Quality Externalities and the Limits of Reputation in Two-Sided Markets*

Chris Nosko†

Steven Tadelis‡

PRELIMINARY AND INCOMPLETE

April 1, 2014

Abstract

Buyers in two-sided marketplace platforms may draw conclusions about the quality of the platform from any single transaction. This induces an externality across sellers that reputation mechanisms will not alleviate. Furthermore, buyers who abandon the platform without leaving feedback will cause seller reputations to be biased. Using data from eBay, we document this externality and argue that platforms can mitigate it by actively screening sellers and promoting the prominence of better quality sellers. Exploiting the bias in feedback, we create a measure of seller quality and demonstrate the benefits of our approach through a controlled experiment that prioritizes better quality sellers to a random subset of buyers. We thus highlight the importance of externalities in two-sided markets and chart an agenda that aims to create more realistic models of two-sided markets.

*We thank many employees and executives at eBay whom without this research could not have been possible. We have benefited from feedback...

†University of Chicago and eBay Research Labs. Email: cnosko@chicagobooth.edu

‡UC Berkeley and eBay Research Labs. Email: stadelis@haas.berkeley.edu

1 Introduction

Decentralized marketplaces constitute some of the most fundamental building blocks of economic activity. Ebay, one of the first internet success stories, morphed from a used-goods auction site into one of the largest decentralized ecommerce platforms with over sixty billion dollars of merchandise changing hands in 2012. The growth of online anonymous two-sided markets is continuing rapidly with eBays global expansion, Amazons Marketplaces, and non-US based sites such as Taubau.com in China to name a few.

The anonymity of traders on two-sided platforms raises concerns about asymmetric information that may cause market failures. To overcome this problem, ecommerce marketplaces use some sort of decentralized "reputation (or feedback) mechanism. Unlike a large seller who bares the full impact of how its current performance affects its reputation and future sales, small decentralized sellers in a two sided market do not internalize the impact of their actions on the marketplace as a whole. In particular, if buyers are uncertain about the seller-quality distribution on a platform, then following a poor outcome, a buyer may update beliefs about the quality of *all* sellers on the platform. Furthermore, if buyers choose to leave the platform after a disappointing transaction without leaving feedback, the platform's reputation mechanism will not convey the actual history of sellers, being positively biased.

We study the challenges faced by market platforms when sellers exert externalities on the platform and when the reputation of these sellers does not accurately convey past performance, thus contributing to the literatures on two-sided markets and on reputation mechanisms. We explore the limits of reputation mechanisms in the face of these externalities, their adverse impacts on the marketplace, and ways in which a platform designer can mitigate some of these adverse impacts. We consider the situation where sellers have some agency over whether a transaction goes well. By providing higher (lower) quality to buyers, either through effort or innate ability, a seller increases (decreases) the likelihood that the buyer will come back and purchase from other sellers on the platform. Hence, a seller's effort generates reputational returns to the platform above and beyond the direct benefits that he internalizes, resulting in an externality that sellers exert on the platform.

This paper offers three contributions. First, using data from eBay, we demonstrate that the externality we describe exists and that feedback is disproportionately absent for transactions that cause buyers to leave the platform. Buyers are not only unwilling to repeatedly transact with a seller who executed a bad transaction, but they are also less willing to deal with the platform as a whole. Second, we propose a mechanism to mitigate the externality problem in which good-quality sellers are prioritized in search results. To better identify higher quality sellers, we exploit the absence of negative feedback to create a new measure of seller quality. This approach suggests that the design of online marketplaces may benefit from an "intermediate" ground between a *laissez faire* marketplace in which every seller is treated equally, and a heavily regulated marketplace that aggressively screens sellers. Last but not least, we conduct a field experiment where we change search result for a randomly chosen subset of buyers using our quality measure and we find that this approach increases the quality of transactions and the retention of buyers.

We begin our analysis by suggesting a simple conceptual framework of buyer behavior in online marketplaces. We then proceed to construct a longitudinal dataset using eBay transactions that follow a cohort of new buyers who joined eBay in 2008 and include all their transactions through August 2013. The data include every transaction made by this cohort, including characteristics of the item, it's price, and characteristics of the item's seller. The goal is then to proceed and measure how the quality of any particular transaction affects the future behavior of buyers on the platform.

We then show that the standard measure used to determine a seller's quality, his reputation feedback, is highly skewed and omits valuable information. The "percent positive (PP) measure for each seller is computed by dividing the number of transactions with positive feedback by the number of all transactions with any feedback for that seller. In our dataset, PP has a mean of 99.3% and a median of 100%, consistent with other studies that use eBay data (see, e.g., Dellarocas and Wood, 2005). This indeed seems highly skewed, suggesting that a central challenge in this paper is to construct a measure of seller quality that buyers cannot self-select on and that more accurately reflects a seller's true quality.[1] We construct a new quality measure that we call "effective percent positive (EPP) that is computed by dividing the number of transactions with positive feedback by the total number of all transactions for that seller, including those with no feedback. In our dataset, EPP has a mean of 69% and a median of 73%, offering much more variability than PP.

To support our approach we use our conceptual framework to empirically analyze the actual behavior of the cohort of buyers with respect to how current transactions affect their future behavior on eBay. In particular, we study the effect of a seller's EPP in a *current* transaction on the buyer's propensity to *continue* buying on eBay. As our framework suggests, a buyer who has a positive experience on eBay will be more likely to continue to transact on eBay again in the future and vice versa. Furthermore, Bayesian buyers should be less sensitive to their experience as they have more experiences on eBay. That is, the response to a negative experience early in a buyer's tenure on eBay should be more severe than later in his purchasing sequence. We confirm these hypotheses and prove that EPP is indeed a better measure of a sellers quality than other available measures.

To further quantify the effect of improving buyer experiences by prioritizing higher quality sellers, we report results from a controlled experiment on eBay that implements our suggested approach by incorporating EPP into eBays search-ranking algorithm. We select a random sample of ebay buyers who, when searching for goods on ebay, will be shown a list of products that promotes the EPP measure of the sellers compared to the control group in which this is not done. The results confirm the conclusions from the regression analysis described earlier and shows that treated buyers who have been exposed to higher EPP sellers are significantly more likely to return and purchase again on ebay compared to the control group of buyers.

Rather than propose to replace PP with EPP for the benefit of buyers to use, we suggest that online marketplaces use measures like EPP in more opaque ways that improve a buyers experience indirectly through the marketplace's search rank algorithm for two reasons. First,

[1]Beyond the observation that feedback is highly skewed, to measure the effect of quality on buyer behavior we cannot use measures of quality that are observable to the buyer because buyers will adjust their purchase behavior, which might affect seller actions. For instance, an observable lower quality seller might list a product for a lower price, and a buyer might knowingly take a risk on that item.

different buyers may interpret the same information in different ways. For one buyer a score of 88% might be satisfactory, while for another it is not, without having a clear understanding of how such a score translates into actual experiences. In theory, every rational expectations model of reputation has buyers being fully informed about the relationship between scores and outcomes, but in practice, and especially for less experienced buyers, such a mapping is unlikely to exist. Second, by using EPP instead of PP, sellers will most likely harass buyers who do not leave feedback in order to manipulate this new measure of seller quality.]

