Discussion Paper Series – CRC TR 224

Discussion Paper No. 267
Project B 04

When and Why Do Buyers Rate in Online Markets?

Xiang Hui [1]
Tobias J. Klein [2]
Konrad Stahl [3]

March 2021

[1] Washington University in St. Louis
[2] Tilburg University and CEPR
[3] University of Mannheim, CEPR, CESifo and ZEW

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

When and Why Do Buyers Rate in Online Markets?*

Xiang Hui† Tobias J. Klein‡ Konrad Stahl§

March 1, 2021

Preliminary. Comments Welcome

Abstract

Anonymous markets would be very difficult to successfully operate without the possibility that buyers rate the seller. Yet many empirical results yield that ratings are non-random and concentrate on extreme experiences. We develop a model of rating decisions in which the buyer is willing to share publicly her opinion about a transaction, if its realized quality differs much from the quality expected by her, where expected quality is influenced by an aggregate of the seller's past ratings. We demonstrate our results empirically using raw data from eBay. In spite of the non-randomness of responses, unweighted rating aggregates appear to rather well reflect reported buyer experience as long as expectations are not extreme.

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

*We thank André Stenzel and in particular Christoph Wolf for constructive discussions and help in the conceptualization of our model, and participants at seminars and workshops. We are grateful to eBay for providing access to the data. Konrad Stahl gratefully acknowledges support by the Deutsche Forschungsgemeinschaft through CRC TR224 (project B04).

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

1 Introduction

Buyers generally agree that previous buyers' ratings are an important source of information about goods and sellers especially in online markets.[1] At the same time, however, many researchers question their informativeness. Tadelis (2016) gives an annotated review of some of the vast literature on rating bias. The quantification and evaluation of that bias necessitates a better conceptual and empirical understanding of when and why buyers do, or do not rate sellers; and whether rating aggregates offered by sellers and platforms adequately reflect the buyers' experiences.

Only a fraction of the buyers rates at all. The conditions are unclear under which buyers decide to do so, and whether the aggregate of such ratings is unbiased (in a sense to be defined below). The typical buyer can thus *a priori* not be sure that ratings provide reliable information about the quality of a transaction offered by the typical seller. We can only understand the meaning of these aggregates when we understand the underlying micro-structure, i.e., when and why buyers rate, and whether these ratings contribute to an unbiased aggregate. Moreover, only a better understanding allows the market designer to improve on rating systems as a reliable source of information to future buyers.

We develop a simple theoretical model of the typical buyer's rating decision, by which the buyer is likely to voice her opinion about a transaction the more, the larger her experience deviates from her expectation about its quality. That expectation is primarily derived from inspecting the seller's public performance score before she decided on the transaction. Technically the deviation is generated by the (absolute) difference between the buyer's Bayesian prior and her Bayesian posterior reflecting her experience from the transaction.

The model can explain a set of new stylized facts we present in this paper. These were obtained leveraging administrative transaction-level data from eBay. In particular, we investigate and explain the buyers' incentive to rate a transaction when their prior belief improves due to the experience gained from it, vs. when their prior belief deteriorates. This includes the complementary decision *not* to rate. We embed within our theory that with on average about 65 percent of transactions rated, the inclination to rate is higher than one would expect in view of the collective goods aspect involved in rating. We then use our theory to explain which factors change this inclination.

We establish the following two stylized facts:

- The likelihood that a buyer leaves a rating decreases in the number of transactions accumulated a seller. In particular, buyers become reluctant to leave a negative rating if a seller enjoys a large and very favorable rating record (which happens frequently).

- The likelihood that a buyer leaves a rating increases when the rating record is diminished

[1] That many buyers prefer to consult with ratings relative to seller recommendations may be due to the suspicion that, as predicted by the theory of signalling under asymmetric information, sellers tend to offer a sugar-coated view of their products and services.

(a rare event). The increase in the likelihood is smaller for a positive than for a negative rating, and smaller if the seller has a longer rating record.

That we can explain these diverse facts consistently within one theory supports our claim that the theory is helpful in explaining rating behavior. Our stylized facts indicate that feedback giving is non-random across buyers, and dependent on the seller's publicly revealed stock of feedback as observed by the buyer. In spite of the non-randomness of the observed rating process, however, *the information sharing process on which our explanation is based may be rather efficient*.

For the purposes of our analysis, we posit that unweighted aggregates of buyer ratings are efficient if they reflect all buyers' opinions in an unbiased manner. Unbiasedness understood this way is relative to the population of buyers. In a world of binary rating, it should just account for the difference in the probability to rate positively vs. negatively if the buyer's experience is positive vs. negative. If that difference would be equal, then the rating would be unbiased –provided unchanging seller behavior and unchanging distribution of the consumer population over time. Unbiasedness would be reflected in convergence to the true state as the number of ratings grows to infinity.[2]

The reason is that buyers simply discontinue to share information about their private experience when their experience corresponds with the rating aggregate publicly reported, e.g., by the platform –and by contrast, share information if it deviates from it. All else given, the observation that a strong rating record discourages rating, calls for an adjustment in market design towards a stronger weight on current vs. past ratings.

More in detail, in our model we look at the rating decision of one typical buyer who has engaged in a transaction with a typical seller, within an anonymous market such as offered by an e-commerce platform. The engagement was based on a prior belief that was formed by observing the seller's public rating record. A motivation to rate of especially the frequent visitor of the platform may be a private concern that purchasing from a platform that houses good sellers is time and effort conserving; or alternatively the motivation may be a public altruistic concern to inform the seller or future buyers. Following either concern, the buyer may, by truthfully rating the transaction, help future potential purchasers from this seller to take a more educated decision; she may also directly incentivize the seller to improve on his performance –or by contrast, help deterring the seller from the platform if deemed a bad type.

Rather than focusing on such motivations we assume a basic one to exist, and concentrate on the self-selection into sharing publicly the experience obtained from a transaction. By our baseline view the decision is motivated by the intensity of the buyer's personal learning expe-

[2] In a world of successive generations of short lived buyers that purchase from long lived sellers that in our view appropriately describes the situation, unbiasedness should rather be relative to the entire population of consumers. Future consumers can learn efficiently from previous buyers in particular, when sellers offer perfect substitutes that are common value at prices unchanging over time, and the buyers are drawn randomly w.r.t. to their taste from the consumer population. In that world the buyers' ratings should converge to a state in which any new generation of buyers identifies with certainty from the previous ratings the seller offering the best deal.

rience. This is summarized by the deviation between her belief formed prior to her transaction about its quality as performed by the seller, and her posterior belief updated by her personal experience from that transaction.

The buyer can keep this deviation private and not rate, or she can share it by leaving a rating. When the difference between posterior and prior belief is large, then the buyer has learned a lot from her transaction, which may benefit others when communicating this learning experience. We posit that she then will be more likely to share the extra information she has gained.

