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

Puneet Manchanda
Grant Packard
Adithya Pattabhiramaiah[1]

April 2010

This version: January 2012

[1] Puneet Manchanda is Isadore and Leon Winkelman Professor and Professor of Marketing, Grant Packard and Adithya Pattabhiramaiah are doctoral candidates, all at the Ross School of Business at the University of Michigan. The authors would like to thank Utpal Dholakia, Pete Fader, Ron Goetler, Anja Lambrecht, S. Sriram, Melanie Zaglia and seminar participants at the University of California, Davis, University of North Carolina, University of Rochester, participants at the 2011 ART Forum, 2010 and 2009 Marketing Science conferences, the 2010 ISMS Practice Conference and the 2010 Marketing Dynamics Conference for valuable feedback and an anonymous firm for providing the data. All correspondence may be addressed to the authors at pmanchan@umich.edu, gpackard@umich.edu and adithyap@umich.edu. The standard disclaimer applies. All correspondence by regular mail should be sent to the first author at Ross School of Business, University of Michigan, Ann Arbor, MI 48109, USA.

## Abstract

Many firms operate consumer-centered social networks or "communities" online. This is motivated by the belief that consumers who join the community become more engaged with the firm, and as a result, increase their economic activity with the firm. We label the revenue generated from the increased engagement as a result of joining the community as "social dollars." This paper tests for the existence and magnitude of social dollars via a difference-in-differences estimator using data from a multi-channel entertainment products retailer that launched an online community. We find that 19% of the post-launch revenue from community customers can be attributed to their joining the community. This result is robust across a variety of tests. In addition, social dollars persist over time, arise in both online and offline channels and affect all product categories sold by the firm. The analysis of community data reveals that social connections – the number and importance of friend ties – and interactions – personal page displays – are positively linked to social dollars.


Keywords: *Virtual Communities, Online Communities, Online Social Interactions, Difference-in-Differences Estimation, User-Generated Content.*

INTRODUCTION

The dramatic rise of social media has seen marketers quickly adapting their strategies to leverage consumer involvement enabled by these media. Americans now spend more of their online time at social network and blog sites (23%) than in any other activity (Nielsen 2011). Firms are following the consumer – over the next five years, social media marketing is expected to represent nearly 20% of marketing budgets (Moorman 2010). A major component of social media marketing is the use of online social networks and brand or product-centered communities to build consumer engagement. For example, a survey of Chief Marketing Officers indicates that they perceive online social networks as the single most promising media opportunity in the coming years, with 60% planning to increase expenditure(s) in this area (Forbes Insights 2009). Analysts project U.S. marketers will spend $1.5 to $2 billion on social media in 2011, with this figure growing to nearly $5 billion by 2016 (Forrester Research Inc. 2011).

While the major share of firm and media attention has focused on marketing via third-party online social networks such as Facebook, many firms have made the choice to build their own networks/communities. Examples include consumer-centered online communities operated by firms such as Amazon, Buy.com, Disney, IKEA, Kraft Foods, Lego, and Procter & Gamble.[2] Depending on the definition of the attributes of a firm-sponsored online community,[3] between 25 and 50 of the top 100 global brands (Interbrand 2011) host their own such community (i.e., not hosted on a third-party network or site). Many more firms appear likely to follow suit—American CEOs and CFOs place online social media/networks in their top five planned technology investments for 2011 (PricewaterhouseCoopers 2011).

The focus of these firm-sponsored social networks is the development of their own "virtual" or online 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 (Bagozzi and Dholakia 2002, Porter and Donthu 2008, Reingold 1993). Common social practices or interactions enabled by online communities include the sharing of product experiences, recommendations or advice, planning for product-centered social interactions (e.g. online or "real-world" group events), and in some cases, the maintenance of a personal profile page intended to

[2] Examples can be found at http://www.amazon.com/gp/help/customer/display.html?nodeId=200280960 (Amazon), http://yub.com (Buy.com), http://disney.go.com/mydisney (Disney), http://www.ikeafans.com (IKEA), www.kraftrecipes.com/community (Kraft Foods), http://mindstorms.lego.com/community (Lego) and http://www.vocalpoint.com (Procter and Gamble).

[3] We looked at three attributes to determine if a firm hosts its own online community. These were (1) the ability for consumers to create and maintain a personal profile page, (2) the ability to create and maintain friend ties, and (3) the ability to interact with friend ties via 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.

convey a user's personality, interests and/or status in the community (Brown, Broderick and Lee 2007, Dholakia, Bagozzi and Pearo 2004, Muniz and O'Guinn 2001, Schau, Muniz and Arnould 2009).

Firms that set up such communities expect to increase customer engagement and loyalty (Fournier and Lee 2009, Porter and Donthu 2008, Williams and Cothrel 2000). The expectation is that this increased engagement and/or loyalty will lead to better economic outcomes for the brand, as exemplified by predictions that firm sponsors of online communities will be "richly rewarded with peerless customer loyalty and impressive economic returns" (Hagel and Armstrong 1997, p. 2). Using surveys and self-report data, previous academic research has reported an increase in purchase intention among online community members (Algesheimer et al. 2005, Porter and Donthu 2008). An analysis of major global brands also found a significant positive correlation between the extent of a firm's consumer engagement through online social media and both revenue and margin growth (Wetpaint/Altimeter 2009). Other researchers have shown that enabling consumer interactions in a firm-sponsored online community is one of seven factors linked to increased future purchase intention and willingness to pay a price premium with online retailers (Srinivasan, Anderson and Ponnavolu 2002).

Two more recent studies have examined the consequences of online community membership using behavioral data. Zhu et al. (2011) found that community membership was linked to greater financial risk-taking as observed in lending (Prosper.com) and bidding (eBay Germany) behaviors. A subsequent lab experiment supported arguments that this behavior was motivated by the perception that other community members will aid them should difficult situations arise in the future. Algesheimer et al. (2010) examined the behavioral consequences of online customer community membership at eBay Germany. They found that eBay bidders and sellers became more selective and conservative in their auction behavior as a result of online community participation, leading to null or negative effects of community participation on individual-level bidding volume, product listings, average amount spent by buyers, and revenue earned by sellers. A unique aspect of the online 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. As we discuss later, 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 direct economic consequences of online communities to the firm or brand. This gap in the literature had been noticed by other researchers as well (e.g., Balasubramanian and Mahajan 2001). Further, while researchers have identified important behaviors or "practices" related to the sustainment of an active online customer or brand community (e.g., Schau et al. 2009), there is little real behavioral evidence regarding which—if any—of these are linked to actual economic outcomes.

This paper addresses these questions and gaps in the literature directly by analyzing the effect of consumer membership in a firm-sponsored online community on actual expenditure (by the member) on the firm's products. We label this incremental expenditure "social dollars." In other words, social dollars represent customer expenditure (firm revenue) that can be attributed directly to the social behavior that consumers engage in by joining the community. We do this by leveraging outcomes of a natural experiment in a retail setting. Specifically, we obtained novel data from a multi-category and multi-channel retailer of experiential products (books, movies, music) that set up its own online community. Our data span consumer activity both before and after the formation of the community. Using difference-in-differences estimation, we pin down the revenue change for those consumers that participated in the firm's online community versus a similar "control" group that did not.

