5

Reputation, Feedback, and Trust in Online Platforms

Steven Tadelis

5.1 INTRODUCTION

As a recent McKinsey article (Briedis et al. 2020) states, "digital marketplaces have been the buzz of the consumer industry for the past several years." Indeed, online marketplaces have grown dramatically over the past two decades and bring value across many areas of our lives. The recent forced lockdowns caused by the COVID-19 pandemic have accelerated the use of online marketplaces even further. Platforms such as eBay, Taobao, Flipkart, Amazon Marketplaces, Airbnb, Uber, Upwork, and many more match consumers and businesses to businesses and individuals and create gains from trade in efficient and effective ways by: (i) allowing businesses to market their goods or get rid of excess inventory; (ii) saving businesses the costs needed to establish their own e-commerce website to generate online consumer traffic; (iii) allowing individuals to get rid of items they no longer need and transform these into cash; (iv) allowing individuals to share their time or assets across different productive activities; and (v) allowing businesses to hire short- and longer-term contractors and employees to perform a variety of remote tasks.

Looking back at the success of these platforms, especially the sharing economy platforms where asset owners allow temporary asset usage by other users, a natural question arises: 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? 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 at the matching stage at multiple sharing economy platforms, which means that users may not be able to verify each other's identities in person when they commit to a transaction. Hence, to many, the rise of multisided online marketplaces, and sharing economy platforms in particular, was not foreseen.

This chapter is adapted from an article titled "Reputation and Feedback Systems in Online Platform Markets," which appeared in the *Annual Review of Economics* in 2016 and is being published with permission.

Indeed, basic economic theory would predict that these platforms should face an uphill battle. In his seminal article "The Market for 'Lemons'", Akerlof (1970) showed how hidden information in the hands of sellers causes markets to fail despite gains from trade. Intuitively, two sources of uncertainty can prevent markets from operating efficiently. First, uncertainty about the quality of a transaction may be inherent to the good or service provided like in Akerlof's adverse selection model, which means that users should be wary of purchasing access to assets they cannot inspect. For example, some hosts on Airbnb may know that the apartments they are offering for a short-term rent are defective, yet they may choose not to reveal the defect and misrepresent their items. Second, quality uncertainty may be a result of unobserved actions that determine the quality of the good or service, what is often referred to as "moral hazard." For example, a host on Airbnb may choose to skimp on cleaning between guest stays and increase the likelihood that the apartment will be in a bad shape when next guest arrives. Of course, both hidden information and unobserved actions might be present simultaneously.

It is therefore 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. This is where new-world online platforms took a page out of an old-world playbook by creating feedback and reputation systems that became central to their operations. The need for reputation-based incentives to foster trust and guarantee successful market operation is an old idea that has been part of commerce for centuries. Just as digital transformation made possible instantaneous matching of users in modern sharing economy markets, the need to coordinate where and when market transactions took place was an important historical innovation. Take, for example, the introduction of trade fairs in medieval Europe (see Grief 2006) in which the successful trade between parties who had never met was supported by governance and reputation mechanisms that gave traders the faith to trade with strangers (see Milgrom, North, and Weingast 1990).

In this chapter, I explain how feedback and reputation systems work in practice, and how they support online markets. While most of the examples I use refer to the e-commerce marketplaces context, the key mechanisms behind interactions between buyers and sellers are similar to the mechanisms driving sharing economy markets (for example, interactions between hosts and guests on Airbnb or between drivers and riders on Uber). Section 5.2 presents the theory behind reputation mechanisms and how they support more efficient trade. Section 5.3 describes the actual working of typical online feedback and reputation systems while Section 5.4 presents findings from a host of empirical papers that explore how reputation works in actual online marketplaces. Section 5.5 highlights some shortcomings of reputation systems and Section 5.6 suggests some considerations for the future design of feedback and reputation systems that can augment their effectiveness. Section 5.7 offers some closing thoughts.

5.2 REPUTATION AND FEEDBACK: THEORY

The difficulty in supporting anonymous online trade can be explained using a relatively simple game-theoretic example.[1] Consider a buyer who finds an online product sold by an anonymous seller. The buyer values the product at \$25, the purchase price is \$15, and the seller has no use for the good, so the seller receives a net value of \$0 if it does not sell the good. The seller's costs of shipping and handling are \$5, so at a price of \$15 the seller will make \$10, and the buyer, paying \$15 for what he values at \$25, is left with a net surplus (or dollar-value utility) of \$10.[2] Further imagine that the buyer must first send money to the seller (as in clicking "buy" online) and then the seller can send the good to the buyer.

The standard assumption in economics is that people are selfish utility maximizers. This would imply that if the buyer clicks "buy" and pays, then the seller is better off just keeping the \$15 and sending nothing, thus saving the \$5 of shipping fees. But we know that many people care about being honest, and hence we need to capture a world where this is the case. To do this, imagine that the seller can be one of two types: one is an honest seller who makes good on promises, and the other is an opportunistic seller who only cares about their own profits. The buyer does not know the type of the seller but does believe that a seller is honest with some well-defined probability $p \in (0,1)$.[3] This simple game is shown in Figure 5.1.

Figure 5.1 describes the story above as follows: "Nature" determines whether the seller that the buyer faces is honest or opportunistic; the reason that both nodes are circled in a dashed ellipse is to indicate that the buyer does not know which type of seller they face, but they do know that the seller is honest with probability p. Hence, the buyer assigns probability p to being on the right side of the ellipse. If the buyer chooses not to trust the seller and not engage in trade, then regardless of the type of seller, there is no trade and both parties get zero. If instead the buyer chooses to trust, then if the seller is honest the game will end in a successful trade, while if the seller is opportunistic then it may go either way, depending on the choice of the opportunistic seller. In the jargon of economics, this game includes both "hidden information" (the type of seller is not known to the buyer) and "hidden action" (the opportunistic type has agency to act in a way that can harm the buyer).