To exemplify the decision to rate as postulated, think first of a starting seller that has a small, if not low rating record. In such a situation an additional positive personal experience and rating generates additional information, and thus is likely to be implemented.

By contrast, think of an established seller with 1000 positive ratings in a row on which our buyer bases her decision to purchase, and an ensuing additional personal positive experience. As the resulting posterior will be almost equal to the prior, an additional positive rating would not contain much new information. By our model, the buyer is unlikely to rate. In turn, prices and the probability of selling are unlikely to react to the 1001st positive rating. By a similar argument, however, a negative rating based on a negative experience after 1000 positive ratings is unlikely as well.

Finally, think of a negative personal experience by our buyer after she has observed only a few positive ratings. This will lead to a substantial downwards adjustment of the buyer's posterior relative to her prior belief. By our conceptualization, that large difference is likely to induce a negative rating. It is empirically well established that this leaves a strong negative imprint on the seller's revenues.

With that simple sequential Bayesian updating exercise we can rationalize all the results we develop empirically, from the simplest one involving decreasing additional information generated from a sequence of identical signals, to intricate ones involving the increased informativeness of rating responses that result in a pattern corresponding to a herding effect –yet with a drastically different interpretation.

We conduct an empirical in-depth analysis to understand what drives these stylised facts. We use rich eBay administrative data at the level of transactions. Nevertheless, testing these empirical predictions and quantifying the effects is not straightforward, for two main reasons. First, as in most trading situations, the *objective* transaction quality for a given seller remains unobserved. Second, the transaction quality could change over time, for instance because a seller learns or because his incentives change. Third, buyers typically choose the seller they buy from. All of this could be influenced by the rating record. A buyer who is new on eBay could be reluctant to buy from a new seller with a small rating record if any, while a buyer experienced in transactions on eBay may feel more at ease doing so –e.g. by expecting a particularly good introductory deal. Conversely, the seller could shirk on the basis of a good rating record.

For our purposes, the ideal situation would be one in which buyers were randomly allocated to sellers, and in which transaction quality would not vary over time for the same seller. Then, the variation in the rating record that buyers see before leaving a rating would be exogenous, which would allow us to regress the rating on the rating record. Our empirical strategy is to control as much as possible for transaction quality and the endogenous matching between buyers and sellers, and to exploit quasi-random variation that is related to the timing of transactions and ratings.

For this, we create a balanced panel of starting sellers whom we follow for one year. The panel structure allows us to control for unobserved differences in seller quality that are time-invariant, and the balanced sample allows us to eliminate the dynamic selection problem in which gradual seller exits cause bias in the estimated effect. We first relate the likelihood to receive a rating to the number of transactions a seller was involved in by that time. The underlying idea is that the more transactions a seller has been involved in, the more ratings she has already received. This should have an effect on the inclination of other buyers to leave an additional rating. We show that this inclination changes from around 70% to 66% from the first to the 86th transactions that a seller has conducted.

This decrease, however, could be partly driven by differences across sellers, time, products and buyer characteristics. To address this concern, we analyze the same correlation within a regression with successive controls, including seller, calendar months, and product fixed effects, and buyer experience as well as her historical tendency to leave any feedback. Controlling for all these factors in a linear probability model, including robustness checks by using eBay's Detailed Seller Ratings (DSR) indices, we find consistently that later transactions are less likely to be rated: the probability of receiving a feedback for the 10th transaction is 0.325 percentage points lower than the first transaction. Based on this estimate, as a starting seller moves from 1 to 86 transactions, her probability of receiving a feedback decreases by 2.8 percentage points (0.325*86), compared to the 4 percentage points decrease in the raw correlation.

We then test a second prediction from our model, namely that a negative feedback increases the probability that any feedback is left, and this increase is smaller for a positive than for a negative feedback. We face three challenges when conducting this test. First, a correlation between a negative feedback and subsequent feedback may come from changes in seller behavior in response to a negative feedback, as argued in Cabral and Hortaçsu (2010). Second, even without changes in seller behavior, subsequent negative feedback can be due to a serially-correlated negative quality shock on the seller side. Third, seller behavior can change over time.

To eliminate the first concern, we use the fact that buyers do not immediately receive the item and hence cannot immediately rate it after their purchase. Hence by the time t a negative feedback is left, more transactions have taken place, with the seller's effort decision taken before t and hence independent of the negative feedback. However, the buyer may have observed the negative feedback when she receives her item and rates. By comparing rating outcomes of these instances to those from transactions that both took place and were rated before t, we are able to hold constant the seller's effort.

To mitigate the second concern, we break the linkage between a negative feedback and seller quality by exploiting instances where negative feedback were mistakenly given by buyers and are retracted later. As before, let such a mistaken negative feedback be given at t. Our results from this exercise confirm our model's prediction: a negative feedback leads to a a one percent higher probability of subsequent negative rather than positive feedback, which is a large effect given the baseline likelihood of 1.5%. This effect is driven primarily by inexperienced buyers, and on sellers with a smaller number of previous positive feedback. Both effects are consistent with our model of rating behavior based on Bayesian updating.

To mitigate the third concern of changes in seller behavior over time, we further control for seller-by-transaction-date fixed effects. In doing so, we compare transactions that took place on the same day before the arrival of the mistaken negative feedback, but some receive their feedback before this arrival while others receive their feedback afterwards. The estimate from this analysis is statistically indistinguishable from the estimate from the main specification. This suggests that seller learning does not drive the herding effect to the extent that the amount of learning in one day is small. Lastly, we perform placebo analyses, in which we replace the arrival of mistaken negative feedback with the arrival of negative consumer reports that are not observable (or as observable) to buyers, namely buyer claims and DSRs. We do not expect any herding effect because buyers should not update their prior after these unobservable negative reports. This is indeed what we find in the data.

Related Literature We do not review the substantial literature on the impact of rating on online markets. Bajari and Hortaçsu (2004), Einav et al. (2016), and Tadelis (2016) provide good surveys. We rather concentrate on theoretical and empirical contributions we directly relate to and build on, with focus on the buyer's rating decision.

As to theoretical contributions, Acemoglu et al. (2019), Hoerner and Lambert (2020), Stenzel et al. (2020) or Vellodi (2020) consider infinite horizon models where every buyer rates, and rating is based on realized utility. We consider a static framework and focus on the decision of whether to rate at all based on juxtaposing expectation and realization.

Our theoretical approach is akin to Hart and Moore (2008) who argue that *contracts may provide a reference point for parties' feelings of entitlement. A party's ex post performance depends on whether he gets what he is entitled to relative to outcomes permitted by the contract.* In our approach the buyer decides to engage in a transaction on the basis of a seller's rating index, that reflects the seller's past performance. The transaction can be considered an implicit contract based on this index, and therefore the index can be considered the typical buyer's reference point. The buyer stays silent when her experience is close to the reference point, and expresses via rating only her strong satisfaction or dissatisfaction relative to that.