Our data and research approach adds to the literature on online communities on multiple dimensions. First, and perhaps most important, we use actual behavioral data to investigate the economic impact of setting up and operating such a community. Second, the availability of consumer panel data before and after the formation of the community allows us to control for selection effects. In other words, we are able to account for the possibility that consumers who join the community are systematically different from consumers that do not. Third, the long time series of our data allows us to investigate whether the change in purchase behavior that results from joining the online community is a short-term effect driven by the novelty of the event (the formation of the community). Fourth, given the multi-channel and multi-category nature of our data, we are able to test whether the formation of the online community affects behavior differentially across channels and product categories. Finally, while the main objective of this paper is to document the existence and magnitude of the social dollars, we also use the observed actions and interactions among community members to isolate behaviors that are correlated with higher expenditures. This allows us to speculate on the source of the social dollars observed.

Our results suggest that social dollars account for 19% of the revenue from members post their joining the community. In other words, this is the incremental revenue that the firm obtains from customers who join the online community, over and above their pre-existing purchase behavior with the firm (and relative to a control group). This magnitude is economically significant 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 estimate of the social dollars to multiple robustness checks and demonstrate that it is indeed robust. For example, the estimate is robust to analysis group composition, to the timing of joining the community (after the community has been launched) and to differences in sample size across comparison groups. An interesting finding is that we do not find evidence of channel cannibalization—we observe positive social dollar effects in both online and "offline" retail channels. Finally, we find that both the quantity and quality of engagement with friend ties established in the community's social network matters. Customers that have ties with more important and/or prominent customers are likely to be more economically beneficial for the firm.

The rest of the paper is organized as follows. We first discuss the research setting and the data. The subsequent sections describe our modeling and analysis strategy, including detailed robustness checks. We next look at both the behavior of community members and their structural position in the community's social network of friend ties to uncover behaviors related to higher expenditure growth. Finally, we discuss the managerial implications of our findings and then conclude.

## RESEARCH SETTING AND DATA

Our data come from a large North American retailer of entertainment and informational media products (e.g. books, movies, music).[4] 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 community in September 2007. The formation and existence of this community was communicated via mass marketing to consumers and current customers. This advertising comprised billboards in stores, banner advertising on the firm's website, print advertising in newspapers, and the firm's house opt-in email list. Advertising announcing the launch of the online 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.

The structure of the community can best be summarized as a Facebook-like website. Upon sign-up, community members manage a profile page that enables the social presentation of a variety of personal and product-related content. Members post textual and graphical information about themselves and their product preferences and/or recommendations, graphically display a variety of personal and firm product items, post on chat boards and establish friend ties. Beyond the profile page, members may contribute a variety of user-generated content (UGC) for the consumption of others who are either within their own social network of friend ties (i.e. "private" content) or for the online community at large (i.e. "public" content). These options include targeted (peer-to-peer) product recommendations, the initiation and management of exclusive (by-invitation) special

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

interest groups (e.g. "Vampire Movie Lovers Club"), the publication of "top 10 lists" and product reviews that can be read by members of the community as well as visitors to the website.

We should note at this stage that the definition of online customer community we rely on nests within it the definition of "brand community" (Muniz and O'Guinn 2001). Muniz and O'Guinn define a brand community as a specialized, non-geographically bound community that engages in a structured set of social interactions pertaining to their admiration for branded goods or services in either face-to-face or computer-mediated environments. While this definition does not impose any constraints that the brand sponsoring the community should be identical to the brand name of the good(s) purchased, in general, research on brand communities has focused on goods sold under a single brand name. In contrast, while many of the social interactions in the community we observe pertain to the retailer's brand (e.g., admiration of a particular retail store location), the primary topics around which community members engage are authors or artists (e.g. Nora Roberts, Coldplay), products (e.g. specific book titles or series), and genres (e.g., comedy movies). The distinction that these brands do not carry the same brand as the retailer is not crucial to our analysis of consumer behavior in online communities. We discuss this in detail later in the paper.

The data used in our analysis were extracted in January 2009. Using an "*n*th-select" random sampling procedure, the firm generated records for 26,624 community members (from about 260,000 such members) for analysis. The firm provided us 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.[5] Each record in the transactional data includes the date of the customer's first purchase, his/her first name, his/her geographic location (in one of the firm's four operating regions) and details on each purchase event. Each purchase event indicates the channel and date of purchase, the specific product (SKU) purchased, revenue net of any standing or promotional discounts received for each product within the transaction, and each SKU's product category classification.

The community data we observe includes the date members joined the community, friend ties they have established in the social network and a selection of the social behaviors in which they may participate, such as posting UGC on their profile page, the website at large, or directly to their network ties. Specifically, we observe the friend and group ties possessed by each member, and the volume of several types of UGC—peer-to-peer product recommendations, product reviews written,

[5] The firm tracks 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 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 total sales revenue.

top 10 lists published, and the number of products (e.g., book cover graphics) displayed on their personal profile page. The time stamp and content for each of these variables is also observed.

There was a difference of fifteen months between the data pull (January 2009) and the formation of the online community (September 2007). We therefore also asked the firm to provide fifteen months of data *before* the launch of the community for these customers. 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), for the full 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 in order 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 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.[6] Of these accounts, 2,352 were also loyalty card holders, which provides full visibility of their purchase behavior with the firm (online and 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-Q5) and the period after the launch of the community as T2 ("post-community," October 2007 to January 2009; quarters denoted as Q6-Q10). Note that while the exogenous change (the launch of the brand 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 when we discuss our analysis and results. Taken from the full sample described above, our analysis sample includes customers (who were also loyalty card holders) for whom we observe behavior across both sales channels 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 potentially differential entry and exit patterns in the treatment and control groups. This has the added benefit of making our results as conservative as possible (discussed in detail in the next section). The application of this criterion limits the analysis 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 analysis groups respectively. As is to be expected, the 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. However, the total expenditure between the analysis sample for the two groups is not different in the T1 period (see

[6] The size of the full treatment group sample (26,624) is larger than the full control group sample (6,091). This is because the firm was principally interested in our analysis of the treatment group's behavior and wanted us to have the flexibility of being able to create reasonably large sub-samples from this group to investigate specific issues. We show in our robustness checks that this does not impact our results in any meaningful way.

Table 1). Note also that in one of our robustness checks, we relax the use of this criterion to test for its impact on our findings.

## EXISTENCE AND MAGNITUDE OF THE SOCIAL DOLLAR

Our objective in this section is to estimate the overall effect of customer participation in an online community on aggregate purchase behavior. Our analysis strategy is as follows. First, we describe our modeling approach to estimate the magnitude of social dollars (if they exist). As noted earlier, we exploit the exogenous formation of the online community and availability of a control group. However, given that our data come from a natural experiment (and not a field experiment), we need to ensure that our findings are not confounded with other explanations. Chief among these is self-selection (or non-random assignment to treatment and control groups). We run a series of robustness checks to first control for self-selection using observables – demographics, spending trends etc. We then also run another series of robustness checks to control for self-selection using unobservables. Having demonstrated the robustness of social dollars to self-selection, we next examine other possible explanations. These include the novelty effect i.e., social dollars may arise because the online community is new and different but may then dissipate over time. We finally examine the possibility that social dollars may arise because of changes in the product mix purchased by treatment and control groups and/or channel switching between online and offline channels. In the section that follows this one, we examine the behavioral sources of the social dollar, which we describe as the extent of a customer's connections with and interactions in the online community.

