Reputation and Feedback Systems in Online Platform Markets*

Steven Tadelis

UC Berkeley

Haas School of Business

NBER and CEPR

February 8, 2016

Abstract

Online marketplaces have become ubiquitous as sites like eBay, Taobao, Uber and AirBnB are frequented by billions of users regularly. The success of these marketplaces is attributed not only to the ease in which buyers can find sellers, but also to the fact that they provide reputation and feedback systems that help facilitate trust. I begin by briefly describing the basic ideas of how reputation helps facilitate trust and trade, and offer an overview of how feedback and reputation systems work in online marketplaces. I then describe the literature that explores the effects of reputation and feedback systems on online marketplaces, and highlight some of the problems of bias in feedback and reputation systems as they appear today. I discuss ways to address these problems in order to improve the practical design of online marketplaces and suggest some directions for future research.

*I am grateful to Oren Reshef for very helpful research assistance and to Tim Bresnahan, the Editor, for helpful feedback on an earlier draft.

1 Introduction

Online marketplaces are clearly one of the greatest success stories of the internet over the past two decades. Marketplaces such as eBay, Taobao, Flipkart, Amazon Marketplaces, Airbnb, Uber, Taskrabbit and many others are booming and providing businesses and individuals with previously unavailable opportunities to profit and succeed. These online marketplaces help match demand with supply in efficient and effective ways: they offer an effective means for companies to market their goods or get rid of excess inventory; they save businesses the extra costs needed to establish their own e-commerce website to generate online consumer traffic; they allow individuals to get rid of items they no longer need and transform these into cash; and more recently, the so called "sharing economy" marketplaces allow individuals to share their time or assets across different productive activities.

The amazing success of online marketplaces is taken for granted today and it is all but impossible to imagine a world without them. Less than two decades ago, however, the rapid success of eBay was a surprise to many early skeptics of online anonymous trade. How is it that strangers who have never transacted with one another, and who may be thousands of miles apart, are willing to trust each other? Any kind of transaction requires some level of trust between the buyer and seller, usually in the shadow of some institutional support like the law or other enforcement mechanisms. Unlike a physical transaction in a store, where the buyer can touch and feel the good he or she is buying, this close contact is absent in electronic commerce and the buyer may not be able to verify the seller's identity. Hence, to many, the rise of ecommerce in general, and of two-sided online marketplace in particular, was a surprise.

The early skepticism was well supported by economic theory. In his seminal article "The Market for Lemons," Akerlof (1970) showed how hidden information in the hands of sellers could hinder the operation of markets to the possible extreme of markets failing to operate despite gains from trade. The literature developed to classify two sources of uncertainty that hinder markets from operating efficiently. First, quality uncertainty may be a result of hidden information that determines the quality of the good or service in the spirit of Akerlof's adverse selection. For example, a seller on eBay or Etsy may know that the good they are selling is defective, yet they may choose not to reveal the defect and misrepresent their item. Second, quality uncertainty may be a result of hidden actions that determine the quality of the good or service, what is often referred to as "moral hazard". For example, a seller on Amazon marketplaces or Flipkart may choose to skimp on wrapping material and increase the likelihood that the good arrives damaged. Of course, both hidden information and hidden action might be present simultaneously.

For a marketplace to flourish, therefore, it is necessary that both sides of the market feel comfortable trusting each other, and for that, they need to have safeguards that alleviate the problems caused by asymmetric information. It is largely understood today that eBay's success was not only due to the relative simplicity and transparency of it's auction format, but also to a brilliant innovation introduced first by eBay and later copied in one form or another by practically every other marketplace: the use of a feedback and reputation mechanism. Indeed, feedback and reputation systems are central to the operations of every ecommerce marketplace and trace some of their heritage to ancient ancestor institutions that were used in the physical marketplaces of the Middle Ages.

Indeed, the need for reputational incentives to foster trust and guarantee successful trade is an old story and has been part of commerce for centuries. Historically, buyers and sellers would meet at centralized marketplaces to search for their trading partners. Coordinating where and when trade took place was an important historical innovation, which can be seen in the introduction of trade fairs in medieval Europe (see Greif (2006)). And what's more, the successful operation of these trade fairs, where people were expected to trade with counterparts whom they had never met, rested on governance and reputation mechanisms that gave people the faith to trade with strangers (see Milgrom et al. (1990). These trade fairs represent one of the very first examples of two-sided markets, of which online marketplaces such as eBay, Taobao, Amazon Marketplace, Airbnb, Uber and others are the modern reincarnation. The medieval European trade fairs offered buyers and sellers a coordinated location in which to meet, just like online marketplaces coordinate buyers and sellers on a single online platform, where buyers can search for the products (or services) they are looking for. And just as governance and keeping track of past transactions were needed to support trade in the trade fairs, so do reputation and feedback systems offer the trust needed to lubricate the online anonymous markets that have emerged since the Internet became widely available in the mid 1990s.

Until the rise of online marketplaces, it was quite difficult for economists to study the workings of real world feedback and reputation systems. Much of the "economics of reputation" literature was constrained mostly to theoretical papers that elaborated on the hazards of hidden information and hidden action models, and studied conditions under which reputation mechanisms can overcome the problems of asymmetric information described above. In general, it was a challenging task to find data that would speak directly to the role of reputation in supporting market transactions.[1] Thanks to the data made available by online marketplaces over the past two decades, many economists have studied a variety of interesting questions related to the operation of online feedback and reputation mechanisms.

In this paper I explain how feedback and reputation systems work in practice, and how they support ecommerce in online marketplaces. Section 2 starts with a brief explanation of the theory behind reputation mechanisms and how they can be designed to support more efficient online trade. Section 3 describes the actual working of typical online feedback and reputation systems. Section 4 presents findings from a host of empirical papers that have been written over the past fifteen years relate how reputation seems to work in actual online marketplace to the theory. Section 5 highlights some of the shortcomings of reputation and feedback systems that have been explored by some recent research, and Section 6 suggests some considerations for the future design of feedback and reputation systems that can augment their effectiveness. Section 7 offers some closing thoughts.

[1] Two notable exceptions that I am aware of are Jin and Leslie (2009) and Hubbard (2002). There is a larger empirical literature that explores the effects of adverse selection in markets, such as Bond (1982), Genesove (1993) and Hendricks and Porter (1988), but these papers do not speak to the role of reputation incentives in markets.

2 Reputation and Feedback in Theory: A Primer

The difficulty in supporting anonymous online trade can be easily explained using a simple game theoretic example, which is a version of the well known "trust game". Consider a buyer who identifies a product listed online by some anonymous seller. Imagine that the buyer values the product at \$25, the purchase price is \$15 and the seller has no alternative use of the good, implying that not selling it results in the seller receiving a net value of \$0. If, say, the seller's costs of shipping and handling are \$5, then at a price of \$15 the seller will be left with a net surplus (or profit) of \$10, and the buyer, paying \$15 for what he values at \$25, is also left with a net surplus (or dollar-value of his happiness) of \$10.

Now imagine that the seller can be one of two types: an honest seller or an opportunistic seller. An honest seller will always ship the item, while an opportunistic seller will maximize his or her expected payoff. The buyer does not know the type of the seller, but does know that a seller is honest with probability $p \in (0, 1)$. Assume that the buyer needs to send payment to the seller first in order to transact, implying that the buyer must choose to trust that the seller will ship the item, or not trust and have no cost or benefit. If trusted, the opportunistic seller chooses whether to ship the good and honor trust, or whether to renege and abuse trust. This simple game is shown in Figure 1.

