Marketing Science
Vol. 29, No. 5, September–October 2010, pp. 805–814
ISSN 0732-2399 | EISSN 1526-548X | 10 | 2905 | 0805

DOI 10.1287/mksc.1100.0560

Growing Two-Sided Networks by Advertising the User Base: A Field Experiment

Catherine Tucker, Juanjuan Zhang
MIT Sloan School of Management, Cambridge, Massachusetts 02142 {cetucker@mit.edu, jjzhang@mit.edu}

Two-sided exchange networks (such as eBay.com) often advertise their number of users, presumably to encourage further participation. However, these networks differ markedly on how they advertise their user base. Some highlight the number of sellers, some emphasize the number of buyers, and others disclose both. We use field experiment data from a business-to-business website to examine the efficacy of these different display formats. Before each potential seller posted a listing, the website randomized whether to display the number of buyers and/or sellers, and if so, how many buyers and/or sellers to claim. We find that when information about both buyers and sellers is displayed, a large number of sellers deters further seller listings. However, this deterrence effect disappears when only the number of sellers is presented. Similarly, a large number of buyers is more likely to attract new listings when it is displayed together with the number of sellers. These results suggest the presence of indirect network externalities, whereby a seller prefers markets with many other sellers because they help attract more buyers.

Key words: two-sided markets; information disclosure; competition; inference; entry; field experiment
History: Received: November 19, 2007; accepted: January 12, 2010; processed by Yuxin Chen. Published online in *Articles in Advance* March 10, 2010.

1. Introduction

Two-sided exchange networks have been a magnet for entrepreneurs, some having been able to expand at a furious pace. Transactions on eBay.com exploded from 0 in 1995 to 340,000 auctions closing per day by 1999 (Lucking-Reiley 2000). Match.com started from scratch in 1994 and as of 2009 boasts listings for more than 12 million men and 8 million women. However, there are also many well-funded two-sided exchange networks that never gained traction. Chemdex.com pioneered the business-to-business (B2B) portal model for the chemical industry and raised $112.5 million through its initial public offering, but it never accumulated enough clients to make a profit. Knowing how to grow network participation helps firms avoid costly flops in this high-stake game. In this paper we examine the extent to which two-sided network operators can grow their user base by advertising the size of the user base itself.

Many two-sided exchange websites try to encourage user participation by advertising their user base but through markedly different formats. Table 1 presents the messages used by different websites when attracting new listings. These websites span a variety of categories, including medical equipment, restaurants, vacation rentals, and employment. One key observation is that these websites differ in which side of the network to focus on. Some highlight the number of buyers (#*Buyers*), some emphasize the number of sellers (#*Sellers*), and some display information on both. For example, among the three online marketplaces for used medical equipment, Kitmondo.com highlights that it has "109,046 used equipment listings"; MEDmarketplace.com advertises that "thousands of medical equipment buyers come ... daily"; and DOTmed.com claims that it has "more than 100,000 registered users from countries around the world" and "more than 150,000 listings on any given day."

It is not clear which of these formats is the most effective. For example, the claim that there are many sellers on the website can be double-edged. On the one hand, the knowledge that there are many other sellers may deter future sellers if they anticipate stiff competition. On the other hand, the presence of many peer sellers may signal that this is a successful exchange network that hosts a large number of buyers.

We study the effect of these different information release formats using data from a field experiment. The field experiment was conducted by a B2B website that brings together sellers and buyers of various categories of goods and real estate properties. Before each potential seller decided whether to post a listing, the website randomized whether to highlight the number of buyers and/or sellers and the exact number to state. We find that when the website displays information on both the number of buyers and

Table 1 Different Formats Used to Advertise the User Base When Attracting New Users

Website name	Category	Message	*#Sellers*	*#Buyers*
Kitmondo.com	Medical equipment	109,046 used equipment listings.	Yes	No
MEDmarketplace.com	Medical equipment	Thousands of medical equipment buyers come to MEDmarketplace.com daily.	No	Yes
DOTmed.com	Medical equipment	More than 100,000 registered users from countries around the world. More than 150,000 listings on any given day.	Yes	Yes
Foodler.com	Restaurant	Foodler works with thousands of restaurants.	Yes	No
Campusfood.com	Restaurant	Access to millions of customers.	No	Yes
OpenTable.com	Restaurant	More than 10,000 restaurants have traded in their pen and paper for the OpenTable System and collectively seated more than 100 million guests.	Yes	Yes
Homeaway.com	Vacation rentals	5 × more traffic than our nearest competitor with more than 75 million traveler visits a year.	No	Yes
VRBO.com	Vacation rentals	More than 65 million traveler visits each year and has more than 120,000 listings worldwide.	Yes	Yes
SFadclub.com	Employment	Over the last 7 years, we've posted more than 2,500 jobs.	Yes	No
Monster.com	Employment	More than 1,300,000 job searches per month in your area!	No	Yes
HudsonValleyHelpWanted.com	Employment	1,120 jobs posted, 171,207 registered job seekers.	Yes	Yes

the number of sellers, the presence of a larger number of sellers reduces further listings. However, when the same number of sellers is highlighted on its own, future sellers are not deterred. Similarly, publicizing a large number of buyers on its own is not particularly effective at attracting sellers to the website.