Alternative theoretical explanations for the rating pattern we model and observe are considered by Brandes et al. (2019). They distinguish between a base rate explanation according to which the cause of extreme reviews would be the self-selection of buyers, and a utility-based explanation, by which consumers receive greater utility from conveying extreme opinions. Our explanation could classify under the latter. By their explanation, buyers with more extreme experiences exit the pool of potential appraisers with lower probability than buyers with less extreme experiences, and thus have more time to write about them before they become inactive.[3] By our contrasting, presumably more natural explanation, buyers do not rate because their experience corresponds more or less to that conveyed by the actual rating index, and are tempted to rate only if not.

There is a long tradition in the marketing and information systems literature reported, e.g., in Ho et al. (2017) that conceptually involves rating as based on some disconfirmation index. In none of the explanations that index is predicated on the individual's prior formed by her perception of platform and seller quality, updated by her private experience from the transaction. This has consequences on the information content to be expected from that index, an aspect that is not discussed in this string of the literature.

We contribute to the social learning literature (see Mobius and Rosenblat, 2014, for a survey). By the concept of social learning, an agent learns from transactions previously conducted by other agents. In our model, that agent learns from her personal transaction and selectively decides to convey her experience to others. The social learning literature focuses on exogenously specified signals. We contribute to that literature by endogenizing the the decision to provide such a signal. While one can easily rationalize within a standard social learning framework our second stylized fact that a negative rating becomes more likely after a first one, it would be difficult to rationalize our first empirical result without endogenizing the learning signal.

Finally Hopenhayn and Saeedi (2019) or Saeedi and Shourideh (2020) look at a rating system that maximizes information content. We do not optimize the rating system, but in our policy recommendations suggest elements fostering the provision of unbiased information.

As to the empirical literature, we contribute to the rich evidence surveyed by Schoenmüller et al. (2019) or Brandes et al. (2019), by which ratings are based on extreme experiences –yet we explain them consistently with our model.

By our conception, buyers rate truthfully if at all. Our sample does not allow us to identify whether amongst the raters or the non-raters there is individual bias towards positive or negative experiences. To some extent, we can dismiss with rating bias as discussed by Mayzlin et al. (2014) and Luca and Zervas (2016). They use data from platforms in which ratings can be given without even engaging in any transaction. By contrast, buyers' ratings on eBay cannot be given without the transaction that is reviewed, which, by substantially increasing the cost of fake rating, to some extent shields the rating system from fake reviews. While we cannot exclude fake ratings from our analysis, we suggest that at eBay, fake ratings that are rather costly in the present case, are initialized, if at all, by starting sellers in order to become operative on the platform.

Nosko and Tadelis (2015) take issue with the disproportionate share of positive ratings and,

[3] They support that view by an experiment involving reminders to review. In view of the selectivity of responses to such reminders, their explanation of rating bias needs to be taken with caution.

based on claims of buyers not reflected by a negative rating, interpret buyer silence as a signal of dissatisfaction with the transaction preceding exit. In this paper, we argue that the buyer may simply not rate once the transaction experience does not deviate much positively or negatively from her expectations. At the same time, as we will discuss below, a large number of positive ratings accumulated by the seller may prevent the buyer from voicing a negative experience, which would induce a negative bias in the reported ratings. Indeed, that may balance a positive –call it a behavioral– bias indicated by Nosko and Tadelis (2015) when presenting evidence on negative buyer claims reported to the platform, but not expressed publicly. Overall, it could be that the bias in rating aggregates is not too big.

Another bias that has been discussed in the literature is reciprocation bias generated by buyers' fear of retaliation, see for instance Resnick and Zeckhauser (2002), Klein et al. (2009), Dellarocas and Wood (2008), or Bolton et al. (2013). Klein et al. (2016) develop on the positive consequences of eBay's dismissal of the negative rating option by the seller.[4] Within a field experiment on Airbnb, Fradkin et al. (2020) show within an experiment that when excluding reciprocation by revealing ratings simultaneously, negative ratings increased, but by only a marginal 0.25 percent. As in Brandes et al. (2019) the experiment involved a reminder to buyers. This again gives rise to the question of buyer self-selection into a response.

A paper closely related to ours is Cabral and Hortacsu (2004). They also use mistaken negative feedback and find that the propensity of subsequent negative feedback increases afterwards. However, they attribute this to changes in seller behavior after observing a negative feedback rather than to changes in the rating buyer's behavior. Our enhanced empirical strategy allows us to separate changes in raters' behavior from changes in seller behavior. Furthermore the use of internal data from eBay allows us to observe seller entry and to track the life cycle of their ratings. This brings us closer to testing our simple but general Bayesian decision framework.

Filippas et al. (2018) argue that ratings have been inflated over time because raters feel pressured to leave above-average rating. This further pushes raters to rate positively because they do not want to harm the rated seller. Our model explains why ratings becomes more positive over time, but with a non-behavioral explanation.

Finally, by looking at a rating system that maximizes seller effort, Hoerner and Lambert (2020) arrive at a conclusion to one we can also derive from our analysis, namely that rating indices should focus on recent ratings. Their rationale is that this prevents seller shirking. While this applies to our situation as well, we show that an additional benefit to basing aggregate ratings more on recent ratings is that buyers are incentivized by a small rating record to leave a rating.

[4]In their sample scraped from the net, they were unable to fully identify seller exit in response to that dismissal. Accounting for that, Hui et al. (2018) show that seller exit and a shift in market share from low-quality to high-quality sellers also contributed to the large improvement of ratings.

2 Model

Consider an anonymous market offered by a platform such as eBay. We focus on one typical buyer's rating decision that is based on a transaction with a typical seller. When rating, the buyer is assumed to truthfully summarize her experience.