### *Modeling Approach*

Given our dataset, we use a difference-in-differences (DD) estimator to help us obtain the magnitude of the social dollars. The use of this econometric estimator helps us rule out alternative explanations such as selection (via the observation of transactions of the same customers before and after) and exogenous factors (via the control group). We aggregate the detailed purchase data to these two periods rather than leveraging a more fragmented time series form to mitigate potential serial correlation and grouped error term effects (Bertrand, Duflo and Mullainathan 2004). The specification we estimate is

$$\Delta R_{ig} = b_0 + b_1 I_g + e_{ig} \qquad (1)$$

where $\Delta R_{ig}$ is the difference between the post and pre period values in the outcome of interest (the total customer revenue for most of our analyses) for consumer $i$ in group $g$, $b_0$ represents baseline purchase behavior, fixed effect $I_g$ captures possible differences between the treatment and control groups prior to the community introduction and $e_{ig}$ is the error term. Note that by focusing on the difference in revenue over the two periods, we control for the influence of aggregate factors that could cause changes in customer *i*'s expenditure from time period T1 to T2 in the absence of the online community, on our result (see Chevalier and Mayzlin, 2006 for a similar specification). The coefficient we describe as the "social dollar" is $b_1$, 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.

As mentioned previously, differential group entry and/or exit represents significant threats to the assumption of no sample composition changes across groups in DD estimation (Blundell, Duncan and Meghir 1998). Our analysis sample, based on the criterion that each consumer transact at least once in T1 *and* T2, ensures that we include only "active" customers in both groups. This resolves uncertainty in the causes of any differential entry or exit by fixing group composition over time. As we discuss later, this analysis sample also provides us the most conservative estimate of the treatment effect of the online community.

In order to carry out our basic DD 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, thus invalidating our identification strategy. It is also crucial to note that our current 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. In a later section on robustness checks, we exploit this feature of our data (that not everybody in the treatment group joins at the same time) to also rule out selection on unobservables.

A quick look at the data descriptives (Table 1) shows that there isn't much difference across the control and treatment groups in total mean expenditure per customer, average purchase size (per order) and the number of orders per customer in the fifteen month period before the community was launched. After the community was launched however, the total mean expenditure and the number of orders goes up for both the treatment and control groups while the average purchase order changes very little for both groups.

--------------------------------------------

Insert Table 1 about here

--------------------------------------------

*Results*

We now turn to our basic DD analysis. As shown in Table 2, our results show that the social dollars, as represented by $b_1$ (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 \$541.35 (in T2) for the treatment group, social dollars are estimated to be 19% of all expenditure post the launch of the community. To the best of our knowledge, this is the first empirical result documenting that an online community can lead to a direct increase in total customer-level revenue for its firm sponsor.

-------------------------------------------

Insert Table 2 about here

-------------------------------------------

Two aspects of this result are noteworthy. First, as noted earlier, the brand name under which this 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 brand community in the context of Lego's toy building-brick product only. Therefore, if anything, we expect that the social dollars effect in an assorted product online community would be *weaker* than that of a single product brand community. Second, our results are quite different from the two other studies that quantified changes in customer behavior as a consequence of joining an online community (Zhu et al. 2011, Algesheimer et al. 2010). The most direct comparison is with the latter study, which found that participation in eBay Germany's online 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 marketing behavior, leading to fewer listings. We propose that the firm-sponsored online social network we investigate may be more representatitive of firm-sponsored online 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 or loan interest rates. 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 community).

Applying our DD analysis to each of order size and order frequency as dependent variables, we find a significant increase in both due to membership in the online brand community (Table 2). The increase in average purchase size, while statistically significant, is relatively small in magnitude

(4.8%). 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 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 firm's website, as well as the conversion rate per visit (Brown, Tilton and Woodside 2002; Holland and Baker 2001; McWilliam 2000), leading to an increase in order frequency.

*Robustness Checks*

We now demonstrate the robustness of our findings. We first seek to replicate our results after accounting for several observables. These include minor demographic differences across analysis groups, cross-sectional differences in expenditures prior to the introduction of the community, the possibility of differential temporal trends across groups, and differences in the sample sizes of the comparison groups. Next, as we do not randomly assign consumers to treatment and control, we conduct three robustness tests to examine the possibility that unobservable differences between the comparison groups give rise to our results.

*Demographics*. First, we compare the treatment and control samples on observables—demographic and past transaction behavior. Recall that we had only first names for each customer in our data. Using a standard "genderizer"[7] list, we were able to infer gender for 82% of the sample. As shown in Table 3, group comparison t-tests indicated that there were slightly fewer males in the treatment than control group with the difference being statistically significant.[8] There were also minor differences in geographic distribution with the treatment group comprising a few less (more) customers drawn from Regions 1 (4).

--------------------------------------------

Insert Table 3 about here

--------------------------------------------

We therefore replicate our main analysis with explicit controls for the minor demographic differences across the treatment and control groups described above. Specifically, to the main specification (equation 1), we add a vector of demographic controls including gender, observed tenure with the firm, and geography (regional location of the customer). For each of these, we add a main effect ($b_2$) and two-way ($b_3$) interaction independently and simultaneously for the demographic variables ($I_d$).

[7] 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).
[8] The gender group comparison is based on the classified names only (82% of analysis sample).

$$\Delta R_{ig} = b_0 + b_1 I_g + b_2 I_d + b_3 I_g I_d + e_{ig} \qquad (2)$$

As shown in Table 4, we observed an interaction of analysis group for the tenure variable only, which reveals a negative relationship such that customers with longer (shorter) tenure show lower (higher) social dollars. This result is noteworthy in that it is inconsistent with a selection explanation that would expect strongly engaged, loyal (i.e. highly tenured) customers to join the community. After accounting for these covariates, we find that the absolute value of the social dollar in our basic DD analysis above is lower than all but one of the social dollar estimates with demographics i.e., it is overall a more conservative estimate. As a proportion, our estimate of social dollars is 19%, which falls within the range of 16-24% when demographics are included as controls. These results suggest that our estimate of the social dollar is robust to minor compositional differences across analysis groups in gender, tenure and geography. Thus, our basic results cannot be driven by these differences.[9]

--------------------------------------------

Insert Table 4 about here

--------------------------------------------

*T1 Expenditure Level.* While it is true that DD estimation is commonly carried out at the group level and within group analysis is usually not performed, it is important for us to ensure that the existence and magnitude of the social dollars is not driven by outliers. In order to check this, we divide our treatment and control groups into expenditure quartiles using their total purchases in T1 as the baseline. We then carry out a separate DD analysis for each quartile. As shown in Table 5, 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 biggest difference is for the middle two quartiles, approximating the “average” customer for the firm. The statistically significant, but relatively weaker effect in the top quartile is driven by the variance in this high purchase volume group. Further, it is also likely that the firm already accounts for a higher share of wallet for these 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

[9] An alternative analysis strategy would be to control for selection using the demographics, followed by the computation of the social dollars via a selection model (Heckman 1979). We attempted this using the observed demographics in the selection equation and the expenditure in T1 in the outcome equation as the independent variables. We found that the model structure was not supported as the coefficient of the Inverse Mills ratio term was not statistically significant (*t-stat* = 0.17). If we ignored the lack of statistical significance for this term and computed the implied social dollars, they account for 34% of all T2 expenditures (a less conservative result than the 19% reported in our DD analysis).

minority of customers, and especially not only by customers who spent heavily with the firm prior to the launch of the community.[10]