This game involves both hidden information (the type of seller) and hidden action (the opportunistic type's choice). It is easy to see that if this game is played only once, then the opportunistic seller would always abuse trust. It follows that the buyer will trust the seller if and only if the likelihood of an honest seller is high enough. The expected benefit from either getting \$10 of value or losing the \$15 charge must not be negative, i.e., $10p + (-15)(1 - p) \geq 0$, or $p \geq 0.6$.

If this game is played more than once then future rewards can discipline an opportunistic seller to actually honor trust. Imagine that $p > 0.6$ so that buyers would be happy to transact once, and that one more transaction opportunity will present itself again in the future. Imagine that the seller discounts future payoffs at a discount factor of $\delta \in (0, 1)$. I'll

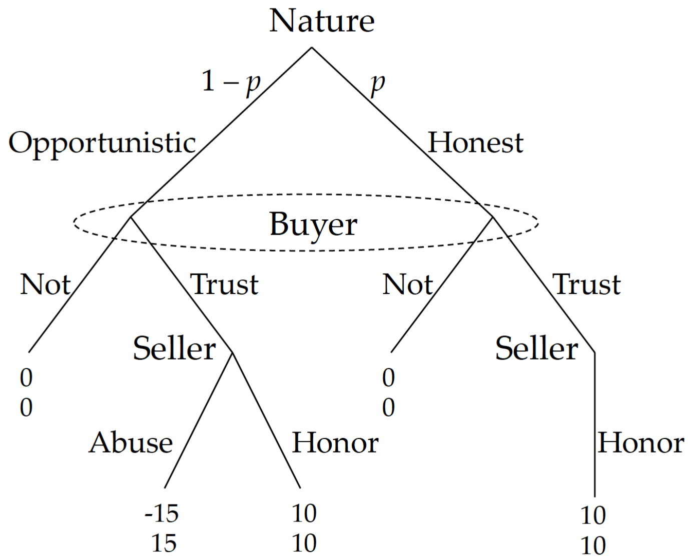

Figure 1: A Trade Game with Asymmetric Information

establish now that if δ is not too small (the future is important enough), then it will no longer be true that an opportunistic seller will choose to abuse trust in the first transaction.

To see this, imagine that the buyer expects an opportunistic seller to abuse trust in the first transaction. If this were the case, then the buyer can use the seller's performance in the first transaction to form accurate beliefs about the type of the seller: if the first transaction was successful, then the seller *must* be honest, and should be trusted again; if the first transaction failed then the buyer *knows* that the seller is opportunistic, and hence chooses not to trust him a second time. But, equipped with these beliefs, if the future is important enough then an opportunistic seller would not find it in his best interest to abuse trust in the first transaction. To see this, first observe that such behavior would result in a payoff of \$15 for the abusive opportunistic seller. If he instead chooses to honor trust, then he receives only \$10 in the first transaction. But then the buyer, given his beliefs, will infer incorrectly that the seller is honest and will trust him again, allowing the seller to then abuse trust in the second transaction and acquire another \$15 then. If the added value of \$15 in the second transaction outweighs the loss of \$5 in the first transaction, which happens if the per-period discount factor is at least $\frac{1}{3}$, then the opportunistic seller will choose to behave honestly in the first transaction. It follows, therefore, that if $p > 0.6$ and the future is important enough, the opportunistic seller will find it beneficial to honor trust in order to get access to the money he can obtain from the second transaction. This is the unique sequential equilibrium under these conditions.

What is more interesting is that with a two-stage game like the example above, trade can occur even when the buyer is less optimistic about the seller's honesty, i.e., even if $p < 0.6$. To see this imagine that the buyer knows that this trading game will be played twice, and imagine that he believes that an opportunistic seller will abuse trust always, which are the most pessimistic beliefs he can have. If he trusts the seller in the first transaction, then with probability p he will obtain a payoff of \$10, and then he will know for sure that the seller is honest and he will obtain another payoff of \$10 in the second stage. If instead he is abused in the first trade then he will opt out of transacting again. Assuming that the buyer too uses δ as his discount factor, with these beliefs he will be happy trusting the seller if and only if,

$$p(10 + 10\delta) + (1 - p)(-15) \geq 0 \quad \Longleftrightarrow \; p \geq \frac{15}{25 + 10\delta}.$$

Note that if δ becomes infinitesimally small, then for the buyer to trust the seller it must be that $p \geq 0.6$ because it effectively becomes like a one-stage game for the buyer. If on the other hand the buyer is extremely patient so that δ approaches 1, then for the buyer to trust the seller it must be that $p \geq \frac{3}{7}$. In this case the buyer will be happy to trust the seller, and if the discount factor is large enough, then the analysis performed earlier implies that the opportunistic seller would rather imitate the honest type and cooperate in the first stage of trade. Hence, with a high enough discount factor and two stages we get more trade in the sense that it is supported for lower values of p. And if we add more potential trade periods in the future, then trade will occur for even lower likelihoods of the seller being honest.[2]

[2] This is an example of a game in the spirit of the seminal work by Kreps et al. (1982). For some values of $p < \frac{3}{7}$ the equilibrium involves some mixed strategies by the seller in the first stage and by the buyer in the second stage. This is beyond what I wish to highlight here as the key insight is that a potential future creates incentives to behave honestly and not abuse trust.

The main idea in the example above is that the presence of honest sellers, together with the prospect of having a valuable future from trading, provide incentives even to opportunistic sellers to behave well. It is well known from the folk theorem for infinitely repeated games that trade can be supported even if the seller is opportunistic for sure ($p = 0$). If the buyer and seller would play this one-stage-game for infinitely many periods, and if the future is important enough (i.e., δ is close enough to 1) then they can use "trigger" strategies to support trade: in the first period the buyer chooses to buy and the seller chooses to ship. In every period thereafter, the buyer buys and the seller ships only if this is what the buyer and seller did in all previous periods, while if the buyer did not buy or the seller did not ship in any period, then they revert to the unique Nash equilibrium of the stage game forever after.

The key idea here is that today's actions will lead to future consequences that affect the prospects of the seller, hence keeping him in check. As it turns out, this powerful mechanism may even work if the seller does not necessarily interact repeatedly with the same buyer. If the seller understands that his current actions will be revealed to all future buyers and not just his current buyer, and that his good behavior today will be rewarded by future business just as bad behavior will be penalized by a lack of future business, then the seller will have an incentive to act in good faith (in this case, ship the item). Knowing that, the buyer will have a good reason to trust the seller to act honestly and will therefore choose to trust and transact with him. This important insight (see Kreps (1990)) sheds light on the powerful role that reputation and feedback systems provide.

Namely, if some sort of public reputation repository allows all future potential buyers to track a seller's past performance, then reputation becomes an important incentive mechanism that facilitates trust in anonymous markets. And herein lies the role of reputation and feedback mechanisms: to provide future buyers with a window into a seller's past behavior with previous buyers in anonymous marketplaces.

There is a vast theoretical literature on the economics of seller reputation (see Bar-Isaac and Tadelis (2008) for a survey) with empirical implications that are rather intuitive. First, sellers with better reputations should attract more potential buyers, and command higher prices for their goods and services. Second, as sellers' reputations get better (or worse), their economic returns and growth will also get better (or worse). These simple yet powerful implications of reputation models can be taken to data, and the recent rise of "big data" online platforms have proven to be a fertile ground to test these implications.

3 Reputation and Feedback in Practice

As mentioned earlier, many have attributed the success of ebay to the advent of its reputation and feedback mechanism (see, e.g., Resnick et al. (2000)). As Dellarocas (2003) put it, "eBay's impressive commercial success seems to indicate that its feedback mechanism has succeeded in achieving its primary objective." (p. 1411.) eBay's reputation mechanism is often described as a resounding success for two reasons. First, eBay exists as a successful business despite the complete anonymity of the marketplace. Second, practically every online marketplace has adopted some form of a reputation or feedback system that is closely related to the one that eBay had introduced back in 1995. It is therefore illustrative to describe in some detail how eBay's feedback system works.