We conclude by discussing a more general agenda for studying reputation and quality in two-sided markets. We argue that factors such as seller externalities play a large and growing role in managing these platforms, and hence delivering welfare to consumers. The standard model that Rochet and Tirole (xxxx) propose for two-sided markets does not allow for externalities between agents on the same side of the market and concentrates on the binary decision of joining a platform or not. While these models may be appropriate for industries such as credit cards (the classic example used in many of these papers), the models cannot capture the complexity of relationships that exist on a large marketplace platforms. In light of the growing importance of two-sided online markets in the economy, we advocate shifting focus to more general models that can incorporate more complex market setups.

2 Conceptual Framework

We wish to distinguish between two possible scenarios for a platform like eBay. The first is that buyers see the platform as a means of gaining access to sellers, but they neither consider characteristics of the platform itself, nor do they believe that sellers on the other side of the platform represent of the platform as a whole. In this case, there are no externalities across sellers. A buyer updates on the quality of the seller that he interacted with, and if the transaction goes badly, he may not deal with that seller again, but this does not affect the buyer's willingness to transact with other sellers on the platform.

In the second scenario the buyer uses outcomes of individual transactions to form beliefs about the whole platform. To consider this, imagine a dynamic Bayesian decision problem of

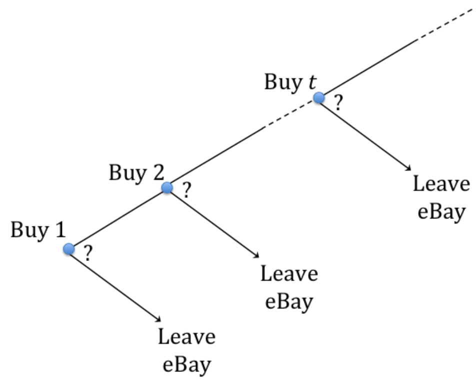

Figure 1: A Buyer's Dynamic Bayesian Decision Problem

a buyer who arrives at the ecommerce marketplace for the first time and is contemplating whether or not to purchase an item. His decision to purchase will rely on three basic elements: first, how much does he enjoy the site (user) experience; second, what are his expectations about the quality of the transaction; last, conditional on his belief, how price competitive is the site compared to other comparable marketplaces. If he decides to purchase, then after he receives the item he will update his beliefs about the quality of the site, and decide whether or not to purchase again, and so on, as depicted in Figure 1 .

Buyers can use a seller's feedback to form expectations about the quality of the seller, and by association, the marketplace overall. Every time the buyer makes a purchase, he collects an observation through which he updates his prior belief about the site's and the seller's expected quality. If the experiences were bad enough, he will update his belief about quality downward enough so as to decide to leave the site altogether. If, however, his experience was good, he will update his posterior in a positive way and continue to purchase from other sellers on the marketplace platform.

This framework of Bayesian updating also implies that the more transactions a buyer has made, the tighter will be his posterior, and this in turn implies that the effect of early experiences will be much more influential on the next purchase decision than an additional later experience. [2] It follows, therefore, that if a buyer experiences a relatively bad transaction earlier in the dynamic decision problem, then he is more likely to leave the marketplace than if he experiences the same transaction after several good experiences. This simple observation will form the basis for the central analysis on buyer behavior in Section 5.

3 Reputation and Transaction Quality at ebay

A large literature argues that reputation mechanisms mitigate inefficiencies in markets with asymmetric information.[3] By publicly revealing ratings from past transactions, sellers are punished for delivering bad quality through the loss of future business from other market participants. Sellers therefore have an incentive to not defraud buyers and ensure that transactions go smoothly even if they are unlikely to interact with that specific buyer again. Mechanisms that operate with these principles in mind have been credited with sustaining markets such as long distance trade during the Middle Ages (Greif, 1989) and are often cited as reasons that online anonymous markets, such as ebay, AirBnB and others, were able to come into existence in the first place (Dellarocas, 2003).

With a well-functioning reputation system, buyers can correctly infer the likelihood of a transaction going well without having interacted with that particular seller in the past, and no other measure of transaction quality. In practice, however, the extent to which reputation systems reveal seller quality depends on two important assumptions. First, that the public information correctly mirrors the quality of past transactions, and second, that buyers are able to correctly interpret reputation information. If either of these assumptions fail then buyers will have information that is an inaccurate measure of seller (or product) quality.

Ebay's reputation system is often described as a resounding success for two reasons. First, is the simple fact that eBay exists as a successful business despite the anonymity of the marketplace. The reputation system seems to be all that is standing in the way of a collapse

[2]This heuristic framework can easily be formalized using a standard dynamic model of a Bayesian decision maker that faces a distribution of quality with a well defined prior on the distribution of quality. Due to the well-understood nature of this dynamic problem, it would be redundant to offer the formal model.

[3]See Bar-Isaac and Tadelis (2008) for a survey.

of trust.[4] Second, many observable reputation characteristics correlate with our prior notions of outcomes that these measures should induce. For instance, sellers with higher reputation scores and more transactions receive higher prices for their products. Similarly, reputation seems to matter more for higher priced goods than for lower priced goods.[5]

Rather than explore the returns from reputation on ebay, we focus attention on the extent to which the observable reputation measures are a good indication of transaction quality, and, if not, to what extent can a better measure of seller quality be created. Once a more accurate measure of quality is established, we will explore whether the second function of a reputation system – to give ex-ante incentives to sellers to ensure high quality – is operating in the current system (i.e., externalities across sellers are insignificant), or whether a platform should intervene and use levers other than the reputation system to increase platform quality.

When buyers complete a transaction, they are given the option of leaving either a positive, negative, or neutral feedback score, or leave no feedback. About 65% percent of buyers leave feedback on a transaction. Ebay uses this information to provide two observable seller reputation measures. The first, percent positive (PP), is defined as the seller's number of positive feedbacks divided by the sum of his number of positives, neutrals and negatives.[6] The second is feedback score, and is a summed value of the number of positive feedbacks minus the number of negative feedbacks. Both of these measures are displayed when a user views detailed item information. Figure 2 shows what this looks like for two different sellers. Seller A as a percent positive score of 96.9 and a seller feedback score of 317, while seller B has a percent positive of 99.5 and a seller feedback score of 44949.

Using the observable eBay reputation measures to examine the effect of seller quality on future outcomes is problematic for two reasons. First, buyers select the sellers they purchase from, leading to bias in any estimates of long term benefits from self-selecting into different

[4] As Dellarocas (2003) puts it, "eBays impressive commercial success seems to indicate that its feedback mechanism has succeeded in achieving its primary objective."

[5] See Bajari and Hortacsu (2004) and Hortacsu and Cabral (2010) for more on these facts.

[6] To be precise, these numbers only look back at the last 12 months of a transaction for a seller and exclude repeat feedback from the same buyer for purchases done within the same calendar week.

