a traditional GLM repeated measures model for both the linear (Huynh-Feldt adjusted $F(3.5,\ 32355) = 0.81$, $p = 0.51$) and quadratic interaction terms (Huynh-Feldt adjusted $F(2.0,\ 18740) = 1.76$, $p = 0.17$). We were also able to replicate this finding at the individual level (as opposed to the group level) using a probit model. Details on these analyses are available on request from the authors.

treatment and control groups after the launch of the community are driven by differential trends in behavior prior to the community launch.

4.3.4 Outlier Months

The majority of our analyses have been carried out comparing customer behavior in the after period to the before period (in aggregate) as well as at the quarter level. However, it is possible that the social dollar is driven essentially by month to month variations in expenditure. Note that the control group allows us to account for basic seasonality patterns. This analysis demonstrates that the effect is present consistently in each month across our data period. In other words, it is not the case that the social dollar exists only for a small number of months.

We conducted two analyses to test this – one where we pooled the data and one where we ran it month by month. In the first test, we regressed monthly transaction expenditures for the treatment and control group users with demographics and month fixed effects on the right hand side (see table 10). The social dollar estimate using this specification is significant and at 22.49% ($ 9.40/$ 41.79). In the second test, we ran our main panel data regression model using only data for each month with demographics (see table 11). We find that the social dollar is significant in 11 out of 12 months (the month of May is marginally significant at t = 1.87) and ranges from 15.47% to 30.21%. Both these analyses show that the social dollar is persistent at the monthly level.

4.3.5 Joining for a Reason Unrelated to the Online Community

Another possibility is that customers may join the community for a reason that is not driven by an affinity for the retailer and/or its community. For example, if a customer was dissatisfied with a purchase, s/he may be driven to join the community and warn others not to buy. Another situation could be if a customer joins essentially to obtain more information about a product recently purchased from the site in terms of functionality, usage, etc. These are likely to be specific situations where the customer pursues immediate action. In these cases, we should expect to see no effects of joining (or even a negative effect if the customer joins due to an adverse reaction to a purchase), especially in a short time frame after joining the community. We therefore shorten the time frame of our main analysis to 2, 3 and 4 weeks. We find evidence for the social dollar in each of these situations (ref. last three rows of Table 9), suggesting that customers are not joining the community for reasons other than affinity to the retailer and/or the community.

4.4 Robustness Checks vis-a-vis Sample Construction

In all the analyses reported so far, we have imposed the requirement that all customers in our sample must be loyalty card holders and that they need to transact at least once in each of the pre- and post-periods. In order to check that our sampling strategy is not driving our result, we relax this requirement.[11] To do this, we restrict our analysis to online channel shoppers who do not have a loyalty card. As there is no way to track offline/retail sales without the use of a loyalty card, this restriction ensures that these shoppers only transact online. The online channel represents approximately 10% of customer expenditure with the firm during the observation period.

For online channel shoppers who are not loyalty card holders, the estimate of social dollars is \$ 55.37 (SE = 16.35, $p < 0.001$) or 20.81% of post-period expenditure if we require at least one transaction in the pre- and post-periods. If we relax the entry/exit restriction (for the online channel only), the estimate of social dollars is \$ 78.43 (SE = 5.75, $p < 0.001$) or 46.52% of post-period expenditure.

Next, we relax the entry/exit restriction for the loyalty card sample that enables observation of both online and offline channel sales. In this case, the estimate of the social dollar is \$ 170.96 (SE = 22.09, $p < 0.001$) or 28.85% of post-period expenditure. The much larger magnitude of the latter two numbers is not surprising as these capture the expenditure of customers who transacted in the post-period but not in the pre-period (these could be new customers or current customers that did not transact in the pre-period). Overall, these results suggest that our primary analysis sampling strategy (loyalty card holders AND no entry/exit [at least one transaction each in the pre- and post-periods]) provides a conservative estimate of the social dollar (18.77%).

[11] As described earlier in § 3, another aspect of our sample is that the size of the treatment group (7,909 customers) is much larger than the control group data (1,255 customers).To ensure that this asymmetry did not bias our results, we randomly sampled 1,255 customers from the analysis treatment group ten times (with replacement) and re-estimated the model on these datasets. We found that the statistical significance and magnitude of the social dollar effect was robust to group sample size differences in all ten cases (these results are available from the authors on request).

5 Moderating the Social Dollar

In the previous sections, we have documented the existence and magnitude of social dollars in the context we examine and demonstrated that they are robust to self-selection concerns and alternative explanations. We now turn our attention to assessing mechanisms that should be related to the observed increase in expenditure by customer community participants. Based on the discusson in § 2, our predictions are as follows: (a) we expect a negative (positive) effect of lurking (posting), (b) a bigger effect as the preference heterogeneity for a product category increases, and (c) a positive effect of the volume of friend ties on the size of the social dollar at the individual member level.

