When and Why Do Buyers Rate in Online Markets?*

Xiang Hui[†] Tobias J. Klein[‡] Konrad Stahl[§]

February 4, 2022

Abstract

Online ratings play an important role in many markets. We study the often disputed information content of these ratings, by proposing a reduced-form Bayesian model of the typical buyer's rating decision. Our empirical evidence based on eBay raw data is in line with even intricate predictions from it. We thus have good reasons to calibrate the model to moments of the data. Our simulations suggest that the rating record reveals the seller's type after about 100 transactions, or 65-70 ratings.

JEL-classification: D83, L12, L13, L81.
Key words: Online Markets, Rating, Reputation

*We are grateful to comments we have received from Liran Einav, Andrey Fradkin, Ali Hortaçsu, Ginger Jin, André Stenzel, as well as seminar participants at Chicago, Johns Hopkins, Rochester, the European Commission/DG Comp, Bocconi, Washington University, the ISMS Marketing Science conference, and the ZEW ICT conference. We would like to thank in particular Christoph Carnehl for constructive discussions and help in the conceptualization of our model. We are grateful to eBay for providing access to the data. Konrad Stahl gratefully acknowledges support by the Deutsche Forschungsgemeinschaft through CRC TR 224 (project B04).

[†]Washington University in St. Louis
[‡]Tilburg University and CEPR
[§]University of Mannheim, CEPR, CESifo and ZEW

1 Introduction

Online ratings play an important role in many markets.[1] In most online markets seller quality is unobserved *ex ante*. While often the sellers themselves offer the option to rate them, the communication of aggregate rating results are under the sellers' control and thus subject to potential communication bias. Online platforms may provide a less biased possibility to learn from online ratings. Offline transactions are also often rated online. Well known examples include hotels, restaurants, and services such as plumbing, construction, medical, and legal services.

At the same time, many researchers question the information content of ratings even if organized by platforms, on many grounds (Resnick and Zeckhauser, 2002b; Dellarocas and Wood, 2008; Nosko and Tadelis, 2015). To recount some of them: only a fraction of all transactions is rated; the ratings may be selective, so positive experiences are more likely to result in a rating than negative ones, and the seller may exploit on this by providing lesser service; or the buyer population may change over time, as the seller matures with his activities on the platform. By all this, aggregate rating scores are considered subject to selection bias.

In this paper, we study how informative the rating aggregate is in the presence of such selection bias. We establish a basis by developing a simple reduced-form Bayesian model of a typical buyer's rating decision. Its central feature is that the rating decision depends on *the buyer's belief about the seller's quality before, and its change induced by the transaction experience*. While the difference to the rather popular "surprise hypothesis" appears subtle, it leads to conflicting predictions. We show that the predictions from our model hold, rather than those from the surprise hypothesis.[2]

Towards testing predictions from our model, we leverage "natural experiments" on eBay, where negative ratings are left because buyers left a wrong feedback. Such wrong negative feedback affects future buyers' beliefs uncorrelated with seller quality, thereby creating ideal variations to identify the causal effect of pure belief updating on rating behavior. The estimates based on these natural experiments provide strong support for our model. This encourages us to calibrate our model to key moments in the data. Our simulations suggest that buyers can learn sellers' quality quite fast even if ratings are biased.

To be more specific, in the model, the buyer has reduced-form Bayesian beliefs about the quality of the seller that are informed by the public rating record. Her experience in the ensuing transaction is private information and informative about the seller type. The buyer is the more inclined to share this information by leaving a rating, the more she has learned during the transaction about the quality of the seller.[3] Technically, how much the buyer has learned is given by the absolute difference between the buyer's Bayesian posterior and her prior about the

[1] See Dellarocas (2003), Bajari and Hortaçsu (2004), Cabral (2012), Tadelis (2016) for surveys on this topic.

[2] See the next section for details.

[3] When the buyer rates, she can only share her transaction experience, but not how much she has learned. This assumption is based on how rating systems are designed in many e-commerce marketplaces, including eBay.

quality of the seller. The model also features a baseline inclination to share an experience even when no learning took place. With this, we capture the empirical regularity that even sellers with many online ratings keep receiving new ratings.

While very simple, the model is based on rather general assumptions. Furthermore, it has a number of what we believe are realistic properties. First, the transaction experience is not fully revealing of the seller's type: transactions with both a high quality and a low quality seller can go wrong. Second, buyers are more inclined to share information about their private experience when they found this information more valuable themselves. Third, the likelihood that an experience is shared is dependent on whether the experience was good or bad, even when the baseline inclination to rate is the same for positive and negative experiences. The reason for this is that the amount of updating is asymmetric. For instance, if a seller's quality record is strong, a negative experience may be more informative about seller quality than another positive one. This already leads to rating bias because it implies that ratings are given selectively.

We derive two testable predictions from the model. The first is that over time it is less and less likely that a seller receives a rating. The reason for this is that buyers learn more and more from the rating record the longer it is; and therefore, in relative terms, less and less from their own transaction. Consequently, they are less and less likely to share their experiences by leaving a rating.

The second prediction is that when a seller has a strong rating record, a negative rating provides additional motivation for future buyers to leave a rating. The reason for this is that a negative rating challenges future buyers' prior beliefs about the seller's quality. Consequently, buyers learn more from their own experience, both when it is positive and negative, and are more likely to share this. This is especially true for a rare negative experience and therefore, the likelihood that an ensuing rating is negative increases.[4]

We present empirical results that are consistent with both predictions –even the rather intricate second one. For this, we have to overcome a number of important empirical challenges. The first challenge is that as typical in both on- and offline markets, seller quality is unobserved and may change over time. The second challenge is that there may be selective matching, which means that for a given seller early buyers are different from later buyers. As a consequence, the likelihood that a transaction is rated –and if so, how it is rated– could change over time. The third challenge is that transaction quality as perceived by the buyer may change over time, for instance because the seller learns, or because he changes prices. The fourth challenge is that there could be transient shocks to transaction quality. This would also make it more likely that a negative rating is followed by another negative rating, because both share the same shock. The

[4]By the popular "surprise" hypothesis, such as in Anderson and Sullivan (1993), the buyer is more likely to rate if the difference between her prior and the transaction experience increases, rather than her prior and her posterior as in our model. Contradicting the predictions from our model, (i) the probability to share a negative experience should increase the buyer's prior belief. Furthermore, any negative rating affecting the buyer's prior should decrease the likelihood that another negative experience is rated, rather than increase it as in our model. Our empirical results support our predictions rather than those from the surprise hypothesis.

fifth challenge is that sellers may react to both receiving many positive ratings, or a negative rating, which could also have an effect on future ratings.

We leverage access to the administrative data from eBay to address these challenges. These data contain a wealth of information at the level of individual transactions. To address the first challenge, we construct a balanced panel of sellers and control for seller fixed effects. To mitigate concerns related to the second challenge, we use the rich information in the data to construct measures of buyer behavior. These are, for instance, the baseline inclination that a given buyer leaves a rating. We then use these measures as controls. To address the third challenge, we use alternative measures in both auction and posted-price formats to show that consumer surplus does not change much over time—which would be an alternative explanation for observing fewer ratings over time, as argued in Acemoglu et al. (2019). Regarding the fourth challenge, we use an empirical strategy based on negative ratings that are left because the buyer accidentally left the wrong feedback. As mentioned before, these wrong negative feedback affect future buyers' beliefs –yet are not driven by negative shocks in seller quality. By isolating the pure effect of previous rating on the buyer's belief updating from seller behaviour, we can study the effect of belief updating on the likelihood to leave positive or negative feedback in later transaction. To address the last challenge, we exploit transactions that took place before the negative feedback arrives, but the rating of this transaction happened after the arrival. These instances allow us to examine changes in rating behavior as a result of changes in belief before the seller had a chance to respond.

With this, we felt encouraged to quantify the speed of learning about seller quality. Because our analysis so far does not include the buyer's search and purchasing decisions, we abstained from developing a structural model, but isolated the very rating decision by calibrating the present model to moments of the data. Our simulations suggest that buyers learn from ratings about the quality of a seller, even if ratings are given selectively and the average rating is biased. We find that after about 100 transactions, buyers have a very good idea whether a seller is of high or low quality.

We proceed as follows. In Section 2 we discuss the related literature. In Section 3 we present our model and our theoretical results. We introduce our data set in Section 4. Sections 5 and 6 contain our empirical results. The simulation study can be found in Section 7. We summarize and conclude in Section 8.

2 Related Literature

We propose a model of selective rating and use it to quantify the speed of learning from online ratings. The selectivity of rating may be influenced by the selectivity of buying which is not part of our present analysis. Nevertheless we briefly review literature we relate to, followed by a review of the literature on the selectivity of rating, and ensued by a number of special topics we relate to, namely the informativeness of missing ratings, strategic rating, rating dynamics, and social learning.

Selective buying Only a selected group of consumers, relative to the consumer population, ends up buying a product, and therefore aggregate ratings are biased by reflecting only buyers' preference. Hu et al. (2017) and Li and Hitt (2008) provide empirical evidence supporting this view by showing that ratings provided by early buyers are based on the product quality and their taste, which is different from taste of later buyers. Acemoglu et al. (2019) provide conditions for ratings to be informative in spite of this, and derive the speed of asymptotic learning when selection takes place at the purchase stage.[5]

In contrast, in our model, we study the rating decision of a typical (randomly assigned) buyer. We focus on the case where a vertical attribute is rated (say transaction quality rather than the taste for the product). In doing so, we focus on the rating decision and how it is influenced by beliefs. We have in common with Acemoglu et al. (2019) that future buyers are able to learn about seller quality despite selective ratings. But the mechanism is very different.

Rating decision We first report on analyses of the individual rating decision, and then briefly on that of aggregates. We see the rating decision as reference-dependent in the spirit of Hart and Moore (2008), who argue that contracts –in our case previous ratings of a seller– may provide a reference point for parties' feelings of entitlement. Specifically, Anderson and Sullivan (1993) provide evidence that buyer satisfaction (and whether to rate) is a function of perceived seller quality before the transaction and "disconfirmation" after the transaction.

This is also called the "surprise" hypothesis: by our Bayesian terminology, the buyer is more likely to rate if the difference between her prior belief and the transaction experience increases, rather than the difference between her prior and her posterior as in our model. This difference is less subtle than one might think: Opposite to the predictions from our model, the probability to share a *negative* experience should *increase* in the buyer's prior belief, while it decreases in our model. Correspondingly, by the surprise hypothesis any negative rating affecting the buyer's prior should decrease the likelihood that another negative experience is rated, rather than increase it as in our model. Our empirical results support our predictions rather than those from the surprise hypothesis.

Few more authors empirically study the selection process at the review stage. (Dellarocas et al., 2006) develop a statistical measure for the propensity to leave a rating. Moe and Schweidel (2012) study how previous ratings affect whether and what buyers rate.[6] They do not observe whether a consumer has bought the product and therefore estimate a structural model with a selection equation. The same is true for Dai et al. (2018), who construct a weighted average of feedback that allows reviewers to differ in their rating stringency, influence from existing review, and temporal changes in product quality. In contrast to these authors, we observe whether

[5] Stenzel et al. (2020) study sellers' pricing decision anticipating its effect on buyer selection and hence rating.
[6] Moe and Trusov (2011) estimate the value of this in terms of extra sales.

ratings have been given and use more direct, classic econometric methods and focus more on causality.

Within a Bayesian estimation framework, Ho et al. (2017) study how the discrepancy between the expected and experienced assessment of the product affects buyers' rating behavior and their subjective perception of the efficacy of the entire rating system over time. Unlike in our model, their posterior is formed based on some long-term average ratings for all product, rather than, much more naturally, buyers' prior based on the seller's rating index, updated by her private experience from the transaction. However, the buyers' view on the efficacy of the review system does not figure in our analysis.

Chakraborty et al. (2020) propose a model with a monopolist, early adopters, and future buyers. They show that in equilibrium ratings are left when they influence the purchase decision of future buyers. The motivation to rate in our model can be reinterpreted in their spirit. We add an empirical analysis.

In the context of eBay, authors typically think of reviews as either positive or negative.[7] On other platforms, consumers evaluate products or experiences on a finer scale.[8] Following up on this, Brandes et al. (2019) ask whether users are more likely to share extreme experiences either because they derive more utility from this, or because buyers with more extreme experiences are less likely to forget extreme experiences. We propose that there is a benefit from leaving a rating. That benefit is the bigger the more buyers learn from a transaction. The amount of learning is bigger for extreme experiences. Therefore, our model provides an alternative explanation for the J-shaped distribution of reviews.[9]

Luca and Reshef (2020) show that the likelihood to leave a rating is influenced by the price consumers pay for a product. The underlying idea is that higher prices signal higher quality, which can affect the degree of dis-confirmation that consumers experience.

Another central theme in the literature is how in the aggregate, the distribution of ratings relates to the distribution of consumer experiences. The interest is triggered by the well-documented stylized fact that aggregate ratings are usually very favorable. This is the case for many platforms (e.g., Chevalier and Mayzlin (2006) and Schoenmüller et al. (2019)), as well as eBay (e.g., Resnick and Zeckhauser (2002a)). We incorporate that rating asymmetry in our analysis.

Informativeness of missing ratings In another strand of the literature the informativeness of ratings is studied by looking at the exit behavior of buyers. Nosko and Tadelis (2015) show that some buyers, especially inexperienced ones, respond to low-quality transactions from a

[7]Neutral reviews are seen as similar to negative ones.

[8]Hu et al. (2009) document that this leads to a J-shaped distribution for the reviews on Amazon in the sense that almost all products have an asymmetric bimodal distribution with more positive than negative reviews.

[9]To be precise, a generalization of our model provides that explanation, where buyers share experiences on a scale with more than 2 points, where experiences are drawn from a parametric distribution with seller-specific parameters, and where learning is about the parameters of that distribution. Appendix **??** sketches such a generalization.

seller by abandoning the platform instead of leaving a negative review. They then propose a new quality measure that incorporates this "silence"—the number of positive feedback that a seller received divided by the total number of transactions—and show that incorporating it in the search ranking algorithm increases the retention of buyers. However, the effect size they find is small. The control group returns to the platform 61.55% of the time, while the treatment group with the new search algorithm returns 61.85% of the time. This suggests that knowing about a seller that he has more unrated transactions is a negative signal, albeit a relatively noisy one. This finding is in line with one of our findings from the simulation study, presented in Figure 6 in the appendix. The figure shows how much future buyers learn from a positive rating, a negative rating, and no rating. They learn the most from a negative rating, and more from a positive rating than from no rating.

Even though the focus of the two papers is different, the empirical analysis in Nosko and Tadelis (2015) nicely complements ours. Both are based on administrative data from eBay. Nosko and Tadelis (2015) follow a cohort of entering *buyers* on the platform and show that buyers are less likely to abandon the platform when they have a negative experience as they gain tenure on the platform, because they have learned about platform quality over time. We instead track a cohort of entering *sellers* on the platform and show that young sellers are more likely to get rated because buyers learn more from the transaction about seller quality when they have not had so many transactions yet, and have therefore not received as many ratings.

Importantly, we allow for rating bias in our model and our finding that the rating system is informative is fully compatible with Nosko and Tadelis (2015)'s argument that their proposed reputation measure contains more information than the commonly used one.

Strategic rating Many platforms (such as TripAdvisor or Jameda) offer the possibility that agents rate without purchase, which facilitates strategic rating induced by competing firms (see, e.g., Chevalier and Mayzlin, 2006; Mayzlin et al., 2014; Luca and Zervas, 2016). eBay does not allow ratings that are not based on a transaction, so we do not have to account for bias generated this way.

Early designed online rating systems often gave rise to incentives not to share negative experiences, because buyers and sellers would rate one another sequentially and ratings would be immediately observable. Therefore, the threat of retaliation led to less negative ratings. This was pointed out by Resnick and Zeckhauser (2002b) and later studied by Klein et al. (2009), Dellarocas and Wood (2008), Bolton et al. (2013), Klein et al. (2016), Hui et al. (2018), and Fradkin et al. (2020), among others. Nowadays, ratings are usually either anonymous or revealed with a delay, so that this incentive is not present anymore.

Rating dynamics Cabral and Hortacsu (2004) follow sellers over time and study reputation dynamics after a negative feedback is received. They describe these dynamics, but do not attempt to quantify how much of them are driven by changes in seller behavior vis-à-vis changes in rating behavior. While changes in seller behavior are likely at play, we can employ with internal data from eBay various approaches to control for time-varying quality provision and still observe a decreasing probability of leaving feedback, which we attribute to changes over time in buyers' intensity to learn from transactions relative to the public rating record.

Filippas et al. (2018) study whether rating behavior is subject to trends over time. They argue that ratings have been inflated over time because raters feel pressured to leave above-average ratings. This further pushes raters to rate positively because they do not want to harm the rated seller. Our model offers an alternative explanation for this phenomenon: the more ratings a seller has acquired, the less buyers learn from a transaction and the more likely it will be that a rating is positive, because the baseline inclination to share a positive experience is generally higher.[10]

Social learning Finally, our paper is loosely related to the social learning literature, as summarized in, e.g., Mobius and Rosenblat (2014). In that literature, future buyers learn from choices made by previous buyers, which were influenced by their respective beliefs. An important difference is that in our case, buyers share their experience in the transaction when leaving a rating, and not the belief that made them buy in the first place. We do *not* model the buyer's purchase decision. Instead, we model the likelihood that a signal is shared, which then provokes social learning.

3 Model

The goal of this section is to propose a simple model in which buyers learn about seller types from the publicly available information and their own experience with a seller. The more they learn from the their own interaction with a seller, the more likely they are to share their experience by leaving a rating. We show that already a simple model with two types of sellers and two types of signals can produce rich predictions. In the following sections we take these to the data.