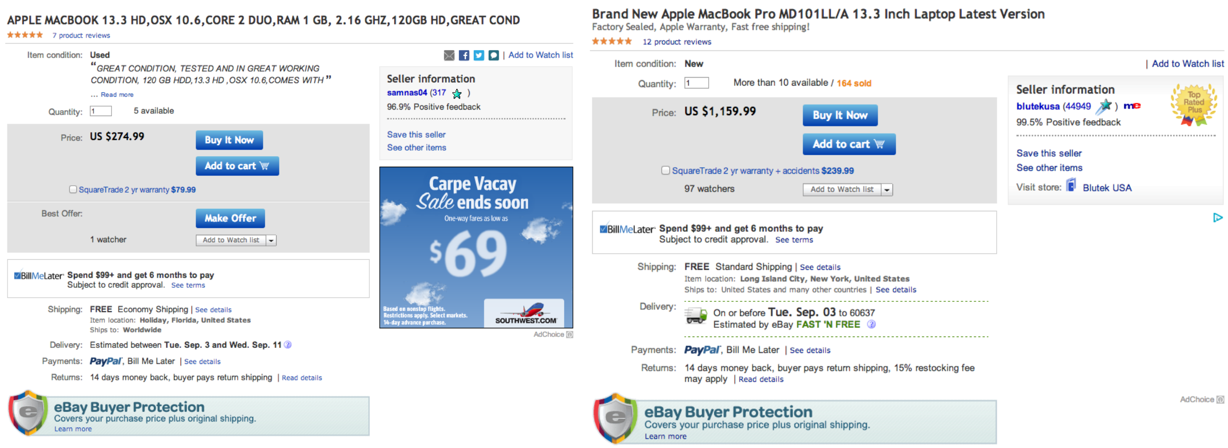

Figure 2: Seller reputation information as displayed to buyers

quality sellers. One clear example is that price may adjust to reflect lower quality reputations. Indeed, there is much evidence that this happens.

Second, the PP measure is highly skewed and hard to parse. Figure 3, displays the histogram of seller PP from a cross section of sellers in October 2011.[7] The X-axis starts at 98.21%, which is the tenth percentile.[8] The median seller has a score of 100%. This could be indicative of a reputation system that works extremely well – bad sellers are selected out when their score falls low enough, leading to a high positive selection. Unfortunately, this is not the case. Using data from ebay's U.S. marketplace, out of over 44 million transactions completed in October of 2011, only 0.39% had negative feedback, while at the same time, over 1% had an actual dispute ticket opened within the eBay system (a step that takes substantially more effort on a buyer's part than leaving negative feedback). This indicates that there are a substantial number of transactions that went badly for which negative feedback was not left.[9]

Another problem is that many buyers may have trouble interpreting the numbers they are presented with. Naively, one may think that a score of 98% is excellent (in some sort of

[7]Our main dataset contains a panel of buyers and tracks their behavior over time. Here, we display information from a separate dataset of all transactions that occurred on the U.S. site in October of 2011. This demonstrates a cross sectional view at one point in time.

[8]There are some newer sellers with a percent positive of 0. Leaving them in would skew the graph such that the detail at the top end could not be seen.

[9]This in fact proves the central tenet in Dellarocas and Wood (2005) who claim that silence in the feedback system includes many transactions for which buyers had bad experiences but chose not to report them.

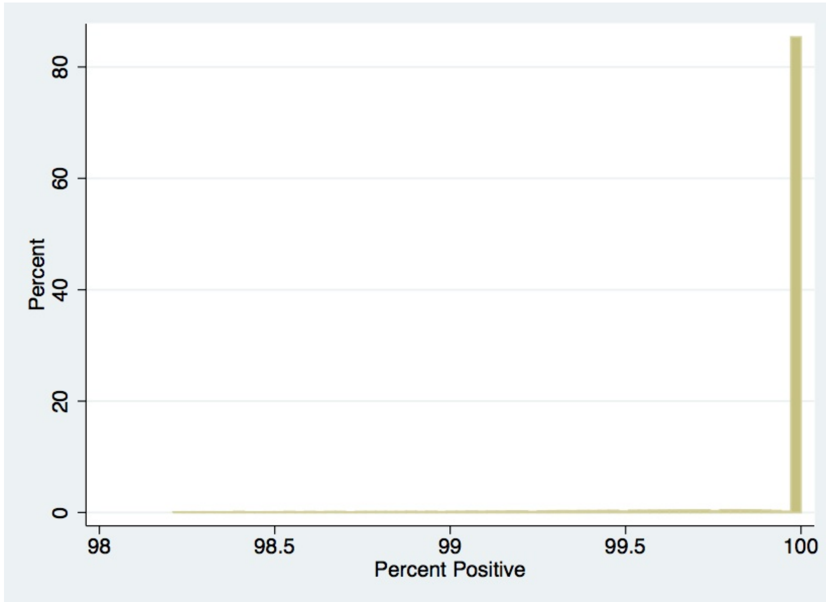

Figure 3: Percent Positive of Cross Section of Sellers from Oct 2011

absolute scale). In reality, a score of 98% places a seller below the tenth percentile of the distribution. This effect may be especially pronounced for new users who may not have seen enough sellers to be able to judge the scale accurately, a point to which we return below.

Arguably, buyers do not leave negative feedbacks because it is not anonymous and sellers historically reacted by reciprocating.[10] Anecdotal evidence shows that sellers sometimes react badly to negative feedback, often harassing buyers in an attempt to get them to change it.[11] In part because of this, a social norm has developed around not leaving negative feedback.

We proceed to construct a measure of *u*nobservable seller quality based on the idea that buyers who experience a bad transaction are less likely to leave negative feedback, and are silent instead, while buyers who experience a good transaction are more likely to leave positive

[10]Up until 2008, both parties could leave negative feedback, and after that sellers can only leave positive feedback or no feedback. There is a long history of reciprocal feedback behavior before the 2008 change as documented by Bolton et al. (2013).

[11]In one case, a seller called the buyer and threatened him after his negative feedback. ("eBay Shopper Says He Was Harassed By Seller," http://www.thedenverchannel.com/lifestyle/technology/ebay-shopper-says-he-was-harassed-by-seller). In another case, a buyer was sued for leaving negative feedback ("eBay buyer sued for defamation after leaving negative feedback on auction site," http://www.dailymail.co.uk/news/article-1265490/eBay-buyer-sued-defamation-leaving-negative-feedback-auction-site.html.)

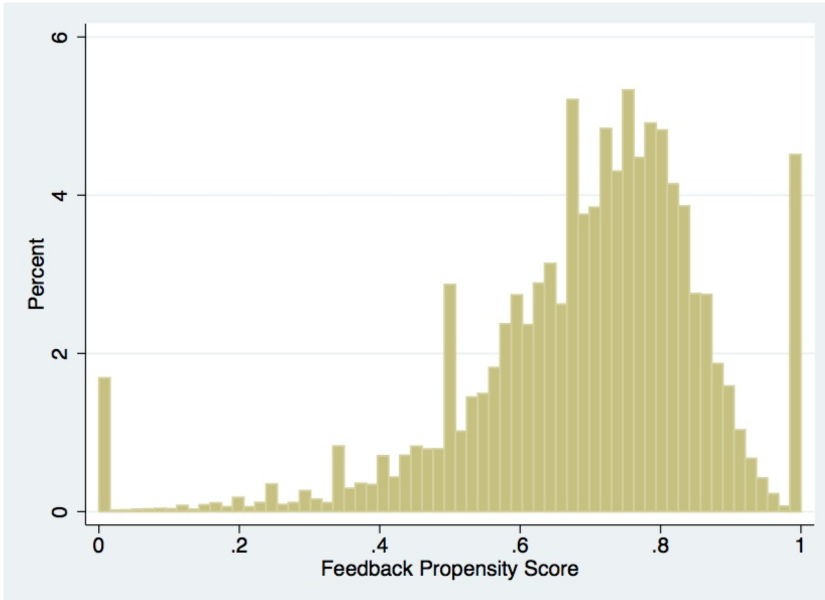

Figure 4: Histogram of Sellers' Effective Percent Positive Scores

feedback.[12] To operationalize this, we measure the propensity of positive feedback to be left for any given seller. Controlling for observable feedback measures (PP and feedback score), we conjecture that a seller with a lower propensity to collect positive feedback will be more likely to deliver a worse experience.