5.1 Data and Model

In order to develop our measure of product preference heterogeneity, we restrict this analysis to the book category (ie., excluding music, movies and other products) as the sub-category classification data that is needed for this analysis is only available for this category. Books are the firm's largest product category, representing 77.3% of treatment and 76.9% of control group expenditure in the 15 month pre-period (t-test = 0.69, $p > 0.45$). 516 Amazon Mechanical Turk workers residing in the United States with a performance rating of 95% or higher were each paid a nominal fee to rate the perceived preference heterogeneity of three randomly selected book sub-categories out of 53 available from the firm (computers, fiction, religion, etc.). This resulted in approximately 30 ratings per book sub-category. Specifically, participants were asked to indicate the extent to which they agreed with five previously validated seven-point scale items (see Feick and Higie, 1992).[12]

Scale reliability was above the desired threshold ($\alpha = 0.74$) and the items converged on a single factor solution. The mean ranking of book categories was 4.64 on a seven-point scale, with a standard deviation of 0.47.[13] The top three book sub-categories in terms of preference heterogeneity were art, fiction and poetry books. The bottom three categories were math textbooks, general reference and language (instruction) books. Sub-categories closest to the preference heterogeneity index mean were game-related, general psychology and true crime books. These results show face

[12]The scale items were - (1) "Most people want the same things from [sub-category] books," (2) "Personal tastes and preferences are not important in how people choose [sub-category] books," (3) "People look for different things when it comes to [sub-category] books," (4) "Whether people will enjoy a particular [sub-category] book is very much an individual, personal matter" and (5) "People can generally agree on what makes a [sub-category] book good or bad."

[13]A detailed list of book sub-categories and their ratings is available from the authors on request.

validity as customer preferences are likely to be more heterogeneous for poetry books than for math textbooks. The analysis that follows incorporates each product sub-category's preference heterogeneity index as a normalized continuous measure.

Consistent with prior research (Schlosser, 2005), our criterion for lurkers is that the community participant has not posted any content at the community website during the observation period. Based on this definition, the number of lurkers in the book purchaser data used for this analysis is 3,326, representing 72.2% of the sample. We use a dummy variable to capture lurkers versus posters (where lurker = 1 and poster = 0). The moderating effect of this social benefit of community participation is therefore the coefficient of the dummy variable

The friend ties information provided by the firm is a count variable representing the number of member-member ties established by a given community member during the observation period. In the data used for this analysis, 1,138 (24.7%) community members had at least one friend tie. Among those with a friend tie, the mean number of ties was 7.34 (SD = 17.77, max = 217). We use the count of friend ties as a continuous measure.[14]

Our analysis strategy for understanding the moderation of the social dollar by posting versus lurking (*LURK*), informational benefits (product preference heterogeneity; *HETE*), and friend ties (*TIES*) is as follows. We incorporate each of these terms as a moderator of the treatment-time $[I_g \times I_t]$ interaction in the basic model (Equation 1), resulting in a three-way term indicating the extent to which the third-level moderator enhances or attenuates the social dollar effect. This can be specified as follows:

$$R_{igt} = \beta_1 I_g + \beta_2 I_t + \beta_3 I_g I_t + \beta_4 (I_g I_t \,.\, I_{LURK}) + \beta_5 (I_g I_t \,.\, HETE) + \beta_6 (I_g I_t \,.\, TIES) + \beta_7 X_{ig} + \varepsilon_{igt} \quad (3)$$

The parameters of interest are the coefficient of the three-way interaction term for each of the lurker dummy (I_{LURK}), thethe product preference heterogeneity term ($HETE$) and the friend ties term ($TIES$).[15]

[14]Seperate models treating preference heterogeneity and friend ties as dichotomized (*HETE*) or dummy (*TIES*) variables replicate the result of the continuous measures in Table 12.

[15]We explored alternative models where we incorporated higher-order interactions and found that the four-way

5.2 Results

We estimate two models (Table 12). Model 1 replicates our main result for the book category alone, with the coefficient of the [Treat × Time] term being positive and significant, representing a 25.3% ($47.27/$186.84) estimate of the social dollar for the books category. Model 2 adds the three community participation moderators. The [Treat × Time × *LURK*] interaction is negative and significant, while the [Treat × Time × *TIES*] term is positive and significant signifying a positive benefit of greater connectedness in customer communities. The [Treat × Time × *HETE*] interaction is non-significant.