3.1 Setup

Consider an online marketplace such as eBay. We focus on one typical buyer's rating decision that is based on a transaction with a typical seller. The transaction experience can be either good or bad. The signal $s \in \{g, b\}$ summarizes her experience with the seller, where $s = g$ if the transaction experience is good, and $s = b$ if it is bad. Sellers can be of high or low quality $q \in \{q_h, q_l\}$. The seller type is unknown to the buyer, but the likelihood that the buyer has a good experience is higher for high quality sellers. Before the transaction she forms a prior belief about seller quality, receives a binary signal about it when performing the transaction, and then

[10] Figure D.4 in the Online Appendix shows that we can generate this in our simulation study.

decides whether or not to truthfully rate the seller by sharing the binary signal. We posit that the likelihood to rate depends on the absolute difference between prior and posterior belief about the quality of the seller. When rating, the buyer reveals her transaction experience.

More specifically, the buyer has initial belief $\lambda \in (0,1)$ that sellers entering the platform are of good quality. She updates this belief using the specific seller's public reputation record, which is summarized by the index y, with $y \in \mathbb{R}$. That index could be any function of information available to date on the seller –in particular of previous ratings. It could either be computed by the platform or by the buyer herself.[11] It could be thought of as a known function that is increasing in the number of positive, and decreasing in the number of negative ratings. Examples are the percentage positive feedback, or an estimate of the likelihood to have a positive experience with the seller, or the fraction of transactions with a positive rating.[12]

To update her initial belief λ in a Bayesian sense, the buyer specifies probability $\sigma_h(y) = \Pr(y|q = q_h)$ that our seller's reputation score is y, given that he is of high, and probability $\sigma_l(y) = \Pr(y|q = q_l)$ that the reputation score is y, given that he is of low quality. Our buyer then forms a Bayesian belief $\mu(\lambda, y)$ that the transaction offered by our seller will be of good quality, and conducts the transaction with our seller if that belief is high enough. Applying Bayes' rule shows that this belief is given by

$$\mu(\lambda, y) = \frac{\lambda \sigma_h(y)}{\lambda \sigma_h(y) + (1 - \lambda)\sigma_l(y)}.$$

Taking the derivative with respect to λ shows that $\mu(\lambda, y)$ is increasing in λ, which means that a consumer patronizing the platform is more inclined to interact with our specific seller if her initial belief is high that sellers entering the platform are of good quality. Another desired property of $\mu(\lambda, y)$ is that a better rating record should increase that belief, while more negative ratings should decrease it. Technically, this is the case under the monotone likelihood ratio property (MLRP) commonly used in models of Bayesian updating,

Assumption 1 (MLRP). $\sigma_h(y)/\sigma_l(y)$ *is strictly increasing in* y*.*

Intuitively, it means that a rating record is the more likely to be from a good seller the better it is. To simplify notation, from now on we keep the dependence on λ and y implicit unless it is needed for the argument and write μ for the prior belief about seller quality before conducting the transaction.

When our buyer conducts a transaction with the seller, she receives the binary signal s about the seller's quality. Let $\rho^h \equiv Pr(s = g|q = q_h)$ and $\rho^l \equiv Pr(s = g|q = q_l)$. The buyer perceives that signal as informative in the sense that it conveys information about the quality of the seller. Hence

[11] If generated by the platform, it could also include the number of unrated transactions. Unless needed, we henceforth talk about y as an aggregate of the seller's rating record.

[12] At any rate, it is a reduced form we take as a starting point to study the rating decision. In Section 7 we study in addition what buyers can learn from a rating record that also contains information on ratings that were not left.

Assumption 2 (Signal)**.** *(i)* $\rho^h > \rho^l$ *and (ii)* $\rho^h > 1/2$.

It follows from Assumptions 1 and 2 that $\mu(y)$ is strictly increasing in y. Thus, for any $\varepsilon \in (0,1)$ there is a realization $\bar{y}(\varepsilon)$ of the index y such that for all y with $y > \bar{y}(\varepsilon), \mu(y) > 1-\varepsilon$, and a realization $\underline{y}(\varepsilon)$ of the index such that for all y with $y < \underline{y}(\varepsilon), \mu(y) < \varepsilon$.

The buyer uses the signal $\rho^i, i \in \{h,l\}$ to form a posterior belief μ^s about seller quality. She makes public her experience with the transaction by leaving a rating if her benefit $b(d)$ from doing so exceeds her cost. The benefit is assumed to strictly increase in the absolute value of the difference $d \equiv |\mu^s - \mu|$ between posterior belief μ^s and prior belief μ. The cost c is drawn from a uniform distribution on the unit interval. The buyer leaves a rating if

$$u(d,c) \equiv b(d) - c \geq 0. \tag{1}$$

The symmetry assumption behind the absolute difference reflects the idea that the intensity of learning drives the rating decision. The higher d, the more the buyer has learned from the transaction, and the more she may expect others to learn from her rating. This rationalizes the positive dependence of $b(d)$ on d. In addition, b can capture a baseline inclination to leave a rating, even if there is no learning.[13] In all, the likelihood that a buyer leaves a rating is given by $b(d)$.[14]

3.2 Updating of beliefs

By Bayes' rule, the buyer forms a *posterior belief* μ^s that the seller is of high quality,

$$\mu^g \equiv Pr(q = q_h | s = g) = \frac{\mu\rho^h}{\mu\rho^h + (1-\mu)\rho^l} \tag{2}$$

and

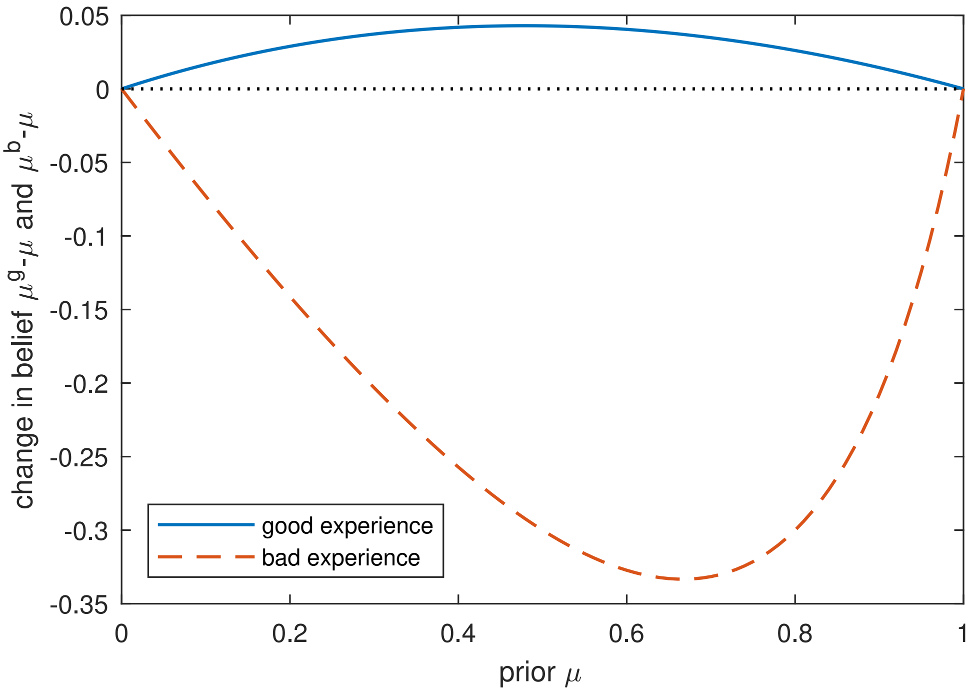
$$\mu^b \equiv Pr(q = q_h | s = b) = \frac{\mu(1-\rho^h)}{\mu(1-\rho^h) + (1-\mu)(1-\rho^l)}. \tag{3}$$

Figure 1 illustrates this. Here, we have chosen parameters that could be realistic for an online marketplace. High quality sellers provide a good experience $\rho^h = 95\%$ of the time, and low quality sellers provide a good experience $\rho^l = 80\%$ of the time.

Starting with $\mu(\lambda, 0)$, the figure shows patterns that are specific to the parameter values we have chosen, and other patterns that hold more generally. First to the more general patterns. It follows directly from (2) and (3) and that there is no updating when the prior is either $\mu = 0$ or $\mu = 1$ and there is some updating if $0 < \mu < 1$. A positive signal always implies that beliefs

[13] See Section 7 for an example.

[14] For general functions $b(d)$ it is a normalization that c is uniformly distributed. Then, we have $\Pr(\text{rating is left}) = \Pr(c \leq b(d)) = b(d)$.



Notes: Figure shows the change in beliefs that is associated with a positive and a negative signal, respectively, with $\rho^h = 0.95$ and $\rho^l = 0.8$.

Figure 1: Change in beliefs

increase, while a negative signal always implies that beliefs decrease. Moreover, (2) and (3) are continuous in μ.

For the specific parameter values that we have chosen for the figure, positive signals are less informative about seller quality than negative signals. Therefore, for any prior μ the buyer learns more from a negative experience, and most when priors are between 0.5 and 0.8. We come back to this example in Section 7, when we calibrate our model to moments of the data and carry out our simulations.

3.3 Dependence of the change in beliefs on the rating record

In our model, the rating decision depends on how much buyers learn from the transaction. We now study the dependence of the amount of learning on the seller's rating record.

The rating record affects the amount of learning through its effect on prior beliefs μ. Recall that it follows from Assumption 1 that prior beliefs μ are strictly increasing in the public rating record y. Therefore, we can focus on the dependence of learning on the level of prior beliefs μ.

Given the parameters used to construct Figure 1, the maximum of the solid line is attained at a value of μ that is below the value at which the minimum of the dashed line is attained. In this example buyers learn the most from a positive experience for a value of μ that is lower than the value for which they learn the most from a negative signal. In the following proposition, we establish for the general case that these extrema are unique and that the one for a negative signal is always obtained at a higher value of μ.

Proposition 1. *Let Assumptions 1 and 2 hold, and consider an ex ante increase in y.*

(i) If the transaction experience was positive, then there exists a unique $\hat{y}$ such that the difference between posterior and prior beliefs increases in y if $y < \hat{y}$, and decreases in y if $y > \hat{y}$. Furthermore, $\mu(\hat{y}) \equiv \hat{\mu} < 1/2$.

(ii) If the transaction experience was negative, then there exists a unique $\check{y}$ such that the difference between prior and posterior beliefs increases in y if $y < \check{y}$, and decreases in y if $y > \check{y}$. Furthermore, $\mu(\check{y}) \equiv \check{\mu} > 1/2$.

(iii) $\hat{y} < \check{y}$.

Proof. To prove (i) we need to show that given the signal $s = g$ the buyer receives from the transaction, the difference between the buyer's posterior and her prior beliefs

$$\mu^g - \mu = \frac{\mu\rho^h}{\mu\rho^h + (1-\mu)\rho^l} - \mu$$

must first increase, and then decrease in y. Hence we are interested in conditions under which the derivative $\frac{d}{dy}(\mu^g - \mu) \gtreqless 0$. Taking that derivative, rearranging and simplifying, we obtain that

$$\frac{d}{dy}(\mu^g - \mu) \gtreqless 0$$

is equivalent to

$$\rho^h\rho^l \gtreqless \mu^2(\rho^h)^2 + 2\mu(1-\mu)\rho^h\rho^l + (1-\mu)^2(\rho^l)^2.$$

Rearranging and simplifying further, we obtain that this is equivalent to

$$\frac{\rho^l}{\rho^h} \gtreqless \frac{\mu^2}{(1-\mu)^2}. \tag{4}$$

We know that $\mu \in [0,1]$. By Assumptions 1 and 2, and the discussion ensuing the latter, there is a $\hat{y}$ such that $\mu'(\hat{y}) = 0$, and $\mu'(\hat{y}) > 0$ for all $y < \hat{y}$ and $\mu'(\hat{y}) < 0$ for all $y > \hat{y}$. This implies that $\hat{y}$ is unique.

Furthermore, by Assumption 2,

$$\frac{\rho^l}{\rho^h} < 1.$$

Hence equality in (4) holds for prior beliefs

$$\hat{\mu} < \frac{1}{2}.$$

To prove (ii), we follow an argument very similar to the proof of (i). In response to the signal $s = b$, we study the derivative of

$$\mu - \mu^b = \mu - \frac{\mu(1-\rho^h)}{\mu(1-\rho^h) + (1-\mu)(1-\rho^l)}.$$

Indeed, $\frac{d}{dy}(\mu - \mu^b) \gtreqless 0$ if and only if

$$\frac{1-\rho^l}{1-\rho^h} \gtreqless \frac{\mu^2}{(1-\mu)^2}.$$

By Assumption 2, the left hand side is larger than 1, so that equality holds for a value $\check{\mu} > 1/2$. This is associated with the unique value $\check{y}$. Again, by a similar argument as in the proof of (i), the value of the difference between the buyer's prior and her posterior is increasing below that value, and above it, it is decreasing.

To prove (iii), recall that we have shown that $\hat{\mu} < 1/2$ and that $\check{\mu} > 1/2$. By Assumption 1, μ is monotonically increasing in y for given λ, and therefore we have $\hat{y} < \check{y}$. □

For an interpretation, recall our central posit that a buyer tends to rate a transaction, when she has learned a lot from it relative to the quality she has expected beforehand from interpreting the seller's performance score. Formally, the probability to rate increases with the absolute difference between prior and posterior beliefs. By the Proposition, if that performance score was not very favorable, another favorable rating shifts our buyer's posterior belief sufficiently much to eventually contribute another rating. But if the performance score was already sufficiently favorable, another favorable rating does not shift the belief that much anymore to induce a rating.

This holds for all transactions conducted by our buyer, no matter her experience. However, the difference between positive and negative experiences is one of detail contained in part (iii) of the proposition: The cutoff in terms of the rating record y between when the rating probabilities increase and decrease is lower for positive experiences than for negative ones –with the difference between the two increasing in the difference between ρ^h and ρ^l. Rather intuitively, the probability that our buyer rates negatively continues to increase with further positive ratings, when that of rating positively already decreases.[15] These results are useful to study rating dynamics for a high quality seller, which we turn to now.

3.4 Rating dynamics for a high quality seller

The vast majority of the transactions on eBay is done by sellers who have a favorable rating record y. These are likely high quality sellers. The reason is two-fold: many low quality sellers

[15] We abstain from studying theoretically the consequences of an increasing number of negatives on the buyer's rating decision. They are empirically irrelevant, as an increasing number of negatives leads the seller to exit the market.

are likely to be selected out by both, eBay, and by early buyers that rate poorly. Here, we follow a high quality seller from the first transaction and study how the inclination of buyers to leave a rating changes over time.

A high quality seller is characterized by a large number of positive ratings. Proposition1 tells us that when that number is sufficiently large, the buyer learns less and less from any kind of messages. This is made precise in the following corollary.

Corollary 1 (Empirical prediction 1)**.** *Let $\mu > \check{\mu}$ and consider a sequence of increases in y. Then the likelihood that a buyer derives a benefit from rating the transaction decreases and shrinks to $b(0)$ as $\mu \to 1$.*

The proof follows directly from Proposition 1. The dynamics characterized in the Corollary are non-trivial. We study them in more detail in our simulations in Section 7. Before then we use it as our first empirical prediction: We will construct a sample of sellers who succeed in the market in the sense that they conduct many transactions in their first year on eBay. For those sellers, we predict that on average across sellers the likelihood that they receive a rating is lower for later transactions.

3.5 Interim ratings and their effect

Our model allows buyers to have bad experiences with a high quality seller. We now study the effect a negative feedback has on the likelihood that ensuing buyers share positive, and negative experiences. By assuming that the seller cannot react strategically to a negative rating, we can study in isolation the interaction between the buyers' rating decisions.

In this context it is important to recognize that rating decisions are almost always taken with a delay after the purchase. In the interim period between the time at which the buyer decided to buy from the seller and the time at which she decides whether to leave a rating, new ratings may arrive. Since we look at a high quality seller, most of these ratings will be positive. However, the case in which they are negative is particularly interesting, because this will go against the general dynamics of increases in y over time.

To study this in more detail, index beliefs and rating records by time t and consider the following observation and decision taken by the buyer:

1. Buyer performs the transaction based on the prior belief $\mu_t \equiv \mu(y_t)$
2. Buyer observes a negative interim rating shock $y_{t+1} < y_t$, resulting in a posterior belief $\mu'_{t+1} \equiv \mu(y_{t+1})$
3. Buyer rates if $b(d'_{t+1}) \equiv b(|\mu'_{t+1} - \mu_t|) \geq c$

By Proposition 1 the negative rating shock induces an increased probability to rate when $y > \check{y}$, and in particular, y approaches $\bar{y}$ from below. In order to determine whether in this case a negative experience is rated with higher or lower probability than a positive one, let $f_1(y) \equiv \mu^g(y) - \mu(y), f_2(y) \equiv \mu(y) - \mu^b(y)$, and $z(y) \equiv f_1(y) - f_2(y)$. $z(y)$ is the difference in the probability of rating after a positive and a negative experience, respectively.

Proposition 2. *Let Assumptions 1 and 2 hold. Let $y > \check{y}$. Then*

$$\frac{d}{dy} z(\mu(y)) > 0$$

as y gets large, and thus $\mu(y)$ approaches 1.

Proof. From Proposition 1 it follows that $\hat{y} < \check{y}$. From Assumption 1 it follows that

$$\mathrm{sgn}\left\{\frac{d}{dy} z(\mu(y))\right\} = \mathrm{sgn}\left\{\frac{d}{d\mu} z(\mu(y))\right\}.$$

Then,

$$\mathrm{sgn}\left\{\frac{d}{d\mu} z(\mu(y))\right\} = \mathrm{sgn}\left\{\frac{\rho^h \rho^l}{[\mu\rho^h + (1-\mu)\rho^l]^2} + \frac{(1-\rho^h)(1-\rho^l)}{[\mu(1-\rho^h) + (1-\mu)(1-\rho^l]^2} - 2\right\}. \tag{5}$$

From the discussion ensuing Assumption 2, there is a $\bar{y}(\varepsilon) > \check{y}$ such that $\mu(y) > 1 - \varepsilon$ for all $y > \bar{y}(\varepsilon)$, implying that μ approaches 1 as y becomes sufficiently large. For simplicity, evaluating (5) at $\mu = 1$, we obtain $\frac{\rho^l}{\rho^h} + \frac{1-\rho^l}{1-\rho^h} \gtreqless 2$. Expanding and simplifying, we get $\rho^h > \frac{1}{2}$. Hence $\frac{d}{dy} z(\mu(y)) > 0$ as y gets large, and thus $\mu(y)$ approaches 1. □

From Proposition 2 it follows that the likelihood that the buyer rates negatively increases relative to the likelihood of rating positively, when interim signals lead y to decrease from y_t to $y_{t+1} < y(t)$, as long as y is sufficiently large.