These results are consistent with the notion of "indirect network externalities" from the two-sided networks literature (Bucklin and Sismeiro 2003, Ellison and Ellison 2005). The existence of many sellers is more likely to attract a heavier traffic of buyers who value a wide selection of sellers. The presence of this buyer traffic, in turn, attracts more sellers. Meanwhile, knowing that there are many buyers encourages future listings, especially when new sellers are assured that the existence of many buyers does not imply the presence of many sellers. These findings are also related to the retail colocation literature. Dudey (1990) and Wernerfelt (1994) show that competing retailers may choose to concentrate in the same shopping mall as a commitment not to raise prices to exploit consumers' sunk costs of traveling. Iyer and Pazgal (2003) demonstrate that the gathering of competing sellers through "Internet shopping agents" can mitigate price competition as a larger number of competitors decreases each seller's chance of winning price-sensitive shoppers. Our results share a theme with this literature that sellers may prefer competition, although we focus on the role of competition as a signal of underlying demand.

To quantify the impact of using these various information display formats, we project the likely returns to the website in terms of saved seller acquisition costs. We find that the most profitable format is not always the most informative one (which discloses information on both sides of the market). Instead, highlighting either seller-only information or buyer-only information is more effective, saving the firm a total of $27,992 and $32,245, respectively, if implemented across the entire website. Moreover, the efficacy of these display formats vary by category. In particular, when only seller information is displayed, there is a significant positive correlation between the number of new sellers acquired and buyers' browsing time within a category, possibly because buyers' need for comparison shopping strengthens indirect network effects.

These findings are timely because there is no research guiding how two-sided networks should advertise their user bases in practice. Conversations with marketing management of the websites listed in Table 1 suggest that this important format choice is often left to the discretion of website designers as an appearance element, rather than being deliberated over as a strategic marketing variable. For example, one website for pharmaceutical buyers and sellers delegated that part of the website design to their summer intern. Another website hosting job seekers and employers told us that their "policy" was an automated legacy from a Web redesign nine years before and consequently was perceived as being "out of [their] control." Our results show that what information to disclose to recruit new users to a website is not a straightforward decision; choosing the wrong display format can hinder growth.

To our knowledge, this paper is the first investigation of how firms can use strategic information revelation as a marketing tool to build two-sided networks. There is a body of research on factors that affect network participation. For example, Fath and Sarvary (2003) find that one growth strategy for B2B exchanges is to subsidize buyers. Chen and Xie (2007) discover that the lack of customer loyalty can end up benefiting firms in the presence of cross-market network effects. Our findings contribute to this literature by showing that the almost costless display of the number of network users can be an effective growth tool if implemented appropriately.

The rest of this paper is organized as follows. Section 2 describes the field experiment and presents the data. Section 3 presents the empirical estimation of the experimental effects. Section 4 projects the aggregate impact of pursuing each of the information display formats. Section 5 summarizes and discusses the findings.

2. Field Experiment and Data

2.1. Business Background and the Field Experiment

We use field experiment data from a B2B website that in appearance resembles craigslist.org.[1] The website, which receives 240,000 clicks per day, provides a common platform for sellers of various types of goods and real estate properties to advertise these items in multiple specific categories and for buyers to read the listings. The target customers are one-person businesses and small-time entrepreneurs. More than 40 major metropolitan areas are served, but there is little cross-geographical browsing (less than 3%, based on an analysis of the IP addresses in our sample). The website draws revenues from banner advertisements on its main page; it does not charge sellers for using its listing service or buyers for browsing listings.

Although a fee is not charged, a seller must register and log in to an individual user account on the website and subsequently fill out a "listing form" to be able to list an item for sale. Sellers will post a listing if their expected return from the listing exceeds the opportunity cost of time spent filling out the forms, any future time costs of monitoring transactions on this website, and any switching costs (Fath and Sarvary 2003). The seller attrition rate from starting a listing to submission is 16% prior to the experiment. The fact that there is significant attrition suggests that the costs of completing a listing are nonzero.[2] Once submitted, the listings appear reverse chronologically on the Web page. Buyers, on the other hand, can view listings without signing up for the website.

As shown in Table 1, many websites advertise the user base as a way to attract more users, but the format varies markedly. In response to this trend, the website we study conducted a field experiment to investigate how disclosing the number of users on either side of the platform affects future seller listings. The website randomly varied whether to display the number of sellers and/or buyers to each potential seller, and if so, how many sellers/buyers to state.[3]

The field experiment ran from November 29, 2006 to January 15, 2007 in the largest city market the website serves. A total of 3,314 attempted listings across 15 categories were exposed to the experiment.[4] "Rentals" was the most popular category, with 837 attempted listings exposed to the experiment. "Fabric-Attire" was the smallest category, with 32 attempted listings subject to the experiment.

The experimental treatment was implemented as follows. Immediately after choosing the product category he or she intended to list in, and before continuing on to the next Web page to fill out the listing form, a potential seller was exposed to an "information page." The text content displayed on the information page was randomly drawn from the following four treatment conditions:

(1) "Presently, there are [#*Sellers*] listings and [#*Buyers*] users viewing these listings in the [category name] category of [city name]."