Walking through a specific example, consider two sellers, A who had 120 transactions and and B who had 150. Assume that both received one negative feedback, and 99 positive feedbacks. Using eBays PP measure, both have a PP of 99% ($\frac{99}{99+1}$). However, seller A had only 20 silent transactions with no feedback while seller B had 50 silent transactions. We define "effective PP (EPP) as the number of positive feedback divided by total transactions, in which case seller A has an EPP of 82.5% while seller B has an EPP of only 66% and is a worse seller on average. Importantly, eBay does not display the total number of transactions a seller has completed and buyers cannot therefore back-out a seller's EPP score.

[12]For example, consider a set-up in which there is a distribution of "public mindedness" among individuals that compels them to enjoy leaving feedback for the benefit of future buyers. If the costs and benefits of leaving feedback would not depend on the quality of the transaction, then the feedback left should be unbiased. However, if the cost of leaving truthful feedback is higher for bad transactions due to the harassment costs, then such a skew in feedback will result.

Figure 4 displays a histogram of EPP scores for the cross section of sellers mentioned above. The spikes at simple fractions (0, .5, 1) come from newer sellers with small number of transactions. The mean of this distribution is .69 and the median is .72. Importantly, unlike percent positive, there is substantial spread in the distribution.

To verify that EPP contains information about buyers' experiences, we define a bad buyer experience (BBE) as one in which the buyer either left negative feedback, opened a dispute with eBay, or left low stars on the detailed seller ratings. Out of the 16,062,135 transactions in our full dataset, 487,780 (3.04 percent) of them resulted in BBEs. Despite the belief that buyers are reluctant to report bad experiences either because of seller wrath or because of the time cost of going through the eBay dispute process, table 1 explores whether their is a relationship between BBEs and seller quality as measured both by observables and by EPP. We run a probit regression of BBE on seller quality scores and controls for price, category, and purchase type (auction or fixed price). All of the coefficients are highly statistically significant. The coefficient of interest is the one on EPP, which is negative and highly significant, indicating that transacting with higher EPP sellers decreases the probability that a BBE will occur, consistent with EPP being a measure of seller quality.

Table 1: Probit of Bad Buyer Experiences and EPP

BBE Flag (0/1)	
Percent Positive	-2.554***
	0.0136
Seller Feedback Score	-0.000000126***
	4.44e-09
EPP	-0.711***
	0.00517
Item Price	0.104***
	0.00119
cons	1.130***
	0.0122
N	15,654,528

Even though EPP is unobservable to buyers, perhaps buyers observe signals that are correlated with EPP, questioning its exogeneity. To explore this we consider whether more experienced buyers differentially select into transacting with higher EPP sellers. Table 2 shows the results of an OLS regression where EPP is the left hand side variable. The variable Buyer Transaction Number is the number of transactions that a buyer has completed on eBay. If more experienced buyers were transacting with higher EPP sellers, we would be worried about the selection story. That does not appear to be the case. Although the variable is statistically significant, its magnitude is negligible.

Table 2: Selection into EPP

	b/se
Percent Positive	1.520***
	0.00100
Seller Feedback Score	-6.01e-08***
	1.74e-10
Buyer Transaction Number	-0.000000391***
	2.29e-08
Item Price	-0.0000176***
	0.000000180
cons	-0.794***
	0.000996
N	15,803,078

4 Data

We collected data for a cohort of new users to the U.S. site of ebay.com who joined the platform anytime in 2008 and tracked all of their usage behavior through time. Our selection process users who both signed up for a new eBay account and purchased an item within 30 days of setting up that account.[13] The 2008 cohort includes over seven million new users,

[13] We have also replicated the analysis in this paper for the 2009 and 2010 cohorts with very similar results.

making their whole transaction history too large for meaningful analysis. As a result, we take a 10% random sample and analyze the behavior of 729,945 buyers.

For each buyer we track all transactions starting with their initial sign up until August 1, 2013. That results in 16,062,135 observations – on average 22 transactions per user. Each observation contains rich information about the transaction, including but not limited to price, item category, title, the seller, whether it was an auction or fixed price, and quantity purchased. There were a total of 2,327,283 sellers associated with all of the buyer transactions. We also collect information on the seller that each buyer transacted with.

Basic seller information such as the feedback score displayed to the user, percent positive, the number of transactions the seller had in the past, are available in the eBay data-warehouse. For each transaction we look backward at the point right before the transaction took place and construct a seller EPP measure for that particular seller. We do this by looking back at all of the seller's transactions on the site (capped in January of 2005, the earliest the eBay data-warehouse stores) up to the point right before the transaction and then dividing the number of positive feedbacks by the total number of transactions for that seller. This generates a complete snapshot of the information structure at the point when the buyer was making his decision and, as such, we do not include the focal transaction in the measure. Recall that the buyers do not observe and cannot back out the EPP measure.

Figure 5 is a histogram of the total number of transactions by an individual buyer over the course of his tenure, truncated at 30 transactions. A large percentage of eBay buyers make very few purchases over their life-cycle – a full 35% of new buyers purchase once and never purchase again, with an additional 13% purchasing twice and moving on. On the other end of the spectrum, there is an extremely large right tail. While the median number of transactions is 3, the mean is 22, the 95th percentile is 88, and the max is 26,921.

5 Externalities Across Sellers

Recall that we are first interested to separate between two scenarios: in the first, buyers use eBay merely to connect with certain sellers, in which externalities across sellers are not

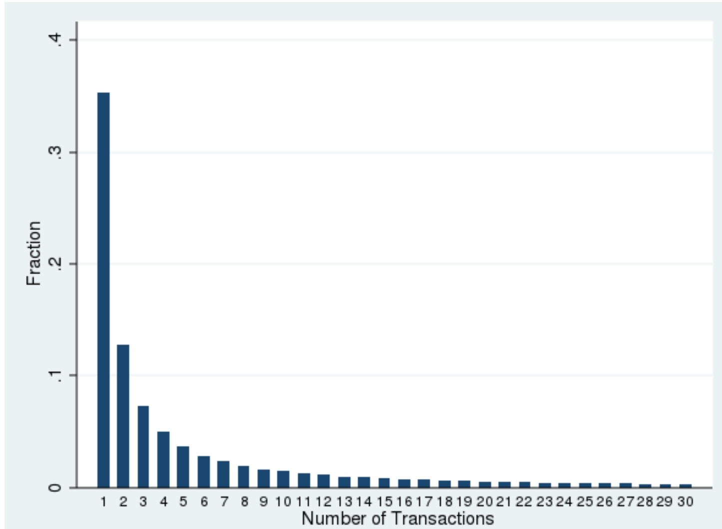

Figure 5: Histogram of Total Transactions by Buyer

present. In the second, buyers consider eBay as a provider of quality services in and of itself, in which case they infer the quality of the platform from individual transaction and an externality across sellers exists.

5.1 Buyer Behavior

Table 3 is a cross tabulation by buyer of how the total number of transactions relates to the total number of sellers that a buyer interacted with. Out of 32,016 buyers who had a total of 20-29 transactions during our sample period, 20,742 of them bought from between 20 and 29 different sellers while 12 of them bought all their transactions from a single seller.[14] This shows that buyers tend to deal with large numbers of sellers and therefore suggests that externalities may indeed exist.