This leads to our second empirical prediction:

Corollary 2 (Empirical prediction 2)**.** *Let y_t get large. Suppose that interim ratings lead y_t to decrease from y_t to y_{t+1}. Let the buyer incorporate the interim decrease in the rating score in her decision. Then*

(i) the likelihood increases that she rates the transaction, relative to not incorporating that shock in her rating decision

(ii) the likelihood that a given rating is negative increases relative to the likelihood that it is positive.

Proof. By Proposition 1, the absolute difference between prior and posterior beliefs increases with a decrease in y when y is sufficiently large, and $\mu(y)$ approaches 1. By Proposition 2, this increase is larger for $\mu - \mu^b$ than for $\mu^g - \mu$. □

To summarize, our second empirical prediction is that an interim negative rating will make it more likely that another buyer shares her experience, and that this effect is stronger for negative experiences than for positive ones. Hence an interim negative rating will increase the likelihood that a future rating is negative, and this more than the likelihood that the future rating is positive.

4 Empirical setting, data, and descriptives

4.1 Empirical setting

The goal of our empirical analysis is to establish the two stylized facts that can be explained by the model in Section 3. We use rich eBay data at the level of transactions. Nevertheless, testing these empirical predictions and quantifying the effects is not straightforward. As already mentioned in the introduction, there are five empirical challenges: unobserved quality, selective matching, time-varying transaction quality, quality shocks, and seller reactions. All but the second challenge are related to the fact common to many analyses not only of online, but also of offline market transactions that transaction quality for a given seller is unobserved. It could change over time, for instance because a seller learns or because his incentives to provide high transaction quality change over time.

The second challenge arises because buyers choose among the sellers the one they buy from. This is likely to be influenced by the seller's competitive position on the platform, and in particular by his rating record. For instance, it could be that buyers who are new on eBay are reluctant to buy from new sellers, while buyers who experienced many positive transactions on eBay feel more at ease doing so.

For our purposes, the ideal situation would be one in which buyers are randomly allocated to sellers, and in which a seller's transaction quality varies neither across sellers and nor over time. Then, the variation in the rating record that buyers see before leaving a rating would be exogenous, which would allow us to regress an indicator for leaving a rating on the rating record.

Our reaction to the challenges above is to control for transaction quality and the endogenous matching between buyers and sellers as much as possible and to exploit quasi-random variation that is related to the timing of transactions and ratings. For this, we create a panel of starting sellers whom we follow over time. This already allows us to control for unobserved differences in seller quality that are time-invariant. We next provide details and summary statistics.

4.2 Sample

We use data from eBay in the U.S. Our starting point is the set of sellers who had their first listing ever in March 2011 (sample 0).[16] From this we construct two subsamples. The first

[16]We chose 2011 because it is a year without changes to the reputation mechanism and March because it is the first month after the winter holiday season.

subsample consists of the set of sellers who have at least 86 transactions in the first year (sample 86). The second subsample consists of the set of sellers who have at least 338 transactions in the first year (sample 338).[17] These are, respectively, the top 5% and the top 1% of the sellers in terms of the number of transactions in the first year.

For these sellers, we construct an unbalanced panel with all transactions based on sample 0, and balanced panels with the first 86 transactions for sample 86, and the first 338 transactions for sample 338. The transactions we use are for the so-called core products, which means that for instance real estate and cars are not in our data.

4.3 Variable definitions and summary statistics

With the considerations in Section 4.1 in mind we construct a number of variables. We report summary statistics for all three samples in Table 1.

There are three panels. Panel A contains seller characteristics. We first create one observation per seller. Then, we report the average across sellers.[18] For all the 141,138 sellers in sample 0 who had their first transaction in March 2011, sales volume in the first year is $1,218. In that year they have 24 transactions on average and sell products in 5 so-called leaf categories. Each listing on eBay has a category attached to it, which is determined using a hierarchical system. A leaf category is the finest level at which products are categorized.[19] The eBay percentage positive is one of the two rating aggregates that is displayed next to a seller name on the eBay platform.[20] It is calculated as the number of positive feedback the seller has received, relative to the number of feedback that were either positive or negative. This means that neutral feedback are discarded. Here we report the percentage positive at the end of the first year.

As samples 86 and 338 contain the top 5% and top 1% of the sellers in terms of the number of transactions within the first year, it is not surprising that these sellers have a higher sales volume and more transactions in more leafs in that year, and that their percentage positive feedback is higher by 6 and 7 percentage points, respectively. The reason for this is that some of the sample 0 sellers who started in March 2011 will stop being active on eBay. Those sellers will be a negative selection in the sense that will more likely receive negative feedback (see for instance Cabral and Hortaçsu, 2006, 2010).

The buyers who bought from these sellers are characterized in Panel B. Our starting point here are all transactions in the first year for sample 0, the first 86 transactions for sellers in sample 86, and the first 338 transactions for sellers in sample 338. From these transactions, we obtained the set of buyers, and for those we calculated three measures. There are 1,792,076

[17]To be precise, here and in the following this means until the end of February 2012.

[18]We use all transactions for this, so not only the first 86 for panel 86 or the first 338 for panel 338. Later we will construct balanced panels with only data on the first 86 and 338 transactions, respectively.

[19]Examples of leaf categories are Boys' Outerwear (newborn-5T), LED Light Key Chains, and Circuit Breaker & Fuse Boxes.

[20]The other number is the feedback score, which is the number of positive ratings minus the number of negative ratings a user has received.

Table 1: Summary statistics

	(1) sample 0 (unbalaced)	(2) sample 86 (balanced)	(3) sample 338 (balanced)
Panel A: Seller characteristics (one observation is one seller)			
sales volume in the first year (USD)	1,218	9,983	27,210
number of transactions in the first year	24	324	983
number of unique leafs in the first year	5	37	61
eBay percentage positive (pos/(pos+neg)) in the first year	0.917	0.978	0.987
observations	141,138	7,085	1,412
Panel B: Buyer characteristics (one observation is one buyer)			
number previous transactions	53	85	84
buyer experience (registered before 01 March 2009)	0.713	0.818	0.824
buyer inclination to leave feedback	0.640	0.678	0.667
buyer criticalness	0.021	0.020	0.020
observations	1,792,076	397,009	303,783
Panel C: Transaction characteristics (one observation is one transaction)			
buyer has bought repeatedly from same seller before	0.150	0.142	0.179
share of transactions with any feedback	0.622	0.670	0.655
share of transactions with neutral or negative feedback	0.016	0.011	0.007
share of transactions with low DSR	0.022	0.019	0.013
share of transactions with a claim	0.020	0.012	0.006
days between transaction and feedback	13.2	12.5	12.9
observations	3,413,354	609,310	477,256

Notes: Averages for three different samples. Sample 0 is the sample of all sellers who had their first listing ever in March 2011. Sample 86 contains the top 5% of those sellers in terms of transactions, and sample 338 contains the top 1%. In Panel A, one observation is one seller and we use all transactions for those sellers. In Panel B, one observation is a buyer for a seller in the respective sample. In Panel C, we use all transactions for sample 0, the first 86 transactions for sample 86, and the first 338 transactions for sample 338. See text for further details and variable definitions.

distinct buyers who bought from the 141,138 sellers who had their first transaction in March 2011.

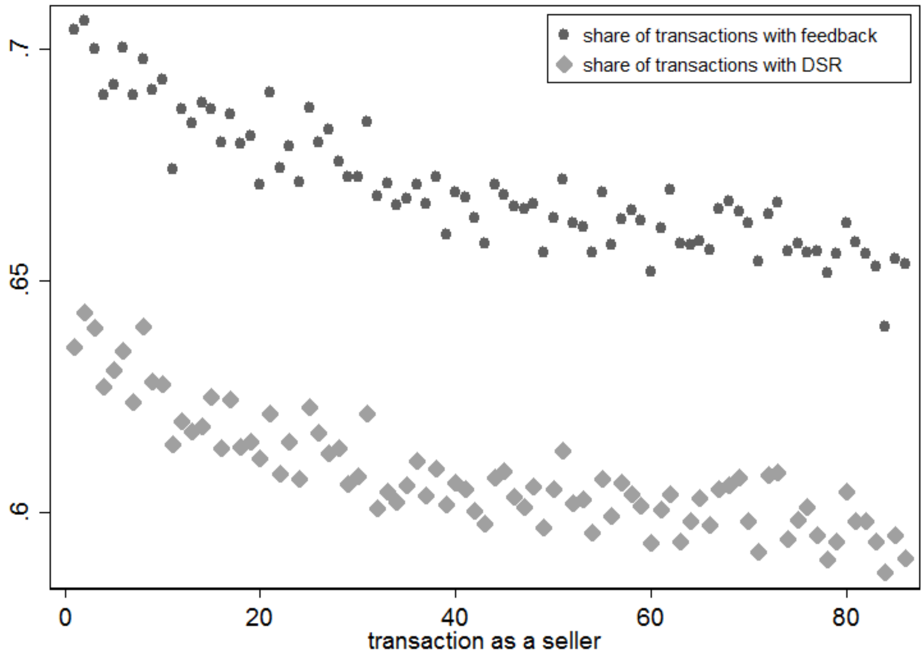
We calculated two measures of buyer experience. The first measure is the number of transactions buyers conducted in the year before March 2011 (the month in which our sellers had their first listing), or by March 2011 if they started after March 2010. For these transactions, we can also calculate how often they left feedback and call this the inclination to leave feedback, and we calculate the share of negative feedback, which we refer to as buyer criticalness. The second measure of buyer experience we mainly use is whether they have registered before March 2009.[21] This is the case for 71.3% of the buyers in sample 0.

Recall that sellers in sample 0 are rated worse on average, as compared to those in sample 86 and sample 338. Interestingly, Panel B shows that both measures of buyer experience for the sellers rated in sample 0 are lower than for those in sample 86 and sample 338. At the same time, the propensity to leave feedback and buyer criticalness is remarkably similar across samples. On average feedback is left for about two-thirds of the transactions, and buyer criticalness is about 2%. This corresponds well to the eBay percentage positive between 98% and 99% that we report in Panel A for samples 86 and 338.

Panel C reports summary statistics for transactions in the first year. We use all transactions for sample 0, the first 86 transactions for sample 86, and the first 338 transactions for sample 338. The first statistic shows that buyers and sellers do not interact repeatedly in the vast majority of transactions. Less than 20% of the transactions are with a seller from which the buyer has bought before. The remaining statistics relate to feedback. For the transactions in our sample, feedback is left for about two out of three transactions. This corresponds well to the numbers reported in Panel B. The share of feedback that are neutral or negative is 2.6% in sample 0. Putting this side-by-side the eBay percentage of 92% positive feedback in Panel A, we can see that sellers with a higher number of transactions have a higher eBay percentage positive. This is also in line with the lower shares of neutral or negative feedback in sample 86 and 338, which are 1.6% and 1.0%, respectively.

The next measure is the share of the transactions with low detailed seller ratings (DSR). Buyers on eBay are asked to provide a DSR after providing a classical feedback. DSR's are ratings in four dimensions, item description, communication, shipping time, and shipping and handling charges. Ratings are left on five-point scales. We define a DSR to be low if at least one of the 4 DSR dimensions has 1 or 2. DSR's are anonymous. Lastly, the gap between the date on which a transaction took place and the date on which the transaction is rated (if at all) is around 13 days on average, and is similar across the samples. We will leverage this gap to control for seller behavior in our later empirical analyses.

[21] In general results are very similar when we use the first measure. See Tables B.2 and B.3.



Notes: Share of transactions with feedback and DSR, respectively, against number of the transaction. Based on sample 86.

Figure 2: Probability to receive feedback

5 Inclination to leave a rating

5.1 Baseline results

Our first empirical prediction in Corollary 1 is that the likelihood that a given transaction is rated decreases when the rating index increases, so that buyers learn less and less from a transaction and therefore are less and less inclined to leave a rating. We now use our balanced panel of transactions to take this prediction to the data.

The advantage of using a balanced sample is that we can plot the likelihood to receive a feedback against the transaction number, and interpret the results as if we followed sellers over time.[22] In Figure 2, we plot the share of transactions with feedback and DSR against the number of transactions performed by the sellers. Both the share of transactions with feedback and the share of transactions with DSRs are lower for later transactions. The underlying idea here is that for our sample of sellers, higher transaction numbers are associated with a higher value of the rating index y—and thus, in the model, with higher priors $\mu(y)$ for the typical buyer. We discuss this in more detail below and provide empirical support for it.

In this figure we do not control for differences across sellers, calendar time effects, and differences across products. Moreover, we do not control for buyer characteristics. A related concern could be that the pattern in Figure 2 is driven by particular types of buyers who buy

[22] By using a balanced sample we circumvent the econometric problem of dynamic selection. Otherwise, sellers who have more transactions are a selected sample, which would lead to confounding.

from young sellers, and buyers that are more likely to leave feedback irrespective of the number of feedback a seller has received already. We address these concerns in the regressions reported in Table 2.

The dependent variable is 100 times an indicator for receiving feedback. The key independent variable is the transaction number divided by 10. Specification (1) corresponds directly to the figure. We find that the likelihood to receive a feedback for the 10th transaction is lower by 0.491 percentage points as compared to the first transaction. We successively add controls in the ensuing columns. In column (2), we control for seller fixed effects and calendar month fixed effects. In column (3), we control for product type using information on the leaf category. This addresses the concern that the likelihood to leave a rating depends on the type of product. Results suggest that the likelihood to leave a feedback is lower by 0.389 for the tenth transaction, and lower by 3.89 percentage points for the 100th transaction.

We control for buyer experience in columns (4) and (5). The estimated coefficient on the interaction term between buyer experience and the transaction index in column (5) suggests that the dependence of the likelihood to receive a feedback does moderately depend on buyer experience: the effect of 10 additional transactions is different by -0.178, from a baseline of -0.211. In column (6) we control for the buyer inclination to leave feedback and find that it does have an effect. Controlling for all of the above factors, we find that later transactions are less likely to be rated. We present results for DSR's in Table B.1 in the appendix. They are similar.

Our model predicts that the likelihood that a transaction is rated decreases in the transaction index, because in later transactions consumers learn less about seller quality from the transaction. Above we show that this prediction is in line with our empirical findings.

Table 2: Probability to receive feedback

	(1) feedback	(2) feedback	(3) feedback	(4) feedback	(5) feedback	(6) feedback
transaction number/10	-0.491***	-0.362***	-0.389***	-0.325***	-0.211***	-0.225***
	(0.0316)	(0.0529)	(0.0523)	(0.0556)	(0.0743)	(0.0759)
buyer experience				2.823***	3.588***	3.894***
				(0.190)	(0.331)	(0.322)
trans. num/10 $\times$ buyer exp.					-0.178**	-0.184**
					(0.0762)	(0.0753)
buyer inclination to leave feedback						26.90***
						(0.275)
trans. num/10 $\times$ buyer inc. to leave fdbk						0.141**
						(0.0570)
seller FE	No	Yes	Yes	Yes	Yes	Yes
month FE	No	Yes	Yes	Yes	Yes	Yes
leaf category	No	No	Yes	Yes	Yes	Yes
adj R-squared	0.000671	0.0621	0.0712	0.0656	0.0656	0.131
observations	609310	609310	607135	515978	515978	515978
number of clusters	7085	7085	7085	7083	7083	7083

Notes: Results of regressions of 100 times an indicator for receiving feedback on the transaction number divided by 10, as well as other controls and interaction terms. One observation is a transaction. Standard errors are clustered at the seller level and account for heteroskedasticity. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

An alternative explanation for these findings could be that earlier consumers select themselves into transactions because they expect to receive a higher surplus, and are therefore more likely to leave a rating. This mechanism plays an important role in the theoretical contribution by Acemoglu et al. (2019). In order to assess this, we investigate the dependence of two measures of consumer surplus on the transaction index. For the first measure we leverage that for auctions we know the winning bid in an auction and the price the winner actually paid, which is the second-highest bid plus a small, fixed increment. The theory on optimal bidding in second price auctions suggests that the difference between the winning bid and the second-highest bid is a good proxy for consumer surplus. Besides auctions, which accounts for around 15% of sales on eBay, we also look at transactions in posted price format. Specifically, the second measure is the price consumers paid for new items, when we control for product ID's. This measure should be inversely related to consumer surplus. Appendix B contains details. For neither of the two measures we find that consumer surplus has decreased in the transaction index.[23] Therefore, controlling for either of the two surplus measures does not change the estimates in Table 2 by much.

5.2 Effects of feedback

The central idea behind our model is that the likelihood to leave a rating is higher when the buyer has herself learned more from the transaction. Our samples 86 and 338 consist of sellers who have established themselves on eBay. This means that their rating record has been improving over time. One of the implications of the model is that with an improving rating record there is less and less room for a deviating posterior, and thus information to share. Therefore it is less and less likely that a rating is left.