--------------------------------------------

Insert Table 5 about here

--------------------------------------------

*T1 Expenditure Trend.* Another factor that could potentially be driving the social dollars estimate is a differential temporal trend in expenditure between the treatment and control groups in T1. To address this concern, we check to see if trends in transaction behavior (total purchases) over time (but within T1) can explain our findings rather than the fact that customers in the treatment group joined the community. For example, it could be the case that while the total expenditure in T1 for the treatment and control groups is not statistically different (as shown in Table 1), there is an increasing trend in the expenditure for the treatment group relative to the control within the T1 period. As time goes by, this 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

$$R_{iq} = X_i\beta + Z_q u + (X_i\beta Z_q u) + \varepsilon_{iq} \tag{3}$$

where $R_{iq}$ is a 5 x 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 the error term ($\varepsilon_{iq}$). 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 ($t(45820) = .50$, $p = .62$). Given the quarterly purchase trend approximates an inverted-U shape (see Table 6), we also specified a model adding a quadratic main effect and interaction for time. The quadratic interaction term was also non-significant ($t(45820) = 1.47$, $p = .14$), failing to support a difference in curvilinear

---

[10] We also carried out another outlier analysis. We trimmed outliers (on total expenditure in T1) that were approximately five standard deviations or more from the mean (we determined this cutoff by examining skewness in a QQ plot). This resulted in the exclusion of 30 customers in the treatment group and 2 customers in the control group. The social dollar estimate without these outliers is \$116.83 (18%) with a significance level of $p < 0.001$.

trends.[11] As an additional test, we present below (Table 6) 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 treatment and control groups after the launch of the community are driven by differential trends in behavior prior to the launch.

--------------------------------------------

Insert Table 6 about here

--------------------------------------------

*Sample Size and Composition*. We next test whether size differences in the treatment and control groups may be impacting our results. As mentioned previously, the firm's research motivations supported the extraction of a larger treatment than control group. To control for differences in sample size, we randomly sampled 1,255 customers from the analysis treatment group (n = 7,909) ten times (with replacement) to precisely match the size of the analysis control group (n = 1,255). We then ran our basic DD analysis on each reduced sample. As shown in Table 7, the statistical significance and magnitude of the social dollar effect is robust to random reduction in the treatment group sample to the same size as the control group.

--------------------------------------------

Insert Table 7 about here

--------------------------------------------

Our final robustness check relaxes the sample inclusion criterion that a customer had to transact (purchase) at least once in T1 and T2 to be included. Recall that our objective in using this criterion was to control for differential entry/exit across comparison groups. The social dollar estimate ($b_1$) when this restriction is relaxed is \$170.96 (SE = 13.18, $p < .001$), or 29% of expenditures post the community. This less conservative estimate of the social dollar is generated because of many community members transacting in T2 (after the community is launched) but not in T1. In short, the difference between this and our basic DD estimate (\$127.01) represents the potential new customer acquisition (entry in T2) benefit that could be attributed to the community. However, in order to keep our identification of social dollars as robust as possible and to report the most conservative magnitude, we restrict our analysis to customers that transacted at least once in each of the two periods.

*Customer-level Unobservables*. So far, we have demonstrated the existence (and magnitude) of the social dollars and also provided some evidence that this finding cannot be explained by

---

[11] 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, $p$ = .51) and quadratic interaction terms (Huynh-Feldt adjusted $F$(2.0, 18740) = 1.76, $p$ = .17).

differences between the treatment and control group due to attributes observed by the analyst (such as demographics and past expenditures). While an experimental design involving random assignment to conditions is ideal, most field applications of DD estimation observe non-random group assigment (e.g., Bertrand, Duflo, and Mullainathan 2004, Donald and Lang 2007, Goldfarb and Tucker 2011a). Given that customers in our data are not randomly assigned to treatment and control groups, it is possible that the two groups could differ on unobservables that drove treatment group members to join the community. While the DD model reduces these selection concerns by design,[12] we pursue additional robustness checks to more thoroughly assess a self-selection explanation for our results. Given that we do 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 brand 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 online communities have a much larger effect on more engaged customers (Algesheimer et al. 2005).

While it is impossible for us to rule all the role of all such unobservables with certainty, we present reasonable evidence that they played no part in our estimation of social dollars.[13] To do this, we exploit a feature that of our data that we had discussed earlier. 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.[14]

Once the community was available to customers, if there is a differential interaction between the availability of the community and the unobservable attribute(s), then a change in transaction behavior should manifest itself even if a (future) member did not yet join the community. We present a series of analyses in which we show that the change in behavior is not driven by *whether* some customers join the community but *when* they choose to join the community. In other words, it is the act of joining and not the mere availability of the treatment (community) that impacts transaction behavior. In fact, the temporal joining data discussed above already suggest that the availability of

[12] Difference-in-differences designs 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).

[13] Note that from the firm's (very pragmatic) point of view, this isn't a serious concern. In the worst case i.e., if systematic differences in one or more unobservable attributes across the treatment and control groups drove our estimate of social dollars, the role of the community can be see as a sorting mechanism that is able to differentiate between similar looking (on observable attributes) customers.

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

the 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 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 definition of control groups for these 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 i.e.,, they joined the community between Q7 and Q10, inclusive (Table 8, column 1). If our prediction (that behavior changes when, not whether, they join) is correct, until these customers actually join the community, we should be able to treat them as control group customers. This analysis can therefore be seen in the same spirit as the falsification analysis carried out in Goldfarb and Tucker (2011b).

We perform the difference-in-differences 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 8, columns 2-4). As shown in the 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.

--------------------------------------------

Insert Table 8 about here

--------------------------------------------

We next focus on a second cohort-based analysis of customers who join the online 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 9) 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 9, nine out of the ten quarterly comparisons in the table are statistically significant. These results also 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.

--------------------------------------------

Insert Table 9 about here

--------------------------------------------

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. In 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, Nair and Narayanan 2011). We examine behavior in the shortest possible window we observe – a day – as well as two, three, four, five, six and seven days.[15] As expected (Table 10), 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 the difference is statistically significant for all of the windows we consider.

--------------------------------------------

Insert Table 10 about here

--------------------------------------------

Overall, the three analyses presented above suggest that it is unlikely that a differential interaction between unobservable attributes and the treatment was the main driver of the social dollars effect. This further reduces the likelihood that there exists an alternative explanation – selection on unobservables – for our findings.

*Temporal, Channel and Category Descriptives*

After demonstrating that the existence and magnitude of social dollars is robust to alternative explanations, we turn our attention towards providing a richer description of the emergence of these dollars. We first examine the possibility that the social dollar is driven by the novelty of the online community. In other words, it is possible that customers respond positively to the community as soon

[15] 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.

as it is set up but then lose interest and revert back to their normal (pre-community) transactional behavior.

To test this, we estimated the social dollar using our DD 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 online 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 11, 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 seems 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.

--------------------------------------------

Insert Table 11 about here

--------------------------------------------

Given that the firm operates both online and offline, a reasonable hypothesis could be that the social dollars arise from differential revenue generation across sales channels. Specifically, the online nature of the community may cause social dollars to be generated more from the online channel or, in an extreme case, to originate entirely from channel switching. In order to investigate this, we first check to see if there is any significant difference in the share of revenue dollars between the online and offline channels before and after the community is formed. We carry out a DD analysis for the proportion of total sales for a given customer that comes from the online channel. We find (Table 12) that the online proportion goes up by 13 points, suggesting that the community does shift purchase behavior towards the online channel.

--------------------------------------------

Insert Table 12 about here

--------------------------------------------

To assess the economic size of this shift, we replicate our basic DD 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 online 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 increases revenues in the offline channel as well. 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 online social networks and communities. This finding thus adds to the literature by documenting an economic measure of positive channel spillover for online social interactions.