The transaction may be of quality $q_i, i \in \{h, l\}$, reflecting high vs. low quality. The buyer does not know that quality. Before the transaction she forms a prior belief about its quality, learns about it when performing the transaction, and then decides whether or not to rate it. We posit that the likelihood to rate depends on the difference between prior and posterior belief about the transaction quality.[5]

More specifically, without seller specific information, the buyer is endowed by an initial belief $\lambda \in (0,1)$ that a transaction on the platform is of good quality. Before engaging in a transaction with a specific seller, the only seller-specific signal she receives is the seller's reputation score. That score is condensed in an index $y(p,n) \in R_+$, where p denotes the number of positive and n the number of negative ratings given by earlier anonymous buyers, that are assumed to be binary.[6]

We suppose that our buyer specifies a probability $\sigma_h(y) = \Pr(y|q = q_h)$ that our seller's reputation score is y given that the seller offers a high-quality transaction, and a probability $\sigma_l(y) = \Pr(y|q = q_l)$ that the reputation score is y given that he offers a low-quality transaction. Based on Bayes' rule, our buyer then forms a prior belief $\Pr(q = q_h|y) \equiv \mu(y)$ that the transaction offered by our seller will be of good quality, and conducts her transaction with our seller if that belief is high enough. This belief is given by

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

Note that $\mu(\lambda, y)$ varies positively with λ, so that a consumer patronizing the platform is more inclined to interact with our specific seller if her initial belief is high that *any* transaction on the platform is of good quality.[7] To simplify notation, we disregard the dependence on λ until needed. Observe, however, that λ may reflect our buyer's experience with trading on the platform, so by self-selection, more experienced buyers may be characterized by a higher λ, and thus, a higher belief.

An additional desired property of $\mu(\lambda, y)$ is that it is increasing in y. Thus, more positive ratings should increase that belief, while more negative ratings should decrease it. $\mu'(y) > 0$ is

[5] We can only observe the buyer's (or seller's) opinion she forms after the transaction and this only when rating it, and never its objective quality. Her opinion may be influenced by the match value she enjoys relative to her prior, that was formed on the basis of the seller's rating as well as his description of the product characteristics, and finally the seller's effort when executing the transaction. These are all common value components under the seller's control.

[6] Good examples are the *percentage positives* $\frac{p}{p+n}$, or the *feedback score* $p - n$ employed by eBay.

[7] In the sense of Bayesian updating, this is already a posterior belief by which the buyer updates her initial belief λ on the basis of the seller-specific rating information.

positive if $\frac{\sigma_h'(y)}{\sigma_h(y)} - \frac{\sigma_l'(y)}{\sigma_l(y)} > 0$. This is equivalent to $\frac{d}{dy}\frac{\sigma_h(y)}{\sigma_l(y)} > 0$ which corresponds to the Monotone Likelihood Ratio Property (MLRP) commonly assumed in modelling Bayesian updating. By this assumption, our buyer considers a good reputation score, that must be based on good ratings, as good news about the underlying quality of a transaction conducted with the respective seller. As $\frac{\sigma_h(y)}{\sigma_l(y)}$ is increasing, the posterior belief is invertible. Thus there is a unique cutoff $\tilde{y}$ such that for the belief to exceed a threshold, the reputation score observed by our buyer has to exceed $\tilde{y}$.[8]

We now suppose that our buyer has conducted a transaction with our seller, and turn to her decision of whether or not to rate it. Our buyer makes public her experience with the transaction if her benefit from doing so exceeds her cost. Her benefit b is assumed to strictly increase in the absolute difference d between prior and posterior belief, i.e. in the degree of positive surprise or disappointment she has experienced in her personal transaction relative to its quality expected *a priori*: she is inclined to rate positively when the posterior increases, and negatively when the posterior decreases relative to her prior belief. She compares the benefit to an idiosyncratic rating cost c drawn randomly from a distribution normalized to the uniform distribution on the unit interval. The buyer decides to rate if the net benefit from doing so, $B(d,\tilde{c}) \equiv b(d) - \tilde{c} > 0$ where $\tilde{c} \equiv b(0) > 0$.

We rationalize these assumptions as follows. When the absolute difference between prior and posterior belief is large, then the buyer has learned a lot from her transaction. In turn, she may expect that also others benefit from her making public this learning experience. The more information there is to share –in the sense that the posterior differs from the prior belief– the higher the likelihood that she will leave a rating.[9] We incorporate the common observation that a sizeable proportion of the buyers rates anyway, by specifying a proportion $\tilde{c}$ of buyers willing to incur the (low) cost of rating no matter the prior-posterior difference b.

To formalize the differences in prior and posterior beliefs, let our buyer's experience with her transaction be summarized in a private signal $s \in \{g,b\}$, where $s = g$ if the transaction is perceived to be of high, and $s = b$ if perceived of low quality. The buyer perceives that signal as informative in the sense that it clarifies the quality of the experienced transaction. Hence she perceives it as having precision higher than $\frac{1}{2}$. Let $Pr(s = g|q = q_h) \equiv \rho^g$ and $Pr(s = b|q = q_l) \equiv \rho^b$. On this basis, the buyer forms a *posterior belief* that the transaction was of high quality

[8]Note that y and λ are substitutes in the formation of a given level of μ.

[9]The temptation to share surprising positive, or negative experiences with others may be enhanced by other public concerns. For instance, a buyer frequently purchasing on the platform may have platform quality concerns, because the presence of only good sellers facilitates the conduct of business on the platform. Or she may be altruistic and care about how much others can benefit from her ratings. In both cases, the more information there is to share, the higher the likelihood that she will leave a rating. Such collective concerns, that could arise randomly across buyers, could easily be embedded within our utility index.

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

and

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

In order to simplify the analysis, we assume in what follows that good and bad signals are perceived with the same precision, so that $\rho^g = \rho^b \equiv \rho \in (\frac{1}{2}, 1)$. We comment on generalizing this assumption at the end of this section. In the ensuing Proposition, we study how the valuations of the seller's performance score prior to the transaction influence the buyer's decision to rate her private positive vs. a negative experience.

Proposition 1. *Let the buyer's prior belief be formed on the seller's public rating record y at the time of her transaction. Consider an ex ante increase in y generated by an increase in the number of positive ratings.*

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

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

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

Proof

(i) We need to show that given the transaction signal $s \in \{g, b\}$ is $s = g$ with precision $\rho = \Pr(s = g | q = q_h)$, the difference between the buyer's posterior and her prior belief

$$\mu^g(y) - \mu(y) = \frac{\mu(y)\rho}{\mu(y)\rho + (1-\mu(y))(1-\rho)} - \mu(y) \tag{4}$$

first increases and then decreases in y. Hence, we are interested in conditions under which

$$\frac{d}{dy}(\mu^g(y) - \mu(y)) = \{-\mu'(y)(2\rho - 1)\} \frac{(1-\rho)(2\mu(y) - 1) + (2\rho - 1)[\mu(y)]^2}{[\mu(y)\rho + (1-\mu(y))(1-\rho)]^2} \tag{5}$$

is positive vs. negative. The first term (in curled brackets) is negative, since by assuming MLRP $\mu'(y)$ is positive, and $\rho > \frac{1}{2}$ so $2\rho > 1$. Furthermore, the denominator of the second term is strictly positive. Hence (5) is positive (negative) if the numerator of the second term is negative (positive).