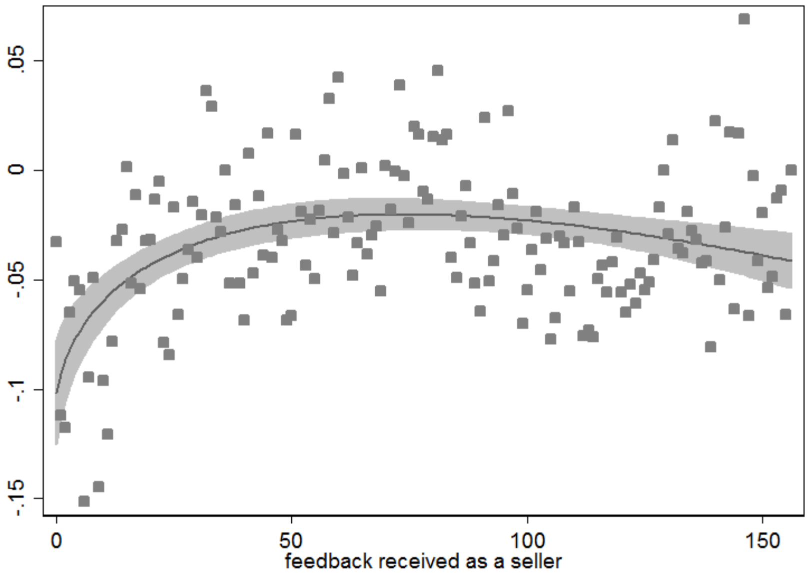
The receivers of the information communicated through ratings are not only other potential buyers, but it is also the seller. Indeed, the seller could use the improvement in his reputation to increase the price of the products sold. And if it is indeed the case that additional positive ratings contain less and less information, then the price the seller sells his products for and the likelihood to sell them should depend positively, but less and less so, on the rating.

We now study the effect of the length of the feedback record on the price and the likelihood of selling. At the same time, we control for the percentage negative feedback. This means that we measure the effect of the amount of information. Figure 3 shows the result. Both the price and the likelihood to sell depend positively on the number of feedback received. The relationship is concave. Figure 3a shows that prices are about 7% lower for the first transactions. Figure 3b shows that the likelihood to sell increases steeply as a result of the first feedback that are received, and less and less so for later feedback. The figures suggest indirectly that the first 50 feedback are particularly informative for the seller.[24]

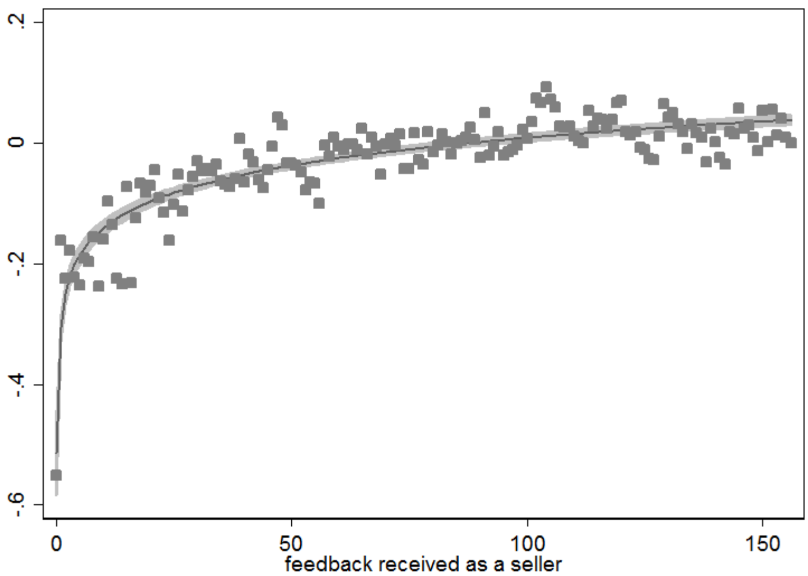
[23]If anything, the second measure suggests that it has slightly increased, as prices were slightly lower for later transactions, but this effect is not economically significant.

[24]Figure 2 shows that for the first 86 transactions the likelihood to leave a rating is about 62 percent on average.

(a) Effect of positive feedback on price



(b) Effect of positive feedback on probability of selling



Notes: Sample 338 restricted to transactions with a product ID. First 156 feedback from sellers who had at least 156 feedback in the restricted sample. 156 feedback chosen because they involve the 90th percentile of feedback index in the restricted sample. Figure 3a shows the result of regressing the logarithm of price plus shipping fee on dummy variables of feedback indices, percentage negative to date, controlling for seller fixed effects and product ID fixed effects. In the regression, we further restrict the sample to new items in the posted price format. The dummy for feedback index = 156 is dropped as the benchmark. Figure 3b is constructed as follows. We first get all listings with product ID from sellers in Sample 338 in their first year. Then we regress the dummy variable for whether a listing sells (1 item or more) on dummy variables of feedback indices, the logarithm of listed price, percentage negative to date, controlling for seller fixed effects and product ID fixed effects. In the regression, we further restrict the sample to items in the posted price format.

Figure 3: Effects of a positive feedback

Table 3: Transaction quality

	(1) claim	(2) neutral or negative	(3) low DSR
transaction number/10	0.107***	0.0718***	0.0501***
	(0.0146)	(0.0119)	(0.0154)
seller FE	Yes	Yes	Yes
month FE	Yes	Yes	Yes
leaf category	Yes	Yes	Yes
adj R-squared	0.100	0.0296	0.0276
observations	607135	607135	607135
number of clusters	7085	7085	7085

Notes: Table shows results of regressions of 100 times an indicator for a buyer claim, a neutral or negative feedback, and a low DSR on the transaction number divided by 10. One observation is a transaction. Standard errors are clustered at the seller level and account for heteroskedasticity. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

5.3 Effects of seller learning

In our analysis we so far control for time effects, the matching between sellers and buyers, and unobserved quality differences across sellers, as long as they are constant over time. We show that the likelihood to leave a rating decreases in the number of transactions. Our explanation is that there is less and less information to communicate and therefore feedback are less likely to be given.

One may wonder whether the same empirical pattern could be driven by seller learning. Suppose first that a sellers learns over time to improve on transaction quality. Consequently, buyers would be positively surprised, implying that they would be more inclined to leave a rating. The pattern documented in Figure 2 and Table 2 would be a lower bound of the effect (as it captures the true effect plus a positive time trend). An indicator for this could be fewer buyer claims recorded by eBay, a lower likelihood that a rating is neutral or negative when given, and a higher DSR when given on average.

By contrast, suppose second that sellers learn to exert moral hazard without being punished for it by negative ratings. This would be a situation in which transaction quality decreases over time for a given seller. We should see an increase in the number of buyer claims and in the likelihood to receive a neutral or a negative rating, and the DSR should be lower.

To investigate this, we regress an indicator for a buyer claim, an indicator for a neutral or negative rating, and an indicator for a low DSR for a given transaction on the transaction variable. We use the same specification as in column (3) of Table 2. As before, the dependent variable is respectively an indicator for the event, times 100. This means that the units of

This means that the 86 transactions in Figure 2 are comparable to the 50 feedback mentioned here.

coefficient estimates are percentages. Table 3 shows the result. The likelihood that a claim is filed is higher for the tenth transaction by 0.107 percentage points. The likelihood that a neutral or negative feedback is left for the tenth transaction is 0.0718 percentage points higher. The likelihood that a low DSR is left for the 10th transaction is higher by 0.0501 percentage points, compared to the overall average of 1.9 percent reported in Table 1.

These effects point in the direction that in our balanced sample, sellers transaction quality slightly decreased over time. However, the order of magnitude of the effects is between half a percentage point and one percentage point per 100 transactions. Recall from Table 2 that the likelihood to leave a feedback decreases by about 3.8 percentage points. Taken together, our interpretation is that part, but not all of this can probably be attributed to a lower transaction quality due to lacking seller effort.

6 Effect of a negative rating on subsequent ratings

Our second empirical prediction in Corollary 2 is that a reduction in the rating index y induced by a negative rating will increase the likelihood that another buyer rates. This effect is stronger (in relative terms) for negative experiences by that buyer, and therefore the likelihood that a feedback is negative increases even if seller quality stays the same. This effect is the stronger, the earlier the negative rating is received by a seller.

We first briefly review the argument behind these theoretical predictions. Suppose a buyer had a negative experience. If her prior was sufficiently high that the transaction was of high quality, then her posterior will not differ much from it, and therefore her inclination to leave a negative rating will be low. Now suppose that after purchasing but before rating, she sees that another buyer has left a negative rating. The buyer will now start from her initial prior based on which she has bought the item, form an updated prior using the rating that has been shared by the other buyer, and then form her posterior using her own experience in the transaction. Her new update will *increase* the absolute difference to her prior—and this even more when her experience was negative rather than positive, because the updated prior attaches a lower probability to the seller being a good type. This increases the probability that she will rate. As the difference between the buyer's prior and her posterior naturally shrinks in the intensity of the experience communicated by earlier buyers, this sequence of events is likelier to be provoked in the seller's earlier than in his later career.

It is not easy to test this prediction, because we cannot simply relate the likelihood to receive a negative rating to the number of negative ratings a seller has already received, even if we control for differences in transaction quality across sellers that do not vary over time. The reason is that transaction quality may not stay constant over time and could have been low for both transactions that led to negative feedback. That is, both feedback could be confounded by a negative transitory quality shock.

Our approach is instead to use negative feedback documented to be given by mistake. We can identify these feedback in the raw data because buyers went through a procedure to have these feedback removed at a later point in time.[25] Unlike us, a later buyer does not know that such a negative feedback was given by mistake, because it has not been removed as yet. Therefore, these feedback should generate the effect we want to quantify, namely changing buyers' prior belief about seller quality. At the same time, the aforementioned problem that two feedback are confounded by a negative shock is not there, because the feedback was not meant to be negative.[26]

With this in mind, we now estimate the the effect of a negative feedback that was given by mistake on the likelihood that a rating is given for later transactions, and on the likelihood that a positive or negative experience is shared, respectively.

6.1 Main specification

Our data is at the level of transactions t for each seller i. We code $rating_{it} = 1$ if a rating is left for seller i on transaction t, and $rating_{it} = 0$ if no rating is left. We also define similar indicators for the event that a positive and a negative rating is left, respectively.

We distinguish between sellers i who at some point receive a negative rating by mistake and those who don't. We call those who receive at some point a negative rating by mistake the treated sellers and indicate this using the variable $treated_i = 1$. Those sellers who never receive a negative by mistake are the control group and for them we code $treated_i = 0$. Finally, for the treated sellers we code $post_{it} = 1$ for all transactions that take place after the wrong negative feedback was received.[27]

We then estimate the linear probability model

$$rating_{it} = \alpha_t + \beta \cdot treated_i + \gamma \cdot post_{it} + \varepsilon_{it}.$$

Our parameter of interest is γ, which is the effect of a negative feedback that was given by mistake on the likelihood that a rating is given for transaction t by seller i. We control transaction fixed effects α_t. They capture the dependence of the inclination to leave a rating on the length of the feedback record. β captures differences between sellers who receive at some point a

[25] A seller can initiate a request for feedback change. When doing so, he chooses one of three reasons for the request: 1. I resolve a problem the buyer had with this transaction. 2. The buyer confirmed that he or she had accidentally left the wrong feedback. 3. Other. See https://www.techjunkie.com/retract-feedback-ebay/. We consider feedback given by mistake only one where the seller indicated reason 2 and the buyer confirmed this.

[26] One may be worried that there was a negative feedback due to a negative quality shock retracted later because the seller bullied the buyer or bought her out. We do not think that this is too common because a seller can only request a revision of feedback received in the last 30 days, and he can do it only once per transaction. That is, once the buyer rejects it he cannot request it again. Beyond that, the seller can only request up to five revisions for every 1,000 pieces of feedback he receives. Moreover, if the seller bought out the buyer, then we would expect that sellers choose reason 1 or 3 in order not to unnecessarily suggest that buyer is at fault. We have selected only feedback that were removed with reason 2.

[27] With this we tend to underestimate the effect of that wrong negative feedback, as there will be transactions for which $post_{it} = 1$ while the negative feedback was already removed.

negative rating by mistake. This could be, for instance, because they sell in different categories.

This is a differences-in-differences model. The identifying assumption, often termed "parallel trends", is that the likelihood that a feedback is left changes with the transaction number in the same way for sellers who at some point receive a negative feedback by mistake, as it does for sellers who never receive a negative feedback by mistake.[28]

We report the results in Table 4. The results in columns (1) to (3) are based on the larger Sample 0. We see that the probability that any feedback is left increases by 4.19 percentage points, the probability that a positive feedback is left increases by 3.86 percentage points, and the probability that a negative feedback is left increases by 0.335 percentage points after the occurrence of a wrong negative feedback. The standard errors are clustered at the seller level. Note that the treatment group of sellers that receive wrong negative feedback received less feedback before the event as compared to the other sellers, perhaps because they sell in markets in which feedback giving is less common.

In columns (4) - (6), we repeat the analyses with Sample 86. Because the sample size is much smaller, we report robust standard errors, but do not cluster at the seller level.[29] The probabilities are qualitatively the same—except that the probability to leave negative feedback increases.

These results relate directly to our second empirical prediction summarized in Corollary 2. By statement (i) in this corollary, the likelihood increases for both types of feedback.[30] By statement (ii) in the corollary, the relative likelihood increases more for negative rather than positive experiences. We need to show that

$$\frac{\Pr(\text{negative feedback after the shock})}{\Pr(\text{negative feedback before the shock})} > \frac{\Pr(\text{positive feedback after the shock})}{\Pr(\text{positive feedback before the shock})}$$

Inserting from the summary statistics for Sample 86 in Table 1, Panel C, column (2), and from the results presented in Table 4, columns (4) and (6),[31]

$$\frac{.011+.00724+.00966}{.011+.00724} > \frac{(.670-0.0666+.0468)-(.011+.00724+.00966)}{(.670-.0666)-(.011+.00724)} \equiv 1.5296 > 1.0634. \tag{6}$$

[28]We provide supportive evidence for this assumption through an leads-and-lags analyses in Table B.5 in the appendix.

[29]For Sample 86, cluster robust standard errors are higher than robust standard errors. The latter are 0.0439, 0.0506, and 0.0099, respectively, for the estimates of the coefficient on post.

[30]The effect increases when controlling the baseline probability for the the treated group that received a negative feedback signal by mistake, by accounting for the reduction in the probability by 0.0666 (Sample 86).

[31]The probabilities are all calculated for the treatment group that received a negative feedback signal by mistake. By example, the numerator of the LHS of the inequality is calculated as follows: 0.011 is the share of transactions with neutral or negative feedback from the total sample. Added to this is 0.00724, the correction of that share for the treated group. Further added to this is 0.00966, the additional share of negatives left by the treated group after the shock. The calculations on the RHS of the inequality all consist of probabilities involving all minus negative feedback received and given by the treated group.

Table 4: Effect of a negative rating on subsequent ratings

	Sample 0			Sample 86		
	(1)	(2)	(3)	(4)	(5)	(6)
	leave any	leave pos.	leave neg.	leave any	leave pos.	leave neg.
post	0.0419**	0.0386**	0.00335**	0.0468**	0.0371*	0.00966*
	(0.0200)	(0.0195)	(0.00144)	(0.0191)	(0.0193)	(0.00559)
treatment group	-0.0398**	-0.0383**	-0.00156	-0.0666***	-0.0738***	0.00724*
	(0.0177)	(0.0172)	(0.00125)	(0.0160)	(0.0161)	(0.00428)
transaction index	Yes	Yes	Yes	Yes	Yes	Yes
adj R-squared	0.00759	0.00732	0.00174	0.000750	0.000872	0.000195
observations	3412510	3412510	3412510	609310	609310	609310
number of clusters	141138	141138	141138	.	.	.

Notes: post dummy = 1 if transaction happens after the date on which the wrong negative feedback was received. In columns (1) - (3), standard errors are clustered at the seller level. In columns (4) - (6), we report robust standard errors. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Thus, our empirical results correspond exactly to the statement of the corresponding Corollary 2 in our theory, which is, we should emphasize, formulated under rather weak conditions. Indeed, the probability that a negative feedback follows an erroneous negative feedback is almost 50% higher than the probability that a positive one follows! We will comment on this in the concluding remarks.

6.2 Event study

We complement our direct main analysis with an event study. It is based on the larger Sample 0. We consider as the event a negative feedback that was given by mistake; we then exploit the randomness in the time it takes until a transaction is rated to identify the effect of that negative rating. We see the event study as a complement to our main analysis, as it allows us to control for transaction quality as explained below, and to obtain additional results on the impact of a wrong negative feedback on transaction quality.

We denote the time at which the first negative feedback was received by t_0 and consider all transactions that took place in a time window that starts 30 days before t_0 and ends 30 days after it. We chose 30 days because Table 1 shows that for our sample the average number of days between the transaction and the feedback is 12.5.[32]

For this set of transactions we define three classes of transactions using both the transaction and the feedback time:

- class 1: transaction and feedback no later than t_0

[32] We also tried specifications without a time window and with a time window only after the first negative feedback was given. The results were very similar.

- class 2: transaction no later than t_0 and feedback after t_0
- class 3: transaction and feedback after t_0.

Thus, class 2 contains transactions that could not be influenced by the seller in response to the negative feedback given by mistake.

We regress an indicator for receiving a negative feedback on the class dummies, omitting always the dummy for class 1. The coefficient on class 2 is the difference in the probability to receive a negative feedback after a negative feedback was received that was later retracted, relative to class 1. Likewise, the coefficient on class 3 is the difference of that probability between class 3 and class 1.

We control for seller fixed effects and transaction number. We also introduce interaction terms with, and control at the same time for buyer experience (whether a buyer has registered no later than 1 March 2009), whether the product is new and has a product ID, and the number of previous positive feedback a seller had.

Before we discuss the results, it is useful to relate this specification to our theoretical model. If it is indeed the case that updating takes place before a rating is left, and if this indeed leads to an increased likelihood to leave a negative rating, then the coefficient on the class 2 indicator will be positive and significantly different from zero. Again, the coefficient on class 2 cannot be affected by changes in seller behavior due to observing the (wrong) negative feedback, because the transaction happened before the event. The coefficient on class 3 will capture both buyer and seller reactions to the negative feedback, and in particular whether sellers react to a later retracted negative feedback by offering higher quality.