(2) "Presently, there are [#*Sellers*] listings in the [category name] category of [city name]."

(3) "Presently, there are [#*Buyers*] users viewing these listings in the [category name] category of [city name]."

(4) (A blank page.)

The number of listings #*Sellers* and the number of buyers #*Buyers* were further randomly drawn for each potential seller. Individual-level randomization ensures that the correlation between listing propensities and #*Sellers* or #*Buyers* is not confounded with unobservable variables. Based on the long-term website traffic, both #*Sellers* and #*Buyers* were drawn from a uniform distribution between 1 and 200. Furthermore, by using the ambiguous word "presently," which allows some vagueness as to the time frame in the website's original language, management intended to avoid deceiving customers through the randomization procedure.[5] We run a series of regressions to ensure that the website implemented the randomization procedures correctly. Nonetheless, we will cluster standard errors at the category level in model estimation and include category, week, day, hour, and

[1] The website's name and location are protected because of confidentiality agreements.

[2] If the total cost of listing were negligible, potential sellers would have posted a listing regardless of the experimental manipulation. In that case, the experiment can be seen as a conservative test of the information display formats.

[3] The website targeted its experiment toward sellers as opposed to buyers, as they hoped to eventually generate extra revenue from paid listings.

[4] We removed listings we judged to be automated or "spam." See §2.2 for details.

[5] Although we use "presently" as the closest American English language approximation to the word that was actually used, the most accurate translation of the word used would be "in very recent times."

number-of-visits fixed effects in our specifications to control for any departure from full randomness.

Prior to the launch of the field experiment, no information was displayed regarding the number of buyers. Meanwhile, the website divides listings into multitiered categories, which obscures the actual number of sellers. Also, the categories we study are almost entirely for sellers with a single unit of a good for sale. These features maximize the chance that each potential seller receives one randomized information exposure. However, unlike lab experiments, a Web-based field experiment may not strictly follow a pure between-subjects design, because it is infeasible to prevent users from returning to the website at a later date or in a different category. A seller could potentially visit the site multiple times and see implausibly large shifts in the number of buyers or sellers highlighted. Therefore we retain data on each seller's first visit on each day but remove data for subsequent visits on the same day. In addition, we include controls in subsequent analyses to capture multiple visits by the same seller.

Finally, after being presented with the information page, a potential seller could either quit listing or proceed to the next page, fill out the listing form, and complete the listing process. Once the seller had submitted the listing form, their item appeared on the website immediately. We do not have access to the listing content (such as prices) because of confidentiality concerns. It is plausible, however, that listing decisions already reflect the prospects for postlisting profits. We therefore focus on the effects of prelisting perception of supply and demand induced by the experimental treatment.

2.2. Data

We use two data sets in our analysis: a click-stream data set and a treatment data set. Each entry in the click-stream data consists of a time stamp, the user's IP address, a record of all Web page requests, and an error code. These click-stream data allow us to determine whether a potential seller actually made a listing. We also use them to draw parallels with browsing behavior in §4.

The treatment data set records the information page that each potential seller was exposed to. Each entry in the treatment data contains an IP address, a time stamp, the product category the potential seller intended to list in, whether information on the number of buyers and/or sellers was displayed, and the actual number of buyers and/or sellers drawn if applicable. This treatment data set spans all potential sellers, including those who decided not to continue listing after receiving the treatment information. Such data on sellers who opted out after having arrived at the website help us circumvent the endogeneity problems surrounding seller-side network participation (Bradlow and Park 2007).

We match the click-stream data and the treatment data using the IP address and the time stamp. We are unable to match 128 observations that contain errors, generally caused by time-outs or Web browser incompatibility. We exclude these 128 observations from our empirical analyses. Reassuringly, however, there is no statistically significant relationship between our ability to match the data and the treatment condition.

One data problem is the presence of spammers, who employ automated listing tools that produce a large number of repeat listings. For example, one user (or bot) made 735 listings during the experiment, most of which were in the used computer equipment category. Because spammers post their listings regardless of the information displayed about the number of buyers and sellers, we exclude them from the analyses. We defined a spammer as a seller who has submitted more than 10 listings within the same category during the experimental period, and we removed 1,509 listings as a result. Our findings are qualitatively unchanged with alternative thresholds to define spammers.

Table 2 summarizes the data. Altogether, we have 3,314 observations of sellers who started the listing process and were exposed to one of the four conditions. On average, 84% of them went on to make a listing. In particular, average listing rates are 85% in each of the conditions where information is displayed, and 84% in the condition where no information is shown.