Now ask yourself what would happen if this game is played only once? Naturally, the opportunistic seller would choose to abuse trust because it saves money and offers a higher profit. In turn, a rational buyer would anticipate this behavior and

[1] I understand that not all readers may be familiar with the tools of game theory and hence, I tried to keep this example as simple as possible and minimize the use of jargon.

[2] This scenario can be thought of as follows: If the buyer is offered to choose freely between two options, one is getting the good, and the other is getting \$25, then this buyer is indifferent between the options – they are equally good. Hence, if the buyer is offered to pay \$15 for the good, it is as if the buyer is getting \$10 worth of net value.

[3] You may wonder where p comes from. In game theory we assume that people have correct beliefs about the environment, possibly from their experience or possibly from trusted sources of information.

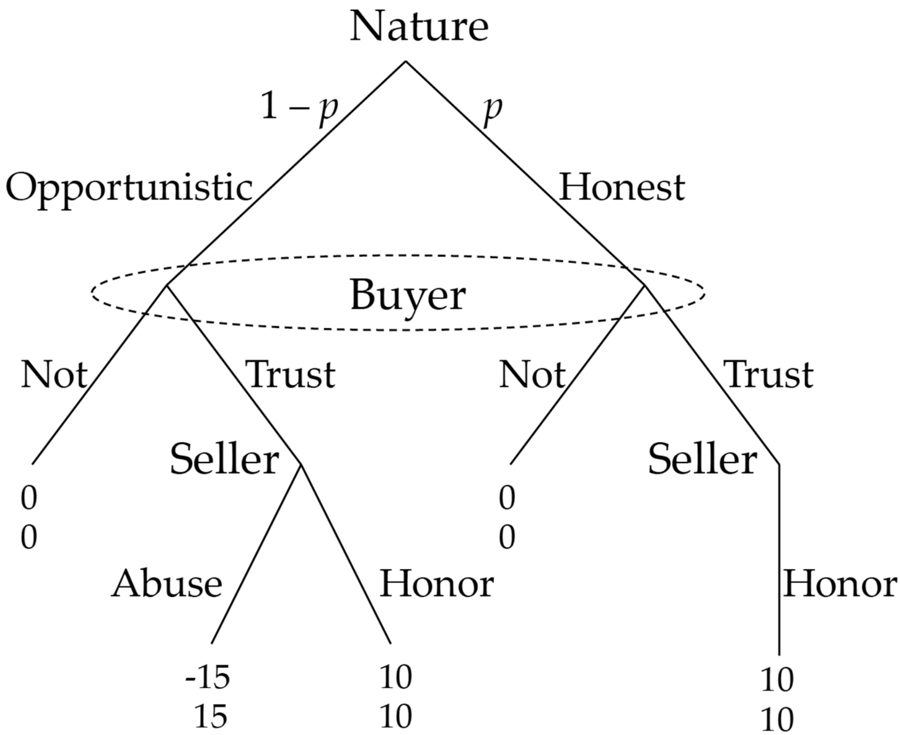

FIGURE 5.1 Trade game with asymmetric information.

will choose to trust the seller if and only if the risk is worthwhile. Assuming that the buyer is risk neutral (i.e., maximizes expected value), the risk is worthwhile if and only if the expected value from trusting the seller is larger than zero (the value of not trading), i.e., $10p+(-15)(1-p)\geq 0$ or $p\geq 0.6$. Intuitively, if the likelihood of an honest seller is high enough (greater than 60 percent) then the risk is worth taking.

Now assume that most people are known to be honest so that $p>0.6$ and any buyer would be happy to transact, and risk being cheated because the likelihood of successfully completing a trade is high enough. Imagine now that our seller will be able to transact with the buyer twice consecutively, across two periods – say in month 1 and month 2. An honest seller will always act honestly, while an opportunistic seller wishes to maximize its expected profits. Being forward looking, assume that at the beginning of month 1 the opportunistic seller discounts future profits in month 2 at a discount factor of $\delta\in(0,1)$. This means that \$1 in month 2 is worth \$$\delta$ at the beginning of month 1.[4]

It is clear that in month 2, when the opportunistic seller is facing its last transaction (there are no future trade opportunities), then it will choose to abuse trust just

[4] This is just like a business that can borrow today against future profits, but must pay some interest rate, making the amount it must pay back higher (it can borrow \$$\delta$ today and repay \$1 next month).

https://doi.org/10.1017/9781108865630.006 Published online by Cambridge University Press

as it would when there is only one opportunity to trade as shown previously. The question is, what would an opportunistic seller do in month 1 when there is a future trade opportunity? As we will now show, if δ is not too small (the future is important enough), then an opportunistic seller will no longer choose to abuse trust in the first transaction and instead will want to "build a reputation" of being honest.

The argument is a bit subtle. Imagine that the buyer in month 1 believes that an opportunistic seller will abuse trust in the first transaction. The buyer is still willing to trust the seller because $p > 0.6$, but in month 2 the buyer will expect to update beliefs about the type of seller they face. If the buyer believes that an opportunistic seller always abuses trust, it follows that in month 2, the buyer can use the seller's performance in the first month to form updated beliefs about the type of the seller: If the first month's transaction was honored then the seller must be honest, and should be trusted again; if the first transaction failed then the buyer infers that the seller must be opportunistic, and hence would choose to not trust the seller in month 2. With these beliefs in place, however, if the future is important enough then an opportunistic seller would not find it beneficial to abuse trust in the first month. To see this, first note that abusing trust would result in a payoff of \$15 for the opportunistic seller because they will receive \$15 in month 1 from abusing trust once, and they will not be trusted a second time and receive \$0 in month 2. If instead the opportunistic seller chooses to honor trust, then they receive \$10 in the first month, and because the buyer will (incorrectly) infer that the seller is honest, then the buyer will trust the seller in month 2 allowing the seller to then abuse trust in the second transaction and acquire another \$15. If the added value of \$15 in the future transaction outweighs the loss of \$5 in the first transaction, then pretending to be honest will pay off. This happens if and only if $\$10 + \$15\delta > \$15$ or $\delta > \frac{1}{3}$. Hence, if honesty is common enough ($p > 0.6$) and the future is important enough ($\delta > \frac{1}{3}$), then the opportunistic seller will choose to honor trust in order to get access to the money he can obtain from the second transaction.[5]