To further investigate how buyers interact with sellers, column 1 of table 4 explores the relationship between a buyer reporting a BBE (bad buyer experience) and the probability that the buyer purchases from the same seller again. We control for the sale type (auction or fixed price), the category, the item price, and the number of repeat transactions the buyer

[14] The total number of buyers adds up to 666,772 instead of 729,945 because we have truncated at buyers with 49 total transactions or less for compactness.

Table 3: Cross Tabulation of Total Transactions by Total Number of Sellers for that buyer

	00-01	02-05	06-09	10-19	20-29	30-49	Total
Tot Trans	Total Number of Sellers						
00-01	258,078	0	0	0	0	0	258,078
02-05	18,219	192,402	0	0	0	0	210,621
06-09	1,041	13,226	50,410	0	0	0	64,677
10-19	527	2,141	11,843	54,289	0	0	68,800
20-29	120	351	547	10,256	20,742	0	32,016
30-49	78	206	229	1,527	9,158	21,382	32,580
Total	278,063	208,326	63,029	66,072	29,900	21,382	666,772

had with the seller up to the focal transaction (standard errors are clustered at the individual level). The coefficient from a probit estimate of BBE on repeat purchase indicates that experiencing a BBE with a seller leads to a smaller chance that the buyer will purchase from the same seller in the future.

The concern with using whether the buyer experienced a BBE as a regressor is that it is a self reported variable, implying that it is not a true reflection of transaction quality. If someone reports a BBE then it is natural to assume that the transaction went poorly, but if they do not report a BBE then it is not clear that the transaction went well. We therefore use EPP to obtain an unselected measure of seller quality. Column 2 of table 4 replaces the BBE variable with EPP . The coefficient is positive and significant, indicating that sellers with higher unobservable quality do generate more sales from a given individual. This indicates that sellers have an incentive to execute good transactions. The evidence in Table 3, however, shows that repeat purchases from the same buyer are relatively rare, which in turn suggests that externalities across sellers exist and the incentive to execute good transactions is muted.

5.2 Externalities across Sellers

In order to explore the extent of seller externalities, we ask how transaction quality affects the probability that a buyer will transact on the platform as a whole again. If, as we conjectured

Table 4: Likelihood of Returning to the Same Seller

	b/se	b/se
Seller Feedback Score	0.000000162***	0.000000196***
	3.00e-09	2.87e-09
Percent Positive	2.238***	1.482***
	0.0835	0.0714
Experienced BBE	-0.114***	
	0.00584	
EPP		0.468***
		0.0159
Item Price	-0.110***	-0.102***
	0.00445	0.00441
Total Transactions 2-5	1.170***	1.166***
	0.00347	0.00346
Total Transactions 6-9	2.133***	2.125***
	0.00601	0.00606
Total Transactions 10-19	2.527***	2.520***
	0.00792	0.00792
Total Transactions 20-29	2.875***	2.871***
	0.0133	0.0132
Total Transactions 30-49	3.087***	3.084***
	0.0140	0.0141
Total Transactions 50+	3.423***	3.428***
	0.0186	0.0190
Cons	-3.592***	-3.164***
	0.0830	0.0688
N	13065655	12928971

earlier in Section 2, buyers learn about the platform from the quality of any given transaction, then a higher quality transaction will cause more purchases overall for the platform and create a disconnect between the incentives for an individual seller and the platform as a whole.

Our econometric specifications include probit regression of the following form,

$$y_{ijt} = \alpha_0 + \alpha_1 EPP_{jt} + \beta \cdot \bar{b}_{it} + \gamma \cdot \bar{s}_{jt} + \delta \cdot \bar{d}_t + \varepsilon_{ijt}, \quad (1)$$

where y_{ijt} is an indicator equal to 1 if buyer i buys again on ebay after purchasing transaction t from seller j, EPP_{jt} is the EPP of seller j at transaction t, $\bar{b}_{it}$ is a vector of buyer characteristics such as how many transactions they completed so far, $\bar{s}_{jt}$ is a vector of seller characteristics such as PP, and $\bar{d}_t$ is a vector of transaction characteristics.

Because buyers do not observe EPP and do not act on it (recall Table 2), it can be considered as an exogenous shock to the quality of the seller. The higher the EPP, the more likely it is that the transaction goes well and therefore the more likely the buyer is to return and purchase on eBay – consistent with our conceptual framework outlined in Section 2.

Table 5: Baseline EPP Regressions

	60 Day Return b/se	180 Day Return b/se	Any Return b/se
Seller Feedback Score	2.22e-10	2.14e-08***	-1.78e-08***
	2.94e-09	3.58e-09	4.37e-09
Percent Positive	-0.709***	-0.588***	-0.975***
	0.0200	0.0219	0.0263
EPP	0.482***	0.461***	0.826***
	0.00576	0.00607	0.00659
Item Price	-0.157***	-0.141***	-0.108***
	0.00119	0.00120	0.00125
Total Transactions	0.0158***	0.0229***	0.0268***
	0.0000326	0.0000539	0.0000680
Constant	0.588***	0.597***	1.054***
	0.0188	0.0205	0.0248
N	7734067	7734067	7734067

Table 5 is our baseline regression table. In column, $y_{ijt} = 1$ if the buyer returns within 60 days of a given purchase, in column 2 it is within 180 days, and in column 3 is if they had any repeat purchase in the sample period. All of the standard errors are clustered at the individual level and all of the regressions include controls for transaction type and category, although the coefficients are not shown for brevity.

The coefficient estimates are stable across the probit specifications and most results confirm in the expected relationship. Buyers who end up purchasing more (total transactions) are more likely on any given purchase to return.[15] Our measure of seller quality, EPP, is highly statistically and economically significant (magnitudes will be discussed below), indicating that when buyers end up with a higher quality seller, they stick around and purchase more both in terms of frequency and revenue.

Interestingly, the observable seller reputation measures are either unstable (Seller Feedback Score) or negative (Percent Positive). We interpret this as selection: Once we control for seller quality (by including EPP), buyers self select into different types of sellers based on their ability to interpret quality or on their initial intentions. Perhaps, someone who knows eBay and is planning on sticking around may be more willing to take a risk with a lower Percent Positive seller, whereas someone who is weary about ebay and does not plan to return may only choose to buy from 100% PP sellers.

5.3 Buyer Learning About the Platform

While there are many reasons that a buyer may make a purchase on eBay and not return in the future, the above regressions help us distinguish between two important ones. The first is one of selection: People come to eBay looking for a specific item, purchase that item and then have no need to return in the future. Such might be true for a rare antique or collectible. The second story, and the one that is new to this paper, is that buyers come to eBay with limited knowledge about the platform, and update beliefs over time. Transacting with a low quality seller influences their decision to come back to the platform as a whole.

[15] We include total transactions as a regressor to control for the selection of who ends up staying on eBay for long periods of time because of their taste for what ebay offers.

The dynamic updating framework described in Section 2 implies that transaction quality should matter less for more experienced buyers relative to newer ones. Consider a buyer who first comes to eBay with limited knowledge about (1) How much he is going to like the overall experience of eBay (site design, checkout system, etc.), or (2) what the distribution of seller quality is on the site. In either of these cases, experience is going to help a buyer learn about his idiosyncratic match value with the site and understand that a bad draw from the seller distribution is one draw of information. As such, as a buyer becomes more experienced he is more likely, on average, to return to the site *both* because of selection (people who had low match values have left) and because a bad draw later will have less of an influence on his beliefs because of a more refined posterior belief.