The prior belief $\mu(y)$ is bounded at $[0,1]$. At $\mu(y)=0$ the numerator is $-(1-\rho)<0$. At $\mu(y)=1$ the numerator is $(1-\rho)(2-1)+(2\rho-1)=\rho>0$. As $1-\rho>0$ and $2\rho>1$, that numerator is a continuous and strictly increasing (convex) function of $\mu(y)$. Hence there exists a unique $\hat{y}$ at which the numerator of the fraction and with it, the derivative of the difference is zero at $\hat{y}$. Furthermore, the derivative of the difference is positive if $y<\hat{y}$, and negative if $y>\hat{y}$.

(ii) Similarly, we need to show that given the transaction signal $s\in\{g,b\}$ is $s=b$ with precision $\rho=\Pr(s=b|q=q_l)$, the difference between the buyer's prior and her posterior belief

$$\mu(y)-\mu^b(y)=\mu(y)-\frac{\mu(y)(1-\rho)}{\mu(y)(1-\rho)+(1-\mu(y))\rho} \tag{6}$$

increases vs. decreases in y. Hence, we are interested in conditions under which

$$\frac{d}{dy}(\mu(y)-\mu^b(y))=\{\mu'(y)(2\rho-1)\}\frac{\rho-2\rho\mu(y)+(2\rho-1)[\mu(y)]^2}{[\rho-(2\rho-1)\mu(y)]^2} \tag{7}$$

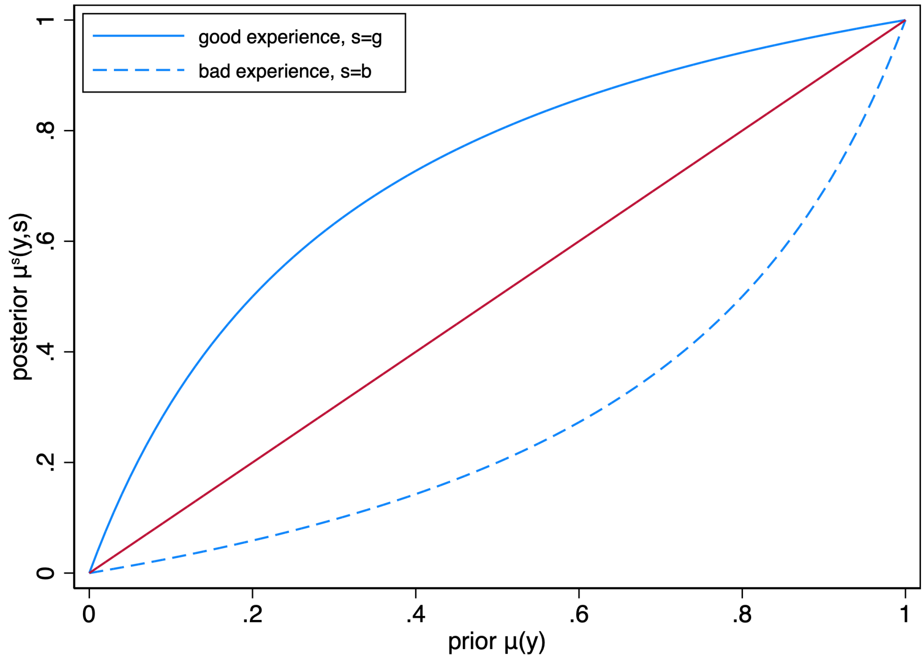
is positive vs. negative. The first term (in curled brackets) is now positive, and as heretofore the denominator of the second term is strictly positive. Hence (6) is positive (negative) if the numerator of the second term shows the same sign.

At $\mu(y)=0$ the numerator is $\rho>0$; at $\mu(y)=1$ the numerator is $\rho-1<0$. The numerator is strictly decreasing in the interval $\mu(y)\in[0,1]$. Hence there exists a unique $\check{y}$ at which the numerator of the fraction and with it, the derivative of the difference is zero at $\check{y}$. Furthermore, the derivative of the difference is positive if $y<\check{y}$, and negative if $y>\check{y}$.

(iii) We can solve for the beliefs at the critical cutoff points $\hat{y}$ and $\check{y}$. Excluding the roots involving $\mu\notin[0,1]$, we obtain $\mu(\hat{y})=\frac{\rho-1+\sqrt{\rho(1-\rho)}}{2\rho-1}$ and $\mu(\check{y})=\frac{\rho-\sqrt{\rho(1-\rho)}}{2\rho-1}$. Indeed, for the signal to be informative, i.e. $\rho\in(\frac{1}{2},1)$, $\mu(\hat{y})$ is an increasing function of ρ since $\frac{d\mu(\hat{y})}{d\rho}=\frac{1}{4\rho(1-\rho)-2\sqrt{\rho(1-\rho)}}$ with $\mu(\hat{y})\in(0,\frac{1}{2})$, and $\mu(\check{y})$ is a decreasing function since $\frac{d\mu(\check{y})}{d\rho}=\frac{1}{4\rho(1-\rho)+2\sqrt{\rho(1-\rho)}}$ with $\mu(\check{y})\in(\frac{1}{2},1)$.

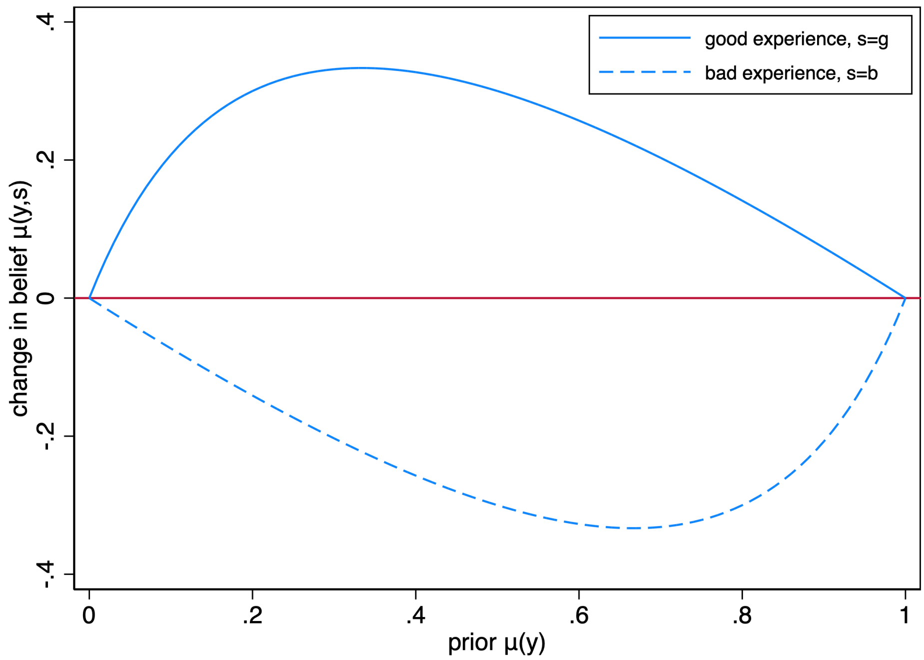
□

Towards an interpretation, recall our central hypothesis 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. Thus the probability to rate increases with the absolute difference between prior and posterior beliefs. From the proposition we infer that 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 doesn't shift the belief that much anymore to induce a rating.