What is more interesting is that with such a two-period game, where early behavior of the seller influences future buyers to update their beliefs, trade can occur even when buyers are less optimistic about the seller's honesty ($p < 0.6$). To see this, imagine that the buyer believes that an opportunistic seller will abuse trust always, which are the most pessimistic beliefs a buyer can have. If the buyer trusts the seller in the first month, then with probability p, the buyer faces an honest seller and will obtain a payoff of \$10, and then will know for sure that the seller is honest, and she'll obtain another payoff of \$10 in the second month. If instead trust is abused in the first month, then the buyer will opt out of transacting again. Assuming that

[5] In fact, in the jargon of game theory, the unique sequential equilibrium (and the unique perfect Bayesian equilibrium) of this game with these assumptions has the opportunistic seller behaving honestly in the first period. For more on these concepts and for a more formal treatment of the material see Tadelis 2012 (Part V).

https://doi.org/10.1017/9781108865630.006 Published online by Cambridge University Press

the buyer also uses δ as their discount factor, then with these beliefs the buyer will choose to trust in month 1 if and only if,

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

Note that if δ becomes infinitesimally small, then for the buyer to trust the seller it is necessary that $p \geq 0.6$ because it effectively becomes a one-time game for the buyer (as δ becomes infinitesimally small, it means that the payoffs from the second month become insignificant and can be ignored). If, on the other hand, the buyer is extremely patient so that δ approaches 1, then the buyer will trust the seller if $p \geq \frac{3}{7}$. The analysis performed earlier implies that the opportunistic seller would rather imitate the honest type and cooperate in the first month of trade. Hence, with a high enough discount factor we get "more trade" in the sense that trust is supported for lower values of p (a lower propensity of honest sellers). And if we add more potential trade periods in the future, then trade will occur for even lower likelihoods of the seller being honest.[6]

The key idea is that a seller's actions today will lead to future consequences, which keep him in check. This mechanism even works if the seller does not interact repeatedly with the same buyer as long as the seller understands that his current actions will be revealed to all future buyers 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. What's more, the value of the business itself will depend on the seller's past performance (see Kreps 1990; Tadelis 1999), an insight that sheds light on the powerful role that reputation and feedback systems play in fostering trust. A public reputation repository allows all future potential buyers to track a seller's past performance, and reputation becomes an important incentive mechanism that facilitates trust in anonymous markets.

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 put to empirical scrutiny and tested using market-level data. In fact, the recent rise of online platforms with "big data" on buyer and seller behavior have proven to be a fertile ground to test these implications.

6 This is an example of a game in the spirit of the seminal work by Kreps et al. (1982). For some values of $p \geq 0.6$ the equilibrium involves the seller using "mixed strategies" in the first stage (i.e., the seller randomizes between honoring and abusing trust), as well as mixed strategies for the buyer in the second stage (i.e., the buyer randomizes between trusting and not trusting). 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. See chapter 17 in Tadelis (2012).

5.3 REPUTATION AND FEEDBACK: PRACTICE

Many have attributed the success of eBay, the very first online marketplace that grew rapidly to mediate the trade of scores of transactions, to its reputation and feedback mechanism (see, e.g., Resnick et al. 2000 and Dellarocas 2003). Indeed, eBay exists as a successful business despite the complete anonymity of the marketplace. Following eBay's lead, practically every online marketplace, including modern sharing economy markets, has adopted some form of a reputation or feedback system. Herein I will describe in detail how eBay's feedback system works, which should give the reader an idea of how these kinds of systems work elsewhere.

A well-functioning reputation system provides future buyers with information about each seller's past behavior, information that is generally produced by the voluntary input of buyers. After completing a transaction on eBay, a buyer has sixty days to leave either a positive, negative, or neutral feedback score for the seller. 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 automatic positive feedback under the assumption that silence is most likely a sign of buyer satisfaction.[7] As I explain in Section 5.5, this may be far from the truth. Also note that leaving feedback requires some time, so a buyer may selfishly choose to leave no feedback at all. Interestingly, back in 2016 about 65 percent of buyers left feedback on eBay, and an even higher fraction (more than 80 percent) left feedback in eBay's earlier days.[8]

Figure 5.2 shows how a seller's feedback, which I refer to as the reputation measure of the seller, is calculated and displayed on eBay. A new Apple MacBook is 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.[9]

To learn more about the seller's history, buyers can click on the feedback score (the number 21,814 in Figure 5.2), which directs them to a detailed feedback profile page that is shown in Figure 5.3, which displays how many positive, neutral, or negative feedback reviews the seller received in the past month, six months, or twelve months.[10] At the bottom of the page there is a rolling list of verbal comments left

[7] See Li et al. (2020) for more on Taobao's reputation system.

[8] 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).

[9] Notice the "Top Rated Plus" badge at the upper right corner of Figure 5.2. 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 feature.

[10] While the feedback score is calculated using all past transactions, the percent positive only uses the last twelve months of a seller's transactions.

https://doi.org/10.1017/9781108865630.006 Published online by Cambridge University Press