Overall, this analysis provides support for the social benefits of customer communities as mechanisms related to the social dollars effect we observe. Notably, we are unable to uncover a significant moderating effect of informational benefits (*HETE*), highlighting the potential relative importance of social benefits (*LURK* and *TIES*), at least within this constrained sample of book products. It is also possible that our indirect measure of the informational benefits conceptually linked to customer communities did not strongly capture this potential benefit. These results are novel in the sense that they provide behavioral evidence for the theoretical predictions in the extant customer community literature. From a more practial point of view, our results show an objective and quantifiable link between specific attributes of customer communities and economic outcomes.

6 Managerial Implications

Our results suggest a significant increase in expenditures from customers that joined the firm's customer community, and that this increase is linked to social behaviors related to key attributes of customer communities. Managers can use our results on how customer community attributes and benefits moderated the social dollar to improve community design and interaction mechanisms. For example, as revealed in our analysis in § 5, the difference in posting versus lurking in the community is linked to a significant positive lift in an individual's contribution of social dollars. A subsequent regression analysis revealed that there is a positive and significant marginal effect

interaction terms were non-significant. Finally, as a robustness check, we also regressed purchases in a given month (or quarter) on posting in the prior month (or quarter), including the full set of covariates. Results for both the monthly and quarterly models supported the lagged relationship between posting in the community and subsequent expenditures at $p < .001$.

of each unit of any kind of posting on consumer expenditure (restricting our analysis to posters only). We find that this effect size (at the mean level of posting) accounts for about 3% of the total expenditures post joining the customer community.[16] Thus, before even accounting for the possible impact of a customer's posts on other community members (i.e., word-of-mouth influence), the firm has evidence supporting the pursuit of tactics to encourage posting in the community to enhance the return on their community investment. Our findings also indicate that the social bonds made by establishing friend ties in the community are linked to the customer's economic bond with the firm, representing approximately 16% of expenditures post joining the community at the mean level of ties. As such, developing community functionality that assists customers in connecting with either existing "real world" friends or high potential "virtual" friends may represent a particularly important priority for managers.

Besides the direct economic benefits to the firm from setting up the online customer community, there are also considerable indirect benefits in terms of the information the community generates for the firm. For example, the data produced as a by-product of the customer community offers a more complete picture of each customer's preferences and behavior by integrating pre-purchase, purchase event and post-purchase activities (e.g., community interactions and purchases). The firm providing the data for the present analysis also reports that the massive quantity of user-generated content produced by community members strongly improves the firm's position in organic search results (i.e., the website appears before competitors when its product offering is sought on major search engines).

Finally, from the firm's perspective, an important question is whether the launch of a customer community results in increased customer expenditure and profits sufficient to recoup the investment made in terms of the community's development and ongoing operations. We approached the firm and were able to obtain estimates of community development and operating costs.[17] Based on the estimated social dollars, community costs, and firm-level margin percentages available in public financial statements we estimate that this firm achieved break-even on its investment when 33,000

[16]For the sample of posters (treatment group minus lurkers), we regressed the post-period quarterly expenditure on posting activity (represented as the total count of all user-generated content) in the previous quarter and other controls (including relevant demographics and pre-period expenditure to control for heterogeneity). The marginal effect of posting was positive and significant. At the mean level of posting per quarter (14.3), it represents, on average, 2.89% of all post-period expenditure.

[17]For confidentiality reasons, we are unable to reveal these figures and therefore offer only anecdotal evidence here.

of its current customers (our conservative restriction case) sign up for the community. Given that the firm acquired over 260,000 members within the first fifteen months after community launch, this appears to have been a very profitable investment for the firm, especially as this number is likely comprised of a mix of both current and newly-acquired customers.

7 Conclusion

Our paper adds to the small, but growing literature on the economic impact of online customer communities. While there is much theoretical and survey-based research available on the motivations of consumers who participate in such communities, there is a paucity of research that uses behavioral (market) data to quantify the possible economic benefits to firms that set up these communities and the mechanisms through which these benefits may occur. Using a novel dataset from a firm that operates such a community, we are able to quantify the incremental expenditure resulting from customer engagement in a community. The availability of customer expenditure both before and after the formation of the customer community allows us to create treatment and control groups, helping to rule against multiple selection issues. We find that social dollars represent a double-digit increase in revenue once customers join the community we observe. These social dollars arise primarily via more frequent orders with the firm rather than increased shopping basket sizes, a result consistent with prior theorizing.

As is important for studies that leverage natural events, we carry out a series of robustness analyses and test for alternative explanations to make sure that our estimate of social dollars can be attributed to customer membership in the community. We find that our estimate is robust to the novelty effect, to differences in expenditure levels across customers before they join the community, to temporal trends between the treatment and control groups before they join the community, and to both observable and unobservable attributes that characterize each group. Furthermore, the social dollar persists over time and exists in both the online and offline channels.[18] Investigation of five potential alternative explanations for the effect we observe did not support these alternatives.