Notes: Posterior belief $\mu^s(y,s)$ for good experience ($s = g$) and bad experience ($s = b$) as function of prior belief $\mu(y)$. Calculated using (2) and (3) with $\rho = 0.8$.

Figure 1: Posterior belief



Notes: Prior minus posterior belief $\mu(y,s) - \mu(y)$ after good ($s = g$) and bad experience ($s = b$) as function of prior belief.

Figure 2: Prior minus posterior belief

Table 1: Bayesian updating example

	good experience		bad experience	
prior	posterior	updating	posterior	updating
0.00	0.00	0.00	0.00	0.00
0.10	0.31	0.21	0.03	-0.07
0.20	0.50	0.30	0.06	-0.14
0.30	0.63	0.33	0.10	-0.20
0.40	0.73	0.33	0.14	-0.26
0.50	0.80	0.30	0.20	-0.30
0.60	0.86	0.26	0.27	-0.33
0.70	0.90	0.20	0.37	-0.33
0.80	0.94	0.14	0.50	-0.30
0.90	0.97	0.07	0.69	-0.21
1.00	1.00	0.00	1.00	0.00

Notes: This table shows posterior beliefs $\mu^s(y,s)$ and the change in beliefs $\mu(y,s)$ for good experiences ($s=g$) and bad experiences ($s=b$). Calculated using (2) and (3) with $\rho=0.8$.

This holds for all transactions conducted by our buyer, no matter her experience. However, the difference between positive and negative experiences is one of –important– detail contained in part (iii) of the proposition: The cutoff between when the rating probabilities increase and decrease is (much) lower for positive experiences than for negative ones. Rather intuitively, the probability that our buyer rates negatively continues to increase with further positive ratings when that of rating positively already decreases.

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.[10] Relevant, however, is the construction of the rating index y. To take the example of eBay, consider $y(PP,FS)$, where $PP \equiv \frac{p}{p+n}$ is eBay's *percentage positives*, and $FS \equiv p-n$ is eBay's *feedback score*. Let y strictly increase in both arguments. PP obviously starts at y close to 1 if the first rating received by our seller is positive.[11] Therefore the empirical observations we work with in the ensuing sections are likely to start with a high y. This motivates the following

Corollary 1. *Consider, as before, an ex ante increase in the number of positive ratings. Let $y > \check{y}$. Then the likelihood that a buyer leaves a rating decreases in y, and shrinks to $\tilde{c}$ as $\mu(y) \to 1$.*

The proof follows directly from the Proposition. Note that the corollary directly reflects the context in which we will analyze our data, which is based on all transactions rather than only the rated ones.

[10]The consequences can be read off directly from the proposition, as they are just the mirror image of the results provided therein.

[11]This is reinforced by anecdotal evidence, by which potential buyers observe PP only.

The mathematically trivial result that the likelihood to rate shrinks $\tilde{c}$ if the buyer's prior converges to 1 strikes us as important when thinking about rating bias. It says that the probability that the buyer rate shrinks if her prior belief is very close to unity. This implies in particular that after many positive ratings, the buyer is less likely to rate negatively after a negative experience. The result can also be given a very intuitive behavioral interpretation: When I observe that a seller has accumulated a very good rating record, I am hesitant to perturb this record (or trusting in my own judgement, e.g. as a young buyer).

2.1 Interim ratings and their effect

Rating decisions are almost always taken with a delay after the purchase. In the interim period between the buyer's transaction and her decision to rate, new ratings may arrive with high probability. Let h be the number of positive and l the number of negative ratings, respectively. When they have entered the seller's performance score, i.e. when $y(p+h,n+l)$, the buyer is unable to identify their arrival time, so the ratings become indistinguishable to earlier ratings.

The question arises as to whether the buyer ignores, and if not, how she incorporates these interim signals in her decision to rate. In our Proposition the buyer (implicitly) ignores them. We now discuss the consequences of these interim ratings on her own decision to rate, and show that the consideration of the interim ratings always has a positive effect on the buyer's rating intentions. Consider the following –partially alternative– sequences of observations and decisions taken by the buyer:

1. Buyer performs the transaction based on the prior belief $\mu_t \equiv \mu(y_t)$
2. Buyer forms a posterior belief $\mu_{t+1} \equiv \mu(y_t, s)$ incorporating her personal experience
3. Buyer rates if $b(d_{t,t+1}) \equiv b(|\mu_{t+1} - \mu_t|) \geq \tilde{c}$
4. Buyer observes h additional positives and l additional negatives, resulting in $y_{t+1} \equiv y(p+h, n+l)$
5. Buyer forms a posterior belief $\mu'_{t+1} \equiv \mu(y_{t+1})$ *not* incorporating her personal experience
6. Buyer forms a further posterior belief $\mu''_{t+1} \equiv \mu(y_{t+1}, s)$ incorporating in addition her personal experience
7. Buyer rates if $b(d_{t,t''+1}) \equiv b(|\mu''_{t+1} - \mu_t|) \geq \tilde{c}$
   Alternatively,
8. Buyer rates if $b(d_{t'+1,t''+1}) \equiv b(|\mu''_{t+1} - \mu'_{t+1}|) \geq \tilde{c}$

Decision 3 is used in Proposition (1). Decisions 7 and 8 differ by how the buyer incorporates the interim signals in the evaluation of her transaction. In decision 7, she fully accounts for them and rates based on the difference between that and her prior on which the transaction was based.

In decision 8, she purely accounts for the novelty of her personal experience and rates based on the difference between that and the seller's rating record after the arrival of the interim ratings.

The likelihood that our buyer will rate under these alternatives depends first on whether she ignores the interim ratings or takes them into account, and if so, whether she updates her prior belief to including them or makes them part of her personal experience; and second on whether her experience and the change in the rating index co-move or move against each other in her update. In view of our theory, all alternatives can be interpreted in terms of whether these alternatives increase or decrease the absolute difference between prior and posterior.[12]

The differences between prior and posterior beliefs and with it, the buyer's rating decisions are again just corollaries to the results presented in our Proposition. In previewing our empirical analysis, we compare rating behavior when the prior belief is high that the buyer performs well. If otherwise, an anonymous market could not be operative without exogenous support.