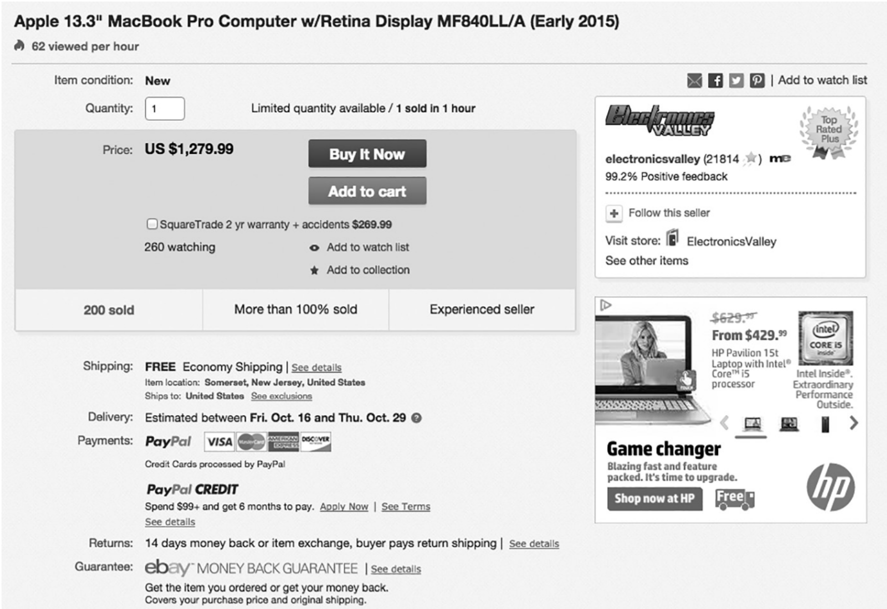

FIGURE 5.2 eBay's view item page displaying feedback.

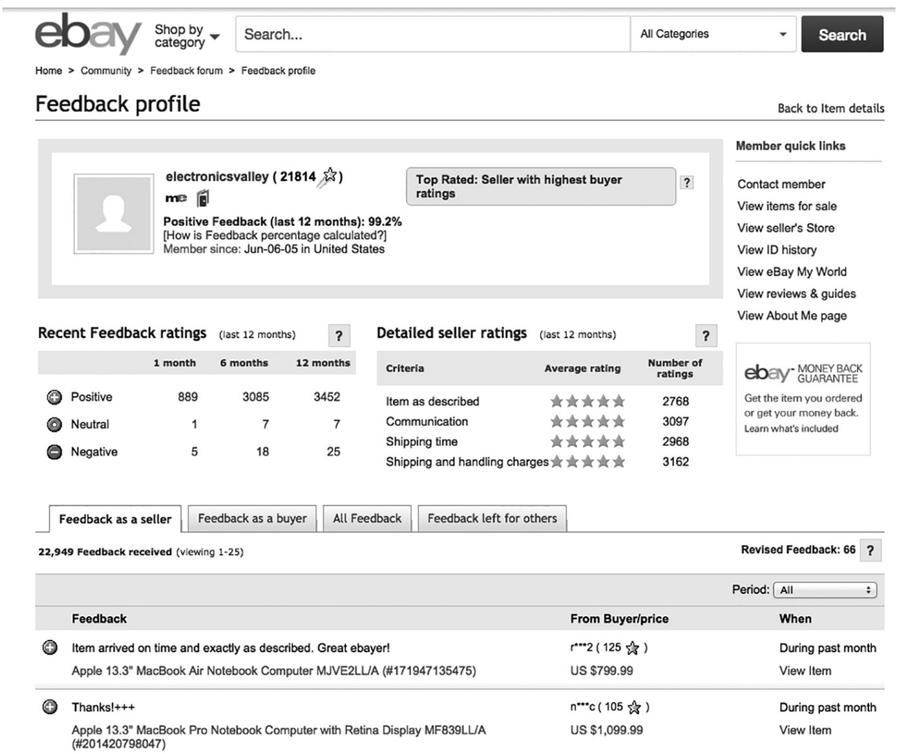

FIGURE 5.3 eBay's display of a seller's feedback profile.

https://doi.org/10.1017/9781108865630.006 Published online by Cambridge University Press

by buyers, 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.

Many ecommerce platforms use a star system (typically one through five stars). It is important to note that reviews may be about the product rather than the seller, a well-known example being the product reviews on Amazon. Platforms must be careful about distinguishing between product reviews and seller reviews in order to avoid confusion. 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.

Before 2008, buyers and sellers on eBay could leave each other positive, negative or neutral feedback with comments. In 2008 eBay changed the feedback system limiting sellers 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; 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 practically "one-sided," like it is currently on eBay and on Amazon's marketplace, is an important design question. In essence, the platform benefits when feedback is informative and does not cause unnecessary friction as part of the user experience. For example, before eBay acquired PayPal's online payment system, buyers would send checks or money orders to sellers, meaning that buyers can abuse trust by not sending payment, just as sellers can abuse trust by not shipping the product. As a result, both buyers and sellers needed some guidance about which counterpart is trustworthy. After eBay acquired PayPal, however, it encouraged sellers to use PayPal as the only form of payment. By doing this, gone were the problems of buyers not sending checks and, for the most part, the problem of buyer abuse was solved, which in turn means that feedback left by sellers for buyers became a lot less valuable.[11]

Still, for some time eBay kept the two-sided feedback system, only to later learn that there is a weakness to two-sided feedback. In collaboration with eBay, Bolton, Greiner, and Ockenfels (2013) used data from eBay during the period when the reputation system was two-sided, and convincingly showed that sellers retaliated against buyers with their feedback. To illustrate their findings, consider pairs of feedback scores, $\left(FB_i, FS_j\right)$ 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 $(+,-)$. The data 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

[11] Buyer abuse still persisted at a very low level: Some buyers would threaten to leave sellers negative feedback for no reason in order to get some partial refund, but the prevalence of this behavior was in the low single-digit percentages during the time I was at eBay.

https://doi.org/10.1017/9781108865630.006 Published online by Cambridge University Press

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 payment is mediated by the site, as it is now, there is still a concern that abuse may occur from either side of the market. The dwelling owners can cause harm to renters in many forms, such as misrepresenting the dwelling, leaving it dirty, not giving the renters a key at the prespecified 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 too can harm the dwelling owner by leaving the space dirty, causing damage, being very noisy, causing the owner trouble, and more. As such, it is imperative that Airbnb continues 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 relative costs and benefits of one- versus two-sided feedback systems.

5.4 HOW WELL DOES REPUTATION WORK?

The data made available by online marketplaces have enabled scholars to study how online feedback and reputation mechanisms work in practice.[12] Most studies used "scraped" data from marketplace webpages and explored whether sellers with higher reputation scores and more transactions receive higher prices and whether reputation seems to matter more for higher priced goods than for lower priced goods.