We then examined theoretically-supported moderators of the social dollar. We find that par-

[18] As reported earlier, the offline channel effect is sensitive to sample specification, ranging from directional ($p = .12$; § 4.1) to highly significant ($p < .001$; § 4.4). At minimum, the results suggest that the online community does not cannibalize the offline channel.

ticipants who partake more in the social benefit of communities by posting user-generated content tend to exhibit higher social dollars relative to participants who are less actively engaged in content-sharing in the community (lurkers). We also find that the extent to which consumers create structured social relationships in the community enhances their contribution of social dollars to the firm. However, we did not find a significant effect for the informational benefit of communities that we expected to observe through increased purchases of products high in preference heterogeneity. Further, we are able to (approximately) document the direct benefit of setting up the community to the firm by reporting the small number of customer participants required to earn a return on this investment. It is important to note that it is not a given that managers will observe such a large direct benefit in other settings and our results must be interpreted in that light. However, besides the direct benefits, there are also many indirect benefits that may be reaped by the firm-operator of an online customer community such as the one we observe in the present research. At a minimum, we hope this research encourages managers to at least consider whether setting up their own online customer community might generate a positive return on investment, and provides guidance in regards to the community attributes they should prioritize to help maximize the community's economic success.

Our analysis suffers from some limitations, primarily due to our data. First, we only examine consumer behavior in a small range of experiential goods categories. Second, our data extends to only fifteen months after the formation of the community, restricting our ability to investigate even longer-term effects on customers and the firm. Third, and as discussed earlier, since we only observe customer-level purchase events in the firm's offline (retail) channel for customers with a loyalty card (as the firm otherwise had no customer-level identifier for in-store purchases), our ability to extrapolate the results to the firm's entire customer base is limited. That said, our supplementary analysis of the social dollar for customers without a loyalty card in the online channel (section § 4.4) reveals that the effect for these non-loyalty customers was even stronger than that of our primary analysis sample. Fourth, given that we do not assign customers to treatment and control groups randomly, we cannot rule out the effect of unobservable attributes (i.e., self-selection) with certainty, although multiple analyses suggest that this is highly unlikely to be an issue. Finally, given that we do not observe customers who also shop for the product categories offered by the firm at its competitors, we cannot pinpoint the source of social dollars precisely (i.e., market growth versus store-switching). We hope that future work might address these limitations.

References

Abadie, A., G. W. Imbens. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica* **74**(1) 235–267.

Algesheimer, R., S. Borle, U. M. Dholakia, S. S. Singh. 2010. The impact of customer community participation on customer behaviors: An empirical investigation. *Marketing Science* **29**(4) 756–769.

Algesheimer, R., U. M. Dholakia, A. Herrmann. 2005. The social influence of brand community: Evidence from european car clubs. *Journal of Marketing* **69**(3) 19–34.

Autor, D. 2003. Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics* **21**(1) 1–42.

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

Bagozzi, R. P., U. M. Dholakia. 2006. Antecedents and purchase consequences of customer participation in small group brand communities. *International Journal of Research in Marketing* **23**(1) 45–61.

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

Bertrand, M., E. Duflo, S. Mullainathan. 2004. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* **119**(1) 249–275.

Blundell, R., A. Duncan, C. Meghir. 1998. Estimating labor supply responses using tax reforms. *Econometrica* **66**(4) 827–861.

Boyd, D. M., N. B. Ellison. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication* **13**(1) 210–230.

Brock, T. C. 1965. Communicator-recipient similarity and decision change. *Journal of Personality and Social Psychology* **36** 650–654.

Bronnenberg, B. J., J. P. Dube, C. F. Mela. 2010. Do digital video recorders influence sales? *Journal of Marketing Research* **47**(6) 998–1010.

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

Brown, S. L., A. Tilton, D. M. Woodside. 2002. Online communities pay. *McKinsey Quarterly* **1** 17.

Chevalier, J., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* **XLIII**(August) 345–354.

Choi, K. 1990. The semiparametric estimation of the sample selection model using series expansions and the propensity score. *Unpublished dissertation* **Chicago: University of Chicago.**

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

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

Dholakia, U. M., S. Vianello. 2009. Effective brand community management: Lessons from customer enthusiasts. Working paper.

Eagly, A. H., W. Wood, S. Chaiken. 1978. Causal inferences about communicators and their effect on opinion change. *Journal of Personality and Social Psychology* **36**(4) 424–435.

Eddy, N. 2012. Social media revenue nears $17 billion in 2012: Gartner. URL `http://www.eweek.com/c/a/IT-Management/Social-Media-Revenue-Nears-17-Billion-in-2012-Gartner-890034/`.

Feick, L., R. A. Higie. 1992. The effects of preference heterogeneity and source characteristics on ad processing and judgements about endorsers. *Journal of Advertising* **21**(2) 9–24.