We report the results in Table 5. Column (1) shows that indeed, the likelihood that a transaction in class 2 receives a negative feedback is one percentage point higher. This is a big effect when we compare it with the likelihood of 2.6% that a rating is neutral or negative (0.016 / 0.622 from Table 1). But it is also surprisingly consistent with the estimates we presented in Table 4: calculating the baseline probability that a given feedback is negative on the basis of Sample 0 analogously to that for Sample 86 in calculation (6), we obtain $0.0144/0.5822$ based on Table 1 and Table 4, and changes to $(0.0144+0.00335)/(0.588+0.0419)$ after the negative feedback signal. The difference between these two numbers is 1.36 percentage points, not much larger than the 1.05 percentage point increase reported in column (1) of Table 5. Lastly, the fact that the coefficient estimate for class 3 is similar to that for class 2 suggests that changes in seller behavior after observing a negative feedback are small.

The results in the following columns show that this effect is driven by inexperienced buyers who registered after 1 March 2009 (column (2)), and that it is not important whether a product is new and standardized (column (3)). Furthermore, the effect becomes smaller when the seller has already more than 73 positive feedback on his record (column (4)).[33] The last result is

[33] 73 is the 75th percentile of the number of positive feedback that all sellers in the event study, i.e., sellers with first negative feedback that was given by mistake, had accumulated.

Table 5: Inclination to leave a negative feedback

	(1) leave neg.	(2) leave neg.	(3) leave neg.	(4) leave neg.	(5) leave neg.
class 2	0.0105** (0.00437)	0.0198*** (0.00726)	0.0109** (0.00457)	0.0410*** (0.0147)	0.0508*** (0.0170)
class 2 × buyer experience		-0.0156** (0.00689)			-0.0164** (0.00691)
class 2 × new product with ID			-0.0100 (0.0103)		-0.0111 (0.0104)
class 2 × >73 previous positive feedback				-0.0353** (0.0154)	-0.0349** (0.0154)
class 3	0.0103** (0.00420)	0.0122** (0.00501)	0.0106** (0.00435)	0.0368*** (0.00981)	0.0390*** (0.0107)
class 3 × buyer experience		-0.00322 (0.00367)			-0.00377 (0.00367)
class 3 × new product with ID			-0.00652 (0.00516)		-0.00595 (0.00484)
class 3 × >73 previous positive feedback				-0.0314*** (0.0108)	-0.0312*** (0.0109)
seller FE	Yes	Yes	Yes	Yes	Yes
transaction index	Yes	Yes	Yes	Yes	Yes
buyer experience	No	Yes	No	No	Yes
new product with ID	No	No	Yes	No	Yes
number previous positive feedback	No	No	No	Yes	Yes
adj R-squared	0.0763	0.0777	0.0762	0.0772	0.0785
observations	20736	20736	20736	20736	20736
number of clusters	187	187	187	187	187

Notes: Sample restricted to all transactions in a time window of 30 days around the time of a negative feedback that was later retracted. Class 1 (omitted) involves transactions with feedback before the feedback of the retracted negative feedback. Class 2 involves transactions that and feedback after it. Class 3 involves feedback and transaction after the retracted negative. 73 is the 75th percentile of the number of positive feedbacks that all sellers in the event study had accumulated, i.e. sellers with first negative feedback that was given by mistake. Standard errors are clustered at the seller level and account for heteroskedasticity. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Table 6: Inclination to leave a negative feedback: additional results and robustness checks

	(1) claims	(2) low DSR	(3) leave neg	(4) leave neut	(5) placebo 1	(6) placebo 2
class2	0.00544**	0.00861**	0.00869*	0.00163	0.00408	0.000553
	(0.00222)	(0.00366)	(0.00455)	(0.00727)	(0.00451)	(0.00159)
class3	0.00441*	0.0117***		0.0118	0.0176**	0.00367**
	(0.00228)	(0.00361)		(0.00719)	(0.00805)	(0.00170)
seller FE	Yes	Yes	No	Yes	Yes	Yes
transaction index	Yes	Yes	Yes	Yes	Yes	Yes
seller $\times$ transaction date FE	No	No	Yes	No	No	No
adj R-squared	0.122	0.0712	0.0926	0.0389	0.0565	0.0420
observations	20736	20736	19526	7184	12764	73904
number of clusters	187	187	162	60	145	714

Notes: Same specification as column (1) in Table 5. Placebo 1 defines the classes based on a claim without a negative or neutral feedback. Placebo 2 defines it as a positive feedback and a low DSR at the same time. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

important in light of our theory, as this also reflects Proposition 1, but within a context that is quite different from that discussed in subsection 5.1.

Table 6 contains a number of additional results and robustness checks. In column (1) we show that a negative feedback also increases the likelihood that a claim is made by another buyer. Column (2) shows that the likelihood that a low DSR is left increases, too.

We conduct one robustness check and two placebo tests. In the robustness check we control for seller-day fixed effects and find a similar effect. Thereby we also control for seller learning. Now, the coefficient on the class 2 indicator is estimated from transactions in a day in which a transaction with later retracted negative feedback was conducted, and for which feedback was given after the later retracted negative feedback.

The two placebo tests define classes based on events that are not observable to others, and which should therefore not have an effect. Placebo 1 defines the classes based on a wrong claim without a negative or neutral feedback. A wrong claim is a claim that was later removed because eBay decided that the underlying issue was either no one's fault or the fault of the buyer. Placebo 2 defines it as a positive feedback and a low DSR at the same time, where the low DSR has later been revised to a high DSR. In both cases, a buyer was not satisfied with the transaction, but this was not due to the behavior of the seller, similar to the negative feedback that were later changed and that we use in our main analysis. Since confounding by quality shocks is not likely and since claims and low DSR's are not observable to other buyers, the estimated effects should not be significantly different from zero. This is what we find.

As another robustness check, we use a differences-in-differences approach for the event study (not to be confused with the more classical differences-in-differences specification underlying Table 4). We select all transactions between the 30th and the 60th day for all sellers,

Table 7: Inclination to leave a negative feedback: event study with differences-in-differences

	(1) leave neg	(2) claims	(3) low DSR	(4) leave neg
class 1	0.000306	0.0234***	0.00339	
	(0.00380)	(0.00469)	(0.00259)	
class 2	0.133***	0.106***	0.0195***	0.162***
	(0.00608)	(0.00680)	(0.00348)	(0.00560)
class 3	0.0249***	0.0409***	0.00959***	0.0510***
	(0.00439)	(0.00525)	(0.00292)	(0.00635)
class 1 × negative	-0.000000686	0.00347	0.000268	
	(0.00000857)	(0.00228)	(0.000890)	
class 2 × negative	0.0338***	0.0624***	0.0455***	0.0321***
	(0.00867)	(0.00862)	(0.00684)	(0.00927)
class 3 × negative	0.00848***	0.00918***	0.00379*	0.00532
	(0.00307)	(0.00350)	(0.00203)	(0.00900)
seller FEs	No	No	No	Yes
trans_index	Yes	Yes	Yes	Yes
adj R-squared	0.112	0.0821	0.0322	0.127
number of observations	63372	63372	63372	63370
number of clusters	3162	3162	3162	3160

Notes: Sample consists of all transactions between the 30th and the 60th day. Classes are defined using the first non-positive feedback. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

and regress an indicator for a negative feedback on class indicators, as defined by the first non-positive feedback. We also define a variable that we call negative that takes on the value 1 if the first non-positive feedback is negative as opposed to neutral, and create interaction terms between the class indicators and this indicator for a first negative feedback.

Results are presented in Table 7. The first outcome is whether a negative feedback is left for a given transaction. For this outcome, the coefficients on the class indicators will be the respective likelihood that a negative feedback is left in the three classes. The likelihood is higher in class 2 than in class 1, meaning that there is some evidence for time varying quality that confounds a first neutral feedback and a subsequent negative one. The coefficient on the class 3 indicator is smaller again, suggesting that quality improves after a first neutral feedback. Importantly, we are not interested *per se* in this pattern, but in the interaction between the indicators for class 2 and negative. This is the additional effect a first negative feedback has because it is observable to future buyers.[34] If it is random whether a first non-positive feedback is negative or neutral, then this is our effect of interest.

If we assume that the feedback is indeed random, then we can interpret our findings causally. We see in column (1) that the effect of a first negative feedback, instead of a neutral one, is a 3.4 percentage point increase in the likelihood that subsequent feedback are negative. The coefficient on the interaction between class 1 and negative can be interpreted as saying that the likelihood that a first feedback is negative is no different from the likelihood that it is neutral.

This is a large effect. The magnitude cannot be directly compared to the effect we estimate in Table 5 because here we use a different sample and a different empirical approach. Nonetheless, the effect size is similar.

Columns (2) and (3) in Table 7 show that, as before, the effect is also present for claims and low DSR's. Column (4) contains a robustness check. Here, we control for seller fixed effects and find a similar effect to that reported in column (1).

7 Simulation study

In this section, we calibrate the model to moments of the data and perform simulations. The goals of these simulations are to assess the impediments to buyer learning due to unrated transactions; and related to that, to determine how fast future buyers are enabled to correctly infer the seller's type from past ratings.

7.1 Simulating paths of ratings

In Section 3, we presented our model of the typical buyer's rating decision. We assumed that before conducting a transaction with a specific seller, our buyer formed a prior belief about that seller's quality based on the initial belief λ that a randomly drawn seller currently on the

[34] To be precise, it is observable because it changes the percentage negative feedback, unlike a neutral feedback.

platform is of good quality, and in view of our particular seller the buyer adjusted that belief by a scalar y that summarized all publicly available information about that seller. The exact shape of the probability functions $\sigma_h(y)$ and $\sigma_l(y)$ used to specify the buyer's prior belief was immaterial and restricted only by the commonly used Assumption 1 that the MLRP holds.

To simulate dynamics, we could adopt an *ad hoc* approach and parameterize these two functions together with a function that would summarize past ratings in that scalar y. We assume instead that our buyer has rational expectations, knows the entire history of ratings, and bases on this the formation of her belief. This allows us to solve for the implied belief at any point in time. It also allows us to compare deviations between the empirically observed rating pattern with that generated under rational expectations based on the buyer's perfect information.[35]

Turning to specifics, index transactions by t and let z_t the state variable containing information about the seller's rating record that is available to the buyer before she receives the private signal $s \in \{g,b\}$ about seller quality from her transaction t. That state variable is empty in $t = 1$ as the seller does not have prior transactions, has one element in transaction $t = 2$, and so on. The elements of z_t are either 0 when no rating has been left, or contain the rating r_t that was left on transaction t.

Denote prior beliefs for the buyer in transaction t by μ_t. These beliefs are informed by λ, which we had defined as the probability that a newly entering seller is of good quality, and z_t. In transaction t, the buyer receives signal s_t and then updates her beliefs to μ_t^s. We linearize the net benefit derived from leaving a rating as

$$u_t \equiv \underline{b}^s + b \cdot |\mu_t^s - \mu_t| - c_t,$$

where the baseline benefit $\underline{b}^s$ depends on the signal s, allowing for an additional, e.g., psychological cost of leaving a negative rating; $b > 0$; and $\mu_t \equiv \mu(\lambda, z_t)$ is determined by Bayes' rule. We retain the assumption that c_t is uniformly distributed (on the unit interval), which means that the probability that a rating is left is equal to u_t.

Our empirical analysis is based on data in sample 86, which includes only sellers who have at least 86 transactions in the first year. We think of them as high quality sellers. Hence we conduct our simulation study for a typical high quality seller. The signal s_t for a transaction conducted by high quality seller is drawn from a Bernoulli distribution with probability ρ^h.

Prior beliefs of the buyer in the first transaction are given by $\mu_1 = \lambda$. In the first transaction, the buyer receives a signal s_1, updates her beliefs to μ_1^s, and then decides whether to leave a rating. Thereafter, the rating record is z_2, the state variable in the second transaction.

Moving to the second transaction, consider first the case in which a feedback has been left in the first transaction. Then, the one element of z_2 is equal to r_1 and therefore, the buyer in the second transaction knows the signal the buyer in the first transaction has received. In

[35]The case in which the buyer only observes the feedback score and the percentage positive feedback while forming rational expectations would be considerably harder to study: we would have to integrate over all possible histories that could lead to the observed rating pattern.

Table 8: Parameter values

parameter	value	explanation	calibrated
ρ^h	0.9500	$\Pr(s=g\|q=q_h)$	fixed
ρ^l	0.8000	$\Pr(s=g\|q=q_l)$	fixed
λ	0.7000	initial belief	fixed
b_0^g	0.6662	baseline probability to rate when signal was good	yes
b_0^b	0.1668	baseline probability to rate when signal was bad	yes
b_1	1.1639	additional benefit from leaving a rating when beliefs change	yes

Notes: The parameter values we used for the simulation study. Appendix C contains details on the calibration procedure.

addition, $\mu_1 = \lambda$ is common information. Therefore, the buyer in the second transaction has all the information to calculate μ_1^s, which is equal to μ_2. Consider now the case in which no feedback has been left in the first transaction. Then, the buyer in the second transaction knows the probability that the buyer in the first transaction leaves a rating for a given signal s. Denote this probability by π_1^s. Using this, she obtains by Bayes' rule

$$\mu_2 = \frac{A}{A+B},$$

where

$$\begin{aligned} A &= \mu_1 \cdot \left\{\rho^h \cdot (1-\pi_t^g) + (1-\rho^h)\cdot(1-\pi_t^b)\right\} \\ B &= (1-\mu_1)\cdot\left\{\rho^l \cdot (1-\pi_t^g) + (1-\rho^l)\cdot(1-\pi_t^b)\right\}. \end{aligned}$$

In words, A is the joint probability that a seller is high quality and no rating was left.[36] $A+B$ is the probability that no rating was left.[37]

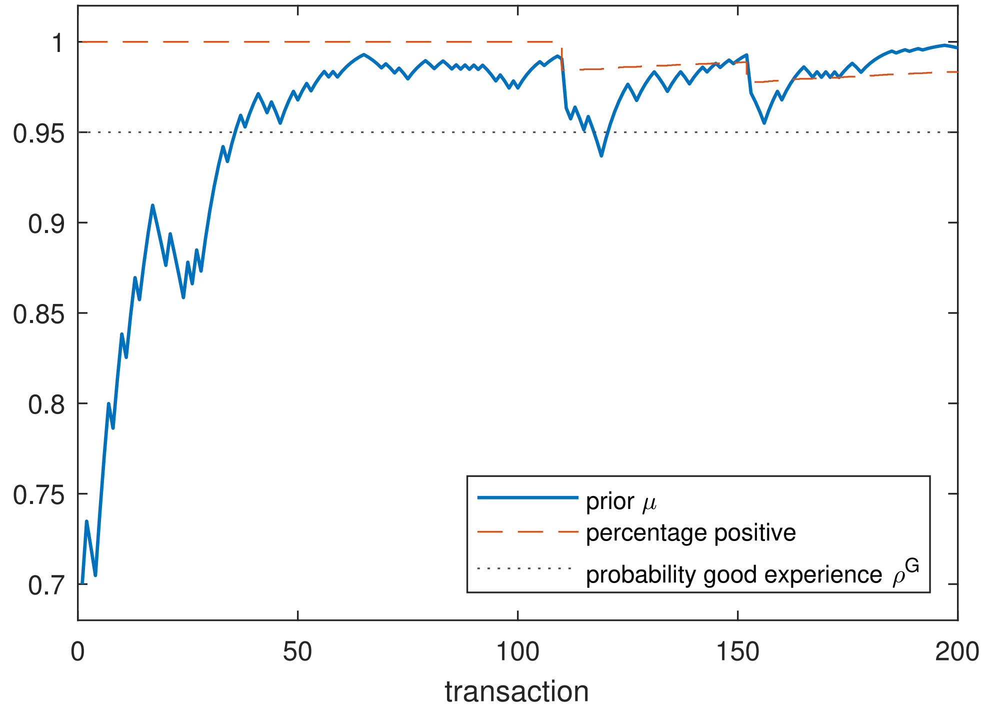
This shows that we can forward iterate to calculate μ_2. It also shows that in the iteration for transaction t, we know from all previous iterations what μ_{t-1} was, and therefore, we also know π_{t-1}^s. From this, we can calculate μ_t, and u_t by drawing s and (c). In our simulation study we forward iterate until $t = 200$.

7.2 Parameter values

Table 8 shows the parameters we used for the simulation. With the eBay context in mind, we set both ρ^h and ρ^l to be of high value, obviously with $\rho^h > \rho^l$. By this, 95 percent of the

[36]This is equal to the probability that the seller in the first transaction was high quality (μ_1) times the probability that no rating was left provided that the seller is high quality (in curly brackets). The expression in curly brackets is the likelihood that a good signal was received from a high quality seller times the probability that no rating was left after a good signal was received, plus the likelihood that a bad signal was received times the likelihood that no rating was given when a bad signal was received.

[37]Figure 6 in Appendix D shows the change in beliefs associated with a missing rating, when $\rho^h = 0.95$ and $\rho^l = 0.8$.



Notes: Evolution of the prior and the percentage positive feedback for one simulation run for a good seller. Horizontal axis = transaction number. Based on setup and procedure described in Section 7.1 and parameter values in Section 7.2.