3.1 Some Basic Features

In theory, the central feature of a well functioning reputation system is to provide future buyers with information about the outcomes of a seller's past behavior. In practice, however, information about past performance needs to be manufactured, and it is generally manufactured using the voluntary input of buyers. For example, after a buyer completes a transaction on eBay, he or she has 60 days to leave either a positive, negative, or neutral feedback score for the seller from whom the product was bought. On the Chinese marketplace Taobao.com, if a seller leaves positive feedback for a buyer but the buyer leaves no feedback then the platform's algorithm leaves an automatic positive feedback under the assumption that silence is most likely a sign of content.[3] As I explain later in Section 5, this may be far from the truth.

A buyer can of course choose to leave no feedback at all, which is what a standard model of selfish behavior would predict because leaving feedback supplies a public good, for which the buyer receives no payoff except for some internally motivated moral satisfaction. Interestingly, such pro-social behavior is prevalent on eBay: today, about 65 percent of buyers leave feedback on eBay, a very high fraction, and an even higher fraction of more than 80 percent left feedback in eBay's earlier days.[4]

The accumulation of feedback that sellers receive is then aggregated by the marketplace platform. Figure 2 shows how a seller's feedback, which I refer to as the reputation measure of the seller, is calculated and displayed on eBay. The webpage in Figure 2 shows a new Apple MacBook being sold by a seller with the username "electronicsvalley" with a *Feedback Score* of 21,814, which is the summed value of the number of positive feedbacks minus the number of negative feedbacks. The page also shows that 99.2% of this seller's feedback was positive, defined as the seller's number of positive feedbacks divided by the sum of his number of positive and negative feedbacks.[5]

If a buyer on eBay wants to learn more about the seller's history, they can click on the feedback score (the number 21814 in Figure 2), which directs them to a detailed *feedback profile* page that is shown in Figure 3. The interested buyer can see how many positive, neutral, or negative feedback reviews the seller received in the past month, six months or twelve months.[6] At the bottom of the page there is a rolling list of comments left by the

[3]See Li et al. (2016) for more on Taobao's reputation system.

[4]This high fraction of feedback may be a surprise to many mainstream economists, but not to Pierre Omidyar, eBay's founder. On his personal profile page it states that "Pierre created eBay in 1995 on the premise that people are basically good" (www.omidyar.com/people/pierre-omidyar).

[5]Notice that there is a badge at the upper right corner of Figure 2 that certifies this seller as an "Top Rated Plus" (previously called an eBay Top Rated Seller, or ETRS). This designation is bestowed on sellers that meet a series of criteria believed by eBay to be an indication of a high quality seller. See Hui et al. (2014) for a lengthy discussion of this program, which is active platform intervention, above and behind the decentralized feedback system.

[6]While the feedback score is calculated using all past transactions, the percent positive only looks back at the last twelve 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: eBay's View Item Page Displaying Feedback

buyers in addition to the feedback itself, and to the right there are stars that indicate the *Detailed seller ratings* (DSRs), which buyers can leave only if they choose to leave feedback first. Unlike the simple positive/negative/neutral feedback that a buyer can leave, which a seller can identify with the buyer who left it (it is not anonymous), the DSRs are anonymous and aggregated so that the seller cannot infer who left them.

On other ecommerce platforms, feedback is often summarized by some star system (usually 1 through 5 stars), and the buyers can typically find out more information by clicking on the scores to see what the most recent ones were, what is the distribution of feedback, and whether there were any meaningful verbal comments. It is important to note that reviews may be about the product rather than the seller, a well known example being the $1-5$ star product reviews on Amazon. Platforms must be careful about distinguishing between product reviews and seller reviews in order to avoid confusion. Multiple review targets may create an inference problem that confuses between the seller's quality of executing the sale

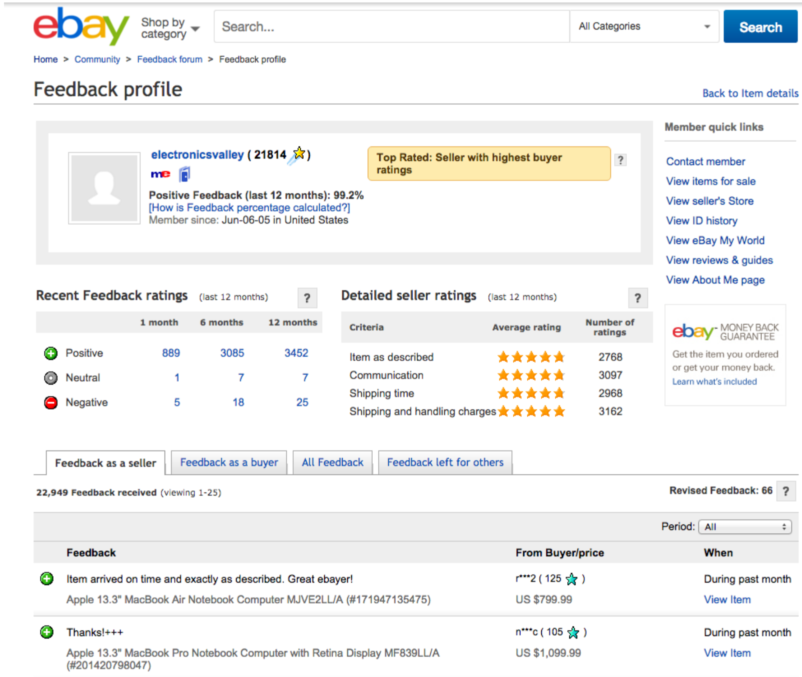

Figure 3: eBay's Display of a Seller's Feedback Profile

and the quality of the product. Many online platforms offer at least one side of the market the ability to make choices that depend on the reputation of the other side of the market.[7]

3.2 One-sided versus Two Sided Feedback

Before 2008, both sides of a transaction on eBay could leave each other positive, negative or neutral feedback with comments. In 2008 eBay changed the feedback system so that sellers are now limited to leave either positive feedback or no feedback at all. On Amazon's marketplace, sellers leave no feedback at all; on Airbnb both owners and renters leave feedback that is then aggregated and publicly observed by future marketplace participants; on Uber both drivers and riders leave feedback, which is not made public, yet drivers see a rider's feedback before accepting a ride and riders see the driver's feedback after the ride was confirmed. Whether reputation should be "two-sided," like it is currently on Airbnb (or on eBay before the 2008

[7]Even on Uber, which according to their site automatically connects riders with the closest driver without making it possible to request a specific driver, a rider can cancel a trip after seeing the feedback rating of the driver who accepted the ride. Also, drivers see the rating of the rider who Uber matched them with and can choose not to accept the ride.

change), or practically "one-sided," like it is currently on eBay and on Amazon's marketplace, is an interesting design question.

During eBay's earlier days, before payments were made through Paypal's online payment system, buyers would send checks or money orders to sellers. As such, just as sellers can renege or under-perform, so could buyers, making it imperative that sellers can choose whether or not to trust a buyer or wait till a check arrives and clears before sending the item to the buyer. However, after eBay acquired PayPal and strongly encouraged sellers to use it as the only form of payment, such problems of buyers not paying have all but disappeared.[8]

The question then is why not just stick to a two-sided feedback system? The answer lies in the problem of retaliation. Bolton et al. (2013) used data from eBay during the period when the reputation system was two-sided, and convincingly showed that sellers wait to get feedback before giving feedback back to buyers. To illustrate their findings, consider pairs of feedback scores, (FB_i, FS_j) left by a pair consisting of buyer B_i and seller S_j who constituted a transaction. For example, a transaction in which both buyer and seller left each other positive feedback is denoted $(+,+)$, while if the buyer left positive feedback and the seller negative feedback, it is denoted $(+,-)$. Bolton et al. (2013) first showed that practically all transactions are either $(+,+)$ or $(-,-)$. They then showed that a vast majority of $(-,-)$ transactions are characterized by the seller leaving feedback on the same day or the day after the buyer does, while the $(+,+)$ transactions happen with less correlation between the buyer's and seller's day of leaving feedback. Hence, sellers' negative feedback scores were primarily retaliatory, which in turn made it painful for buyers to leave negative feedback (a point to which I return later.)