Table 2 Summary Statistics

Variable	Description	Mean	Std. dev.
Dependent variable			
List	Whether a potential seller submits a new listing	0.840	0.367
Explanatory variables			
SellerInfoOnly	Indicator variable for whether the potential seller is in a condition with seller information only	0.250	0.430
BuyerInfoOnly	Indicator variable for whether the potential seller is in a condition with buyer information only	0.260	0.440
SellerInfo&BuyerInfo	Indicator variable for whether the potential seller is in a condition with both buyer and seller information	0.250	0.430
#Sellers	Number of sellers randomly drawn (in thousands)	0.103	0.057
#Buyers	Number of buyers randomly drawn (in thousands)	0.102	0.047
3,314 observations			

3. How Information on the User Base Affects Listings

We first assess how information on the number of buyers and sellers affects a potential seller's probability of listing. To do so, we estimate a probit model, pooling data from all conditions. Equation (1) summarizes the specification, where α, γ, and β are vectors of parameters to be estimated:

$$
\begin{aligned}
&\text{prob}(list_i = 1)\\
&\quad = \Phi(\alpha_i + \beta_0\ X_i + \gamma_i + \beta_1\ SellerInfoOnly_i\\
&\qquad + \beta_2\ BuyerInfoOnly_i + \beta_3\ SellerInfo\&BuyerInfo_i\\
&\qquad + \beta_4\ (\#Sellers_i \mid SellerInfoOnly)\\
&\qquad + \beta_5\ (\#Buyers_i \mid BuyerInfoOnly)\\
&\qquad + \beta_6\ (\#Sellers_i \mid SellerInfo\&BuyerInfo)\\
&\qquad + \beta_7\ (\#Buyers_i \mid SellerInfo\&BuyerInfo)). \qquad (1)
\end{aligned}
$$

In the above specification, Φ stands for the cumulative distribution function of the standard normal distribution. The first term, α_i, is the fixed effect for the category that listing i belongs to. To control for sellers' different arrival time at the website, we include a vector of time effects, X_i, which consists of fixed effects for the hour of the day, day of the week, and the week. The vector γ_i contains a set of dummy variables that capture whether this is seller i's first, second, or up to tenth visit to the website as observed in the data.

We include three dummy variables that measure the level effect of the treatment condition: $SellerInfoOnly_i$ is equal to 1 when only the number of sellers is displayed to seller i, $BuyerInfoOnly_i$ is equal to 1 when the number of buyers is displayed in isolation, and $SellerInfo\&BuyerInfo_i$ is equal to 1 when both the number of sellers and number of buyers are displayed. The continuous variables $\#Sellers_i$ and $\#Buyers_i$ capture the number of sellers and buyers shown to a potential seller, either separately or in combination. The conditional terms $\#Sellers_i \mid SellerInfoOnly$ and $\#Buyers_i \mid BuyerInfoOnly$ capture the effect of the number of sellers and buyers when displayed in isolation. Last, $\#Sellers_i \mid SellerInfo\&BuyerInfo$ captures the effect of the number of sellers if buyer information is also available, whereas $\#Buyers_i \mid SellerInfo\&BuyerInfo$ measures the effect of the number of buyers when displayed together with seller information.[6]

Column (1) of Table 3 shows the estimation results for a null model that includes all explanatory variables except the experimental treatments. This null model provides a benchmark to assess the explanatory power of the other models being estimated. Column (2) of Table 3 shows the results for another nested specification of Equation (1), which focuses on the level effects of the three treatment conditions. Model fit shows little improvement above column (1), and the three condition dummies are all insignificant. It seems that the mere presentation of different formats of information does not significantly affect the probability of listing. This result echoes the raw data, where the average listing probabilities are 85% in each of the conditions where information is displayed and 84% in the condition where no information is shown. This similarity in posting propensities across treatment conditions may arise because the treatment has insignificant impact, or because the effects from high and low numbers of users within each condition have balanced each other out. To investigate this issue, we estimate the full specification suggested by Equation (1).

Column (3) of Table 3 reports the results of the full probit specification. The Akaike's information criterion (AIC) and Bayesian information criterion (BIC) statistics indicate that the full specification fits the data significantly better than both the null model of column (1) and the nested specification in column (2). A comparison of the mean absolute error suggests a 3% improvement of fit in column (3) compared to column (1). There are again no significant level effects from the three condition dummies. This result suggests that any website growth following the highlighting of user base information comes from the exact numbers of users claimed, rather than through aggregate effects, like the perceived informativeness or professionalism of a specific display format. The opposite signs for *#Sellers | SellerInfo&BuyerInfo* and *#Buyers | SellerInfo&BuyerInfo* indicate that the individual-level treatment has counterbalancing effects for contemplating sellers, depending on whether the exact market information each seller receives is favorable or unfavorable. The fact that providing buyer and seller information only increases aggregate listing rates from 84% to 85% further suggests that these effects are likely to be symmetric in magnitude.

The coefficients for *#Sellers | SellerInfoOnly* and *#Buyers | BuysersInfoOnly* are positive but not significant. That is, when highlighted on its own, the number of either sellers or buyers does not affect listing decisions. However, the coefficient of *#Sellers | SellerInfo&BuyerInfo* is negative and significant; when the information about heavy presence of sellers is highlighted in conjunction with buyer information, it discourages future listings. Similarly, *#Buyers | SellerInfo&BuyerInfo* is positive and significant, which

[6] Note that these conditional variables are not interactive terms and are consequently not subject to the Ai and Norton (2003) critique about the interpretation of cross-derivatives in nonlinear models. Nevertheless, we estimate a linear model to ensure that the linearized coefficients are similar to the marginal effects reported.