Forrester. 2012. Forrester marketing trends: What's hot in interaction marketing. Tech. rep. URL `http://www.comptia.org/Libraries/Colloquium/Marketing_Trends_-_Forrester.sflb.ashx`.

Fournier, S., L. Lee. 2009. Getting brand communities right. Tech. Rep. 87 (4), Harvard Business Review.

Gershoff, A., M. Broniarczyk, P. West. 2001. Recommendation or evaluation? task sensitivity in information source selection. *Journal of Consumer Research* **28**(3) 418–438.

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

Goldfarb, A., C. Tucker. 2011. Search engine advertising: Channel substitution when pricing ads to context. *Management Science* **57**(3) 458–470.

Gruner, R. L., C. Homburg, B. A. Lukas. 2014. Firm-hosted online brand communities and new product success. *Journal of the Academy of Marketing Science* **42**(1) 29–48.

Hagel, J., A. Armstrong. 1997. *Net gain expanding markets through virtual communities*. Harvard Business School Press, Boston.

Hartmann, W., H. S. Nair, S. Narayanan. 2011. Identifying causal marketing mix effects using a regression discontinuity design. *Marketing Science* **30**(6) 1079–1097.

Heckman, J., H. Ichimura, J. Smith, P. Todd. 1998a. Characterizing selection bias using experimental data. *Econometrica* **66**(5) 1017–1098.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* 153–161.

Heckman, J. J., H. Ichimura, P. Todd. 1998b. Matching as an econometric evaluation estimator. *The Review of Economic Studies* **65**(2) 261–294.

Holland, J., S. Baker. 2001. Customer participation in creating site brand loyalty. *Journal of Interactive Marketing* **15**(4).

Imbens, G. W., T. Lemieux. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* **142**(2) 615–635.

Interbrand. 2011. Best global brands 2011. Tech. rep. URL `http://www.interbrand.com/en/best-global-brands/Best-Global-Brands-2011.aspx`.

Jang, H., L. Olfman, I. Ko, J. Koh, K. Kim. 2008. The influence of on-line brand community characteristics on community commitment and brand loyalty. *International Journal of Electronic Commerce* **12**(3) 57–80.

Kannan, P. K., A.-M. Chang, A. B. Whinston. 2000. Electronic communities in e-business: Their role and issues. *Information Systems Frontiers* **1**(4) 415–426.

Kozinets, R. V. 1999. E-tribalized marketing? the strategic implications of virtual communities of consumption. *European Management Journal* **17**(3) 252–264.

Kozinets, R. V. 2002. The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research* **39**(1) 61–72.

Laroche, M., M. R. Habibi, M.-O. Richard, R. Sankaranarayanan. 2012. The effects of social media based brand communities on brand community markers, value creation practices, brand trust and brand loyalty. *Computers in Human Behavior* **28**(5) 1755–1767.

Lemke, D., W. Reed. 2001. War and rivalry among great powers. *American Journal of Political Science* **45**(2) 457–469.

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

Mathwick, C., C. Wiertz, K. De Ruyter. 2008. Social capital production in a virtual p3 community. *Journal of Consumer Research* **34**(6) 832–849.

McAlexander, J. H., J. W. Schouten, H. F. Koenig. 2002. Building brand community. *Journal of Marketing* **66**(1) 38–54.

McPherson, M., L. Smith-Lovin, J. M. Cook. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* **27** 415–444.

McWilliam, G. 2000. Building stronger brands through online communities. *Sloan Management Review* **41**(Spring) 43–54.

Muniz, A. J., T. O'Guinn. 2001. Brand community. *Journal of Consumer Research* **27**(3) 412–432.

Paul, S. 2012. CEOs are finally warming up to social media. URL `http://mashable.com/2012/05/24/ibm-ceo-social-media/`.

Porter, C. E., N. Donthu. 2008. Cultivating trust and harvesting value in virtual communities. *Management Science* **54**(1) 113–128.

Puhani, P. 2000. The heckman correction for sample selection and its critique. *Journal of Economic Surveys* **14**(1) 53–68.

Rheingold, H. 1993. *The Virtual Community: Homesteading on the Electronic Frontier*. Addison-Wesley., MA.

Rosenbaum, P. R., D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**(1) 41–55.

Schau, H. J., A. M. Muniz Jr, E. J. Arnould. 2009. How brand community practices create value. *Journal of Marketing* **73**(5) 30–51.

Schlosser, A. 2005. Posting versus lurking: Communicating in a multiple audience context. *Journal of Consumer Research* **32**(2) 260–265.

Schniederjans, D., E. S. Cao, M. Schniederjans. 2012. Enhancing financial performance with social media: An impression management perspective. *Decision Support Systems* **in press**.

Shaver, J. M. 1998. Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Management Science* **44**(4) 571–585.