This fear of retaliation was most likely a central cause behind the fact that almost all buyers left positive feedback on eBay, which in turn caused eBay to switch from the two-sided reputation system to a one-sided reputation system. This is not, however, a good prescription for all online marketplaces. Take the lodging marketplace Airbnb as an example. Even if

[8] Though eBay and Paypal had recently split into two separate corporations in July 2015, use agreements that had been signed between the companies will provide the continuation necessary for the platform to continue operating as it has.

payment is mediated by the site, as it is, there is still a concern that double moral hazard may occur. The owners can misrepresent the home they are renting, leave it dirty, not give the renters a key at the pre-specified time, and more. At the same time, the renter's role on Airbnb is not just to pay like they do on eBay and wait for an item to arrive: they can leave the home dirty, cause damage, be very noisy, etc. As such, it is imperative that Airbnb continue to keep a two-sided reputation system for trust to prevail in their marketplace. In fact, Airbnb even verifies the identity of all parties given the high stakes involved. Each marketplace, therefore, must weigh the costs and benefits from one- versus two-sided feedback systems.

4 Reputation Effects: Empirical Findings

4.1 Observational Studies

As mentioned earlier, the data made available by online marketplaces over the past two decades has made it possible for economists to study a variety of interesting questions related to the operation of online feedback and reputation mechanisms. By and large these studies used what is referred to as "scraped" data, where computer scripts are written to "crawl" the internet and collect data from marketplace webpages. Many of the studies asked whether sellers with higher reputation scores and more transactions receive higher prices for their products, or whether reputation seems to matter more for higher priced goods than for lower priced goods. In what follows I will describe some of the empirical studies that have been published in the past 15 years that speak to these issues.[9]

Early studies have collected what are now considered minuscule datasets. One of the earliest papers by McDonald and Slawson (2002) collected data from 460 auctions completed in 1998 of collector-quality Harley Davidson Barbie dolls. The study recorded seller reputation

[9]Space prevents me from covering al the literature, and as such I chose a sample of papers that are either early papers, or that introduce new directions to the preceding work. My apologies to those authors whose work is not included in this survey. See Bajari and Hortaçsu (2004) for an older survey that references more studies.

scores, sellers experience, auction characteristics, the number of bids and the auction's closing price. Because the closing price and the number of bids are bound to be correlated, they use Seemingly Unrelated Regressions to simultaneously estimating the effect of the reputation score on both the number of bids and the closing price. Their results suggest that eBay's reputation score is positively correlated with both the closing price and the the number of bids.[10]

There are several concerns with the interpretation of McDonald and Slawson (2002), or similar studies that use this kind of scraped data. First is the problem of endogeneity as there is no random assignment. For instance, good reputation may be correlated with a variety of omitted variables influencing the dependent variables. For example, the authors themselves mention unclear language and grammar as possibly confounding factors. Also, experienced sellers may post more appealing photos or product descriptions, and they might be better at recognizing market trends. Second, though statistically significant, the results are not economically meaningful. For example, their analysis implies that a one point increase in reputation corresponds to a 4 cents increase in final price. Given that the median reputation score of the sample is 21 points, this means that moving from no reputation at all to the median increases price by less than 1 dollar, a negligible impact when considering a median price of $275.

Another early paper tat takes a slightly different approach is Livingston (2005), who uses a sample-selection model to challenge the previous findings that sellers' reputation has a small or insignificant effect on the final sale price. Livingston (2005) presents a theoretical model that predicts a positive, yet decreasing returns to reputation. Using data from 861 eBay auctions for Taylor brand golf clubs, he finds a positive effect of reputation on the probabilities of receiving a bid and of a sale happening, as well as on the sale price, especially for the first positive reviews. Because negative reviews are extremely scarce, reputation is defined as the number of positive reviews, using five dummy variables (for no reputation and for each quartile) to allow for a more flexible nonparametric estimation. The main novelty is

[10]Melnik and Alm (2002) follow a similar approach using data from 450 auctions for 1999 $5 U.S. gold coins that were completed in 2000.

that Livingston (2005) observed that the second-price auction design implies that estimating the impact on price should be done using a latent variable model. That is, because the second highest valuation is censored, OLS estimates are downwards biased, which is corrected using a sample selection model. His results suggest that, on average, the first 1-25 reviews increase payment by $20.42 (about 5% of an average price of $409.96). In comparison, moving from any two subsequent quantiles has no significant effect on sale price, consistent with the theoretical prediction of the paper. Note, however, that despite offering results that have more meaningful economic magnitudes, the problem of endogeneity due to a lack of random assignment is still of some concern.

An interesting paper by Jin and Kato (2006) takes a different approach than the papers described earlier. They examine the connection between price, *claimed* quality, reputation and *true* quality, using observational data of baseball card auctions on eBay. The true quality of cards was determined by purchasing actual cards from online auctions and having them examined by professional rating agencies that rank cards on a 10 point scale.[11] Naturally it is all but impossible to verify the quality of ungraded cards online, implying that buyers are exposed to potential moral hazard by sellers who may exaggerate card quality in order to receive higher payments. Jin and Kato (2006) collect data of the five most traded baseball cards on eBay for seven months, resulting in 1,124 auctions, of which 67% were graded (usually scores of 8 and above). Of the full sample, 81% of auctions sold with at least one offer above the reserve price. Buyers in their sample have two signals for the quality of an ungraded card: the rating and claims. The data show that claims of quality for ungraded cards seem suspiciously high, both when observing their distribution and when considering sellers' expected payoffs. The data also show that sellers who claimed extremely high quality had significantly lower ratings. Nevertheless, buyers seem to be willing to pay more for cards with higher claimed quality. On average, a card with a claimed quality 10 is sold for 75.5% more than a card with a claimed quality below 9. Moreover, increase in claimed quality significantly raises the probability of auction completion (one claimed-grade increase

[11] This approach follows the the literature on adverse selection, e.g., Bond (1982), in which the econometrician is in a position of knowing more than the market participants.

increases the likelihood of completion by 4.63%), yet ratings do not affect the price paid for the card, consistent with Livingston (2005) regarding the reputation premium. Surprisingly, interactions between rating and claims are insignificant, suggesting that buyers prefer higher rated sellers, but are willing to trade with low rated sellers and pay more if the claimed quality is high enough. Possibly the most interesting result in Jin and Kato (2006) concerning a seller's reputation is that reputable sellers (using eBay's ratings) are less likely to make extreme claims about their cards' quality (grades of 9 and above). Moreover, reputable sellers are less likely to default or provide counterfeit cards. However, conditional on authentic delivery, higher reputation ratings are not correlated with higher true quality. This may explain why buyers are more willing to trade with reputable sellers, but are not willing to pay more.

Yet another approach is taken by Cabral and Hortaçsu (2010), who collect a much lager set of observations than previous studies, and propose a different strategy for studying the effects of seller reputation on eBay. In particular, they collect a series of seller feedback histories, creating a panel of seller histories, and proceed to estimate the effect of reputation on a seller's sales rate. They find that when a seller receives a first negative feedback rating, his weekly sales growth rate drops from an estimated positive rate of five percent to a negative rate of eight percent. A clever assumption they use to perform their analysis deals with the fact that eBay does not provide information on how many past transactions a seller completed. Cabral and Hortaçsu (2010) propose two assumptions, provide evidence that they are reasonable. First, they assume that the frequency of a seller's feedback is a good proxy for the frequency of actual transactions, and second, they assume that the nature of feedback correlates with buyer satisfaction. A disadvantage of their approach is that they are not able to observe the effect on prices, but the previous literature suggests that the effect on prices is small.