Table 3 Information Display Formats and Seller Listing Propensities

Variable	(1) Null model	(2) Null model + Condition-level effects	(3) Full model	(4) Ratio model I	(5) Ratio model II
SellerInfoOnly (d)		−0.00758	−0.00518	−0.00168	−0.00531
		(0.147)	(0.0332)	(0.0301)	(0.0334)
BuyerInfoOnly (d)		0.0395	−0.00274	0.00787	−0.00282
		(0.119)	(0.00952)	(0.0235)	(0.00983)
SellerInfo&BuyerInfo (d)		0.0607	−0.0140	0.0165	0.0165
		(0.195)	(0.0343)	(0.0386)	(0.0385)
#Sellers | SellerInfoOnly			0.0372		0.0355
			(0.0283)		(0.0283)
#Buyers | BuyerInfoOnly			0.105		0.105
			(0.141)		(0.140)
#Sellers | SellerInfo&BuyerInfo			−0.248***		
			(0.0696)		
#Buyers | SellerInfo&BuyerInfo			0.524***		
			(0.139)		
Seller-to-Buyer Ratio | SellerInfo&BuyerInfo				−0.00130**	−0.00100**
				(0.00061)	(0.00061)
Category fixed effects	Yes	Yes	Yes	Yes	Yes
Hour fixed effects	Yes	Yes	Yes	Yes	Yes
Day of week fixed effects	Yes	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes	Yes
Visit fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	3,314	3,314	3,314	3,314	3,314
Log-likelihood	−1,238.12	−1,237.64	−1,233.60	−1,236.46	−1,236.31
Pseudo-R^2	0.15	0.15	0.15	0.15	0.15
AIC	2,478.24	2,477.29	2,469.20	2,474.93	2,474.61
BIC	2,484.35	2,483.39	2,475.31	2,481.04	2,480.71

Notes. Marginal effects; standard errors are in parentheses. Dependent variable: whether a seller submits a listing. (d) indicates that the marginal effect is calculated as a discrete change of dummy variable from 0 to 1.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

means that a larger number of buyers attracts listings when seller information is also provided.

The coefficients for *#Sellers | SellerInfoOnly* and *#Sellers | SellerInfo&BuyerInfo* are significantly different ($\chi^2 = 35.18$, $p < 0.001$). We interpret this result in the following way. A larger number of competing sellers often implies higher competitive pressure. However, in two-sided networks, it could also suggest the presence of lots of buyers through an "indirect network externality," whereby network users benefit from heavier traffic on the other side of the market (Nair et al. 2004). Potential sellers, when they are uncertain about the number of buyers on the site, could therefore infer a large number of buyers from a high concentration of sellers (Rochet and Tirole 2006). That is, when a potential seller does not know how many buyers are in the market, the exact number of existing sellers is less likely to matter, because the potential seller would assume that the market is in equilibrium. This is especially true if a seller who contemplates posting assumes that existing sellers possess private information about the nature of demand at the website, which justifies their listing decisions. On the other hand, when the buyer-side information is provided as well, such inferences become unnecessary, leaving heavy seller presence as a negative sign of stiff competition.

Similarly, indirect network externalities also help explain why potential sellers do not respond more positively to the mere information that there are many buyers: The concentration of buyers can also imply the presence of many competing sellers. This may happen because potential sellers assume that buyers have private information about the density of sellers on the site, and they consequently rely on buyer participation as a guide to the extent of likely competition. Alternatively, the seller may reason that the large number of buyers itself has already attracted a large number of competing sellers. In both cases, high buyer concentration is less desirable when the seller is uncertain about the number of sellers.[7]

[7] Admittedly, a website user can investigate the number of sellers by counting. However, this practice is costly, as the website presents listings on multiple pages that hinders counting. Consequently, this user might choose the alternative cognitive route—inferring the number of sellers from the information provided to her. Moreover, even if counting occurs, it dilutes the treatment effects, making the experiment a conservative test.

The findings can be interpreted as an "imputation under uncertainty" effect—when only given information on one side of the market, a potential seller imputes a commensurate traffic for the other side. In other words, potential sellers are likely to impute a constant *ratio* of sellers to buyers in the *SellerInfoOnly* and *BuyerInfoOnly* conditions, regardless of the numbers of users displayed. However, when both buyer and seller information is shown, potential sellers are more likely to respond to the ratio computed therein. As a robustness check, we estimate a "ratio model" where we replace the number of sellers and/or buyers with the seller-to-buyer ratio. Column (4) of Table 3 reports the results. In the seller-only and buyer-only conditions, the imputed constant ratio cannot be separately estimated from the condition dummy. However, when both buyer and seller information is shown, a large computed seller-to-buyer ratio deters listings, as expected. We also estimate an alternative specification in column (5) where we include *#Sellers* | *SellerInfoOnly* and *#Buyers* | *BuyersInfoOnly*. Both coefficients are insignificant, whereas *Seller-to-Buyer Ratio* | *SellerInfo&BuyerInfo* remains negative, consistent with the imputation story.

As another robustness check, we conduct variance inflation factor (VIF) tests within a linear probability framework. All the VIF scores are well below the cutoff point of 10 proposed by Hair et al. (2009) (the average value is 2.67, the highest is 6.62), suggesting that multicollinearity is not a significant issue for our specification. We also perform the tests recommended by Belsley et al. (1980) and obtain a condition number of 5.55, which is well below 30, the threshold beyond which multicollinearity may be a problem.