Shriver, S. K., H. S. Nair, R. Hofstetter. 2013. Social ties and user-generated content: Evidence from an online social network. *Management Science* **59**(6) 1425–1443.

Srinivasan, S. S., R. Anderson, K. Ponnavolu. 2002. Customer loyalty in e-commerce: an exploration of its antecedents and consequences. *Journal of Retailing* **78** 41–50.

Thompson, T. J., J. Hertzberg, M. Sullivan. 2011. Social media and its associated risks. *Grant Thornton LLP White Paper* URL `http://insights.grantthornton.ca/i/137039`.

Verbeke, Molenberghs. 2000. *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York, NY.

Vranica, S., S. Raice. 2012. The big doubt over facebook. URL `http://www.wsj.com/news/articles/SB10001424052702304868004577378122958515302`.

Wallace, D., S. B. Green. 2001. *Analysis of Repeated Meausres Designs with Linear Mixed Models*, vol. Modeling Intraindividual Variability with Repeated Measures Data. D. s. moskowitz, s. l. hershberger (eds) ed. Psychology Press, Philadelphia, PA.

Wansink, B., R. J. Kent, S. J. Hoch. 1998. An anchoring and adjustment model of purchase quantity decisions. *Journal of Marketing Research* **35**(1) 71–81.

Watts, D. J., P. S. Dodds, M. E. J. Newman. 2002. Identity and search in social networks. *Science* **296**(5571) 1302–1305.

Williams, R. L., J. Cothrel. 2000. Four smart ways to run online communities. *Sloan Management Review* .

Wirtz, J., A. d. Ambtman, J. Bloemer, C. Horvath, B. Ramaseshan, J. v. d. Klundert, Z. G. Canli, J. Kandampully. 2013. Managing brands and customer engagement in online brand communities. *Journal of Service Management* **24**(3) 223–244.

Zaglia, M. E. 2013. Brand communities embedded in social networks. *Journal of Business Research* **66**(2) 216–223.

Zhu, J., U. Dholakia, X. Chen, R. Algesheimer. 2012. Does online community participation foster risky financial behavior? *Journal of Marketing Research* **XLIX**(June).

Tables

	Control	Treatment	t-stat
T1: 15 Months Pre			
Total Spend	511.38	489.73	-1.12
Average Purchase	49.82	46.24	-3.61***
Purchase Frequency	11.43	11.90	1.12
T2: 15 Months Post			
Total Spend	571.32	676.69	4.25***
Average Purchase	48.05	46.48	-1.80*
Purchase Frequency	12.25	15.60	6.62***
Observations	1255	7909	

*** $p<0.001$, ** $p<0.01$, *$p<0.05$

Table 1: Purchase Statistics by Group - Both Periods

Variable	Control	Treatment	t-stat
% Female[A]	70	73	2.01*
Tenure at launch (months)	38.31	37.98	0.33
Average Household Size[B]	2.92	2.94	2.94**
Median Family Income (x1000$)[B]	70.04	69.29	-1.65
% with Computer Access[B]	74.35	74.09	-1.99*
Education Spending ($)[B]	1367.3	1375.1	0.78
T1 (Pre) Quarter Spend			
Q1	66.68	64.76	-0.41
Q2	121.42	112.26	-1.61
Q3	113.02	107.17	-1.19
Q4	109.23	106.01	-0.65
Q5	101.37	99.96	-0.30
T2 (Post) Quarter Spend			
Q6	163.49	181.28	2.29*
Q7	99.81	119.76	3.91***
Q8	93.54	113.04	3.30***
Q9	85.64	110.05	4.94***
Q10	128.50	152.12	3.31***

*** $p<0.001$, ** $p<0.01$, *$p<0.05$, $^{+}p<0.1$

[A] Gender inferred for 82% of sample using a standard “genderizer” database.

[B] County-level statistics.

Table 2: Summary Statistics by Group

§	Analysis	Objective
4.1	Baseline OLS expenditure regression	Quantifying our treatment effect of interest
4.1	Expenditure regression including demographic controls	Controlling for observed customer heterogeneity
4.2	Expenditure regression with:	Controlling for potential biases due to treatment self-selection
4.2.1	a) Selection on observables – Matching estimators	
4.2.2	b) Selection on unobservables – Heckman/Semi-parametric selection models	
	Expenditure regression using a subset of the treatment as the "control" group	
	Regression-discontinuity style analysis	
Online Appendix	Rosenbaum bounds approach	
	RCR (relative-correlation restrictions) approach	
4.3	Testing for the validity of alternate explanations for Social Dollars:	Assessing whether competing alternative explanations can explain our results
4.3.1	The Role of Outliers	
4.3.2	The Novelty Effect	
4.3.3	Differential Trends in Customer Loyalty	
4.3.4	Outlier Months	
4.3.5	Joining for a Reason Unrelated to the Online Community	
4.4	Robustness Checks vis-a-vis Sample Construction	Testing for the impact of relaxing the 'Loyalty card holder only' constraint and allowing for differential entry/exit