An interesting question is what would be the impact of introducing a reputation system into a marketplace in which one was not previously used. This question is addressed by Cai et al. (2014) who consider the case of Eachnet, a Chinese auction cite, which had nearly a 90%

share of the Chinese market during its years of operation (1999- 2003). Unlike eBay, exchange of products and money was done offline, hence transactions relied heavily on trust between the parties. That said, because the parties eventually met in person for the exchange, there may have been less uncertainty at the time that they met in order to consume the transaction. In 2001 Eachnet introduced a centralized feedback system, which enabled buyers to rate sellers after each transaction, similar to eBay's. The data collected by Cai et al. (2014) comes from the platform and is large: it contains a random sample of 125,135 sellers who posted almost two million listings throughout Eachnet's years of operation. On average, 44.5% of listing were completed with at least one transaction, designated as a successes. Since sellers' feedback scores are obviously not available prior to the introduction of centralized feedback, they are approximated feedback using the cumulative number of successful listings each seller had since joining Eachnet, which turns out to be a good approximation as they show using success rate data for the period after centralized feedback was introduced. Cai et al. (2014) examine how a seller's reputation, as approximated by cumulative success rates, affects his behavior and outcomes, and how this is impacted post feedback centralization. One result demonstrates that an increase in the cumulative success rate of a seller is correlated with a larger fraction of repeating buyers, but this effect weakens after feedback centralization. This result ties nicely with intuition: centralized feedback acts as a substitute to relationships. A second result shows that centralized feedback leads sellers with higher cumulative success rates to sell more products in more regions, suggesting that centralization facilitates the expansion of reputable sellers into new markets. Last, they show that a higher cumulative success rate is generally correlated with a lower hazard rate of exiting the market. Perhaps surprisingly, this effect diminishes after feedback centralization. Cai et al. (2014) also test for the existence of a reputation premium. Though prices appear to be lower for reputable sellers, they do enjoy more listings and higher success rates, in line with the results of Cabral and Hortaçsu (2010).

4.2 Exploiting Exogenous Variation

As argued earlier, one shortcoming of observational studies that every applied economist is aware of is the possible endogeneity concern that results form potential selection or omitted variable biases. Of course, the gold standard strategy used to control for endogeneity problems is either to perform a controlled experiment, or to find some clever exogenous variation and exploit it to identify reputation effects.

Resnick et al. (2006) perform a controlled field experiment by conducting a series of sales of identical items (in their case, collector's postcards) where they vary reputation by randomly assigning items to either an established seller's account with a good reputation, or to a new account with little reputation history. They estimate an eight percent price premium to having 2,000 positive and 1 negative feedback over a reputation of 10 positive and no negative feedbacks.

A more recent paper by Klein et al. (2015) cleverly takes advantage of a change in the way that eBay reports feedback, together with the fact that feedback for sellers has two components: the non-anonymous simple feedback of positive, negative and neutral ratings, and the anonymous feedback of Detailed Seller Ratings (or DSRs, as seen in Figure and described in Section 3.1.) The exogenous change was implemented by eBay in May of 2008, after the company had realized that a fear of retaliation most likely causes buyers not to leave negative ratings. (Recall the discussion of retaliation in Section 3.2.) As a result, starting in May 2008, sellers no longer were able to leave the buyer negative or neutral feedback, but can leave either positive feedback or no feedback for buyers. This change should encourage buyers to report negative feedback following a poor experience, which can leads sellers to respond in two ways: first, inherently bad sellers may be forced to leave eBay following a series of negative ratings, consistent with an adverse selection story. Second, sellers may work harder to improve buyer satisfaction, consistent with a moral hazard story. Klein et al. (2015) scraped data containing monthly information on feedback received from about 15,000 eBay users between July 2006 and July 2009. This period includes the introduction of anonymous DSR ratings (May 2007) as well as the introduction of one-sided feedback (May 2008). They find that the change to one-sided feedback led to a significant increase in buyer satisfaction using the DSR reviews, but did not lead to a change in the exit rate of sellers from the market. The results therefore suggest that the effect is mainly one of reducing moral hazard and not causing a change in the composition of sellers due to reducing adverse selection. This is a clever way in which Klein et al. (2015) exploit the exogenous change and look at the evolution of DSRs, which have always been anonymous and could not have been retaliated against. This observation supports their key assumption that the mapping between seller behavior and DSRs was not affected by the change to one-sided feedback.

Some studies overcome the inability to run field experiments on counterfactual changes to feedback and reputation systems by employing laboratory experiments. Bolton et al. (2004) compare trading with feedback to trading without feedback and find that while feedback induces quite a substantial improvement in transaction efficiency (a lack of cheating and improved trust), it also exhibits a public goods problem in that participants understand that the benefits of trust and trustworthy behavior are enjoyed by everyone and are therefore not completely internalized. On one hand, such laboratory studies exhibit controlled experiments that cleanly identify the effect that is of interest. On the other hand, a weakness of this approach is that the laboratory may simulate a situation that is far from the workings of a real-world marketplace.

5 Biases in online Feedback Systems

The empirical findings from the studies described above suggest that the empirical facts are consistent with the general theory and intuition of what reputation and feedback systems should do in practice. Yet a close look at the literature shows that the effects of reputation are small. An important question follows: how accurately are the reputation measures reflecting the variation in performance?

As described earlier, retaliation on eBay may cause this user-generated feedback to be biased, as buyers will refrain from leaving negative feedback. A growing literature has shown that user-generated feedback mechanisms are often biased, and can be prone to influence by sellers. Dellarocas and Wood (2008) conjecture that the extremely high percent positive reputation measures on eBay are explained by the fact that many buyers who have poor experiences choose to leave no feedback at all.[12] They proceed to suggest an econometric technique to uncover the true percent of positive transactions based on several assumptions, most notably that reputation is two-sided, but their approach can no longer work after the change that eBay made in 2008.

Nosko and Tadelis (2015) expand on Dellarocas and Wood (2008) and directly show how biased reputation measures really are. Using internal eBay data, Nosko and Tadelis (2015) establish that the percent positive measure has a mean of 99.3% and a median of 100%. The distribution of feedback from their study is described in Figure 4, which displays the histogram of seller percent-positive measures from a dataset containing close to two million sellers who completed over fifteen million transactions between June 2011 and May 2014. Notice that the X-axis starts at 98%, which is the tenth percentile. The median seller has a score of 100%. This could be indicative of a reputation system that works extremely well – bad sellers exit when their score falls even slightly, leading to a high positive selection. Unfortunately, as this is not the case – the data show that there are three times as many complaints to customer service as there are negative feedback scores, suggesting that buyers avoid leaving negative feedback and prefer instead to pursue more costly actions of demonstrating their dissatisfaction.

Both the reciprocal findings of Bolton et al. (2013) described in Section 3.2, as well as anecdotal evidence[13] described in Nosko and Tadelis (2015), suggest that it is more "expensive" to leave a negative review than it is to leave a positive review because of seller retaliation

[12] Li (2010) proposes a mechanism designed to solve the problem of missing reports and positive bias. The mechanism provides the sellers with an option for giving rebates to rating buyers. Li and Xiao (2014) extend the model of Li (2010) and conduct a laboratory experiments to test the main hypotheses. The lab results suggest that higher reporting costs decrease buyers willingness to review sellers, leading to a decrease in buyers' trust and sellers' trustworthiness, but these results are not statistically significant. Additionally, since the research design lacks a fear of retaliation, reports are negatively biased.