Table 6 cross tabulates the transaction number of an individual buyer (i.e., the transaction number of the buyer's tenure) against whether or not a buyer returns to eBay and purchases again within 180 days. The table clearly shows that as a buyer becomes more experienced, he is much more likely to return to eBay and purchase, consistent with a story of either selection into who continues on eBay or learning about the idiosyncratic match with the platform.

Next consider the effect that the quality of a transaction has as a buyer makes more purchases. Table 7 extends our baseline regression in Table 5 to include the buyer's experience, measured by how many transaction he completed. As the results on the interaction effects between experience (Transaction Cat) and EPP show, more experienced buyers are less effected by the EPP of the seller, implying that they are not responding as much to new information.[16] This is consistent with the dynamic Bayes framework outlined is Section 2.

5.4 Counterfactual

To quantify the extent to which seller externalities exist, we look at how much consumer behavior would change if ebay could shift buyer purchases from low quality (EPP) sellers to high quality sellers. Specifically, we look at the shift in the distribution of terminal nodes in

[16]The regression includes the uninteracted experience measures but these are not reported for brevity. Not surprisingly, and consistent with Table 6.

Table 6: Cross Tabulation of Transaction Number with Probability of Purchasing within 180 days

	No Return	Return	Total
01-05	736,619	1,410,065	2,146,684
	34.31	65.69	100.00
06-09	140,481	781,441	921,922
	15.24	84.76	100.00
10-19	151,700	1,433,135	1,584,835
	9.57	90.43	100.00
20-29	65,936	1,042,218	1,108,154
	5.95	94.05	100.00
30-49	60,923	1,490,097	1,551,020
	3.93	96.07	100.00
50-99	49,755	2,176,415	2,226,170
	2.24	97.76	100.00
100+	41,393	6,322,824	6,364,217
	0.65	99.35	100.00
Total	1,246,807	14,656,195	15,903,002
	7.84	92.16	100.00

Table 7: Baseline Regressions including Buyer Transaction Number

	60 Day Return	180 Day Return	Any Return	
	b/se	b/se	b/se	b/se
Seller Feedback Score	-6.82e-08***	-5.26e-08***	-6.82e-08***	
	2.67e-09	3.13e-09	2.67e-09	
Percent Positive	-1.554***	-1.414***	-1.554***	
	0.0210	0.0226	0.0210	
EPP	0.909***	0.862***	0.909***	
	0.00798	0.00792	0.00798	
Item Price	-0.162***	-0.153***	-0.162***	
	0.00120	0.00113	0.00120	
Trans Cat 06-09*EPP	0.379***	0.372***	0.379***	
	0.0132	0.0141	0.0132	
Trans Cat 10-19*EPP	0.387***	0.352***	0.387***	
	0.0121	0.0130	0.0121	
Trans Cat 20-29*EPP	0.354***	0.338***	0.354***	
	0.0146	0.0166	0.0146	
Trans Cat 30-49*EPP	0.272***	0.320***	0.272***	
	0.0144	0.0164	0.0144	
Trans Cat 50-99*EPP	0.131***	0.294***	0.131***	
	0.0148	0.0166	0.0148	
Trans Cat 100-199*EPP	-0.00192	0.343***	-0.00192	
	0.0191	0.0206	0.0191	
Trans Cat 200-999*EPP	-0.137***	0.267***	-0.137***	
	0.0272	0.0267	0.0272	
Trans Cat 1000+*EPP	-0.274**	0.0669	-0.274**	
	0.0964	0.0873	0.0964	
cons	1.309***	1.387***	1.309***	
	0.0199	0.0214	0.0199	
N	15654527	15654527	15654527	

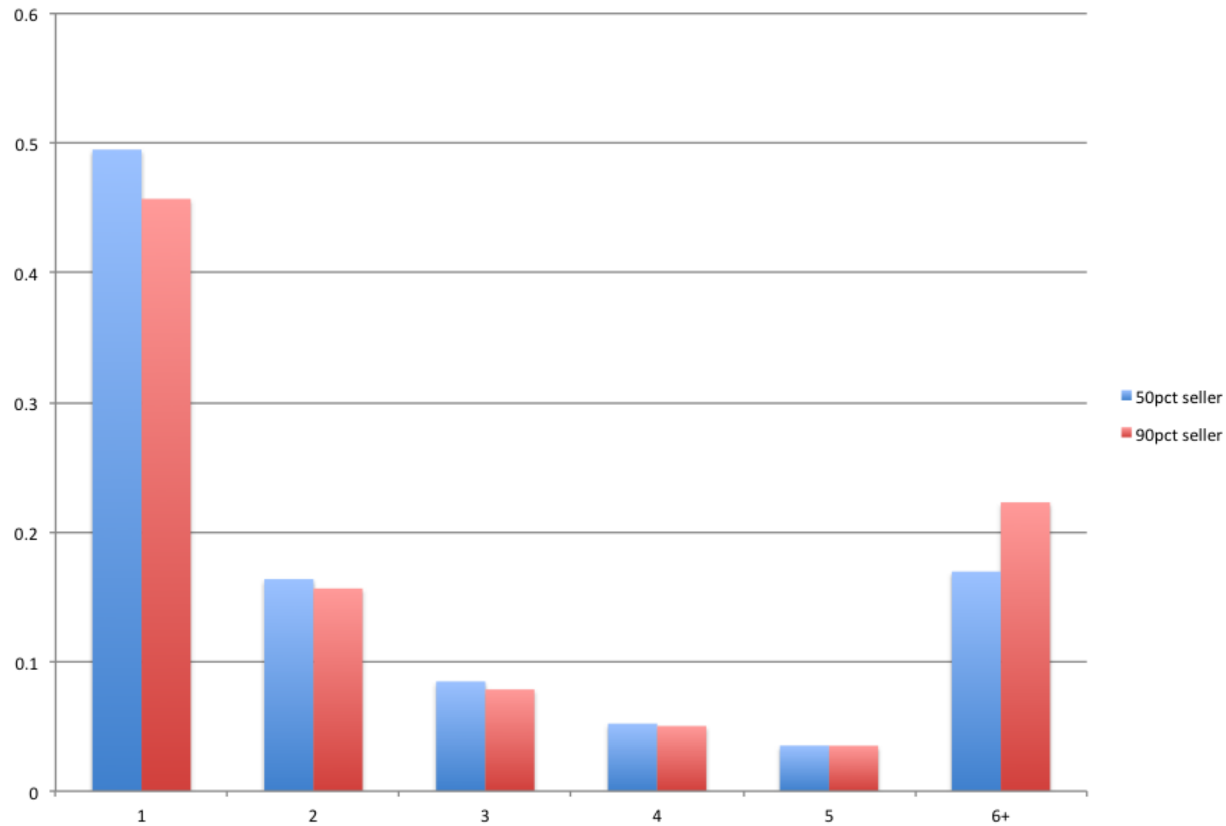

Figure 6: Counterfactual Distribution of Terminal Nodes

Figure 2 (i.e., when a buyer leaves eBay and doesn't return), if we could shift buyers from the median EPP seller to the 90th percentile EPP seller. Figure 6 plots this distribution.