The results of Table 3 suggest that the format in which user base information is highlighted does affect seller participation differently, even in settings such as our experiment where participation costs are low or nonpecuniary. However, it is not clear how large the differences are between the different formats at the aggregate level. We investigate this issue in the next section.

4. Projecting the Aggregate Effects of Different Display Formats

In this section, we quantify the aggregate effects of using different display formats in terms of saved customer acquisition costs. We perform this analysis for the different product categories in the data. These projections aim to translate parameter estimates into metrics that are easily comparable both across product categories and with other marketing techniques. They also allow us to further investigate at the category level the apparent similarity of aggregate listing rates across conditions. Indeed, we find that different display formats achieve distinct effects across categories.

For each display format and each category j, we project the total cost savings in acquiring new sellers ($TotalCostSavings_j$) relative to a baseline case of presenting no information. We use Equation (2) to calculate the cost savings.

$$\begin{aligned} TotalCostSavings_j = {} & AcqCostPerNewSeller_j \\ & \times (E(\#NewSellers_j \mid \text{DisplayFormat}) \\ & - E(\#NewSellers_j \mid \text{NoInfo})), \end{aligned} \tag{2}$$

where

$$\begin{aligned} & E(\#NewSellers_j \mid \text{DisplayFormat}) \\ & \quad = \#PotentialNewSellers_j \\ & \qquad \times \widehat{\text{prob}}(list_j = 1 \mid \text{DisplayFormat}). \end{aligned}$$

We begin by estimating Equation (1) for each category. Using the parameter estimates, we then compute the fitted probability of listing in category j and for each display format ($\widehat{\text{prob}}(list_j = 1 \mid \text{DisplayFormat})$). In doing so, we rely on the actual traffic of the website to yield realistic projections: in place of *#Sellers* and *#Buyers* randomly drawn for the experiment, we use the average actual number of sellers and buyers over the two weeks prior to the experiment to approximate the word "presently."[8] Although compared with other two-week windows, the numbers we use are representative, this specification does represent a substantial simplification to maintain tractability of the projections. For example, our estimates do not reflect the possibility that the policies we study may themselves change the user base over time. In broader contexts, if website traffic is not stationary, we will need a more precise characterization of the traffic dynamics such as an arrival time model.

We need a way to map the listing probabilities into cost savings. We first project how many new sellers each of these display formats would secure for the firm ($E(\#NewSellers_j \mid \text{DisplayFormat})$) over a 12-month period if the format were applied to each category across the entire website (not just the regional market chosen for the experiment, which accounts for 16% of the total traffic). We obtain this number by multiplying the fitted listing probability for a category and management's forecast of the total number of potential new sellers in this category over a 12-month period ($\#PotentialNewSellers_j$). We then convert the benefit into dollar terms by using estimated

[8] Using other time periods, such as a week or a month, does not affect the relative effectiveness of each of the strategies, although it does affect the absolute numbers.

Table 4 Seller Acquisition Cost Savings from Each Information Display Format

	Cost savings relative to the no-information condition					
	Seller information only		Buyer information only		Seller and buyer information	
Category	Mean	[95% CI]	Mean	[95% CI]	Mean	[95% CI]
Commercial properties	2,338*	[2,992, 1,683]	1,080*	[1,372, 788]	927	[2,060, −206]
Computers	−24,876*	[−20,372, −29,380]	1,476	[3,523, −572]	−36,077*	[−31,182, −40,973]
Digital	2,132	[6,477, −2,213]	2,644*	[3,635, 1,652]	−1,944	[4,162, −8,051]
Electronics	792	[1,720, −137]	368	[876, −140]	−4,077*	[−3,460, −4,694]
Furniture	7,338*	[11,095, 3,580]	4,736*	[5,153, 4,319]	−505	[2,188, −3,198]
General	12,624*	[12,834, 12,415]	5,100*	[5,323, 4,878]	1,024*	[2,008, 39]
Home sales	9,395*	[12,439, 6,350]	4,709*	[5,326, 4,092]	−3,756	[16, −7,528]
Media	1,928*	[2,039, 1,817]	407*	[521, 293]	613*	[1,072, 154]
Office supplies	6,100*	[12,186, 15]	5,326*	[6,064, 4,588]	−8,571	[189, −17,332]
Other	−147	[561, −854]	419*	[517, 320]	−2,440*	[−1,657, −3,223]
Rentals	18,379*	[35,756, 1,003]	3,912	[8,708, −884]	−9,145	[11,684, −29,974]
Shared office space	3,492*	[6,408, 576]	5,937*	[6,246, 5,628]	−1,438	[1,534, −4,409]
Tickets	−1,555*	[−284, −2,827]	10	[116, −96]	−6,198*	[−5,993, −6,402]
Transportation	−9,948*	[−7,561, −12,335]	−3,879*	[−2,965, −4,792]	−13,508*	[−11,667, −15,348]

Notes. All numbers are reported in U.S. dollars. An asterisk indicates that the estimate is significantly different from 0 at the $p = 0.05$ level. The "Fabric-Attire" category is excluded because there are too few observations. CI, confidence interval.

costs of acquiring each new seller for the category ($AcqCostPerNewSeller_j$). As a proxy for seller acquisition costs, we use the cost of search advertising for a category in the country where the website is based. These costs range from less than $1 per seller in the "Fabric-Attire" category to around $4 in the real estate categories.[9]

The projections are based on a bootstrapping methodology, whereby we conduct 1,000 replications of a randomized draw from the data with replacement to obtain standard errors for the cost saving estimates. These projections make three assumptions: first, that our estimates for the metropolitan area where the field experiment was conducted also apply to other regional markets the site serves; second, that observed behaviors are not mainly driven by seasonality; and third, that there are no changes to the average size of a category over these 12 months. As a result, we start with a baseline of approximately 160,000 new sellers per year across all categories when no user base information is displayed.