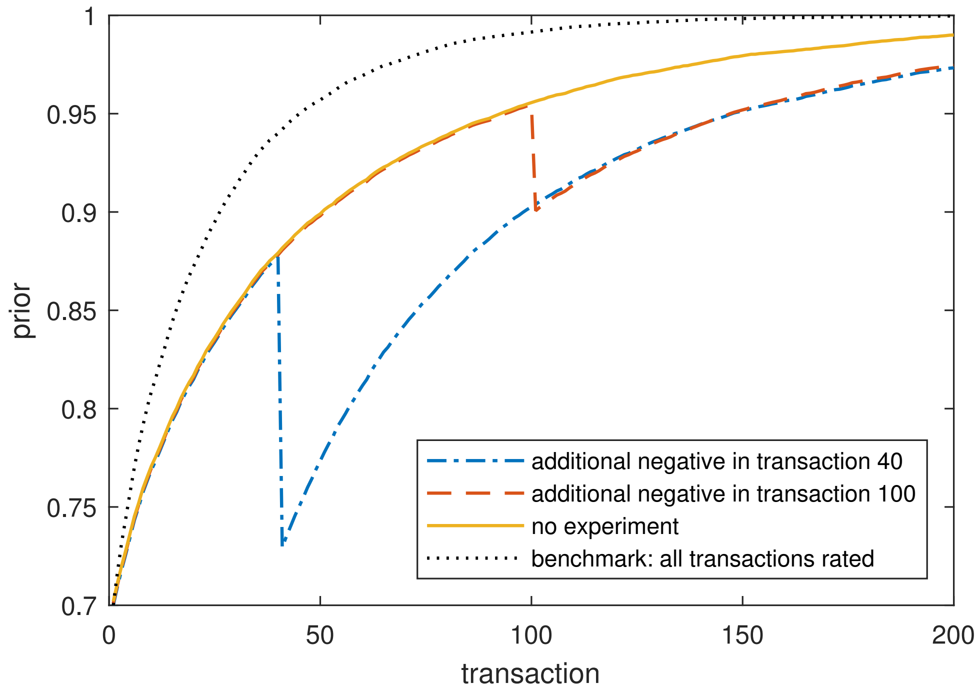
Figure 4: Evolution of prior and percentage positive for one simulation run

time buyers have a good experience with high quality sellers, and 80 percent of the time with low quality sellers. We specified the initial belief that the platform houses good sellers with $\lambda = 0.7$. The remaining parameters were calibrated. For this, we specified a set of moments and then used the L2 norm of those moments to find parameter values. Appendix C contains details on the calibration procedure.

7.3 Results

Figure 4 shows the evolution of the prior and the percentage positive feedback for one simulation run. In total, two negative feedback are left, at the times at which the red line jumps down. Since we model a good seller, the times happen to be rather late. The thick blue line shows the evolution of the prior. We start off at λ. The prior increases with every feedback. The smaller downward movements are related to no feedback being left. Buyers downward adjust the prior, because it is more likely that no feedback was left when a bad signal was received. However, the figure indicates that these events contain relatively little information.

The horizontal black line shows the likelihood that a positive signal is received, which is equal to ρ^h. Interestingly, after about 40 transactions, the percentage positive feedback lies most of the time above that value. This means that feedback aggregates are biased. At the same time, after about 40 transactions buyers are already fairly certain that the seller is of high



Notes: Evolution of the prior for a good seller when we average across 10000 simulation runs. There are three lines. One is the plain average. The other two are for two experiments where we either exogenously leave a negative feedback for transaction 40 or 100. Based on setup and procedure described in Section 7.1 and parameter values in Section 7.2.

Figure 5: Evolution of prior for average simulation run and with two experiments

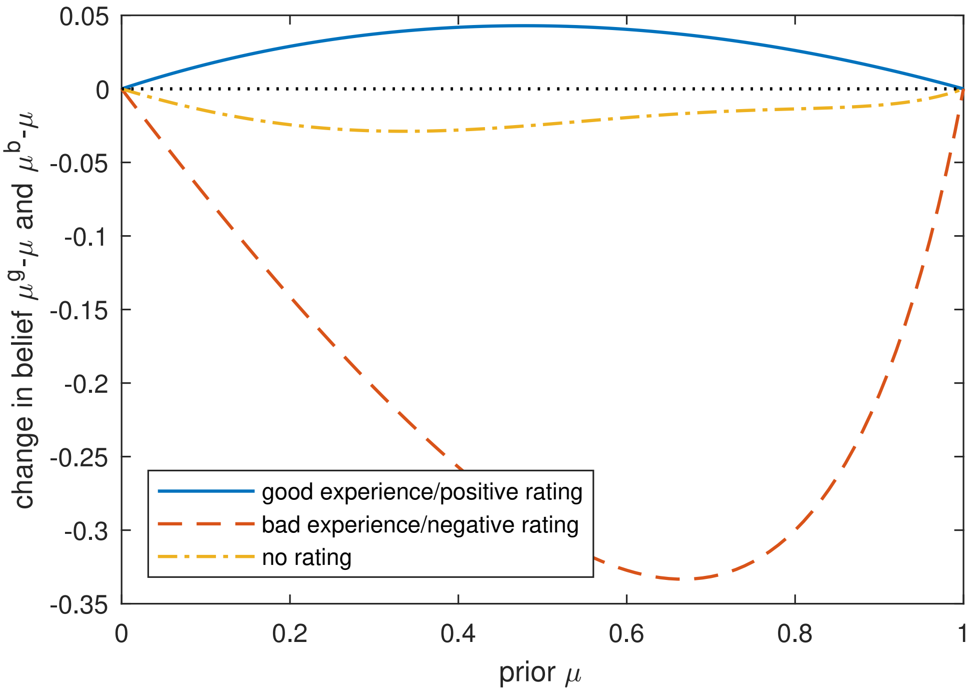
quality.[38] Thus they can learn about seller quality even if aggregate ratings are biased.

We simulated 10000 such simulation runs and then took the average to study the systematic elements. When doing so, we also conducted two counterfactual experiments. In the first experiment we imposed that the experience in the 40th transaction is negative and that it is shared. In the second experiment, we did the same, but for the 100th transaction.

Figure 5 shows the result for the prior.[39] Recall that we undertake the simulations for a good seller. For this seller, the solid line shows that prior converges to one, but that it takes about 100 transactions for the average seller until a buyer attaches a probability of 0.95 to him being of high quality. In comparison, the dotted line shows that it would take a bit less than 50 transactions if all transactions would be rated. If two-thirds of the transactions would be rated, then it would take 75 transactions until a buyer attaches a probability of 0.95 to a seller to be of high quality when he actually is. That is, half of the gap between 50 and 100 transactions can be explained by not all transactions being rated. The other half is explained by ratings being selective, and in particular a lower inclination to leave a rating if the transaction experience was

[38]This is consistent with the finding in Figure 3 that the first 50 feedback are particularly informative about seller quality, as reflected in their impacts on sales price and sales probability.

[39]Appendix D contains additional figures that also show how the percentage positive feedback evolve for good and bad sellers, the evolution of the probability that a rating is left, the evolution of the probability that a rating is negative, how the evolution of the percentage positive feedback for a good seller depends on the arrival of additional negative feedback, and what the effect of ignoring information on missing ratings is on the evolution of the prior.



Notes: Extends Figure 1 and shows change in beliefs associated with a positive signal or rating, no rating, and a negative signal or rating. Based on setup and procedure described in Section 7.1 and parameter values in Section 7.2.

Figure 6: Change in beliefs (extended)

bad.

The other two lines in the figure are for the two experiments we conduct. We exogenously left a negative feedback for either transaction 40 or 100. We see that the prior changes more when this is done for the earlier transaction. The reason for this is that that time, the prior is still further away from 1.[40] It takes about 60 additional transactions until the prior is again at its old level. When an additional negative feedback is left for transaction 100 instead of 40, the prior drops by less, but it takes longer until it reaches its previous level, about 90 transactions.

In Figure 6 we finally show that for our parameter values buyers would learn relatively little from transactions in which no ratings are left. This suggests that our assumption on the formation of the typical buyer's belief is a good approximation, in the sense that providing buyers only with information on the number of ratings and the percentage positive feedback would not provide them with much less information.

8 Summary and policy recommendation

With this paper we wish to contribute to the conceptual and empirical understanding of what can be learned about seller quality by studying when and why buyers leave a rating. We develop a

[40]This is consistent with the finding in Park et al. (2021) that earlier reviews have a larger impact on product sales.

simple theoretical model of a buyer's rating decision. With our model we can explain a number of stylized facts and findings in the literature. We present additional, novel empirical results that are supportive of the model, as they correspond in detail to model predictions. In order to quantify the informativeness of ratings, we calibrate the model to moments of the data to perform a simulation study.

A central element of the model is a buyer's belief about the quality of the seller. This belief is initially informed by the public rating record. The buyer's experience from the transaction generates an additional informative private signal. The buyer is inclined to share this signal the more, the more the experience changes her belief about seller quality, i.e., the more she learns from the transaction. Thus, by this theory, ratings are informative, as they are more likely to be given if a buyer learns more from the transaction.

This holds for both positive and negative buyer experiences. At the same time, we allow for a reluctance in sharing negative experiences that could arise for many reasons, including behavioral ones. This reluctance leads to bias in ratings in the sense that the average rating for a seller is higher than the average transaction experience. We model, and show empirically, however, that a negative rating may not only induce ensuing ratings with a higher probability but that this probability is higher if the rating is based on a negative rather than a positive experience. While one could rationalize this finding *ad hoc* by a behavioral imitation argument, it is explained within our theory by a rational argument, namely the fact that a negative shock provided by a first negative rating increases the buyer's learning process by modifying her prior, and with it the incentive to make public her private experience from a transaction. Since negative ratings tend to be under-represented, inducing negative ratings this way may not hurt the information content of the rating profile.

In all, our simulation results suggest that ratings can be very informative even though they are biased. The rating record reveals the seller's type after about 100 transactions, or 65-70 ratings. This estimate is derived under the assumption that future buyers are aware of the reluctance of past buyers to share negative experiences, which allows them to interpret ratings correctly by placing more weight on ratings that are less likely to be given. A platform could help future buyers by constructing corrected rating aggregates that already put more weight on ratings that are less likely to be given. Our results suggest that this is the case for negative ratings and for ratings that are given later in a seller's career.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2019). Learning from reviews: The selection effect and the speed of learning. *National Bureau of Economic Research*.

Anderson, E. W. and M. W. Sullivan (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing science 12*(2), 125–143.

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

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

Brandes, L., D. Godes, and D. Mayzlin (2019). What drives extremity bias in online reviews? theory and experimental evidence. *Theory and Experimental Evidence (September 6, 2019)*.

Cabral, L. (2012). Reputation on the internet. *The Oxford Handbook of the Digital Economy*, 343.

Cabral, L. and A. Hortacsu (2004). The dynamics of seller reputation: Theory and evidence from ebay. Technical report, National Bureau of Economic Research.

Cabral, L. M. B. and A. Hortaçsu (2006). The dynamics of seller reputation: Theory and evidence from ebay. Mimeograph.

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

Chakraborty, I., J. Deb, and A. Öry (2020). When do consumers talk? Cowles Foundation Discussion Paper No. 2254.

Chevalier, J. A. and D. Mayzlin (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research 43*(3), 345–354.

Dai, W., G. Jin, J. Lee, and M. Luca (2018). Aggregation of consumer ratings: an application to yelp. com. *Quantitative Marketing and Economics*.

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

Dellarocas, C., F. Dini, and G. Spagnolo (2006). Designing feedback mechanisms for e-procurement platforms. In N. Dimitri, G. Piga, and G. Spagnolo (Eds.), *Handbook of Procurement*, Chapter 18. Cambridge University Press.

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

Filippas, A., J. J. Horton, and J. Golden (2018). Reputation inflation. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 483–484.

Fradkin, A., E. Grewal, and D. Holtz (2020). Reciprocity in two-sided reputation systems: Evidence from an experiment on airbnb. 2020.

Hart, O. and J. Moore (2008). Contracts as reference points. *Quarterly Journal of Economics 123*(1), 1–48.

Ho, Y.-C., J. Wu, and Y. Tan (2017). Disconfirmation effect on online rating behavior: A structural model. *Information Systems Research 28*(3), 626–642.

Hu, N., P. A. Pavlou, and J. J. Zhang (2009). Why do online product reviews have a j-shaped distribution? overcoming biases in online word-of-mouth communication. *Communications* *of the ACM 52*(10), 144–147.

Hu, N., P. A. Pavlou, and J. J. Zhang (2017). On self-selection biases in online product reviews. *MIS Q. 41*(2), 449–471.

Hui, X., M. Saeedi, and N. Sundaresan (2018). Adverse selection or moral hazard, an empirical study. *The Journal of Industrial Economics 66*(3), 610–649.

Klein, T. J., C. Lambertz, G. Spagnolo, and K. O. Stahl (2009). The actual structure of ebay's feedback mechanismn and early evidence on the effect of recent changes. *International Journal of Electronic Business 7(3)*, 301–320.

Klein, T. J., C. Lambertz, and K. O. Stahl (2016). Market transparency, adverse selection, and moral hazard. *Journal of Political Economy 124(6)*, 1677–1713.

Li, X. and L. Hitt (2008). Self-selection and information role of online product reviews. *Information Systems Research 19*(4), 456–474.

Luca, M. and O. Reshef (2020). The impact of prices on firm reputation. Technical report, National Bureau of Economic Research.

Luca, M. and G. Zervas (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science 62*(12), 3412–3427.

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

Mobius, M. and T. Rosenblat (2014). Social learning in economics. *Annual Reviews in Economics 6*, 827–847.

Moe, W. W. and D. A. Schweidel (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science 31*(3), 372–386.

Moe, W. W. and M. Trusov (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research 48*(3), 444–456.

Nosko, C. and S. Tadelis (2015). The limits of reputation in platform markets: An empirical analysis and field experiment. Technical report, National Bureau of Economic Research.

Park, S., W. Shin, and J. Xie (2021). The fateful first consumer review. *Marketing Science*.

Resnick, P. and R. Zeckhauser (2002a). *The Economics of the Internet and E-Commerce*. Amsterdam, Elsevier Science.

Resnick, P. and R. Zeckhauser (2002b). Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. the economics of the internet and e-commerce. In M. R. Baye (Ed.), *Advances in Applied Microeconomics*, Volume 11, Amsterdam. Elsevier Science.

Schoenmüller, V., O. Netzer, and F. Stahl (2019). The extreme distribution of online reviews: Prevalence, drivers and implications. *Columbia Business School Research Paper* (18-10).

Stenzel, A., C. Wolf, and P. Schmidt (2020). Pricing for the stars. Technical report, Bocconi University.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics 8*, 321–340.

Online Appendix

A Our model and the J-shaped distribution

We formulated our model for the simple case in which seller quality is either high or low, buyer experiences can either be good or bad, and ratings—if given at all—are positive or negative. This allowed us to highlight the idea that the rating decision depends on the amount of learning. The model also fits well to the eBay context in which we conduct our empirical analysis.

On other platforms, buyers can leave reviews on a finer scale. In principle, one could re-code these reviews on a coarser scale and then our model would still apply. Alternatively, we could generalize our model to understand phenomena such as the frequently observed so-called J-shaped distribution of reviews, by which on a scale between very positive and very negative ratings, many are very positive and a few are very negative, with very few in between.[1]

Here we sketch how this could be done. We could assume that the buyer experience is drawn from a normal distribution with seller-type specific mean and variance. Buyers have a prior about these parameters for each seller. They learn from reviews and their own experience. The rating scale could be described by a set of cutoffs for the experience, so that an experience below the lowest cutoff leads to the lowest possible rating when it is given; an experience between the lowest and second lowest cutoff leads to the second lowest rating if it is given, and so on. The amount of learning could be measured by the (absolute value of the) change in beliefs about the seller-specific mean of the experience distribution, and, as in our model, the benefit associated with leaving a rating increases in the amount of learning.

[1] As but one example, Hu et al. (2009) find that that pattern for reviews provided on Amazon.

B Robustness

This appendix contains additional results and robustness checks. Table B.1 reproduces Table 2 for DSR's as the outcome. In Tables B.2 and B.3 we use alternative definitions of experience. Table B.4 reproduces Table 2 for separate DSR categories. None of these invites a reinterpretation of our results.

In Figures B.1, we study whether changes in consumer surplus over seller life can explain the empirical pattern in rating behavior shown in Figure 2. In Figure B.1a, we study auctions and use the winning bid minus the sales price in an auction as a proxy for consumer surplus. By the auction format on eBay, the dominant strategy is to bid one's own valuation, and therefore the winning bid is a good approximation of the winner's maximum willingness to pay. In Figure B.1b, we use the sales price as inversely related to consumer surplus. For this, we use transactions with posted price where the product had a product ID and was in new condition. We then regress the logarithm of price on dummy variables for the transaction index, controlling for seller fixed effects, transaction month fixed effects, and product ID fixed effects, to account for potential changes in product portfolio as a seller grows. In both figures, we do not observe dramatic changes in consumer surplus; if anything, Figure B.1b suggests a price decrease as transaction index increases, which by Acemoglu et al. (2019)'s model should lead to an increase in the likelihood of rating. Summarizing, the results suggest that changes in consumer surplus over time cannot explain the empirical patterns in rating behavior in Figure 2.

Table B.5 provides supporting evidence for the parallel trends assumption we make to obtain the results in Table 4. We interact the 'wrong feedback' dummy with the dummy variables for the periods preceding the wrong negative feedback in terms of transaction index: The dummies 't-n' refer to (10*(n-1)+1)–10*n th transactions before the wrong negative feedback, where n = 1, 2, ..., 8. We see that in Sample 0 as reported in columns (1) - (3), although there appears to be significant imbalance from the ninth periods and earlier as seen in the coefficient on 'wrong feedback', the coefficients in front the lead terms are mostly statistically insignificant in the most recent eight periods, suggesting that the parallel trend assumption mostly holds in periods just preceding the events. In columns (4) - (6) where we analyze Sample 86, while there are some differences in the average feedback rate across sellers who have received wrong negative versus those that have not, we note that the difference in fact goes against our story: sellers who received wrong negative feedback were actually receiving less feedback in the preceding months, perhaps because of the lower baseline tendency of receiving feedback in certain categories; but despite this, they receive more feedback after their wrong negative feedback.