[13] 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.)

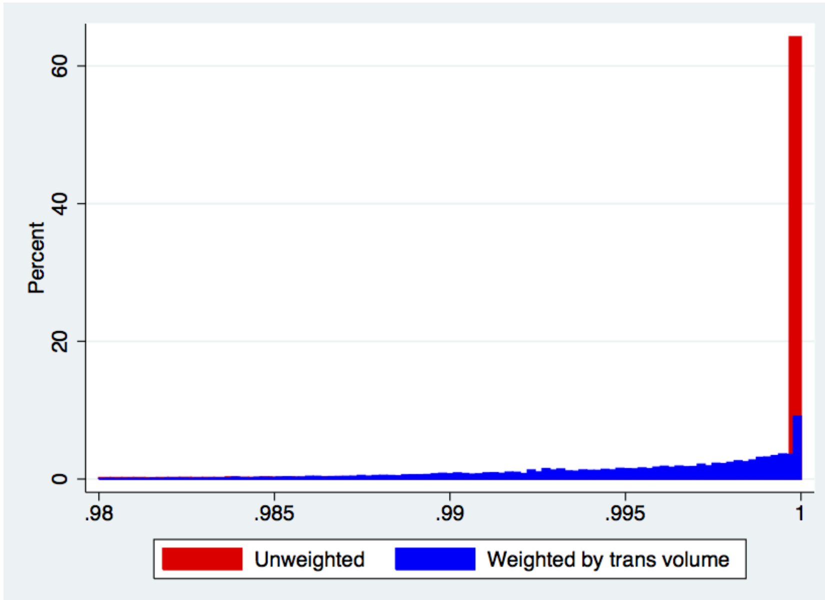

Figure 4: Percent Positive of Sellers on eBay

and harassment. Hence, a central challenge is to construct a measure that more accurately reflects a seller's true quality. Nosko and Tadelis (2015) suggest a new quality measure they call "effective percent positive" (EPP), which is calculated by dividing the number of positive feedback transactions by the total number of transaction, thus penalizing sellers who are associated with more transactions for which the buyers left no feedback. The tenet of their approach is that silence is bad news. Consider, for example, two sellers: Seller A, who had 120 transactions, and seller B who had 150, yet both received one negative feedback and 99 positive feedbacks. Both have a $\frac{99}{99+1} = 99\%$ positive and both have a score of $99 - 1 = 98$. Seller A, however, had only 20 silent transactions while seller B had 50 silent transactions. Hence, seller A has an EPP of 82.5% while seller B has an EPP of only 66% and is assumed (and later verified) to be a worse seller on average.

The distribution of the EPP measure is described in Figure 5 using the same set of sellers for which their percent-positive scores were described in Figure 4. EPP has a mean of 64%, a median of 67%, and exhibits significantly more variability than the percent positive measure because there is a lot of variation in the choice of buyers to be silent across sellers.

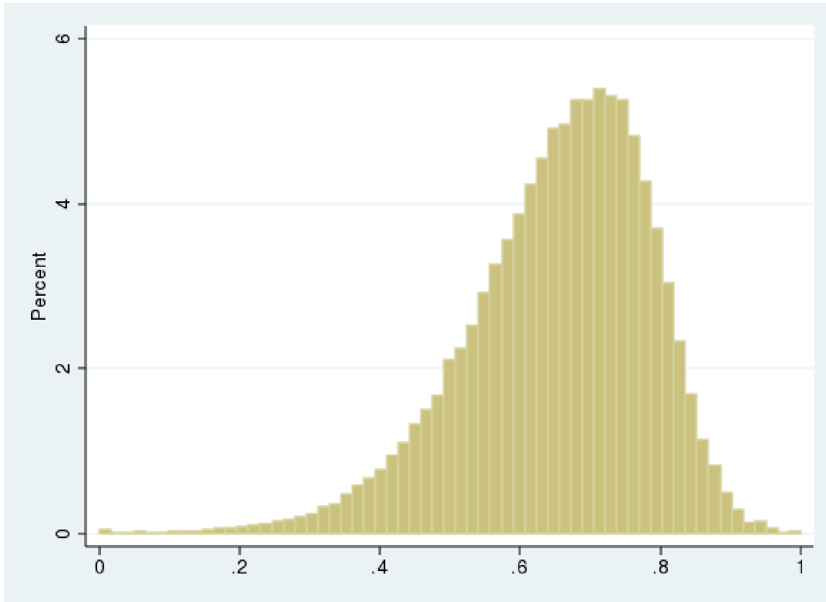

Figure 5: Histogram of Sellers' Effective Percent Positive Scores

Because they have access to internal eBay data, Nosko and Tadelis (2015) use a "revealed preference" approach to study the effect of a seller's EPP on the buyer's propensity to continue buying on eBay after that transaction, which distinguishes their paper from most papers that collect scraped data from marketplaces and are limited to consider only prices and quantities. This approach allows them to get to the heart of the question of whether reputation mechanisms are indeed steering buyers away from low quality sellers. 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.

Nosko and Tadelis (2015) show that a buyer who has a better experience on eBay (indicated by buying from a seller with a higher EPP score) is more likely to continue to transact on eBay again in the future. They then report results from a controlled experiment on eBay that incorporated EPP into eBay's search-ranking algorithm. The treated group was a random sample of eBay buyers who, when searching for goods on eBay, were shown a list that prioritized products from sellers with a higher EPP score compared to a control group. The results show that treated buyers who were exposed to higher EPP sellers were significantly more likely to return and purchase again on eBay compared to the control group of buyers. Hence, they argue that marketplace platforms can benefit from using data in ways that uncover better measures of seller quality. Further implications about the design and engineering of feedback systems are discussed in Section 6.

Bias in reputations is not unique to the eBay marketplace. Mayzlin et al. (2014) exploit different policies about who can leave feedback across several travel sites and show biases in ratings for hotels from the online travel sites that are consistent with strategic feedback manipulation by sellers. In particular, consider a geographic area in which several sellers, in this case hotels, compete for the business of travelers. It will be in the strategic interest of these hotels to inflate their own ratings with fake positive reviews, while at the same time try to negatively impact the ratings of their competitors by submitting fake negative reviews for them. What makes the Mayzlin et al. (2014) paper particularly clever is that they do not attempt to categorize which reviews are fake reviews versus those that are not, which on the face of it is impossible because fake reviews are designed to mimic real reviews. Instead they take advantage of a key difference in website rating systems where some websites accept reviews from anyone while others require that reviews be posted by consumers who have purchased a room through the website. That is, some travel websites are like the rating website Yelp, where anyone can leave a review, while others only let buyers who purchased a stay at a hotel through the website to leave a review on the hotel they actually stayed at. If posting a review requires an actual purchase, the cost of a fake review is much higher. The upshot is then that they measure the differences in the distribution of reviews for a given hotel between a website where faking a review is expensive and a website where faking a review is cheap. The results in Mayzlin et al. (2014) indeed show greater bunching at the extreme ratings for hotels on the sites where posting reviews is cheaper, and this is exacerbated by local competition (more local hotels). Hence, for reviews to be less biased it is critical to impose some kind of cost to prohibit fake reviews by non-purchasers.

Fradkin et al. (2015) study the bias in online reviews by using internal data from Airbnb, and like Nosko and Tadelis (2015) report results from field experiments conducted by the online marketplace. In one experiment they offer users a coupon to leave feedback and show the users who were induced to leave feedback report more negative experiences than reviewers in the control group, suggesting that otherwise they would have probably been silent. In a second experiment they disable retaliation in reviews, similar to what eBay did in 2008, and find that retaliation (or rewards for positive feedback) cause a bias, but that the magnitude of this bias is smaller than that caused by a lack of incentives to leave truthful feedback. Interestingly, using data on social interactions between buyers and sellers on the site, they show that such interactions result in less negative reviews. This result suggests that a challenge for online marketplaces is the potential loss of information following the social interaction of buyers and sellers on the site.