Table 3: Overview of Analyses

	(1)	(2)	(3)	(4)
	OLS		Kernel	Heckman
	Two-periods	Monthly	matching	selection♣
β_3(Social Dollar)	\$ 127.01	\$ 8.07	\$ 142.09	\$ 221.98
	(14.43)	(0.92)	(15.56)	(75.36)
	18.77%	18.34%	20.99%	32.82%

♣ - ignoring the non-significance of the ρ-parameter.
SEs appear in parentheses below estimates (SEs are clustered at the user level for the monthly OLS model). Bootstrapped SEs with 100 replications are reported for the matching estimator.

Table 5: Expenditure regression results assessing Selection

	(1a)	(1b)	(2)	(3)	(4)	(5)	(6)
	Total Spend (Online + Retail)	Total Spend (Online + Retail)	Average Purchase	Purchase Frequency	Proportion Online	Online Spend	Retail Spend
β_1 (Treatment)	-21.65	12.09	-3.57**	0.47	-0.04**	-14.95	-6.69
	(22.20)	(25.46)	(0.93)	(0.47)	(0.01)	(11.78)	(17.91)
β_2 (Post period)	59.94*	33.53	-1.76	0.83	-0.07***	26.96^{+}	32.98
	(29.17)	(33.31)	(1.23)	(0.61)	(0.01)	(15.48)	(23.53)
β_3 (Social Dollar)	127.01***	148.54***	2.01	2.87***	0.130***	87.79***	39.23
	(31.39)	(35.96)	(1.32)	(0.66)	(0.01)	(16.66)	(25.32)
Demographics							
% Female^{A}		33.81*					
		(14.13)					
Tenure (months)		4.80***					
		(0.20)					
Average Household Size^{B}		-17.62					
		(29.83)					
Median Family Income (x1000\$)^{B}		1.4e-3***					
		(4.9e-4)					
% with Computer Access^{B}		1323.8***					
		(306.6)					
Education Spending (\$)^{B}		-0.16***					
		(0.05)					
Intercept	511.38***	-513.34**	49.82***	11.42***	0.39***	137.78**	373.61***
	(20.63)	(192.81)	(0.87)	(0.43)	(0.01)	(10.94)	(16.64)
Observations	18328	13458	18328	18328	18328	18328	18328

Standard errors appear in parentheses below estimates.

*** p<0.001, ** p<0.01, *p<0.05, $^{+}$p<0.1

^{A} Gender inferred for 82% of sample using a standard "genderizer" database.

^{B} County-level statistics.

Table 4: Treatment Effect on Total Spend, Avg. Basket Size and Purchase Frequency and Channel Breakup

	(1)	(2)	(3)	(4)
	Quartile 1	Quartile 2	Quartile 3	Quartile 4
T1 Total Purchases				
Mean	85.05	241.35	460.93	1181.50
Median	83.70	241.35	454.03	928.27
SD	42.51	50.45	82.23	946.86
β_3(Social Dollar)	101.29***	143.30***	143.41***	120.25*
	(19.44)	(21.60)	(24.55)	(57.20)
Proportion of T2 Purchases	0.29	0.32	0.23	0.09
Observations	2291	2291	2291	2291

Standard errors appear in parentheses below estimates.

*** p<0.001, ** p<0.01, *p<0.05

Table 7: Panel regression by Pre-Period (T1) Total Purchase Volume Quartile

Distance from Launch	One Quarter	Two Quarters	Three Quarters	Four Quarters	Five Quarters
β_3 (Social Dollar)	21.62*	44.25***	64.66***	101.87***	127.01***
	(9.06)	(9.40)	(11.09)	(14.43)	(17.24)
Observations	6259	7842	8627	9001	9164

DV=Customer Expenditure. Standard errors appear in parentheses below estimates.

*** p<0.001, ** p<0.01, *p<0.05

Table 8: Temporal Persistence regression

Window	Obs.	\$ Pre	\$ Post	t-stat
		0-1 week		
1 day	7865	4.88	17.25	20.31***
2 days	7859	6.77	20.26	20.81***
3 days	7850	8.42	22.23	20.65***
4 days	7846	10.17	23.91	19.91***
5 days	7844	11.87	25.19	18.44***
6 days	7841	13.36	26.71	17.45***
7 days	7839	15.07	27.86	16.64***
		2-4 weeks		
14 days	7807	21.86	34.81	15.86***
21 days	7760	27.28	39.81	14.64***
28 days	7698	31.69	44.20	13.81***