--------------------------------------------

Insert Table 13 about here

--------------------------------------------

Finally, we check to see if the social dollars reflect a change in customer shopping baskets. While virtually all the product categories the retailer sells are experiential in nature, it is possible that the community may affect purchases in one category more than the other. Given that the firm is known primarily as a bookseller, and the user-generated content in the community is also dominated by this product category, we might expect that the core good around which social relationships are structured will disproportionately benefit from any economic returns they may offer (Balasubramanian and Mahajan 2001, Muniz and O'Guinn 2001). We therefore break down total sales at the firm into four categories – audio, books, video and other. All four categories show a significant increase in sales (Table 13). The source of the revenue increase is primarily books (84% of social dollars), followed by the catch-all "other" category (7%), then video (6%) and audio (3%) products. However, relative to base sales of each category, books do not see a disproportionate share of the social dollar. The volume of T2 purchases within each category attributable to the brand community is 30.0% (audio), 20.0% (books), 32.9% (video) and 9.9% (other). The "other" product category in the present data is perhaps most ancillary to the firm's product mix, containing mostly gift and impulse items. As these products are not likely a focus of community-centered discussion or content sharing, it is unsurprising that the community effect is smaller here than in the other three categories.

## BEHAVIORAL CORRELATES OF SOCIAL DOLLARS

In the previous sections, we have documented the existense and magnitude of the social dollars (in our context). We have also found that the existence of social dollars is robust i.e., can be attributed to customers joining the online community. Social dollars exist in both channels and across all product categories (sold by the firm). Thus, we believe that we have answered the "whether social dollars exist" question.

Given that our data include activities and interactions between customers who are part of the community, we now turn our attention to plausible explanations of "why" we see the existence of social dollars. Towards this end, we carry out an analysis of activity at the individual level and relate it to the extent of the social dollar. In order to understand how activity in the brand community could relate to social dollars, we turn to prior research on behavioral participation in online communities, brand communities, and online social networks. Our review of these literatures suggests that behaviors consistently related to positive economic outcomes for the firm-sponsors of customer-centered communities can be assigned to one of two distinct "buckets." We label these two buckets "community connection" (CC) and "community interaction" (CI) behaviors. CC behaviors can be seen as a representation of the customer's embeddness in the network of social relations within the community. CI behaviors, on the other hand, represent the customer's role in creating and exchanging content generated by her as well as the consumption of the content created by other community users. A sample of the specific behavioral constructs prior researchers have investigated and how they relate to CC and CI behaviors is summarized in Figure 1.

--------------------------------------------

Insert Figure 1 about here

--------------------------------------------

With this rough typology in mind, we examined our data to find behaviors that can be classified into these buckets. In the online community we observe, CC can be captured by two aspects of the data – the number of person-to-person and person-to-group ties established in the community by each member.[16] These statistics represent the two degree values available in the community's bipartite (two mode) social network of individual and group ties (Granovetter 1973). In our context, the former refers to the virtual links established between individuals in the community and latter refers to the number of special interest groups (e.g., "Dumbledore's Army") to which an individual belongs.

For CI behaviors, we use four behaviors tracked in our data. The first of these is the number of peer-to-peer (P2P) product recommendations, in which the contributing member directs a recommendation for a specific product to another community member by name. The second is the number of products for which a community member provides a textual review and/or star rating. These reviews and/or ratings are non-directed, and can be viewed by community members and visitors to the product page of the retailer's website. The third CI variable is the number of "Top Ten" lists the community member has contributed. These lists are aggregations of products into thematic "best of" groups by the community member (e.g., "My top ten crime novels"), and as such are to some extent simply "grouped" product recommendations. The last is the number of product items (e.g. book cover images) the customer has graphically displayed as part of her personal "bookshelf" on the

---

[16] As is typical in analysis of large social networks, we observe a bounded set of the full online community network of friend ties. For the calculation of network statistics (e.g. friend ties, centrality) used later in this paper as predictors of treatment group expenditure, we included a $k$ = 1 (second-order) snowball sample (Carrington, Scott and Wasserman 2005). This approach incorporates in the calculation of network statistics all direct friend ties who were not already among the original random sample.

community member's personal profile page. Table 14 presents summary statistics for both the CC and CI behavior variables that we use. Note that we compute these variables at the end of T2.

--------------------------------------------

Insert Table 14 about here

--------------------------------------------

To carry out our analysis, we regress total expenditure (in T2) on the CC and CI variables at the individual-level for customers in the treatment group only. Thus, this is a cross-sectional analysis. In order to control for heterogeniety, we use the customer's total expenditure in T1 as a covariate. The results from this analysis are presented in Table 15.

--------------------------------------------

Insert Table 15 about here

--------------------------------------------

We find that one CC behavior variable, the number of friend ties, and one CI behavior variable, the number of product items displayed on the profile page, have a positive and significant relationship with T2 expenditure, with the former being more significant than the latter. The results suggest that higher number of friend ties is related to a higher level of embeddedness in the network, leading to higher expenditure (we focus on a deeper understanding of this finding below). The number of product items displayed constitutes an implicit (rather than explicit) recommendation. We speculate that the customers' effort to present a robust personal profile is indicative of their level of social engagement with the community. Supporting this, prior research suggests that individuals are motivated to invest in personal profile management in online social networks when their personal identity is highly engaged in the interaction environment (Dholakia, Bagozzi and Pearo 2004, Manago et al. 2008).

In terms of relative effect sizes, the marginal value of a friend tie is \$12.20 while the average value of friend ties is \$17.32 (\$12.20 X 1.42 friends). In a similar vein, the marginal value of a book that is shelved is \$0.28 with the average value being \$3.94 (\$0.28 X 14.06 items shelved). Thus, on average the CC behavior has a larger economic relationship with the social dollar than the CI behavior – on average it leads to over four times as many social dollars – for our data. This finding corroborates prior research indicating that customer and/or brand communities are inherently social phenomena, with their expected benefits to the firm increasing in step with the customer's embeddness in the community's network of social ties (Balasubramanian and Mahajan 2001, Fournier and Lee 2009, Williams and Cothrel 2000). Finally, as expected, a customer's expenditure in T1 is highly significant (most likely as it acts as a proxy for heterogeneity).

In the above analysis, the friend ties variable represents a simple count of the number of direct (local) ties held by the member, a statistic known as "degree" in the network theory literature (Freeman 1979). This variable explains only the basic, local nature of the customers' community connections. We draw further from this literature to provide a more robust assessment of the relationship between the social dollar and a customer's connections in the online community's social network. We do this by restricting the next analysis to a standardized measure of degree and three other well-established measures of network centrality (cf. Jackson 2008, 38-41 for equations). The first incremental measure included is "betweeness centrality" (Freeman 1979). This statistic describes the proportion of shortest paths between any two other individuals in the social network on which a given customer lies. A customer with high betweenness centrality is one who connects distinct regions or "sub-communities" within the overall network. Early work by Brown and Reingen (1987) revealed that individuals with high betweeness centrality have special access to novel product information, and may be particularly important in transmitting that information *to* others (although not necessarily subject to influence *by* others). The second incremental measure we include is "closeness" (Freeman 1972). Closeness is the inverse of the average geodesic (shortest path) distance of an individual to all other members of a network that may be reached from it, and so is an estimate of how proximate that individual is to the "global" store of information or influence in the community network. The last incremental measure we include is "eigenvector centrality" (Bonacich 1972). This measure is built on a recursive premise that a person's access to information and influence is determined by the same attributes in her neighbors, whose access to information and influence is determined by the same attributes as their neighbors, and so on. Perhaps the most well-known application of a form of eigenvector centrality is Google's PageRank algorithm, which assesses a website's relevance based on the relevance of the sites that link to it. In our context of customer social behavior in an online community, we propose this measure captures the extent to which an individual is connected to well-connected, and thus more informed and/or influential, others. Unlike degree and betweenness, there is virtually no prior work that has examined the relationship between closeness and eigenvector centrality and its relationship to social influence in networks.