Table B.1: Probability to receive DSR

	(1) feedback	(2) feedback	(3) feedback	(4) feedback	(5) feedback	(6) feedback
transaction number/10	-0.966***	-0.608***	-0.814***	-0.725***	-0.540**	-0.634***
	(0.102)	(0.152)	(0.152)	(0.161)	(0.234)	(0.242)
transaction number/10 squared	0.0632***	0.0287*	0.0470***	0.0480***	0.0365	0.0473*
	(0.0113)	(0.0150)	(0.0150)	(0.0162)	(0.0262)	(0.0271)
buyer experience				0.280	1.087**	1.436***
				(0.197)	(0.482)	(0.471)
trans. num/10 $\times$ buyer exp.					-0.295	-0.327
					(0.275)	(0.271)
trans. num/10 sq. $\times$ buyer exp.					0.0186	0.0216
					(0.0334)	(0.0332)
buyer inclination to leave feedback						26.81***
						(0.444)
trans. num/10 $\times$ buyer inc. to leave fdbk						0.606**
						(0.246)
trans. num/10 sq. $\times$ buyer inc. to leave fdbk						-0.0590**
						(0.0288)
seller FE	No	Yes	Yes	Yes	Yes	Yes
month FE	No	Yes	Yes	Yes	Yes	Yes
leaf category	No	No	Yes	Yes	Yes	Yes
adj R-squared	0.000494	0.0510	0.0600	0.0549	0.0549	0.117
observations	609310	609310	607135	515978	515978	515978
number of clusters	.	7085	7085	7083	7083	7083

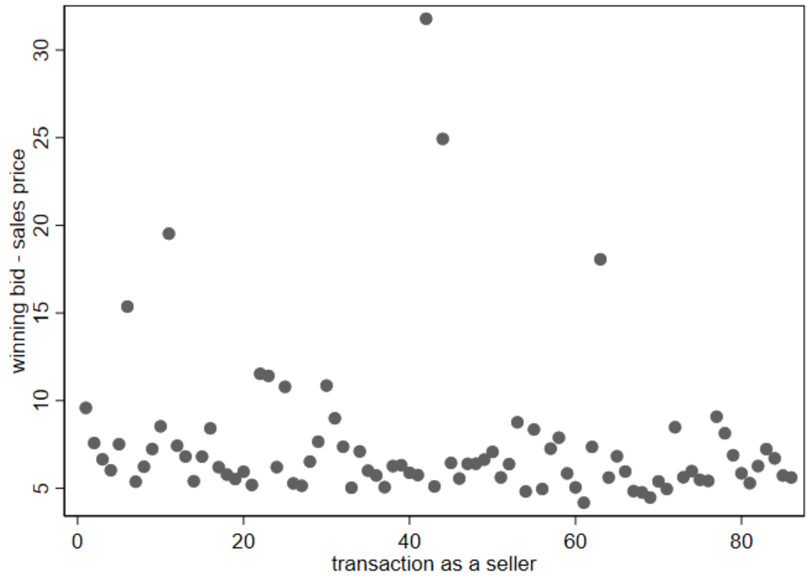
Notes: Table shows results of regressions of an indicator for receiving a DSR on the transaction number divided by 10 and the transaction number divided by 10 squared, as well as other controls and interaction terms. One observation is a transaction. Standard errors are clustered at the seller level and account for heteroskedasticity. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Table B.2: Probability to receive feedback

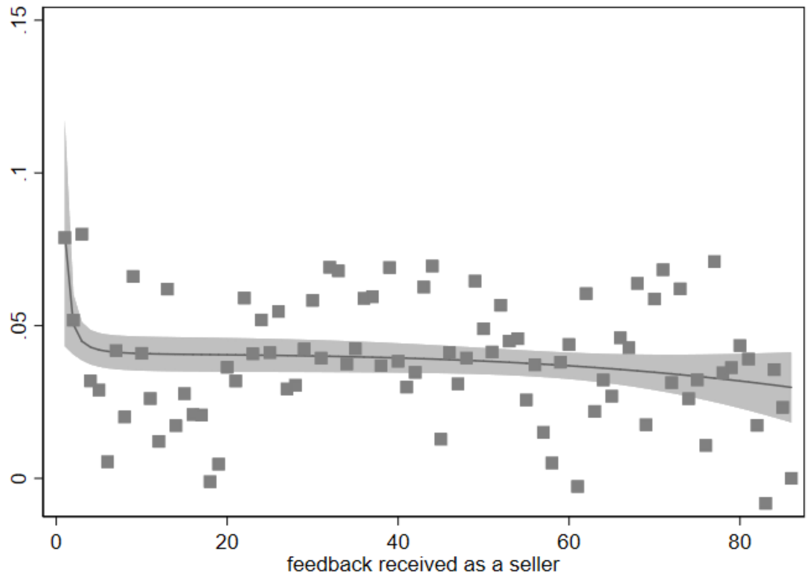
	median split of buyer characteristics			buyer exp. by num. trans in prev yr.		
	(1)	(2)	(3)	(4)	(5)	(6)
	feedback	feedback	feedback	feedback	feedback	feedback
transaction number/10	-0.663***	-0.512***	-0.563**	-0.641***	-0.786***	-0.829***
	(0.154)	(0.196)	(0.228)	(0.154)	(0.175)	(0.188)
transaction number/10 squared	0.0361**	0.0273	0.0421*	0.0350**	0.0495***	0.0541***
	(0.0154)	(0.0211)	(0.0252)	(0.0154)	(0.0180)	(0.0197)
buyer experience	1.159***	2.065***	1.902***	7.228***	6.413***	4.862***
	(0.179)	(0.446)	(0.407)	(0.189)	(0.462)	(0.445)
trans. num/10 $\times$ buyer exp.		-0.317	-0.222		0.454*	0.248
		(0.253)	(0.237)		(0.249)	(0.241)
trans. num/10 sq. $\times$ buyer exp.		0.0186	0.00491		-0.0460	-0.0269
		(0.0307)	(0.0293)		(0.0283)	(0.0274)
buyer inclination to leave feedback			39.47***			25.89***
			(0.412)			(0.396)
trans. num/10 $\times$ buyer inc. to leave fdbk			0.304			0.561**
			(0.224)			(0.223)
trans. num/10 sq. $\times$ buyer inc. to leave fdbk			-0.0390			-0.0496*
			(0.0259)			(0.0264)
seller FE	Yes	Yes	Yes	Yes	Yes	Yes
month FE	Yes	Yes	Yes	Yes	Yes	Yes
leaf category	Yes	Yes	Yes	Yes	Yes	Yes
adj R-squared	0.0649	0.0649	0.242	0.0695	0.0695	0.133
observations	515978	515978	515978	515978	515978	515978
number of clusters	7083	7083	7083	7083	7083	7083

Notes: In columns (1)-(3), both buyer characteristics and defined based on the median values. For buyer experienced and buyer inclination to leave feedback are 01feb2005 and 0.857143, respectively. In columns (4)- (6), we use 75-25 percentile split, and buyer experienced is measured in terms of number of transactions in the previous year. The cutoffs for buyer experienced and buyer inclination to leave feedback are 78 and 0.981982, respectively. Standard errors are clustered at the seller level and account for heteroskedasticity. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

(a) Winning bid minus sales price



(b) log(sales price)



Notes: Figure shows the dependence of two measures related consumer surplus on the transaction index. For both we use Sample 86. In Figure B.1a we use the difference between the winning bid and the price that was paid as the measure. We use all auction sales and plot the average winning bid minus the sales price against the transaction index. In Figure B.1b, we use transactions with posted price where the product had a product ID and was in new condition. We regress the logarithm of sales price on dummy variables for the transaction index, controlling for seller fixed effects, transaction month fixed effects, and product ID fixed effects. The dummy for transaction index = 86 is dropped as the benchmark. The figure plots coefficients on the transaction index dummies and a local polynomial fit.

Figure B.1: Dependence of consumer surplus on transaction index

Table B.3: Inclination to leave a negative feedback

	(1) leave neg.	(2) leave neg.	(3) leave neg.	(4) leave neg.	(5) leave neg.
class 2	0.0105** (0.00437)	0.0156*** (0.00456)	0.0109** (0.00457)	0.0410*** (0.0147)	0.0160 (0.0105)
class 2 × buyer experience		-0.0107** (0.00534)			-0.00543 (0.00605)
class 2 × new product with ID			-0.0100 (0.0103)		-0.00681 (0.0102)
class 2 × number previous positive feedback				-0.0353** (0.0154)	-0.00764 (0.0117)
class 3	0.0103** (0.00420)	0.0152*** (0.00435)	0.0106** (0.00435)	0.0368*** (0.00981)	
class 3 × buyer experience		-0.0101** (0.00434)			-0.00228 (0.00432)
class 3 × new product with ID			-0.00652 (0.00516)		-0.000914 (0.00517)
class 3 × number previous positive feedback				-0.0314*** (0.0108)	0.00632 (0.00474)
seller FE	Yes	Yes	Yes	Yes	Yes
transaction index	Yes	Yes	Yes	Yes	Yes
buyer experience	No	Yes	No	No	Yes
new product with ID	No	No	Yes	No	Yes
number previous positive feedback	No	No	No	Yes	Yes
adj R-squared	0.0763	0.0765	0.0762	0.0772	0.0758
observations	20736	20736	20736	20736	20736
number of clusters	187	187	187	187	187

Notes: Standard errors are clustered at the seller level and account for heteroskedasticity. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Table B.4: Appendix: Imitation effect robustness

	(1) low DSR1	(2) low DSR2	(3) low DSR3	(4) low DSR4
class 2	0.00593**	0.00586**	0.00271	0.00289
	(0.00268)	(0.00241)	(0.00285)	(0.00214)
class 3	0.00702**	0.00584**	0.00329	0.00646**
	(0.00300)	(0.00256)	(0.00286)	(0.00253)
seller FE	Yes	Yes	Yes	Yes
transaction index	Yes	Yes	Yes	Yes
adj R-squared	0.0916	0.0992	0.102	0.0780
observations	20736	20736	20736	20736
number of clusters	187	187	187	187

Notes: DSR1 = item as described. DSR2 = communication. DSR3= shipping time. DSR4= shipping charge. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Table B.5: Appendix: Inclination to leave any feedback robustness

	Sample 0			Sample 86		
	(1)	(2)	(3)	(4)	(5)	(6)
	leave any	leave pos.	leave neg.	leave any	leave pos.	leave neg.
wrong feedback	-0.0414**	-0.0398**	-0.00161	-0.0454**	-0.0456**	0.000131
	(0.0188)	(0.0183)	(0.00132)	(0.0218)	(0.0220)	(0.00460)
wrong feedback × t -8	-0.00822	-0.0134	0.00520	-0.655***	-0.646***	-0.00909*
	(0.0353)	(0.0344)	(0.00485)	(0.0222)	(0.0222)	(0.00464)
wrong feedback × t -7	0.0276	0.0302	-0.00261	-0.0446	-0.0345	-0.0101**
	(0.0292)	(0.0294)	(0.00336)	(0.156)	(0.156)	(0.00463)
wrong feedback × t -6	0.0174	0.0217	-0.00425	-0.145	-0.181*	0.0362
	(0.0280)	(0.0280)	(0.00328)	(0.109)	(0.108)	(0.0447)
wrong feedback × t -5	0.0669**	0.0629**	0.00398	0.128	0.105	0.0233
	(0.0267)	(0.0266)	(0.00456)	(0.0799)	(0.0832)	(0.0330)
wrong feedback × t -4	0.0237	0.0214	0.00229	-0.00157	-0.0601	0.0585
	(0.0326)	(0.0332)	(0.00653)	(0.0761)	(0.0781)	(0.0383)
wrong feedback × t -3	0.0231	0.0281	-0.00500	-0.122**	-0.123**	0.00188
	(0.0299)	(0.0295)	(0.00339)	(0.0585)	(0.0586)	(0.0126)
wrong feedback × t -2	-0.0202	-0.0246	0.00440	-0.0497	-0.0398	-0.00995**
	(0.0268)	(0.0265)	(0.00665)	(0.0505)	(0.0505)	(0.00460)
wrong feedback × t -1	0.00687	0.00676	0.000110	-0.00798	-0.0354	0.0275
	(0.0248)	(0.0244)	(0.00561)	(0.0473)	(0.0480)	(0.0171)
post	0.0435**	0.0401**	0.00340**	0.0256	0.00888	0.0168***
	(0.0206)	(0.0200)	(0.00149)	(0.0242)	(0.0244)	(0.00584)
transaction index	Yes	Yes	Yes	Yes	Yes	Yes
adj R-squared	0.00759	0.00732	0.00174	0.000766	0.000886	0.000222
observations	3412510	3412510	3412510	609310	609310	609310
number of clusters	141138	141138	141138	.	.	.

Notes: The post dummy equals 1 if the transaction happens after the date on which the wrong negative feedback was received. The dummies 't-n' refer to (10*(n-1)+1)–10*n th transactions before the wrong negative feedback. In columns (1) - (3), standard errors are clustered at the seller level. In columns (4) - (6), we report robust standard errors. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Table C.6: Moments

predicted quantity	target value	scale factor	value
probability rating is left for $t = 1$	0.7	0.01	-0.0751
probability rating is left for $t = 86$	0.65	0.01	0.0666
probability rating negative	0.02	0.001	0.2403
increase in probability negative rating after negative in $t = 40$	0.01	0.001	-0.6869
max $\{0, \max_\mu$ probability negative experience shared - 0.5$\}$	0	0.05	1.0552
average transaction number first negative	37	10	2.76831

Notes: Table shows the moments we used in the calibration procedure. See main text in Appendix C for details.

C Details on the calibration procedure

Here we describe the procedure we used to find parameters so that model predictions match a few key patterns in our data. Recall from Table 8 that we have set $\rho^h = 0.95$, $\rho^l = 0.8$, and $\lambda = 0.7$. The parameters we would like to find are the preference parameters b_0^g, b_0^b, and b_1.

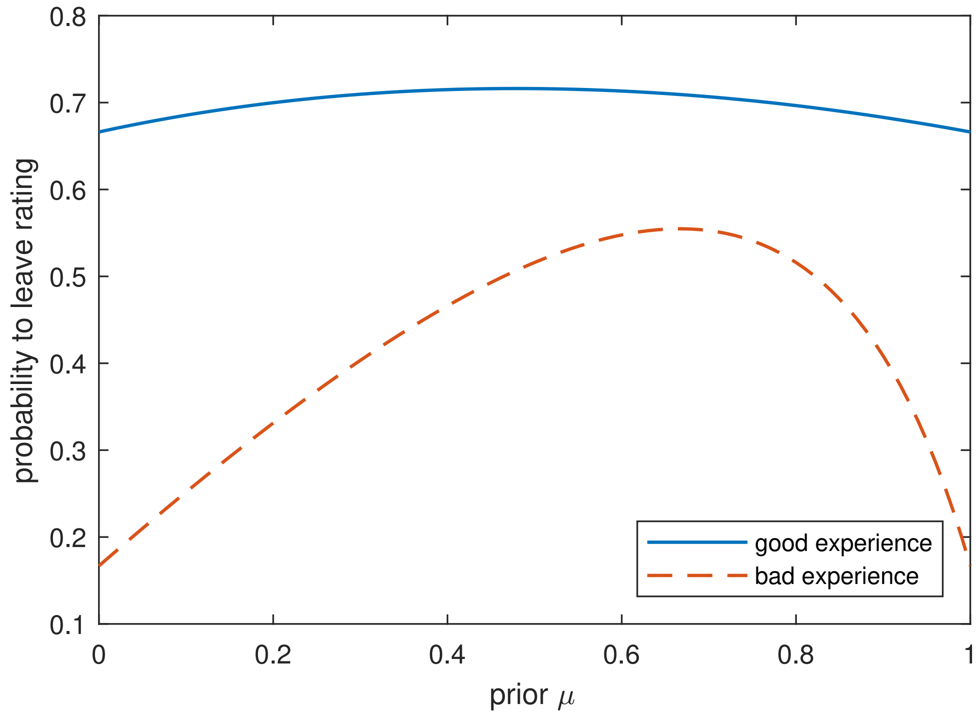
We specify the 6 moments in Table C.6. The first column contains predicted quantities that are calculated as averages over 10000 simulated paths. The first two target quantities are the probabilities that a rating is left for the first and the 86th transaction, respectively. The target values approximate the ones in Figure 2. The third target quantity is the percentage of negative feedback. The number 0.02 is related to the 97.8 percent percentage positive feedback in the first year that are reported in Table 1 for sample 86. Our fifth target quantity is how much the predicted probability that a negative experience is shared, for the prior for which this quantity is biggest, lies above 0.5. The target quantity here is 0, as we would like to find parameters for which that probability is always below 0.5. Finally, our sixth target quantity is the average transaction number for the first negative feedback. The target quantity here is 37 and is directly taken from the data for sample 86.

For each trial value of the parameters, we calculate the 6 predicted quantities in the first column of Table C.6. Then, we subtract the target value in the third column and divide the difference by the scale factor in the second column.[2] Then, we calculate the sum of the squared re-scaled differences. This goal function is then numerically minimized.[3]

The last column in Table C.6 contains the value of the re-scaled difference. We consider a value of less than 1 to be desirable and in that sense, we are able to find parameter values so that our simulations fit the data well.

[2]These scale factors were chosen so that the difference divided by the scale factor would be 1 for a economically significant but not too big difference. For instance, we would consider a difference of 0.01 between the predicted probability that a rating is left for $t = 1$ and the target value 0.7 comparable to a difference of 10 between the average transaction number for which the first negative feedback was received and 37.

[3]As usual, we take the random draws for all 10000 paths before optimizing over the parameters.



Notes: Probability that a rating is left, by experience. Based on setup and procedure described in Section 7.1 and parameter values in Section 7.2.