*** p<0.001, ** p<0.01, *p<0.05

Table 9: "Regression-Discontinuity" Style comparisons of Treatment Group Means

(A) Join Cohort Group Comparisons

	(1)	(2)	(3)	(4)
Treatment Join Quarter Cohort	Q6	Q7	Q8	Q9
T1 Quarterly Mean Spend	72.31	94.04	92.47	92.39
t-stat for T1 Qtrly. Mean versus				
T2 Q7 Spend	13.09***			
T2 Q8 Spend	10.26***	2.46*		
T2 Q9 Spend	11.24***	1.49	2.40*	
T2 Q10 Spend	23.08***	7.42***	5.69***	8.70***

(B) Join Cohort analysis

	(1)	(2)	(3)	(4)
Treatment Join Quarter Cohort	Q6	Q7	Q8	Q9
Control Join Quarter Cohort	Q7-Q10	Q8-Q10	Q9-Q10	Q10
Pre/Post Period Definition	Q5/Q6	Q4-5/Q6-7	Q3-5/Q6-8	Q2-5/Q6-9
β_3 (Social Dollar)	39.76***	41.75***	25.70*	56.00**
	(4.92)	(7.85)	(11.86)	(19.11)
Observations	7909	4406	3069	1975

Standard errors appear in parantheses below estimates.

*** p<0.001, ** p<0.01, *p<0.05

Table 6: "Join Quarter" Cohort Group Comparisons/"Join Quarter" cohort regression

Parameter	Est.
Treatment	1.05
	(0.68)
Post	-3.24***
	(0.89)
Treatment × Post (Social Dollar)	9.40***
	(0.96)
Month_2	-4.47***
	(0.82)
Month_3	-1.97*
	(0.82)
Month_4	-6.03***
	(0.82)
Month_5	-3.45***
	(0.82)
Month_6	-1.73*
	(0.82)
Month_7	-6.94***
	(0.82)
Month_8	-5.61***
	(0.82)
Month_9	-9.00***
	(0.82)
Month_10	-10.86***
	(0.82)
Month_11	-0.22
	(0.82)
Month_12	29.29***
	(0.82)
Avg HHL size	-1.86*
	(0.61)
Med fam income	0.00***
	(0.00)
Education spending	-0.01***
	(0.00)
Computer access	75.09***
	(7.15)
Gender (F=1)	2.11***
	(0.34)
Loyalty Tenure	0.32***
	(0.01)
Intercept (Month_1)	-21.49***
	(4.19)

*** $p<0.001$, ** $p<0.01$, *$p<0.05$, $^{+}$ $p<0.1$; Month_2=Feb, Month_3=March etc. SEs appear in parentheses below estimates.

Table 10: Regression with month fixed effects

Month	Est.	Social Dollar (%)
January	10.39**	24.25
	(3.29)	
February	8.20**	21.94
	(3.02)	
March	10.42**	26.31
	(3.35)	
April	11.04**	31.15
	(3.48)	
May	6.11^{+}	15.80
	(3.26)	
June	8.90*	22.85
	(3.25)	
July	5.59*	15.56
	(2.66)	
August	11.91***	30.21
	(3.13)	
September	7.41**	21.34
	(2.68)	
October	8.93**	23.90
	(2.81)	
November	9.88**	19.52
	(3.67)	
December	14.02**	15.47
	(4.80)	

*** $p<0.001$, ** $p<0.01$, *$p<0.05$, $^{+}$ $p<0.1$;
DV=Customer Expenditure; SEs appear in parantheses below estimates.

Table 11: Panel Regression model for each month

	Model 1	Model 2
n = 5,544 (T=4606, C=938)	Est.	Est.
Treatment	37.83**	38.27**
	(112.34)	(13.77)
Time	141.05***	141.04***
	(13.87)	(17.62)
Treatment × Time	47.27*	126.12***
	(17.76)	(23.34)
Treatment × Time × LURK		-111.35***
		(14.12)
Treatment × Time × HETE		-1.07
		(8.96)
Treatment × Time × TIES		3.96***
		(0.71)
Gender (F)	-1.84	-5.71
	(8.45)	(8.42)
Avg HHL size	4.19	0.84
	(17.35)	(17.24)
Med. fam. income	0.00	0.00
	(0.00)	(0.00)
Comp. Access (%)	557.19**	605.84**
	(179.59)	(178.40)
Education Spending	-0.08**	-0.09**
	(0.03)	(0.03)
Loyalty tenure (months)	0.90***	0.93***
	(0.12)	(0.12)
Intercept	-223.24*	-243.60*
	(112.34)	(111.59)

*** p < .001, ** p < .01, * p < .05, + p < .10; SEs appear in parentheses below estimates; DV = Expenditure on books.

Table 12: Moderating the Social Dollar