Each of the four measures described above is standardized relative to all other individuals in the bounded network we observe. Table 16 presents summary statistics for these measures. We then run a regression analysis with expenditure in T2 as the dependent variable and these four measures and the expenditure in T1 as the independent variables. The results (Table 17) show that, of the network measures, only eigenvector centrality is significant in terms of explaining overall expenditure among community members in T2.

--------------------------------------------

Insert Tables 16 and 17 about here

--------------------------------------------

The addition of these measures allows us to zero in on the relationship between outcomes and network connections. Specifically, we find (Table 17) that it is the number of *well-connected* friends (eigenevctor) that is significantly related to the expenditure outcome rather than merely the number of friends (degree) as we found in our previous analysis.

Note that the two analyses reported in this section were restricted to customers who were community members at the end of our observation period. Thus, our endeavor here was to understand the relationship between the outcome variable and behavioral covariates for customers that have selected themselves into the treatment; i.e., our results are conditional on community membership. Also, the above analysis – regressing the expenditure on CI and CC behaviors – is essentially correlational. Given our data, it is not clear that we can draw out any direct causal relationships. However, we can offer two possible hypotheses with respect to these. First, it seems likely that customers who join the community are more engaged with the firm, its brandmarks and/or products than those that do not. Given that they have joined the community, the measures described above can be seen as proxies for this engagement. This engagement leads these customers to increase their expenditure on the firm's products. This increased expenditure – the social dollars – can arise in a direct manner (i.e., customers increase their *overall* consumption of books, CDs, DVDs, etc.) or an indirect manner (i.e., customers *shift* their expenditure from competing firms to the firm from which we obtained the data). Unfortunately, given that we do not observe all expenditure on these product categories for these customers (i.e., purchases at competing firms), we cannot distinguish between these two sources.

## IMPACT ON FIRM AND MANAGERIAL IMPLICATIONS

From the firm's point of view, the important question is whether the launch of an online customer community results in increased revenue and profits, especially relative to the investment made in terms of the community's development and ongoing operations. In order to quantify this, we approached the firm and were able to obtain estimates of community development and operating costs. Based on the estimated social dollars, these costs[17] and firm-level margin percentages available in public financial statements, we estimate that the firm achieves break-even on its investment when 33,000 of its current customers (our conservative restriction case) sign up for the community. Given that the firm acquired 260,000 members within the first fifteen months after community launch, this was clearly a very profitable investment for the firm, especially as this number is comprised of a mix of both current and newly-acquired customers.

Our results also suggest that the firm can now examine the behavioral correlates of the social dollar in greater detail via experiments and incentives. For example, using experiments that provide consumer incentives to display more personal profile content and/or to develop ties with

[17] Due to confidentiality reasons, we are unable to reveal these figures.

influential or well-connected community members, the firm can directly test for causal relationships between these CC and CI behaviors and social dollars. If these behaviors do lead to higher expenditures, in some ways they can be confirmed as objective measures of engagement that are directly linked to positive economic outcomes.

Besides the direct economic benefits to the firm from setting up the 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 firm-sponsored online community offers a 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), an informational boon for customer relationship management and other life-cycle based marketing strategies. The community interactions content can further be mined to identify products that are trending in terms of social relevance, to identify who discusses them (e.g. network opinion leaders or special-interest groups), and to understand "how" they talk about them (e.g. attribute or lifestyle-related language). This insight can subsequently be used to optimize marketing mix decisions (e.g., promotions). 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). While it is likely that hosting customer communities on third-party websites such as Facebook provides reach to a broader audience, this strategy does not offer the same level of access and control over customer interaction management and data offered by a firm-sponsored social network, nor is the third-party community interaction data commonly available to the firm in a manner that can be easily linked to customer-level purchase behavior.

## CONCLUSION

Our paper adds to the small, but growing literature on the economic impact of online customer and/or brand 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. Using a novel dataset from a firm that operates such a community, we are able to quantify the incremental expenditures resulting from customer engagement in a community. The availability of customer expenditure both before and after the formation of the community allows us to create treatment (community members) and control groups, helping to rule out multiple selection issues. We find that social dollars represent about 19% of revenue once customers join the online community. These social dollars arise primarily via more frequent orders with the firm, rather than increased shopping basket sizes.

As is important for studies that leverage natural events, we carry out a series of robustness analyses to check 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. Investigating the social dollar more carefully, we find that it persists over time, exists in both the online and offline channels and across the broad range of product categories sold by the firm.

We then briefly examined the behavioral correlates of the social dollar. We find that aspects of community connections (the number of friend ties) and community interaction (the number of products the customer displayed on the personal profile page) are both related to the social dollar. Further investigation of the customers' structural position in the community's social network shows that the extent to which a community member is connected to more important (well-connected) friend ties in the network is linked to increased expenditures at the firm. Finally, we are able to 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. As we note, there are also many indirect benefits that can be reaped by the firm-operator of a community such as the one we observe in the present research.

Our analysis suffers from some limitations, primarily due to the data that we leverage. First, we only examine consumer behavior in four experiential goods categories. Second, our data extends to only fifteen months after the formation of the community, restricting our ability to investigate longer-term effects on customers and the firm. Third, given that we do not assign customers to treatment and control groups randomly, we cannot rule against the effect of unobservable attributes with perfect certainty (although multiple analyses suggest that this is unlikely to be an issue). Finally, given that we do not observe customers shopping for these categories with competitors, we cannot pinpoint the source of the social dollars precisely. In addition, the lack of primary data makes it hard to make strong causal arguments. We hope that future work will address these limitations.

REFERENCES

Algesheimer, R., U. M. Dholakia and A. Herrman (2005), "The Social Influence of Brand Community: Evidence from European Car Clubs," *Journal of Marketing Science*, (69) July, 19-34.

Algesheimer, R., S. Borle, U. M. Dholakia and Siddharth S. Singh (2010), "The Impact of Customer Community Participation on Customer Behaviors: An Empirical Investigation," *Marketing Science*, (29) 4, 756-769.

Bagozzi, R. P. and U. M. Dholakia (2002), "Intentional Social Action in Virtual Communities," *Journal of Interactive Marketing*, 16 (2), 2-21.

Balasubramanian, S. and V. Mahajan (2001), "The Economic Leverage of the Virtual Community," *International Journal of Electronic Commerce*, 5 (3), 103-38.

Bertrand, M., E. Duflo and S. Mullainathan (2004), "How Much Should We Trust Difference-in-Differences Estimates?", *Quarterly Journal of Economics*, 119 (1), 249-75.

Blundell, R., A. Duncan and C. Meghir (1998), "Estimating Labour Supply Responses Using Tax Reforms," *Econometrica*, 66 (4), 827–61.

Bonacich, P. (1972), "Factoring and Weighting Approaches to Status Scores and Clique Identification," *Journal of Mathematical Sociology*, 2, 113-120.