Another form of bias is grade inflation. Horton and Golden (2015) document substantial levels of "reputation inflation" on the online labor marketplace, oDesk, that uses a five-star feedback system for freelance employees who bid on jobs that are posted by potential employers. The data show that from the start of 2007 to the middle of 2014, average feedback scores increased by one star. Moreover, in 2007, 28% of contracts ended with a feedback score of less than 4 stars, whereas in 2014 only 9% of transactions commanded these lower ratings. Data from a competing platform, Elance, shows similar patterns. Interestingly, they are able to show that this inflation is not wholly explained by changes in marketplace composition, in that bad sellers exit the marketplace over time. Only half of the increase in scores can be explained by the composition of market participants. Like Bolton et al. (2013) and Nosko and Tadelis (2015), Horton and Golden (2015) conjecture that giving negative feedback is more costly than giving positive feedback due to retaliation. They further argue that what constitutes bad feedback depends on the market penalty associated with that bad feedback. Together, the paper argues that these two factors can create a race of ever-increasing reputations.

Zervas et al. (2015) demonstrate that grade inflation is also severe on Airbnb, where ratings are overwhelmingly positive, averaged at 4.7 out of 5 stars with 94% of property ratings with 4.5 or 5 stars. They show that TripAdvisor ratings are also relatively positive, but to a lesser degree with an average of 3.8 out of 5 stars. Some, but not all, of the differences can be attributed to heterogeneity of properties reviewed. Nevertheless, the positive bias on Airbnb persists even when comparing cross-listed properties, and ranking across platforms is only weakly correlated. They argue that differences are mostly attributed to strategic considerations incentivized by Airbnb's bilateral review system (which do not exist on TripAdvisor).

One more channel through which bias in reputation may occur is by sellers try to fraudulently "buy" a reputation that they do not deserve. Brown and Morgan (2006) show some cases in which this practice happened on eBay's marketplace. Xu et al. (2015) document and explain the rise of a centralized marketplace for fake reputations for sellers on the Alibaba marketplace in China. Hence, it may be possible for sellers to fraudulently acquire a reputation that they do not deserve, and marketplace designers must be aware of such practices and make every effort to detect and punish this kind of behavior.[14]

6 Online Feedback and Market Design

The economic literature takes the view that market participants understand the equilibrium they are playing, and correctly infer information from signals and actions. In theory, there is no difference between theory and practice; but, in practice, there is. A potential problem that may prevent online feedback and reputation systems from properly working is that many buyers may have trouble interpreting the information they are presented with. Naively, one may think that a score of 98% is excellent (in some sort of absolute scale). In reality, as Nosko and Tadelis (2015) show, a score of 98% on eBay places a seller below the tenth percentile of the distribution. And it is unclear that the more informative EPP measure constructed by Nosko and Tadelis (2015) would be interpreted correctly by buyers. For this reason Nosko and Tadelis (2015) choose not to reveal the new measure to buyers, and instead run a controlled experiment that incorporated the information contained in the EPP measure into eBay's search-ranking algorithm.

[14]Not all attempts to purchase a reputation may be fraudulent. Signaling theory suggests that high-quality sellers may pay for honest feedback knowing that the feedback they receive will bode well for them. See Li et al. (2016) for a study of such behavior in Taobao's marketplace.

This approach offers a new direction that does not advocate for a "laissez faire" approach in which buyers are shown information to act on, and instead argues that marketplace platforms can benefit from a more paternalistic, or regulator-like approach, that does not assume that market participants can decipher information effectively. In this sense I very much advocate for the view expressed in Roth (2002) that market designers "cannot work only with the simple conceptual models used for theoretical insights into the general working of markets. Instead, market design calls for an engineering approach." (p. 1341.) That is, market designers can build systems that, rather than present information that buyers must digest and interpret, will instead make recommendations that can rely on underlying data that is not made visible to buyers.

In particular, marketplace platform designers can use data in ways that uncover better measures of seller quality and rather than let buyers select sellers based on measures of reputation, engineering and variants of machine learning can help match buyers to better sellers. The literature on search costs has demonstrated correlation between ranking and purchase (or click-through) behavior (Ghose et al., 2013). Indeed, buyers are more likely to select an item higher up on the search results page, implying some sort of search costs. This implies that platforms can influence a buyer's consideration set of sellers by manipulating the set of sellers that the buyer sees, as well as the order in which they appear.[15]

Marketplaces can rely on a host of other sources of internal data to infer the quality of sellers, and using engineering methods to create better matches between buyers and seller. For example, many marketplaces allow buyers and sellers to exchange messages before and after a transaction occurs. Masterov et al. (2015) showed that text-mining these messages could reveal unhappy buyers even if they chose not to leave negative feedback. This information could also be used to rank sellers by quality, and manipulate the consideration sets of buyers. More advanced implementation of Natural Language Processing can offer deeper insights into how messages translate to experience and buyer satisfaction. Marketplace platforms can then

[15] Promoting seller quality may come at the expense of providing fewer relevant items. The long-term benefit from buyers interacting with better quality sellers and returning to the site must be weighed against the short-term loss of buyers being less likely to purchase because they do not find what they want.

create engineered measures of seller performance that aggregate both what is seen publicly (past feedback) and what is not (messages), to create better measures of seller quality. Then, using their search and presentation algorithm, they can engineer how to promote better quality sellers for the continued health of the marketplace. This would alleviate the need for buyers to decipher what a certain rating means.

One nice feature of using search prioritization to match buyers with sellers is that technology can allow marketplaces to differentially match buyers with different preferences to sellers that best suit the buyers' needs. For example, new buyers may be very sensitive to their first few transactions on a platform. As such, if a platform identifies a new or inexperienced buyer, it may be in the interest of efficiency to match this buyer with the platform's very best sellers. If instead, experienced buyers are more likely to search extensively (because they have lower search costs driven by more familiarity with the marketplace platform), and as a consequence of their experience can better evaluate sellers and take risks, then the platform may wish to expose these buyers to a more diverse set of sellers that may carry a wider range of products.

Still, user-generated feedback will continue to be an important input into the myriad of metrics that marketplaces can and will use to match their buyers with high quality sellers in order to generate good experiences. The challenge will remain how to engineer ways in which more accurate feedback is generated. The experiments and rests described in Fradkin et al. (2015) suggest that a challenge for online marketplaces is the potential loss of information following any social interaction of buyers and sellers on the site. As such, marketplaces may choose some sort of incentives to motivate more truthful feedback from buyers, such as the coupons described in Fradkin et al. (2015) or the seller incentives described in Li (2010).

Another important area for future applied research, which is completely absent from the current literature, is how to combine economic thinking and analyses with human-computer interaction. Namely, it is instrumental for marketplaces to study the role played by a user's experience, such as the the way in which information is displayed, and the way in which users interact with the information and the interface. For example, Amazon and Yelp show a distribution of ratings whereas Uber and eBay only show average scores. Do buyers respond differently to different formats of information? Can too much information cause confusion? Are there formats that help people make better decisions and maybe even encourage people to interact more with the feedback system and leave more accurate feedback? Mainstream economic theory assumes that the way in which information is presented has no impact on decision making, but introspection suggests otherwise. It may therefore be instrumental to crack the mystery of cognitive limitations and bounded rationality as these pertain to the presentation of information in online marketplaces.