Early studies collected what are now considered tiny datasets. For example, McDonald and Slawson (2002) collected data from 460 auctions completed in 1998 of collector-quality Harley Davidson Barbie dolls. Because the closing price and the number of bids is bound to be correlated, they used an approach known as Seemingly Unrelated Regressions to simultaneously estimate the effect of reputation 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 number of bids. However, the interpretation of McDonald and Slawson (2002), as well as similar studies, raises some concerns. First, good reputation may be correlated with a variety of omitted variables influencing the dependent variables. The authors themselves acknowledge unclear language and grammar as possibly confounding factors. Second, though statistically significant, the results are economically very small: for example, a one-point increase in reputation corresponds to four cents

[12] See Resnick et al. (2000) for a more detailed discussion of these studies.

increase in final price. With a median sample reputation score of twenty-one points, this means that increasing reputation from none at all to the median increases price by less than $1, a fraction of a percent of the median price of $275.

Jin and Kato (2006) took a different approach and studied the relationship 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. Jin and Kato (2006) collected data of the five most traded baseball cards on eBay for seven months, resulting in 1,124 auctions, of which 67 percent were graded. Of the full sample, 81 percent 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 made by the seller. The data indicate 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, but buyers were still willing to pay more for cards with higher claimed quality. Furthermore, an increase in claimed quality significantly raises the probability of an auction ending with a sale, but ratings do not seem to create a reputation premium, consistent with Livingston (2005). Perhaps the most interesting result in Jin and Kato (2006) is that reputable sellers (using eBay's ratings) are less likely to make extreme claims about their cards' quality. 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.

Cabral and Hortaçsu (2010) used a different strategy by collecting a series of seller feedback histories, thus creating a panel of seller histories, and proceed to estimate the effect of changes in reputation on a seller's sales rate. They find that when a seller receives his first negative feedback rating, his weekly sales growth rate drops from a positive rate of 5 percent to a negative rate of 8 percent. To overcome the fact that eBay does not provide information on how many past transactions a seller completed, Cabral and Hortaçsu (2010) make two assumptions and provide evidence that they are reasonable. First, they assume that a seller's frequency of feedback is a good proxy for the frequency of actual transactions, and second, that feedback correlates with buyer satisfaction.

An interesting question is what the impact would be of introducing a reputation system into a marketplace that did not have one. To answer this question, Cai et al. (2014) study the case of Eachnet, a Chinese auction cite that had about a 90 percent share of the Chinese market during its years of operation (1999–2003). Unlike eBay, exchange of products and money was done offline, and this face-to-face exchange may create less uncertainty at the time of the transaction. In 2001 Eachnet introduced a feedback system that enabled buyers to rate sellers after each transaction. Cai et al. (2014) received a large amount of data from the platform, containing a random sample of 125,135 sellers who posted almost two million listings throughout Eachnet's years of operation. Because sellers' feedback scores are obviously not available prior to the introduction of centralized feedback, their prefeedback reputation is approximated with the cumulative number of successful listings each seller had since joining Eachnet. Cai et al. (2014) examined how a seller's reputation, as approximated by cumulative success rates, affects the seller's behavior and outcomes, and how things changed after feedback was made available. First, an increase in the cumulative success rate of a seller is correlated with a larger fraction of repeating buyers, but the effect weakens after making feedback available. This is intuitive: centralized feedback is a substitute to the trust built in relationships. Second, centralized feedback leads sellers with higher cumulative success rates to sell more products in more regions, suggesting that formal feedback helps reputable sellers expand into new markets. Last, a higher cumulative success rate is generally correlated with a lower hazard rate of exiting the market, an effect that diminishes after feedback centralization. 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).

A shortcoming of observational studies, like those just described, is the possible "endogeneity" concern from potential selection or omitted-variables biases. In other words, it is possible that sellers with higher reputation scores exhibit other information that causes buyers to be more interested in their products, so that it is not the reputation score per se that is accountable for the observed outcomes. Two ways around this are either a randomized controlled experiment, where similar goods and experiences are sold by sellers that are the same but for their reputation, or some source of exogenous variation in reputation scores that are not correlated with other important drivers of outcomes.[13] Resnick et al. (2006) ran a controlled field experiment by offering a series of sales of identical items (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 estimated an 8 percent price premium to having 2,000 positive and 1 negative feedback over a reputation of 10 positive and no negative feedbacks, which is quite sizable.

Klein, Lambertz, and Stahl (2013) cleverly took advantage of a change in the way that eBay reported feedback, together with the fact that feedback for sellers has two components: the nonanonymous simple feedback of positive, negative, and neutral ratings, and the anonymous feedback of Detailed Seller Ratings (or DSRs, as seen in

[13] This relates to the problem of distinguishing causation from correlation, as described in detail in Angrist and Pischke (2008). A randomized controlled experiment controls for all but one variable of interest, and creates two groups where only the variable of interest differs and all else remains the same. This allows the researcher to measure the causal effects of changes in the variable of interest without the concern that other important variables may also differ across the two groups. Exogenous variation refers to cases where there was no carefully designed experiment, but where for other reasons the researcher can be confident that the changes in the variable of interest are not correlated with other important variables that determine outcomes.

Figure 5.3). In May 2008, after realizing the problem of seller retaliation (recall the discussion of retaliation in Section 5.3), eBay removed a seller's ability to leave the buyer negative or neutral feedback. The belief was that this change will encourage buyers to report negative feedback following a poor experience, which can cause sellers to respond in two ways. First, some really bad sellers may leave eBay following a series of negative ratings. Second, sellers may work harder to improve buyer satisfaction. Klein et al. (2013) scraped data containing monthly information on feedback from about 15,000 eBay users between July 2006 and July 2009, a period that included both the introduction of anonymous DSR ratings (May 2007) and of one-sided feedback (May 2008). They found 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.