Brown, J., A.J. Broderick and N. Lee (2007), "Word-of-Mouth Communication Within Online Communities: Conceptualizing the Online Social Network," *Journal of Interactive Marketing,* 21 (3), 2-20.

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

Brown, S. L., A. Tilton, and D. M. Woodside (2002), "On-line Communities Pay," *McKinsey Quarterly*, 1, 17.

Carrington, P. J., J. Scott and S. Wasserman (2005), *Models and Methods in Social Network Analysis*, Cambridge: Cambridge University Press.

Chevalier, J. and D. Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews", *Journal of Marketing Research,* 43 (August), 345-354.

Dholakia, U. M., R. P. Bagozzi and 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, 241-63.

Donald, S. G. and K. Lang (2007), "Inference with Difference-in-Differences and Other Panel Data," *The Review of Economics and Statistics*, 89 (2), 221-233.

Freeman, L. C. (1979), "Centrality in Social Networks Conceptual Clarification," *Social Networks*, 1, 215-239.

Forbes Insights (2009), "The Role of the CMO: Marketing Strategies for 2010," available at [www.slideshare.net/anatinge/cmo-survey-results-forbes-insights].

Forrester Research Inc. (2011), "US Interactive Marketing Forecast, 2011 to 2016," available at http://www.cmo.com/sites/default/files/us_interactive_marketing_forecast,_2011_to_2016.pdf

Fournier, S. and L. Lee (2009), "Getting Brand Communities Right," *Harvard Business Review*, 87 (4), 105-11.

Goldfarb, A. and C. Tucker (2011a), "Advertising Bans and the Substitutability of Online and Offline Advertising," *Journal of Marketing Research*, 48 (2), 207-227.

Goldfarb, A. and C. Tucker (2011b), "Search Engine Advertising: Channel Substitution When Pricing Ads to Context," *Management Science*, 57 (3), 458-470.

Granovetter, M. S. (1973), "The Strength of Weak Ties," *American Journal of Sociology,* 78 (6), 1360-1380.

Hagel, J. I. and A. Armstrong (1997), *Net Gain: Expanding Markets through Virtual Communities*, Boston: Harvard Business Press.

Hartmann, W., H. S. Nair and S. Narayanan (2011), "Identifying Causal Marketing Mix Effects Using a Regression Discontinuity Design," Marketing Science, 30 (4), 1079-1097.

Heckman J. J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153-61.

Holland, J. and S. M. Baker (2001), "Customer Participation in Creating Site Brand Loyalty," *Journal of Interactive Marketing*, 15 (4), 34-45.

Imbens, G. W. and T. Lemieux (2008), "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142, 615-35.

Interbrand (2011), "Best Global Brands 2011," available at[http://www.interbrand.com/en/best-global-brands/Best-Global-Brands-2011.aspx].

Jackson, M. O. (2008), *Social and Economic Networks*, Princeton, NJ: Princeton University Press.

Manago, A. M., M. B. Graham, P. M. Greenfield and G. Salimkhan (2008), "Self-presentation and Gender on MySpace," *Journal of Applied Developmental Psychology*, 29 (July), 446-58.

McWilliam, G. (2000), "Building Stronger Brands through Online Communities," *Sloan Management Review*, 41 (Spring), 43-54.

Moorman, Christine (2010), "The CMO Survey: February 2010," Duke University Fuqua School of Business and the American Marketing Association, available at [http://faculty.fuqua.duke.edu/cmosurvey/].

Muniz, A. M. Jr. and T. C. O'Guinn (2001), "Brand Community," *Journal of Consumer Research*, 27 (March), 412-32.

Nielsen (2011), "State of the Media: The Social Media Report", available at [http://blog.nielsen.com/nielsenwire/social/]

PricewaterhouseCoopers (2011), "Private Company Trendsetter Barometer," available at [www.barometersurveys.com/vwAllNewsByDocID/54AE305195DC70E08525784C007109A8]

Porter, C. E. and N. Donthu (2008), "Cultivating Trust and Harvesting Value in Virtual Communities," *Management Science*, 54 (1), 113-28.

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

Schau, H. J., A. Muniz Jr. and E. J. Arnould (2009), "How Brand Community Practices Create Value," *Journal of Marketing,* 73 (September), 30-51.

Solis, Brian (2010), *Engage: The Complete Guide for Brands and Businesses to Build, Cultivate, and Measure Success in the New Web*, Hoboken, NJ: Wiley.

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

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

Wallace, D. and S. B. Green (2001), "Analysis of Repeated Meausres Designs with Linear Mixed Models," (pp. 94-121) in D. S. Moskowitz, S. L. Hershberger (Eds), *Modeling Intraindividual Variability with Repeated Measures Data.* Philadelphia, PA: Psychology Press.

Wetpaint/Altimeter (2009), "ENGAGEMENTdb," available at [http://www.altimetergroup.com/2009/07/engagementdb.html]

Williams, R. L. and J. Cothrel (2000), "Four Smart Ways to Run Online Communities," *Sloan Management Review*, 41 (4), 81-91.

Zhu, R., U. M. Dholakia, X. Chen, and R. Algesheimer (2011), "Does Online Community Participation Foster Risky Financial Behavior?," *Journal of Marketing Research*, forthcoming.

**Table 1: Purchase Statistics by Group**

| | Control | Treatment | t-stat |
|---|---|---|---|
| T1: 15 Months Pre | | | |
| Total Revenue | 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 Revenue | 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 2: Main Difference-in-Differences (DD) Estimate**

| | (1) Total Revenue | (2) Average Purchase | (3) Purchase Frequency |
|---|---|---|---|
| $b_1$ (Social Dollar) | 127.01*** | 2.25* | 2.87*** |
| | (17.24) | (1.07) | (0.34) |
| Observations | 9164 | 9164 | 9164 |

Standard errors appear in parentheses below estimates.
*** $p$<0.001, ** $p$<0.01, *$p$<0.05

**Table 3: Summary Statistics by Group**

| Variable | Control | Treatment | t-stat |
|---|---|---|---|
| % Female[A] | 70 | 73 | 2.01* |
| Tenure at launch (months) | 38.31 | 37.98 | 0.33 |
| Geography | | | |
| % Region 1 | 33 | 30 | -2.34* |
| % Region 2 | 55 | 56 | 0.69 |
| % Region 3 | 6 | 6 | 0.15 |
| % Region 4 | 6 | 9 | 2.53* |

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

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

**Table 4: Robustness of Main DD Estimate to Demographic Differences**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| $b_1$ (Social Dollar) | 127.01*** | 163.38*** | 207.62*** | 116.27*** | 141.59*** | 127.23*** | 127.99*** | 224.16** |
| | (17.24) | (34.73) | (30.85) | (20.96) | (25.71) | (17.77) | (17.84) | (85.84) |
| Group x Gender(F) | | -37.55 | | | | | | -56.52 |
| | | (41.35) | | | | | | (43.53) |
| Group x Tenure | | | -1.59** | | | | | -1.61* |
| | | | (0.59) | | | | | (0.64) |
| Group x Region 1 | | | | 36.28 | | | | 56.27 |
| | | | | (36.84) | | | | (82.83) |
| Group x Region 2 | | | | | -26.15 | | | 25.75 |
| | | | | | (34.65) | | | (79.70) |
| Group x Region 3 | | | | | | -38.12 | | -29.39 |
| | | | | | | (67.97) | | (112.59) |
| Group x Region 4 | | | | | | | -15.26 | -- |
| | | | | | | | (69.43) | -- |
| Observations | 9164 | 7573 | 8262 | 9164 | 9164 | 9164 | 9164 | 6812 |

Standard errors appear in parentheses below estimates.