We show first that as long as the prior belief is sufficiently large, the probability to rate positively is less affected than that of rating negatively when the rating index suffers a negative shock. At the outset, let $\bar{y}$ such that $\mu(y) \equiv 1$ for all $y \geq \bar{y}$, and $f_1(y) \equiv \mu(y, s = g) - \mu(y)$, $f_2(y) \equiv \mu(y) - \mu(y, s = b)$, and $z(y) \equiv f_1(y) - f_2(y)$.

Proposition 2. *There exists a unique $\tilde{y} \in (\check{y}, \bar{y})$ such that $\frac{d}{dy}z(y) > 0$ for all $y \in (\tilde{y}, \bar{y})$.*

Proof Recall that $\check{y} = \underset{y}{argmax}\{f_1(\mu(y))\} = \underset{y}{argmax}\{\mu(y) - \frac{\mu(y)(1-\rho)}{\mu(y)(1-\rho)+(1-\mu(y))\rho}\}$. By Proposition 1, $f_1(\mu(y))$ is strictly decreasing for $y > \hat{y}$, and thus in particular at $y = \check{y} > \hat{y}$. Therefore, since $\frac{d}{dy}f_2(\mu(\check{y})) = 0$, we have $\frac{d}{dy}f_1(\mu(\check{y})) < \frac{d}{dy}f_2(\mu(\check{y}))$, which implies that $\frac{d}{dy}z(\mu(\check{y})) < 0$.

Differentiating $z(\mu(\check{y}))$,

$$\begin{aligned}
&sgn\{\frac{d}{dy}z(\mu(\bar{y}))\} = sgn\{\frac{d}{d\mu}z(\mu(\bar{y}))\} = \\
&= sgn\{\frac{\rho[\mu\rho + (1-\mu)(1-\rho)] - \mu\rho(\rho-(1-\rho))}{[\mu\rho + (1-\mu)(1-\rho)]^2} \\
&+ \frac{(1-\rho)[\mu(1-\rho) + (1-\mu)\rho] - \mu(1-\rho)((1-\rho)-\rho)}{[\mu(1-\rho) + (1-\mu)\rho]^2} - 2\}.
\end{aligned}$$

This simplifies to

$$\frac{\rho(1-\rho)}{[\mu\rho + (1-\mu)(1-\rho)]^2} + \frac{\rho(1-\rho)}{[\mu(1-\rho) + (1-\mu)\rho]^2} - 2.$$

Letting $\mu \to 1$ and simplifying further, we obtain $\frac{1}{4} > \rho(1-\rho)$, which holds since $\rho > \frac{1}{2}$. Hence $\frac{d}{dy}z(\mu(\bar{y})) > 0$.

[12] In addition, we could consider that our buyer correctly interprets the interim ratings as more recent than the ratings on which she has based her decision to transact, which she cannot date. Then her update would react more strongly to these interim ratings than to changes in the rating before purchase. The results we use in our empirical analysis remain unchanged with this refinement.

A tedious calculation reveals that for $y \in (\check{y}, \bar{y})$,

$$\frac{d^2 z}{dy^2}(\mu(y))) =$$
$$= 2(1-\rho)\rho(2\rho-1)\left(\frac{1}{(\mu\rho+(1-\mu)(1-\rho))^3}+\frac{1}{(\mu(1^-\rho)+(1-\mu)\rho)^3}\right) > 0.$$

Thus $z(\mu(y))$ is strictly convex, has a unique minimum at some $y \equiv \tilde{y} \in (\check{y}, \bar{y})$, and strictly increases for all $y \in (\check{y}, \bar{y})$.

□

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 $y_{t+1} < y(t)$. We finally show the effect of an interim negative signal on our buyer's decision to give a second negative based on a negative personal experience, and summarize this in

Corollary 2. *Let $y_t > \tilde{y}$. Let the buyer suffer a negative experience. Suppose now that interim signals lead y to decrease from y_t to $y_{t+1} > \check{y}$. Then the likelihood that the buyer rates increases when she incorporates the interim decrease in the rating score in her decision relative to not incorporating it.*

Proof Consider decision 7. We need to show that the difference $d_{t,t''+1}$ between the prior on which the transaction decision was based and the posterior in which the interim signals are incorporated is larger than the difference between that prior and the posterior ignoring the interim signals. But

$$\mu(y_t) - \mu(y_{t+1}, s) > \mu(y_t) - \mu(y_t, s)$$

because $y_t > y_{t+1}$ implies $\mu(y_t, \cdot) > \mu(y_{t+1}, \cdot)$.

Consider now decision 8. Then we need to show that the difference $d_{t,t''+1}$ between the prior formed including the interim signals and the posterior including in addition the signal generated from her personal experience is larger than, again, the difference between the prior on which the transaction decision was based and the posterior ignoring the interim signals. But

$$\mu(y_{t+1}) - \mu(y_{t+1}, s) > \mu(y_t) - \mu(y_t, s)$$

because the difference is a decreasing function of y, and $y_{t+1} < y_t$.

□

Before we move on to the empirical analysis in the ensuing section, we should emphasize that all our results, and in particular those emphasizing the high weight given to negative relative to positive experiences in rating, take even more force when we abstain from assuming $\rho^g = \rho^b$, i.e., the precision our buyer extracts from a good vs. a bad experience conditional on the state that the seller is good vs. bad, and consider instead $\rho^b < \rho^g$. This can be easily rationalized by invoking variations in the sellers' incentives not modelled here for simplicity: A good seller has always the incentive to improve on the buyer's belief that he is good, and thus wants to improve on the impression ρ^b; a bad seller has always the incentive to make the buyer believe that he is a good one, and therefore wants to diminish the informativeness of the signal embedded in ρ^b.

3 Data

3.1 General empirical approach

The goal of our analysis is to establish two stylized facts that can be explained by the model in Section 2: (i) The likelihood that a buyer leaves a rating decreases in the number of transactions accumulated a seller; (ii) the likelihood that a buyer leaves a rating increases when the rating record is diminished (a rare event). The increase in the likelihood is smaller for a positive than for a negative rating, and smaller if the seller has a longer rating record.

Our analysis uses rich eBay data at the level of transactions. Nevertheless, testing these empirical predictions and quantifying the effects is not straightforward. There are two main reasons for this. First, as in any offline market, the transaction quality for a given seller is unobserved. It could change over time, for instance because a seller learns or because the incentives to provide high transaction quality change over time. Second, buyers choose among the sellers the one they buy from. This is likely to be influenced by the rating record. For instance, it could be that buyers who are new on eBay are reluctant to buy from new sellers, while buyers who have already had many 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.

The idea of our empirical strategy 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.

3.2 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).[13] From this we construct two subsamples. The first

[13] 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).[14] 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.

3.3 Variable definitions and summary statistics

With the considerations in Section 3.1 in mind we construct a number of variables. We report summary statistics for all three samples in Table 2.