Table 4 presents the projected seller acquisition cost savings by display format and by category. There are two observations to note. First, the most informative display format, which discloses the number of both sellers and buyers, is not always the most effective one in attracting new sellers. In fact, displaying either seller-side or buyer-side information in isolation is more effective for the focal website. The firm would have saved $27,992 in total if it had displayed seller information in isolation. This display format would have made a significantly positive impact in eight categories but a negative impact in three categories. Alternatively, the firm would have saved $32,245 had it displayed only buyer information. This format would have made a positive impact (although, on average, smaller than the seller-only format) in nine categories and a negative impact in one category. On the other hand, the firm would have lost $85,096 by displaying both seller and buyer information, with positive effects in two categories but negative impact in five. These results suggest that two-sided platforms can actually fare better by allowing potential customers to draw inferences about the user base than by providing full transparency.

A second observation is that the efficacy of each display format also varies across categories. For example, in eight categories (such as "Furniture" and "Home Sales") displaying only seller information significantly outperforms providing both seller and buyer information. One possible explanation is that different categories represent different price levels and degrees of differentiation. If products in a category are highly priced and widely differentiated (which is likely to be true for furniture and home sales), buyers may want to browse many listings, thus strengthening indirect network externalities (Ellison and Ellison 2005). In these categories, the presence of many sellers could signal a high concentration of buyers, whereas heavy traffic of buyers could imply the existence of many sellers. To explore this possibility empirically, we obtain the average time buyers spend browsing listings within a category prior to the experiment. Consistent with our conjecture, we find that the correlation between the number of new

[9] We obtain these statistics by using the "comScore Marketer" database to establish search engine query patterns for the website conducting the field experiment and by quoting data on ad costs from the Google Traffic Estimator for these search queries.

sellers acquired and category browsing time is 0.614 ($p = 0.007$) for the seller-only display format, which reduces to 0.316 ($p = 0.201$) for the buyer-only format. In addition, when both seller and buyer information is displayed, which weakens the need for inferences, the correlation between the number of new sellers and category browsing time is 0.444 ($p = 0.065$).

These results suggest that firms should be strategic about selecting their information display format. As documented by Table 1, two-sided network firms have chosen markedly different information display styles, even when they serve the same product category. Unfortunately, however, these choices have often been made without considering the strategic consequences. Our findings suggest that providing more information does not always translate into attracting more customers and that the efficacy of the display formats varies by category. The choice of the optimal display format therefore should take into account potential customers' knowledge about the network's existing users and category characteristics such as prices, differentiation, and buyers' need for comparison shopping.

Note that the long-run profitability of each information display format may be modified by user learning. The extent to which savings can persist is unclear if potential sellers believe that the display strategies are endogenous and adjust their expectations accordingly. In this sense, what we observe in this study is user-side reactions to exogenous/experimental shifts in display strategies. However, during the two months of the experiment, despite potential suspicion of the endogenous nature of the format choices, users reacted differently to information presented in different ways. This means that the information display format choice does affect user participation in tangible ways, at least in the short run.

5. Conclusion

This paper examines how advertising the size of the user base affects further participation in two-sided networks. We investigate this question using field experiment data from a website that brings together buyers and sellers of various categories of goods and real estate properties. We find that when information about both buyers and sellers is displayed, the existence of many sellers deters further seller listings. However, this deterrence effect disappears when only the number of sellers is presented. Similarly, a large number of buyers is more likely to attract new listings when it is displayed together with the number of sellers. These results may be explained by indirect network externalities: a seller is attracted to posting on a website where there are many peer sellers because their presence helps to attract more buyers to the market; meanwhile, a high number of buyers is not always good news because it may signal the existence of many competing sellers.

We estimate the dollar values of the information display formats in terms of saved customer acquisition costs across different categories. We find that the most informative format, which presents information from both sides of the market, is not necessarily the most profitable one. In fact, for the focal website, displaying either seller or buyer information in isolation is more effective than displaying both concurrently. In addition, the efficacy of the display formats varies across product categories. The number of new sellers acquired is positively and significantly correlated with buyer browsing time within a category when only seller-side information is displayed. These results suggest that two-sided networks' optimal way of advertising the user base should take into account potential customers' knowledge about the network and buyers' need for comparison shopping.