5.5 BIASES IN ONLINE FEEDBACK SYSTEMS

The studies just discussed suggest that reputational forces are at work in online marketplaces, but a question remains: How accurately does feedback capture variation in performance? As suggested earlier, retaliation on eBay may have caused feedback to be biased, as buyers chose to refrain from leaving negative feedback for sellers. A recent literature has demonstrated that user-generated feedback mechanisms suffer from bias. Dellarocas and Wood (2008) conjectured that the extremely high percent-positive reputation measures on eBay may be because many buyers who suffered poor experiences chose not to leave feedback at all.[14] They derive implications from the fact that eBay's reputation system was two-sided (buyers and sellers leave each other feedback) and use these implications to develop an econometric technique that uncovers the true percent of positive transactions. However, because eBay switched to one-sided feedback after 2008, their proposed approach no longer works.

Nosko and Tadelis (2015) use internal eBay data to directly show how biased reputation measures really are. Their data show 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 5.4, 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. One naive conclusion is that the reputation system works exceptionally well because bad sellers (below the high 90s) leave the platform. 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.

[14] Li (2010) proposed 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) extended the model and conducted a laboratory experiment 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.

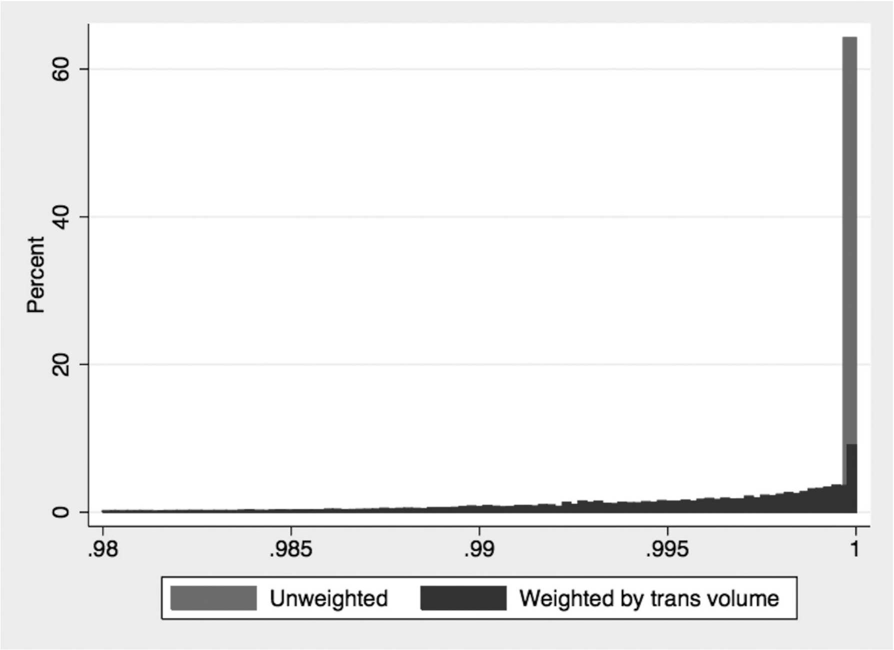

FIGURE 5.4 Percent positive of sellers on eBay.

The findings in Bolton et al. (2013) described in Section 5.3, that sellers retaliated towards buyers who left them negative feedback, and the evidence described in the study, suggest that for buyers, leaving different types of feedback entails different consequences. In particular, buyers find it more "expensive" to leave a negative review than a positive one, which in turn means that with a given propensity to leave feedback, this asymmetry between leaving positive versus negative feedback inherently creates upward bias. Nosko and Tadelis (2015) suggest a new quality measure, "effective percent positive" (EPP), which is calculated by dividing the number of positive feedback transactions by the total number of transactions. This penalizes sellers who are associated with more transactions for which the buyers left no feedback, based on the insight that no feedback includes in it a measure of negative outcomes.

The distribution of the EPP measure is described in Figure 5.5 using the same set of sellers for which the percent-positive scores were described in Figure 5.4. EPP has a mean of 64 percent, a median of 67 percent, and exhibits significantly more variation than percent positive. But can this be verified as a better measure of quality?

To demonstrate this, Nosko and Tadelis (2015) use a "revealed preference" approach to study the effect of a seller's EPP on a buyer's propensity to continue buying on eBay after that transaction. This distinguishes their paper from the papers that collect scraped data and cannot track the behavior of buyers on the site and

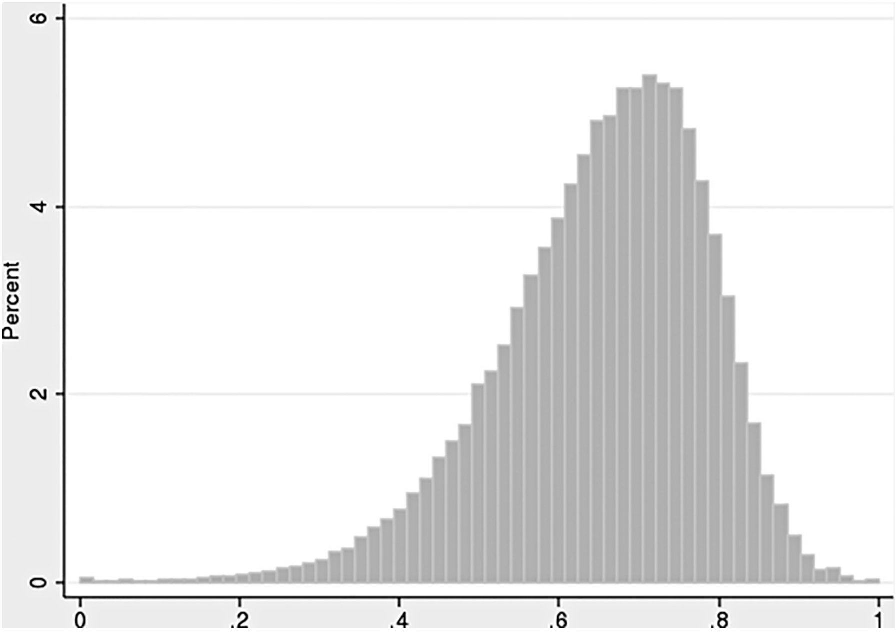

FIGURE 5.5 Histogram of sellers' effective percent-positive scores.

allows them to get to the heart of the question of whether reputation mechanisms help buyers avoid low quality sellers. Importantly, eBay does not display the total number of transactions a seller has completed, and buyers cannot therefore compute a seller's EPP score.