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

**Table 5: DD by Pre-Period (T1) Total Purchase Volume Quartile**

| | (1) Quartile 1 | (2) Quartile 2 | (3) Quartile 3 | (4) Quartile 4 |
|---|---|---|---|---|
| T1 Total Purchases | | | | |
| Mean | 85.05 | 241.35 | 460.93 | 1181.5 |
| Median | 83.70 | 241.35 | 454.03 | 928.27 |
| SD | 42.51 | 50.45 | 82.23 | 946.86 |
| $b_1$ (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 6: Comparison Group Total Purchases Means by T1 (Pre) Quarter**

| T1 (Pre) Quarter | $M_{\text{treatment}}$ | $M_{\text{control}}$ | *t-stat* |
|---|---|---|---|
| Q1 | 64.76 | 66.68 | 0.41 |
| Q2 | 112.26 | 121.42 | 1.61 |
| Q3 | 107.17 | 113.02 | 1.19 |
| Q4 | 106.01 | 109.23 | 0.65 |
| Q5 | 99.96 | 101.37 | 0.30 |

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

**Table 7: Reduced Treatment Group Sample DD**

| Sample | $b_3$ (Social Dollar) | % of T2 | Sample | $b_3$ (Social Dollar) | % of T2 |
|---|---|---|---|---|---|
| 1 | 122.13*** | 18% | 6 | 155.76*** | 22% |
| | (19.34) | | | (22.15) | |
| 2 | 114.30*** | 17% | 7 | 115.29*** | 18% |
| | (22.38) | | | (20.05) | |
| 3 | 140.57*** | 21% | 8 | 120.53*** | 19% |
| | (20.55) | | | (20.38) | |
| 4 | 130.38*** | 19% | 9 | 114.73*** | 18% |
| | (21.70) | | | (18.78) | |
| 5 | 163.08*** | 23% | 10 | 129.54*** | 21% |
| | (22.29) | | | (19.24) | |

Standard errors appear in parentheses below estimates.
*** $p$<0.001

**Table 8: "Join Quarter" Cohort DD**

| | (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 |
| $b_1$ (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 parentheses below estimates.
*** $p$<0.001, ** $p$<0.01, *$p$<0.05

**Table 9: "Join Quarter" Cohort Group Comparisons**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Q6 | Q7 | Q8 | Q9 |
| T1 Quarterly Mean Revenue | 72.31 | 94.04 | 92.47 | 92.39 |
| <u>$t$-stat for T1 Qtrly. Mean versus</u> | | | | |
| T2 Q7 Revenue | 13.09*** | | | |
| T2 Q8 Revenue | 10.26*** | 2.46* | | |
| T2 Q9 Revenue | 11.24*** | 1.49 | 2.40* | |
| T2 Q10 Revenue | 23.08*** | 7.42*** | 5.69*** | 8.70*** |

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

**Table 10: "Regression Discontinuity" Style Comparisons of Treatment Group Means**

| Window | Obs. | $ Pre | $ Post | *t-stat* |
|---|---|---|---|---|
| 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*** |

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

**Table 11: Temporal Persistence DD**

| Distance from Launch | One Quarter | Two Quarters | Three Quarters | Four Quarters | Five Quarters |
|---|---|---|---|---|---|
| $b_1$ (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 |

Standard errors appear in parentheses below estimates.
*** $p$<0.001, ** $p$<0.01, *$p$<0.05

**Table 12: Channel Share DD**

| | (1) Proportion Online | (2) Online Revenue | (3) Retail Revenue |
|---|---|---|---|
| $b_1$ (Social Dollar) | 0.130*** | 87.79*** | 39.23** |
| | (0.01) | (10.71) | (12.96) |
| Observations | 9164 | 9164 | 9164 |

Standard errors appear in parentheses below estimates.
*** $p$<0.001, ** $p$<0.01, *$p$<0.05

**Table 13: Product Category DD**

| | (1) Audio | (2) Books | (3) Video | (4) Other |
|---|---|---|---|---|
| $b_1$ (Social Dollar) | 3.36*** | 106.82*** | 7.43*** | 9.40* |
| | (1.00) | (14.91) | (1.86) | (3.90) |
| Observations | 9164 | 9164 | 9164 | 9164 |

Standard errors appear in parentheses below estimates.
*** $p$<0.001, ** $p$<0.01, *$p$<0.05

**Table 14: Summary Statistics for Community Behaviors (n = 7909)**

| Community Behavior | Mean | SD | Min | Max |
|---|---|---|---|---|
| CC: Friend ties | 1.42 | 8.17 | 0 | 245 |
| CC: Group ties | 0.29 | 1.60 | 0 | 44 |
| CI: P2P recos sent | 0.50 | 13.42 | 0 | 627 |
| CI: Reviews written | 1.31 | 12.75 | 0 | 603 |
| CI: Top 10 lists written | 0.58 | 2.46 | 0 | 100 |
| CI: Products displayed | 14.06 | 62.26 | 0 | 2170 |

**Table 15: Relationship between Community Behaviors and Post-Community Expenditures**

| DV: T2 Expenditures | |
|---|---|
| CC: Friend ties | 12.20*** |
| | (1.64) |
| CC: Group ties | -1.45 |
| | (5.35) |
| CI: P2P recos sent | -0.73 |
| | (0.54) |
| CI: Reviews written | 0.00 |
| | (0.55) |
| CI: Top 10 lists written | -2.32 |
| | (3.00) |
| CI: Products displayed | 0.28* |
| | (0.13) |
| T1 Expenditure | 0.89*** |
| | (0.01) |
| Observations | 7909 |
| $R^2$ | 0.47 |

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

**Table 16: Summary Statistics for Network Centralities**

| Network Centrality | Mean | SD | Max* |
|---|---|---|---|
| Degree | .00026 | .00148 | .04421 |
| Closeness | .00033 | .00068 | .00174 |
| Betweenness | .00004 | .00042 | .01995 |
| Eigenvector | .00364 | .01425 | .30970 |

*All centrality statistics are normalized (range 0 - 1). Community members of maximal centrality in the observed network reside outside the analysis group.

**Table 17: Relationship between Network Centralities and Community Expenditure**

| DV: T2 Expenditures | b |
|---|---|
| Degree Centrality | -18820 |
| | (15550) |
| Closeness Centrality | 9767 |
| | (11980) |
| Betweeness Centrality | -47240 |
| | (33770) |
| Eigenvector Centrality | 4815** |
| | (1237) |
| T1 Expenditure | .895*** |
| | (.011) |
| Observations | 7909 |
| $R^2$ | 0.47 |

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

**Figure 1: Constructs supporting Proposed Behavioral Correlates of Economic Outcomes**

| Community Connections (CC) | Community Interactions (CI) | Citation |
|---|---|---|
| *"Group membership"* | *"Participation"* | Bagozzi & Dholakia, 2002 |
| *"Approval utility"* | *"Focus-related utility"* | Balasubramanian et al. 2001 |
| *"Relationships"* | *"Dialogue"* | McWilliam, 2001 |
| *"Shared consciousness"* | *"Reciprocity"* | Muniz & O'Guinn 2001 |
| *"Community relations"* | *"Asset management"* | Williams & Cothrel, 2000 |
| *"Social networking"* | *"Engagement", "Impression management"* | Schau et al. 2009 |
| *"Connection"* | *"Conversation"* | Solis, 2010 |