Figure D.2: Probability to leave rating

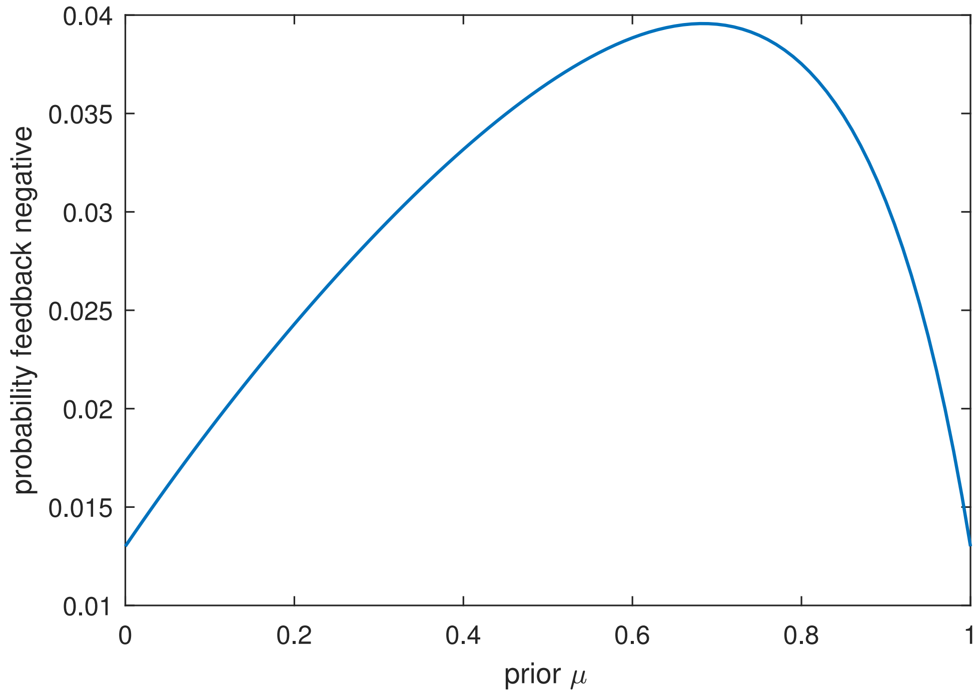
D Additional figures for the simulation study

The first two figures characterize rating behavior as a function of the prior and the experience. They are produced for the parameter values in Table 8. Figure D.2 shows how the probability that a rating is left depends on the prior and the experience. From this we can derive what the likelihood is that a rating is negative. This likelihood is plotted in Figure D.3.

The third figure, Figure 6, shows what future buyers with rational expectations can learn from past ratings. This figure is related to the discussion in Section 7.1.

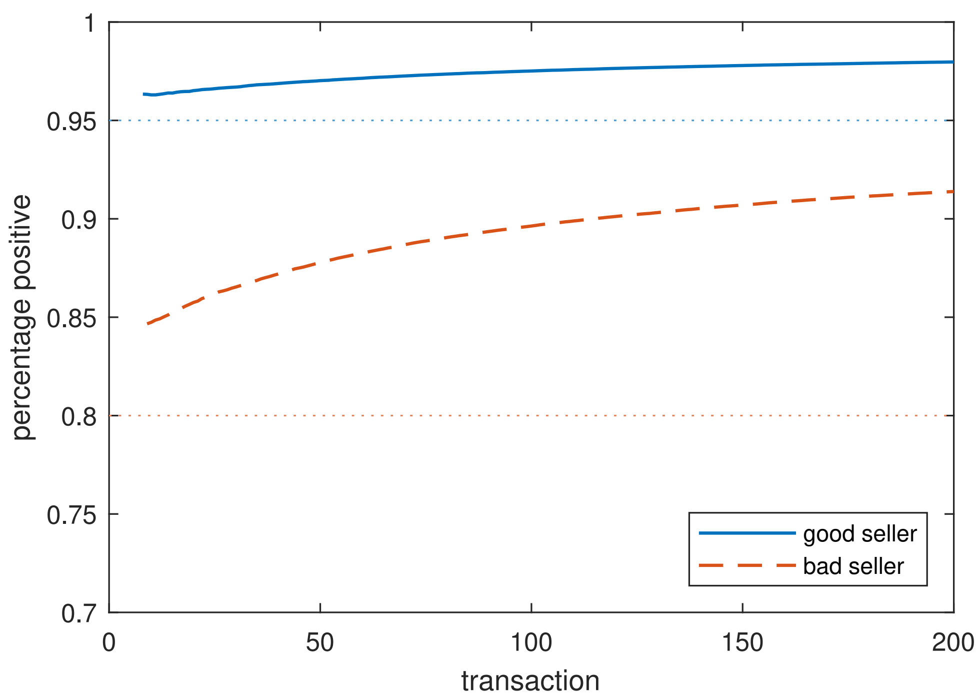
The remaining figures obtained by simulating 10000 paths and taking the average. Figure D.4 shows how the percentage positive feedback to date evolves for good and bad sellers and compares it to the respective likelihood that a transaction experience is positive. This shows that ratings are biased. Figure D.5 shows how the prior evolves for good and bad sellers.

The following three figures show results for good sellers and how they depend on the arrival of additional negative feedback. Figure D.6 shows the likelihood that a transaction is rated, Figure D.7 shows the likelihood that a rating is negative, and Figure D.8 shows the simulated percentage positive ratings to date. Finally, Figure D.9 is also for a good seller and shows what the effect of ignoring information on missing ratings is on the evolution of the prior.



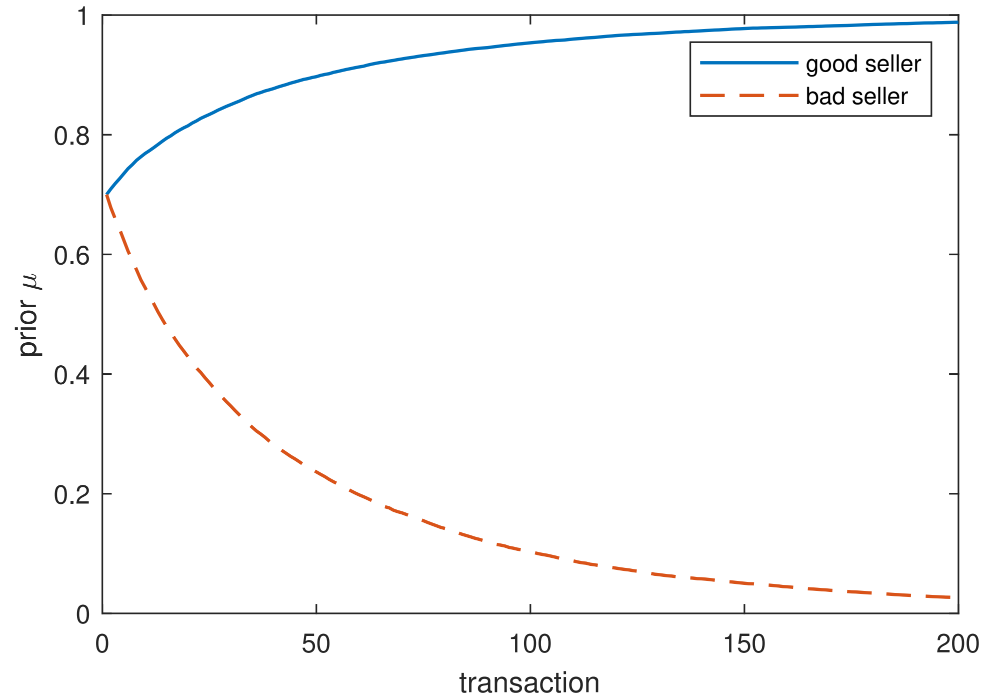
Notes: Probability that a rating is negative provided that it is left. Based on setup and procedure described in Section 7.1 and parameter values in Section 7.2.

Figure D.3: Probability a rating is negative



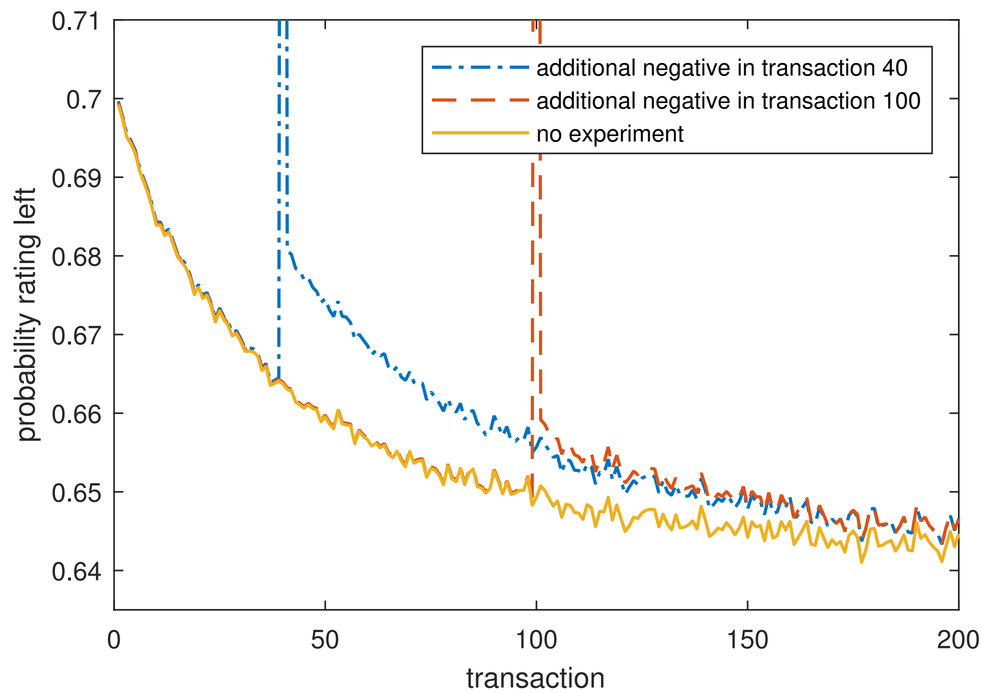
Notes: Percentage positive feedback to date when we average across 10000 simulation runs (solid and dashed line). Also shows likelihood to have a good experience for good and bad sellers, respectively (dotted lines). Based on setup and procedure described in Section 7.1 and parameter values in Section 7.2.

Figure D.4: Evolution of percentage positive ratings



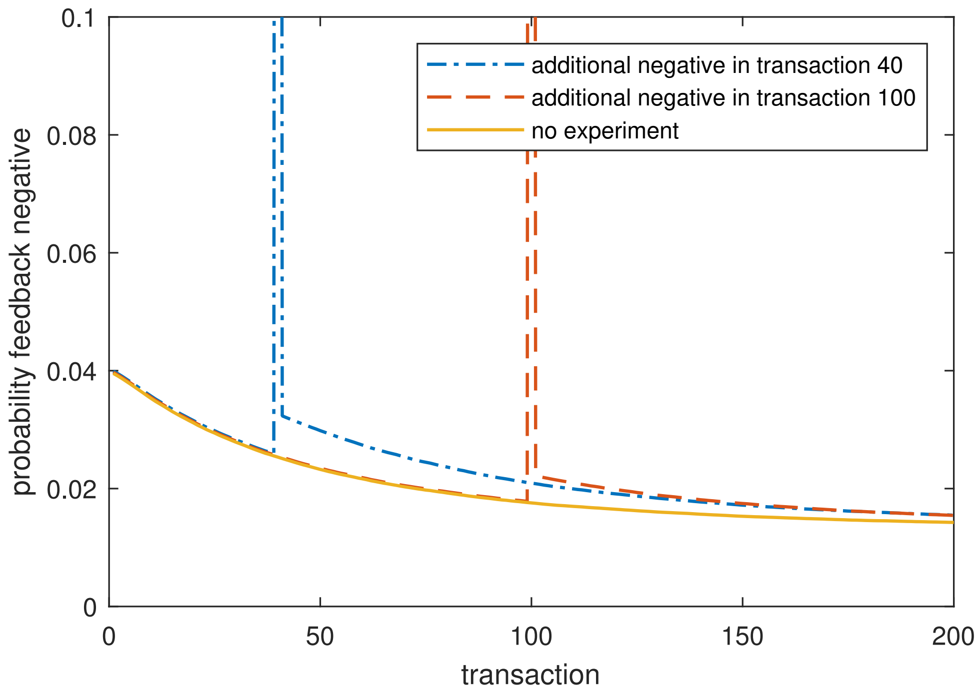
Notes: Prior when we average across 10000 simulation runs (solid and dashed line). Based on setup and procedure described in Section 7.1 and parameter values in Section 7.2.

Figure D.5: Evolution of prior



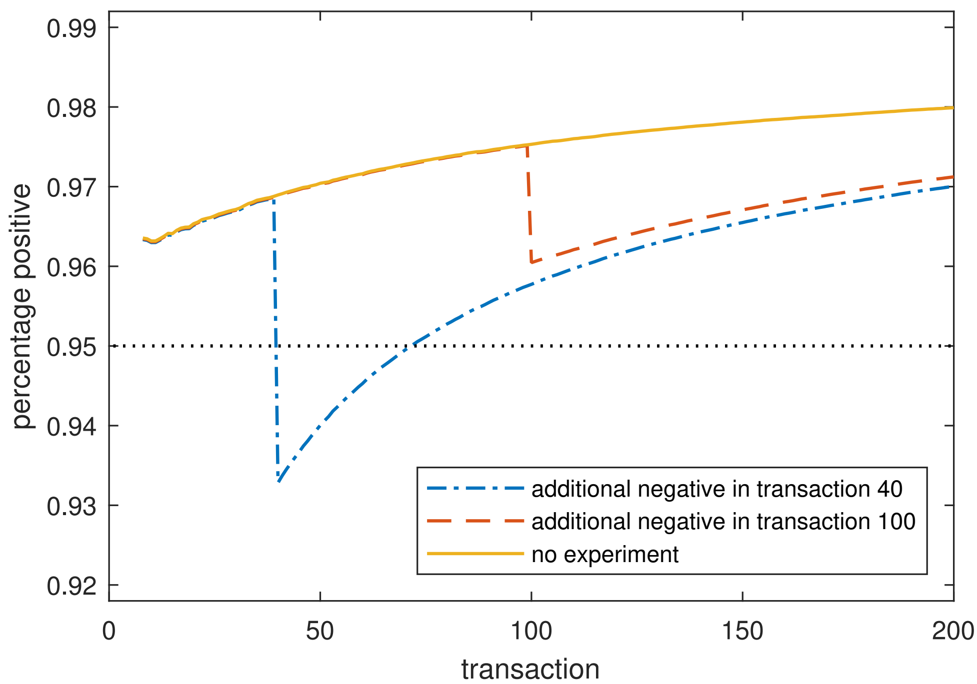
Notes: Probability that a rating is left evolves for a good seller. Average of 10000 simulated paths and for the case that an additional negative rating is left in transaction 40 and 100, respectively. Based on setup and procedure described in Section 7.1 and parameter values in Section 7.2.

Figure D.6: Evolution of probability that a rating is left



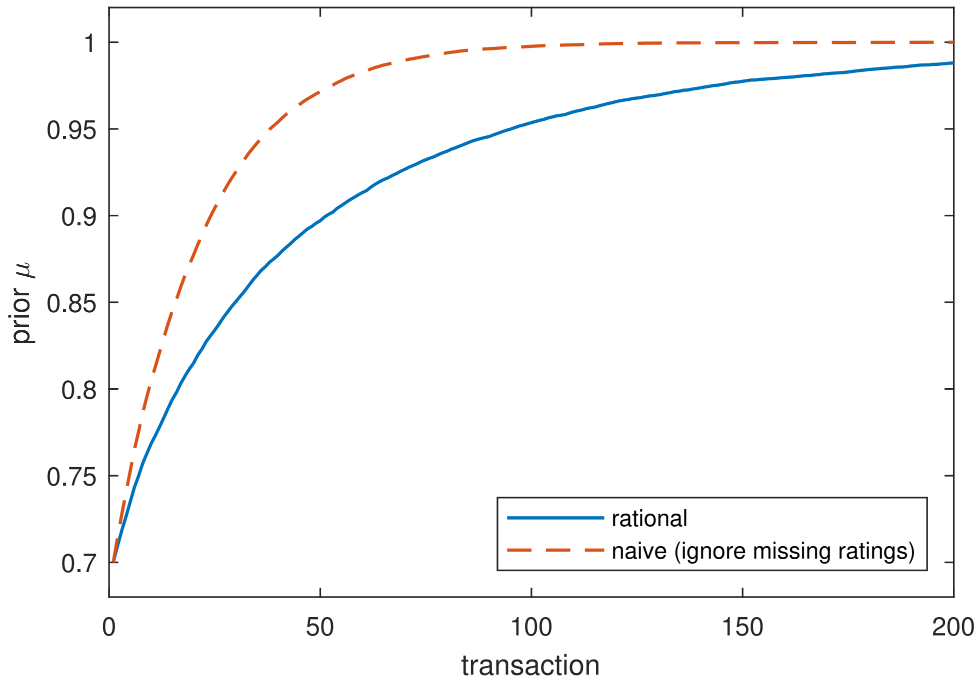
Notes: Likelihood that a given feedback is negative evolves for a good seller. Average of 10000 simulated paths and for the case that an additional negative rating is left in transaction 40 and 100, respectively. Vertical axis truncated at 0.1. Based on setup and procedure described in Section 7.1 and parameter values in Section 7.2.

Figure D.7: Evolution of probability that a rating is negative



Notes: Percentage positive ratings up to a given transaction for a good seller. Displayed for the average of 10000 simulated paths and for the case that an additional negative rating is left in transaction 40 and 100, respectively. Based on setup and procedure described in Section 7.1 and parameter values in Section 7.2.

Figure D.8: Evolution of percentage positive feedback



Notes: Comparison evolution of the prior for a good seller when missing ratings are taken into account and when ignored. Average over 10000 simulated path. Based on setup and procedure described in Section 7.1 and parameter values in Section 7.2.

Figure D.9: Evolution of prior with and without taking missing ratings into account