The blue bars indicate the percentage of buyers that would end up at each of the first five nodes (leaving after 1,2,...5 transactions), and collapsing all future nodes (6+) were they to transact with a median EPP seller; the red bars does the same with the 90th percentile EPP seller. This is created by fitting the regression from column 3 of 5, assuming that all buyers interact with a seller of the given type. This then generates a probability at each node of returning back to eBay. Table 8 lists the numbers that went into this table. Columns 2 and 3 are the fitted values assuming the median seller and the 90th percentile seller, respectively. It's the difference between these values that generates the difference between the blue and red bars. Columns 3 and 4 reverse the probabilities to get the probability of not returning (or churning).

Table 8: **Counterfactual Distribution of Terminal Nodes**

Transaction #	p(buy) Median	p(buy) 90%	p(not buy | trans) Med	p(not buy | trans) 90%
1	0.505	0.543	0.4950	0.4570
2	0.676	0.712	0.1636	0.1564
3	0.752	0.797	0.0847	0.0785
4	0.797	0.837	0.0521	0.0502
5	0.828	0.864	0.0352	0.0351
6+	0.8436	0.8768	0.1694	0.2228

The exercise we are conducting is simple, but we think illustrative.

TO BE COMPLETED

6 Using Search to Internalize Seller Quality

There are many mechanisms that a platform can use to achieve the goal of matching buyers with better quality sellers. At a draconian extreme, the platform can expel any seller from the platform once it learns that a transaction was less than perfect. This would almost surely result in a selected pool of high quality sellers, pushing buyers to move toward transacting with better sellers. In would also strongly incentivize sellers to ensure that transactions went well according to the metric used to determine being kicked off of the site. The downside would be a number of sellers mistakently getting booted off the platform, along with a reduction in the overall site inventory.

Instead, eBay has traditionally adopted a very laissez-faire approach to managing sellers on its marketplace. Indeed, its very ethos rests on buyers and sellers being good "citizens." For many years this ethos, along with the reputation system, was incorporated by corporate eBay into policies that can best be described as Caveat Emptor. This has begun to change recently, with eBay taking a more active role in managing the marketplace.[17] These measures include more actively seeking to weed out bad quality sellers (to a much smaller extent than the extreme described above) and creating a "buyer protection" program that allows buyers

[17]See, for example, "EBay to Get More Involved in Transaction Disputes", http://www.pcworld.com/article/163099/article.html

to voice complaints to the platform directly about a transaction for potential reimbursement, rather than having to go through the individual sellers.[18]

We suggest an intermediate approach and ask to what extent online platforms can leverage their search technology to shift buyer purchase behavior. At one extreme, if a seller never showed up in search results then no buyers would find him and the policy would effectively remove the seller from the site. On the other extreme, a laissez-faire approach is to only take into account seller quality to a very limited extent in prioritizing search results. We report results from an experiment where we incorporate our measure of seller quality, EPP, into ebay's search algorithm so that higher EPP sellers get displayed higher on the search results page (holding everything else constant). This form of intervention into the marketplace provides a large amount of variation in the extent to which sellers are prioritized (by varying the weighting on seller quality).

In order for this strategy to work, buyers must face some search cost – they must actively be more likely to select an item higher up on the search results page. We note that the literature on search costs has demonstrated correlation between ranking and purchase (or click-through) behavior, i.e., (Ghose, Ipeirotis, and Li 2013), but to the best of our knowledge, the experimental evidence we present (although in the context of making a separate point), is some of the first to show how buyers respond to truly exogenous shifts in search rankings.

A potential cost of this strategy is that products may be harder for users to find, i.e., we impose search costs on users that do not exist with the ranking scheme that focuses on item relevance. Thus, one component of this experiment is to estimate the trade off inherent in manipulating search results to prioritize better quality sellers. On the hand, there could be a longterm benefit from buyers interacting with better quality sellers and returning to the site

[18]Every platform implicitly or explicitly takes a stand on how actively they manage their marketplace. While a company like eBay has traditionally been hands off, Amazon has always extensively pruned its seller pool on Amazon Marketplace, making them jump through hoops to join, holding transaction receipts in escrow, and kicking sellers off the site quite quickly. Similarly, Stubhub (now an eBay company) has been much more careful to control the buyer experience. Buyers purchasing on Stubhub are not even aware of which seller they are purchasing from and all disputes are handled with Stubhub directly. These policies completely negate the need for a reputation system and essentially mean there are no externalities across sellers (all of this is internalized by the platform).

more often in the future. On the other hand, buyers may be less likely to purchase because they have a harder time finding the product, or price that they are looking for.

The experiment we conduct allows us to do three things: (1) Answer any lingering doubts about the exogeneity of EPP as an unobserved measure of seller quality; (2) Explore the extent to which consumers respond to search ranking schemes, and hence how effective changes in them might be for platforms wishing to internalize seller quality externalities; and (3) Quantify the downside of using search rankings – the extent to which consumers do not purchase because they are unable to find the product they are looking for.

From December 14th, 2011 though January 2, 2012, 10% of ebay's U.S. site traffic—about 5 million searches per day—was placed into our experimental treatment and exposed to a ranking scheme that differed from the default site rankings. Because of other site considerations, we had limited control over the weighting that the EPP measure received, a point to which we will return below.

The ideal experiment would treat active users, ensuring that an individual user was always in the treatment group whenever he searched for a product on eBay. However, it is not always possible to unambiguously link a site visit to a specific user either because the user visits the site from a computer or browser he has never signed in from before, or because he has deleted his cookies (which is the way sites keep track of users between visits). This creates the potential for leakage between the treatment and control groups, where a user is sometimes exposed to search results in the treatment group and sometimes from the control (normal) group.

In order to understand the magnitude of this problem, it is necessary to understand how eBay (and most online entities) links site visits to users. Ebay stores a Globally Unique Identifier (GUID) in a cookie on browsers that visit the site, allowing ebay to track whether the same browser visits the site in the future. Through a complex algorithm, eBay attempts to match GUIDs to user IDs (UIDs) by tracking whether that browser was used to sign into an eBay account at any time (thereby matching GUID to UID). Multiple GUIDs may be linked to the same UID if a user signs in from multiple browsers on the same computer or from multiple computers. Our experiment was run at the GUID level, meaning 10% of active GUIDs were placed into the treatment group. This means that a user could be placed into the treatment group for one, but perhaps not all, of the GUIDs linked to a their user account. Fortunately, we can track this behavior and observe the number of searches that a user made within the treatment and control groups, allowing for a partial correction to the problem.

We collected data on all searches done during the experimental period (including the search query and the items that were displayed to the user), whether or not the GUID was in the treatment or control group, any other site behavior that the user did (clicks on products, etc.), and purchases both during the experimental period and after the experimental period. It's important that we are able to track purchases beyond the end of the experiment, because this is our measure of buyer satisfaction. In other words, we wish to track whether a buyer is more likely to come back and purchases again if randomly assigned to the treatment group, conditional on purchasing in the experimental period.

Because an individual search returns a number of products (typically 50 on a single page), and each of those items is associated with a seller (and hence an EPP score), we need a way to collapse a whole search into one single measure of the quality of sellers in the search. We choose to do this with a weighted average we call "discounted search EPP", where the discounting comes from weighting by the position of the item is the search results. Specifically, for any given search we weight all of the EPP scores from items that were displayed to the user by the inverse of the item's position on the search results page. This reflects the prevailing belief that items ranked higher up on the page are more visible, and hence play a larger role in the user's decision process.