Nosko and Tadelis (2015) show that a buyer who buys from a seller with a higher EPP score is more likely to continue to transact on eBay again in the future, which by revealed preference suggests a better experience. They also report results from a randomized 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. Jaffe et al. (2019) use data from Airbnb to further explore the revealed preference approach of Nosko and Tadelis (2015) and find similar results. Further implications about the design and engineering of feedback systems are discussed in Section 5.6.

Mayzlin, Dover, and Chevalier (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. What makes that 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

https://doi.org/10.1017/9781108865630.006 Published online by Cambridge University Press

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. 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 Mazylin 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 nonpurchasers.

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) causes 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. Like Bolton et al. (2013), 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 harmful feedback depends on the market penalty associated with that feedback. The paper argues that these two factors together can create a race of ever-increasing reputations. Zervas, Proserpio, and Byers (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 percent of property ratings with 4.5 or 5 stars.

One more channel through which bias in reputation may occur is by sellers trying 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.[15]

5.6 ENGINEERING REPUTATION SYSTEMS

Economic theory takes the view that market participants understand the equilibrium they are playing, and correctly infer information from signals and actions. In practice, however, buyers may not correctly interpret the feedback information they are presented with. Naively, in some sort of absolute scale, a score of 98 percent is considered excellent. But, as Nosko and Tadelis (2015) show, on eBay this score places a seller below the tenth percentile of seller feedback, and it is unclear whether the more informative EPP measure constructed by Nosko and Tadelis (2015) would be interpreted correctly by buyers. For this reason, Nosko and Tadelis (2015) propose not to reveal effective measures to buyers, but instead choose to run a controlled experiment that incorporated the EPP measure into eBay's search-ranking algorithm.

This approach offers a new direction for improved marketplace performance. Instead of showing buyers information on seller quality, platforms can benefit from a more paternalistic, or regulator-like approach, that does not rely on participants correctly deciphering information. 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). Trust can therefore be engineered by way of a process in which recommendations rely on underlying data that is not made visible to buyers. Of course, this requires buyers to trust that the platform is operating in their best interest, a trust that I believe is justified. Just as the motivation of repeat business is at the heart of the value of a good reputation, so does future business motivate platforms to offer buyers a positive experience every time they purchase on a marketplace platform.

Marketplaces can rely on a variety of internal data to infer the quality of sellers. For example, many marketplaces allow buyers and sellers to exchange messages before and after a transaction occurs. Masterov, Mayer, and Tadelis (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] 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. (2020) for a study of such behavior in Taobao's marketplace.

https://doi.org/10.1017/9781108865630.006 Published online by Cambridge University Press

create engineered measures of seller performance that aggregate both what is seen publicly (past feedback) and what is not (messages or customer service complaints), to create better measures of seller quality. Search algorithms can be engineered to promote better quality sellers for the continued health of the marketplace, alleviating buyers from deciphering what a certain rating means.

User-generated feedback will continue to be an important signal that marketplaces will use to match buyers with high quality sellers, and the challenge of engineering ways to procure more accurate feedback remains. The experiments 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 use of coupons to motivate feedback as described in Fradkin et al. (2015) and similarly, the use of rebates for feedback described in Li, Tadelis, and Zhou (2020).

One last issue warrants discussion, especially because of tension it raises with the common belief that a drive for transparency is likely to be central to any regulatory oversight, as well the discomfort it creates for the fundamental approach of game theory. As mentioned earlier, Nosko and Tadelis (2015) propose not to reveal effective measures to buyers to remove the burden to interpret what feedback really means. However, they discuss a second reason not to make new measures of reputation transparent: transparency can reduce the quality of information that the platform can generate from these measures. Take for example the effective percent-positive score developed by Nosko and Tadelis (2015), which counts silence as a negative mark against a seller. If sellers learn that this is part of the way they are evaluated, then they would harass buyers who do not leave them feedback, which in turn can both generate positive feedback when it is not warranted, as well as drive buyers to abandon the platform. This suggests that transparency is not necessarily the best policy. Indeed, Google's famous "Quality Score" is used by the company to rank sponsored search links, and though the company does give guidance on ways to improve the score, they do not reveal the exact way in which it is estimated. This allows them to control the user experience for quality of ads because, after all, the main reason so many people use Google is because of the quality of the search engine and the relevance of ads.

As for the discomfort with game theory, note that at the heart of "equilibrium" analysis is that every player has correct beliefs about everything relevant to their environment. But in the case of platform quality, it may be in the best interest of the platform (and its buyer-side) that sellers are left in the dark about some aspects of how they are evaluated, and only given coarse feedback about how to improve. If the exact formula of evaluation relies on hidden biases in the data, then revealing these hidden biases may cause behavior that will undermine the value of the current formulas, effectively causing the platform designers to play a cat-and-mouse game with abusive sellers. Even though I agree with eBay's founder, Pierre

Omidyar, who was famously quoted saying that "People are basically good," platforms need to safeguard against the few who try to take advantage of others, and this seems to require making sure that bad actors do not have insight into the ways they are being detected.

5.7 CONCLUDING REMARKS

Reputation and feedback systems are critical to foster trust and trustworthiness in online marketplaces. The rise of these platforms and their penetration to practically every household owes much of its success to reputation and feedback systems.

In the past few years there has been a lot of scrutiny by regulators who question whether and how to protect customers from a variety of hazards on sharing economy platforms. The rapid growth of such platforms suggests that feedback and reputation systems do a reasonably good job at policing bad behavior, possibly eliminating the need for onerous rules and regulations. At the same time, several studies described in this chapter document biases in feedback and reputation systems that can be improved upon.