The findings also shed light on some documented ambiguities surrounding the effect of competition on entry. Existing research has emphasized the entry deterrence role of competition. However, this received wisdom has also been questioned by robust findings of "competition neglect" (Camerer and Lovallo 1999, Toivanen and Waterson 2005, Simonsohn 2010), and "competition contagion" (Narasimhan and Zhang 2000, Debruyne and Reibstein 2005), where firms are indifferent to markets that are heavily congested or even more likely to enter these markets. Although many studies rely on bounded rationality, such as a limited capacity for iterative thinking (Camerer et al. 2004), to explain excess entry and the high incidence of postentry failure, our findings suggest an alternative explanation of nonnegative entry correlations within a rational framework. If firms are uncertain about demand, they may react positively to entry of other firms, viewing their entry as a favorable sign of demand. In this sense, our approach echoes Wernerfelt (1995), who reinterprets the seemingly "irrational" compromise effect as consumers' rational inference of product fit from what is available in the market. Our findings suggest that competition neglect can be a rational outcome where entrants infer market potential from existing competition.

There are several ways to extend this research. In the market setting we study, entry costs are not pecuniary. It would be interesting to match optimal information disclosure policies to the price of participating in a network. We also only study participation decisions on one side of a two-sided network. Future research could investigate how a firm's information disclosure strategy affects the combined positive feedback mechanism for both the buyer side and the seller side. Meanwhile, the signaling role of firms' choice of information display formats can be an important topic and deserves full analytical exploration. It would also be useful to further explore the imputation under uncertainty effect. One way to do so is to directly ask experiment subjects their imputed values after provision of partial information. Finally, it would be interesting to investigate other strategic variables such as postentry prices, as discussed by Chen et al. (2002), which are unavailable to us in this current study.

Acknowledgments

The authors thank the editor, area editor, and the reviewers for their insightful feedback. The authors are grateful to John Hauser, Peter Reiss, Duncan Simester, J. Miguel Villas-Boas, Birger Wernerfelt, and participants at the 2008 Marketing Science Conference and the 2008 Summer Institute in Competitive Strategy for thoughtful comments.

References

Ai, C., E. C. Norton. 2003. Interaction terms in logit and probit models. *Econom. Lett.* **80**(1) 123–129.

Belsley, D. A., E. Kuh, R. E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, New York.

Bradlow, E. T., Y.-H. Park. 2007. Bayesian estimation of bid sequences in Internet auctions using a generalized record-breaking model. *Marketing Sci.* **26**(2) 218–229.

Bucklin, R. E., C. Sismeiro. 2003. A model of Web site browsing behavior estimated on clickstream data. *J. Marketing Res.* **40**(3) 249–267.

Camerer, C. F., D. Lovallo. 1999. Overconfidence and excess entry: An experimental approach. *Amer. Econom. Rev.* **89**(1) 306–318.

Camerer, C. F., T.-H. Ho, J.-K. Chong. 2004. A cognitive hierarchy model of games. *Quart. J. Econom.* **119**(3) 861–898.

Chen, Y., J. Xie. 2007. Cross-market network effect with asymmetric customer loyalty: Implications for competitive advantage. *Marketing Sci.* **26**(1) 52–66.

Chen, Y., G. Iyer, V. Padmanabhan. 2002. Referral infomediaries. *Marketing Sci.* **21**(4) 412–434.

Debruyne, M., D. J. Reibstein. 2005. Competitor see, competitor do: Incumbent entry in new market niches. *Marketing Sci.* **24**(1) 55–66.

Dudey, M. 1990. Competition by choice: The effect of consumer search on firm location decisions. *Amer. Econom. Rev.* **80**(5) 1092–1104.

Ellison, G., S. F. Ellison. 2005. Lessons about markets from the Internet. *J. Econom. Perspective* **19**(2) 139–158.

Fath, G., M. Sarvary. 2003. Adoption dynamics in buyer-side exchanges. *Quant. Marketing Econom.* **1**(3) 305–335.

Hair, J. F., Jr., R. E. Anderson, R. L. Tatham, W. C. Black. 2009. *Multivariate Data Analysis*, 7th ed. Macmillan, New York.

Iyer, G., A. Pazgal. 2003. Internet shopping agents: Virtual co-location and competition. *Marketing Sci.* **22**(1) 85–106.

Lucking-Reiley, D. 2000. Auctions on the Internet: What's being auctioned, and how? *J. Indust. Econom.* **48**(3) 227–252.

Nair, H., P. Chintagunta, J.-P. Dubé. 2004. Empirical analysis of indirect network effects in the market for personal digital assistants. *Quant. Marketing Econom.* **2**(1) 23–58.

Narasimhan, C., Z. J. Zhang. 2000. Market entry strategy under firm heterogeneity and asymmetric payoffs. *Marketing Sci.* **19**(4) 313–327.

Rochet, J.-C., J. Tirole. 2006. Two-sided markets: A progress report. *RAND J. Econom.* **37**(3) 645–667.

Simonsohn, U. 2010. eBay's crowded evenings: Competition neglect in market entry decisions. *Management Sci.* **56**(7) 1060–1073.

Toivanen, O., M. Waterson. 2005. Market structure and entry: Where's the beef? *RAND J. Econom.* **36**(3) 680–699.

Wernerfelt, B. 1994. Selling formats for search goods. *Marketing Sci.* **13**(3) 298–309.

Wernerfelt, B. 1995. A rational reconstruction of the compromise effect: Using market data to infer utilities. *J. Consumer Res.* **21**(4) 627–633.