Finally, there is one important and far reaching challenge: how can we aggregate an individual's (or entity's) reputation across different platforms? For example, if I have accounts on Airbnb and eBay, and I own a small restaurant that has reviews on Yelp, how can I leverage these all to create some kind of meta-profile that will let me jump start trust on a new platform? A relatively new company by the name of "The World Table" is making some first steps towards a world in which a person's reputation may be aggregated across platforms, and more importantly, vetted by other people and even tied to their real identities. There are many challenges to such an approach, staring with verifiability and ending with privacy concerns. They claim that using blockchain technology–used by crypto-currencies such as Bitcoin–is the solution to the challenges of a centralized reputation score for individuals and entities. Time will tell whether these ideas will develop to create meta-reputation scores that will give individuals incentives to be diligently trustworthy because their reputation in one platform will depend on their behavior across many others.

7 Concluding Remarks

The role of reputation and feedback systems is to promote trust and trustworthiness in online marketplaces so as to reduce frictions caused by asymmetric information, and in turn increase the efficiency of these markets. The rise of online marketplaces and their penetration to practically every household is, to a large part, attributable to some level of success from these reputation and feedback systems.

With the rise of online marketplaces as a major force in overall commerce, regulators have started considering whether and how to intervene in order to protect costumers from a variety of hazards. On one hand, the huge growth of online marketplaces suggests that feedback and reputation systems do a good job at policing bad actors and minimizing fraudulent behavior, possibly eliminating the need for onerous rules and regulations. At the same time a hist of studies described in this paper show that there are all sorts of biases in feedback and reputation systems that can be improved upon.

I hope that this survey whets the appetite of economists and other researchers, as there are many exciting directions in which they can dig their teeth into new sources of data from online marketplaces to deepen our understanding of how feedback and reputation systems work. There are, I believe, many opportunities to collaborate with other disciplines, such as machine learning and cognitive sciences, to better understand the strengths and limitations of current feedback and reputation systems. It is apparent that the design of feedback and reputation systems will continue to play an important role in the broader area of market design as it applies to online marketplaces.

References

Bajari, P. and Hortaçsu, A. (2004). Economic insights from internet auctions. *Journal of Economic Literature*, 42(2):457–486.

Bar-Isaac, H. and Tadelis, S. (2008). Seller reputation. *Foundations and Trends® in Microeconomics*, 4(4):273–351.

Bolton, G., Greiner, B., and Ockenfels, A. (2013). Engineering trust: reciprocity in the production of reputation information. *Management Science*, 59(2):265–285.

Bolton, G. E., Katok, E., and Ockenfels, A. (2004). How effective are electronic reputation mechanisms? an experimental investigation. *Management Science*, 50(11):1587–1602.

Bond, E. (1982). A direct test of the 'lemons' model: The market for used pickup trucks. *American Economic Review*, 72(4):836–840.

Brown, J. and Morgan, J. (2006). Reputation in online auctions: The market for trust. *California Management Review*, 49(1):61–81.

Cabral, L. and Hortaçsu, A. (2010). The dynamics of seller reputation: Evidence from ebay*. *The Journal of Industrial Economics*, 58(1):54–78.

Cai, H., Jin, G. Z., Liu, C., and Zhou, L.-A. (2014). Seller reputation: From word-of-mouth to centralized feedback. *International Journal of Industrial Organization*, 34(C):51–65.

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

Dellarocas, C. and Wood, C. A. (2008). The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science*, 54(3):460–476.

Fradkin, A., Grewal, E., Holtz, D., and M., P. (2015). Bias and reciprocity in online reviews: Evidence from experiments on airbnb. *16th ACM Conference on Electronic Commerce, June 2015, (EC 2015)*, pages xxx–yyy.

Genesove, D. (1993). Adverse selection in the wholesale used car market. *Journal of Political Economy*, 101(4):644–665.

Ghose, A., Ipeirotis, P. G., and Li, B. (2013). Examining the impact of ranking on consumer behavior and search engine revenue. *Management Science, Forthcoming.*

Greif, A. (2006). *Institutions and the Path to the Modern Economy.* Cambridge Unievrsity Press, Cambridge, UK.

Hendricks, K. and Porter, R. H. (1988). An empirical study of an auction with asymmetric information. *American Economic Review*, 78(5):865–883.

Horton, J. and Golden, J. (2015). Reputation inflation: Evidence from an online labor market. *Working paper, New York University.*

Hubbard, T. (2002). How do consumers motivate experts? reputational incentives in an auto repair market. *Journal of Law and Economics*, 45(2):437–468.

Hui, X.-A., Saeedi, M., Sundaresan, N., and Shen, Z. (2014). From lemon markets to managed markets: the evolution of ebay's reputation system. *Working paper, Ohio State University.*

Jin, G. and Leslie, P. (2009). Reputational incentives for restaurant hygiene. *American Economic Journal: Microeconomics*, 1(1):237–67.

Jin, G. Z. and Kato, A. (2006). Price, quality, and reputation: evidence from an online field experiment. *The RAND Journal of Economics*, 37(4):983–1005.

Klein, T., Lambertz, C., and Stahl, K. (2015). Adverse selection and moral hazard in anonymous markets. *Journal of Political Economy*, forthcoming.

Kreps, D. M. (1990). Corporate culture and economic theory. In Alt, J. E. and Shepsle, K. A., editors, *Perspectives on Positive Political Economy*, pages 90 – 143. Cambridge University Press, Cambridge, UK.

Kreps, D. M., Milgrom, P., Roberts, J., and Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, 27(2):245–252.

Li, L. I. (2010). Reputation, trust, and rebates: How online auction markets can improve their feedback mechanisms. *Journal of Economics & Management Strategy*, 19(2):303–331.

Li, L. I., Tadelis, S., and Zhou, X. (2016). Buying reputation: Evidence from the rebate-for-feedback mechanism in alibaba. *Working paper.*

Li, L. I. and Xiao, E. (2014). Money talks: Rebate mechanisms in reputation system design. *Management Science*, 60(8):2054–2072.

Livingston, J. A. (2005). How valuable is a good reputation? a sample selection model of internet auctions. *Review of Economics and Statistics*, 87(3):453–465.

Masterov, D. V., Mayer, U. F., and Tadelis, S. (2015). Canary in the e-commerce coal mine: Detecting and predicting poor experiences using buyer-to-seller messages. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, EC '15, pages 81–93, New York, NY, USA. ACM.

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–55.

McDonald, C. G. and Slawson, V. C. (2002). Reputation in an internet auction market. *Economic Inquiry*, 40(4):633–650.

Melnik, M. I. and Alm, J. (2002). Does a seller's ecommerce reputation matter? evidence from ebay auctions. *The Journal of Industrial Economics*, 50(3):337–349.

Milgrom, P. R., North, D. C., and Weingast, B. R. (1990). The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. *Economics & Politics*, 2(1):1–23.

Nosko, C. and Tadelis, S. (2015). The limits of reputation in platform markets: An empirical analysis and field experiment. *NBER Working paper*, (No. w20830).

Resnick, P., Zeckhauser, R., Kuwabara, K., and Friedman, E. (2000). Reputation systems. *Communications of the ACM*, 43(12):45–48.

Resnick, P., Zeckhauser, R., Swanson, J., and Lockwood, K. (2006). The value of reputation on ebay: A controlled experiment. *Experimental Economics*, 9(2):79–101.

Roth, A. E. (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, 70(4):1341–1378.

Xu, H., Liu, D., Wang, H., and Stavrou, A. (2015). E-commerce reputation manipulation: The emergence of reputation-escalation-as-a-service. *Proceedings of 24th World Wide Web Conference*, (WWW 2015):1296–1306.

Zervas, G., Proserpio, D., and John, B. (2015). A first look at online reputation on airbnb, where every stay is above average. *Working paper, Boston University*.