There is still much to explore in order to deepen our understanding of how feedback and reputation systems can be improved. It is clear that the design and engineering of feedback and reputation systems will continue to play an important role in the broader area of market design as it applies to sharing economy platforms.

REFERENCES

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics* 84(3): 488–500.

Angrist, Joshua D. and Jorn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Bar-Isaac, Heski and Steven Tadelis. 2008. "Seller Reputation." *Foundations and Trends®in Microeconomics* 4(4):273–351.

Bolton, Gary, Ben Greiner, and Axel Ockenfels. 2013. "Engineering Trust: Reciprocity in the Production of Reputation Information." *Management Science* 59(2):265–285.

Briedis, Holly, Michele Choi, Jess Huang, and Sajal Kohli. 2020. "Moving Past Friend or Foe: How to Win with Digital Marketplaces." McKinsey & Company. www.mckinsey.com/industries/retail/our-insights/moving-past-friend-or-foe-how-to-win-with-digital-marketplaces

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

Cabral, Luis and Ali Hortaçsu. 2010. "The Dynamics of Seller Reputation: Evidence from Ebay*." *The Journal of Industrial Economics* 58(1):54–78.

Cai, Hongbin, Ginger Zhe Jin, Chong Liu, and Li-An Zhou. 2014. "Seller Reputation: From Word-of-Mouth to Centralized Feedback." *International Journal of Industrial Organization* 34(C):51–65.

Dellarocas, Chrysanthos. 2003. "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms." *Management Science* 49(10):1407–24.

https://doi.org/10.1017/9781108865630.006 Published online by Cambridge University Press

Dellarocas, Chrysanthos and Charles A. Wood. 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., E. Grewal, D. Holtz, and M. Pearson. 2015. "Bias and Reciprocity in Online Reviews: Evidence from Experiments on Airbnb." Paper presented at the 16th ACM Conference on Electronic Commerce, June 2015 (EC 2015).

Greif, Avner. 2006. *Institutions and the Path to the Modern Economy*. Cambridge University Press.

Horton, John and Joe Golden. 2015. "Reputation Inflation: Evidence from an Online Labor Market." Working Paper, New York University.

Hui, Xi-Ang, Maryam Saeedi, Neel Sundaresan, and Zeqian Shen. 2014. "From Lemon Markets to Managed Markets: The Evolution of EBay's Reputation System." Working Paper, Ohio State University.

Jaffe, Sonia, Peter Coles, Steven Levitt, and Igor Popov. 2019. "Quality Externalities on Platforms: The Case of Airbnb." Working Paper.

Jin, Ginger Zhe and Andrew Kato. 2006. "Price, Quality, and Reputation: Evidence from an Online Field Experiment." *The RAND Journal of Economics* 37(4):983–1005.

Klein, Tobias, Christian Lambertz, and Konrad Stahl. 2013. "Adverse Selection and Moral Hazard in Anonymous Markets." *ZEW – Centre for European Economic Research Discussion Paper no. 13–050*.

Kreps, D. M. 1990. "Corporate Culture and Economic Theory." Pp. 90–143 in *Perspectives on Positive Political Economy*, edited by J. E. Alt and K. A. Shepsle. Cambridge University Press.

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson. 1982. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory* 27(2):245–252.

Li, Lingfang (Ivy). 2010. "Reputation, Trust, and Rebates: How Online Auction Markets Can Improve Their Feedback Mechanisms." *Journal of Economics \& Management Strategy* 19(2):303–331.

Li, Lingfang (Ivy), Steven Tadelis, and Xiaolan Zhou. 2020. "Buying Reputation as a Signal of Quality: Evidence from an Online Marketplace." *Rand Journal of Economics* 51(4):965–988.

Li, Lingfang (Ivy) and Erte Xiao. 2014. "Money Talks: Rebate Mechanisms in Reputation System Design." *Management Science* 60(8):2054–2072.

Livingston, Jeffrey 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, Dimitriy V, Uwe F. Mayer, and Steven Tadelis. 2015. "Canary in the E-Commerce Coal Mine: Detecting and Predicting Poor Experiences Using Buyer-to-Seller Messages." Pp. 81–93 in *Proceedings of the Sixteenth ACM Conference on Economics and Computation, EC '15*. ACM.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier. 2014. "Promotional Reviews: An Empirical Investigation of Online Review Manipulation." *American Economic Review* 104(8):2421–2455.

McDonald, Cynthia G. and V. Carlos Slawson. 2002. "Reputation in an Internet Auction Market." *Economic Inquiry* 40(4):633–650.

Milgrom, Paul R., Douglass C. North, and Barry R. Weingast. 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, Chris and Steven Tadelis. 2015. "The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment." NBER Working Paper (no. w20830).

https://doi.org/10.1017/9781108865630.006 Published online by Cambridge University Press

Resnick, Paul, Richard Zeckhauser, Ko Kuwabara, and Eric Friedman. 2000. "Reputation Systems." *Communications of the ACM* 43(12):45–48.

Resnick, Paul, Richard Zeckhauser, John Swanson, and Kate Lockwood. 2006. "The Value of Reputation on EBay: A Controlled Experiment." *Experimental Economics* 9(2):79–101.

Roth, Alvin E. 2002. "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics." *Econometrica* 70(4):1341–1378.

Tadelis, Steven. 1999. "What's in a Name? Reputation as a Tradeable Asset." *American Economic Review* 89(3):548–563.

Tadelis, Steven. 2012. *Game Theory: An Introduction*. Princeton University Press.

Xu, H., D. Liu, H. Wang, and A. Stavrou. 2015. "E-Commerce Reputation Manipulation: The Emergence of Reputation-Escalation-as-a-Service." *Proceedings of 24th World Wide Web Conference* (WWW 2015):1296–1306, ACM.

Zervas, Georgios, David Proserpio, and John Byers. 2015. "A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average." Working Paper, Boston University.