Figure 7 shows a kernel density plot of the discounted EPP scores for all searches in the treatment and control groups. The mean in the control group is 59 and the mean in the treatment group is 64. The distributions are statistically different from each other, meaning that the experiment did in fact change rankings, but given the spread, the movement in the mean is not very large. Consequently, simply using the pure experimental treatment doesn't have a significant impact on outcomes. Instead, we look for heterogeneity across search terms.

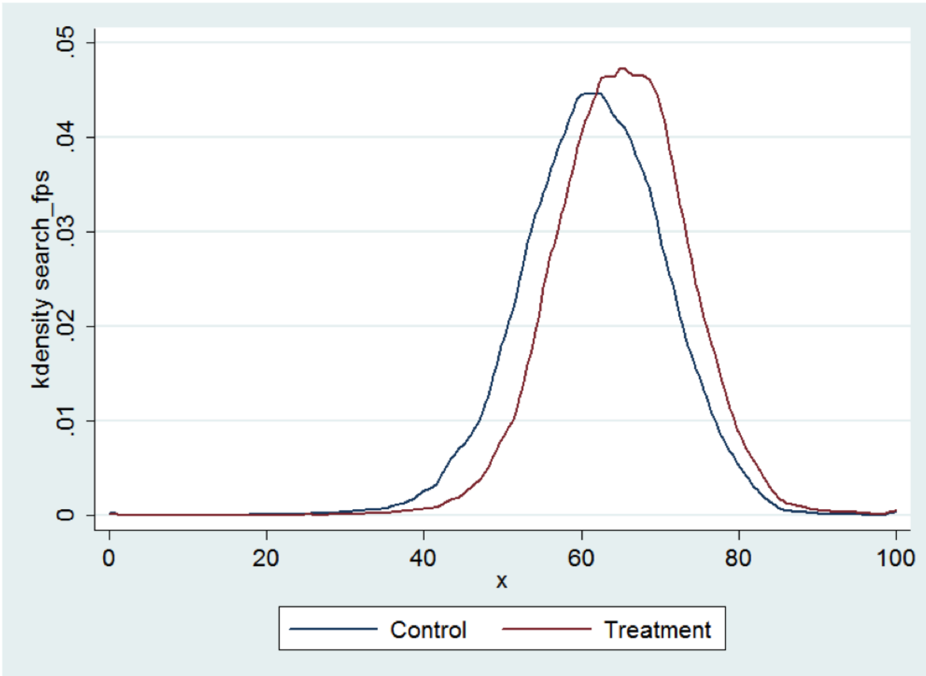

Figure 7: Discounted Seller EPP Scores between Treatment and Control Groups

The idea is that for some search terms EPP was increased idiosyncratically more than for other search terms. This happens because some search terms happen to have sellers of lower or higher quality in them and hence will be more affected by the experimental weighting scheme. To look for this heterogeneity, we match search terms that appeared in both the control and treatment groups and were searched within one day (although we've adjusted this to different lengths and the results are similar).

Figure 8 plots the percentage difference in discounted EPP between the matched queries in control versus treatment. There is substantial heterogeneity across different search terms, giving us hope that we can pick up differences in user behavior even with an experiment where the mean treatment effect was relatively small.

Table 9 regresses probability of returning again on controls and whether or not the user was in the treatment group. We restrict the sample to users whose search term could be matched between treatment and control groups (i.e., someone in the control group searched for exactly the same query as someone in the treatment group) and where there was at least a 4 percent difference in discounted EPP. Importantly the regression conditions on purchasing

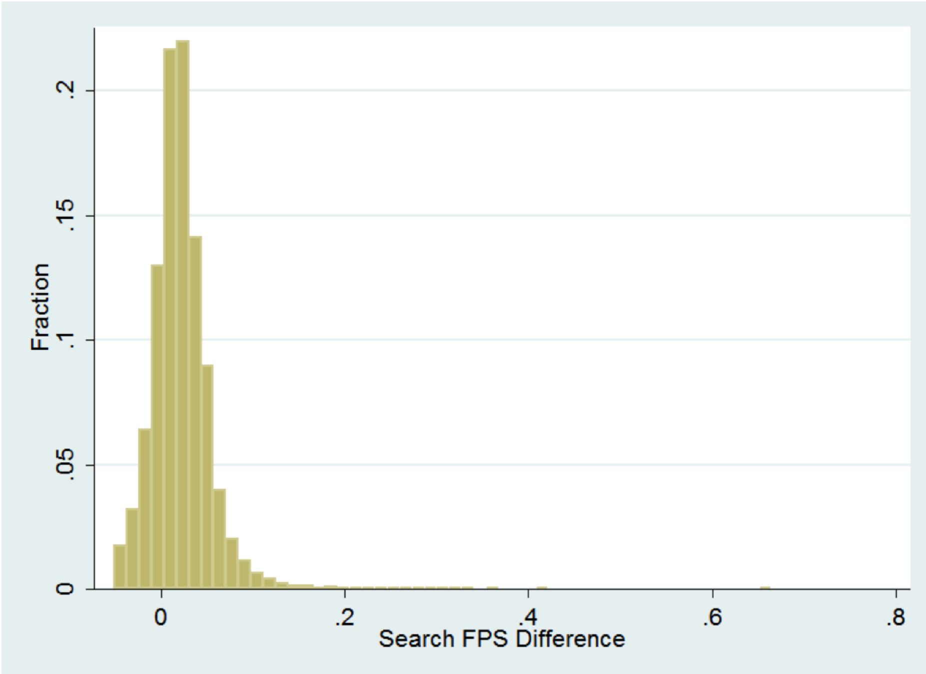

Figure 8: Percentage Difference between Discounted EPP scores for Matched Queries

Table 9: Regression of probability of purchasing again on being in the Treatment Group

Probit regression					Number of obs	24540
					LR chi2(4)	77.74
					Prob > chi2	0
Log likelihood = -12209.055					Pseudo R2	0.0032
Next Trans	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Treatment	0.0436241	0.0183001	2.38	0.017	0.0077566	0.0794916
Slr Feedback	-9.80E-08	1.92E-08	-5.11	0	-1.36E-07	-6.04E-08
% Positive	0.5216369	0.6537514	0.8	0.425	-0.7596923	1.802966
Item Price	-0.0969539	0.0135143	-7.17	0	-0.1234415	-0.0704664
Constant	0.3448969	0.6503971	0.53	0.596	-0.929858	1.619652

within the experimental period and then examines whether the user comes back and purchases after the experiment ended. Because users in the treatment group were exogenously presented with a ranking scheme that prioritized better quality sellers, the probability that they come back to the site should increase. Indeed, that is what we find. The treatment dummy is statistically significant, and, while not nearly as large as the non-experimental EPP results above, still economically meaningful.

We believe the experiment demonstrates three important facts: 1) That transaction buyer satisfaction is an important component of an individual's propensity to return to a platform, over and above an individual's propensity to return to an individual seller. We worked to demonstrate this point with observational data in the first part of the paper and use the experiment to corroborate those results. 2) That platforms can use an intermediate mechanism – search result rankings – to guide buyers to better quality sellers, alleviating some of the externality issues associated with platform transaction quality. We note that search ranking may be one of many levers that platforms might use here. 3) That search ranking causally affects buyer purchase decisions. We do this buy exogenously varying the search ranking of the same product in treatment vs. control groups which gets around the traditional problem of search ranking endogeneity.

7 Discussion

TO BE COMPLETED