There are three panels. Panel A contains seller characteristics. Here, we first create one observation per seller and then report the average across sellers.[15] Sales volume in the first year for all 141,138 sellers who had their first transaction in March 2011 (sample 0) is $1,218. They have 24 transactions on average and sell products in 5 so-called leaf categories in the first year. 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.[16] The eBay percentage positive is one of the two numbers that is displayed next to a seller name on the eBay platform.[17] It is calculated as the number of positive feedbacks the seller as received, relative to the number of feedbacks that were either positive or negative. This means that neutral feedbacks are discarded. Here we report the percentage positive at the end of the first year.

Sample 86 and 338 contain the top 5% and top 1% of the sellers in terms of the number of transactions within the first year, respectively. Therefore, it is not surprising that these sellers have a higher sales volume and more transactions in more unique leafs in the first year. It is also not surprising that the percentage positive is lower for all sellers who started in March 2011, as compared to the sellers in sample 86 and sample 338. The reason for this is that some of the sellers who started in March 2011 will stop being active on eBay. Those sellers will be a negative selection of sellers in the sense that will more likely receive negative feedbacks (see for instance Cabral and Hortaçsu, 2006, 2010). For all sellers who started in March 2011 the percentage positive feedbacks is 91.7%. This is considerably lower than the percentage positive feedback for a randomly drawn seller on eBay. In sample 86, the percentage positive is about 98% and in sample 338 it is about 99%.

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

[15] 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.

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

[17] 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 2: 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
any feedback	0.622	0.670	0.655
share feedbacks neutral or negative	0.026	0.016	0.010
share transactions with neutral or negative feedback	0.016	0.011	0.007
share transactions with neutral feedback	0.006	0.005	0.004
share transactions with negative feedback	0.011	0.006	0.003
share transactions with low DSR	0.022	0.019	0.013
share 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 obervation 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.

Panel B contains information on buyers who bought from these sellers. 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 distinct buyers who bought from the 141,138 sellers who had their first transaction in March 2011.

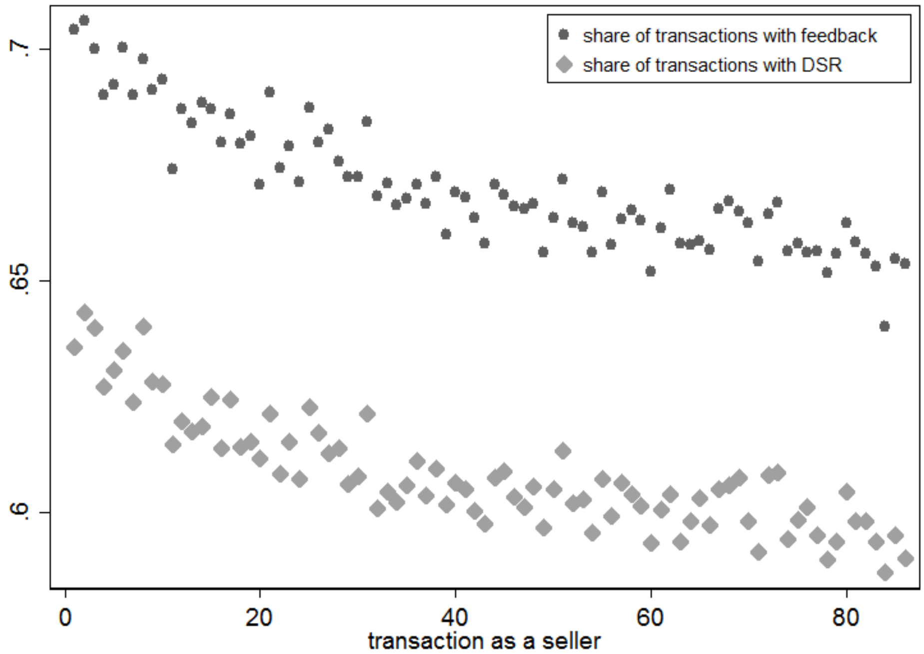
We have calculated two measures of buyer experience. The first measure is the number of transactions buyers had in the year before March 2011 (which is the month in which our sellers had their first listing), or by March 2011 if they started after March 2010. For these transactions, we can also calculate 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. Our main measure of buyer experience that we will also use below is whether they have registered before March 2009.[18] This is the case for 71.3% of the buyers in sample.

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 those sellers that are rated worse in sample 0 is 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 sample 86 and sample 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. Feedback is important in anonymous markets if buyers and sellers do not interact repeatedly. The first statistic that is shown confirms that this is the case for 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 row reports the share of transactions with neutral or negative feedback, as opposed to the share of all feedback.

The next measure is the share of the transactions with a 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 ratings. Panel C shows that the

[18] In general results are very similar when we use the other measure. See Table A.2 and A.3.



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

Figure 3: Probability to receive feedback

share of transactions with a low DSR is similar to the share of the transactions with a neutral or negative feedback.

4 Inclination to Share an Experience

4.1 Baseline results

The basis of our model in Section 2 is that the likelihood to give a feedback increases in the absolute difference between the buyer's prior and posterior beliefs. It is formally stated in Proposition 1 that this difference decreases in the number of ratings. 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 follow sellers over time.[19] In Figure 3, we plot the share of transactions with feedback and DSR against the number of transactions performed by the sellers. It shows that both the share of transactions with feedback and the share of transactions with a DSR are lower for later transactions. The underlying idea here is that for our sample of sellers, higher transaction numbers are associated with better rating indices y –and thus, in the model, with higher priors $\mu(y)$ for the typical buyer. We

[19]Technically, we thereby circumvent the econometric problem of dynamic selection. Dynamic selection means that those sellers who have more transactions are a selected sample of sellers, which leads to confounding.

discuss this in more detail and provide empirical support for this below.

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 3 is driven by particular types of buyers who buy from young sellers, and buyers that are more likely to leave feedback irrespective of the number of feedbacks a seller has received already. We address this concern in Table 3.

The table shows regression results. 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 (that is, 0.00491), 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) suggest that the dependence of the likelihood to receive a feedback do 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 A.1 in the appendix. They are similar.

Table 3: 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: This table shows 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 $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

4.2 Effects of Feedback

The central idea our model rests upon is that the likelihood to rate positively depends on the difference between posterior and prior, which can be interpreted as the amount of information there is to share from the perspective of the buyer.

Our sample consists 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 increasing 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 also the seller. 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.

Above, we have studied the dependence of the likelihood that a rating is left on the transaction number. The underlying idea was that the rating record increases and therefore, buyers' prior increases in the transaction number. To provide empirical support for this, we study the dependence of prices and the likelihood to sell a product on the number of feedback a seller has received. Figure 4 shows the result. Both the price and the likelihood to sell depend positively on the number of feedback received. The relationship is concave. Figure 4a shows that prices are about 7% lower for the first transactions. Figure 4b 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.

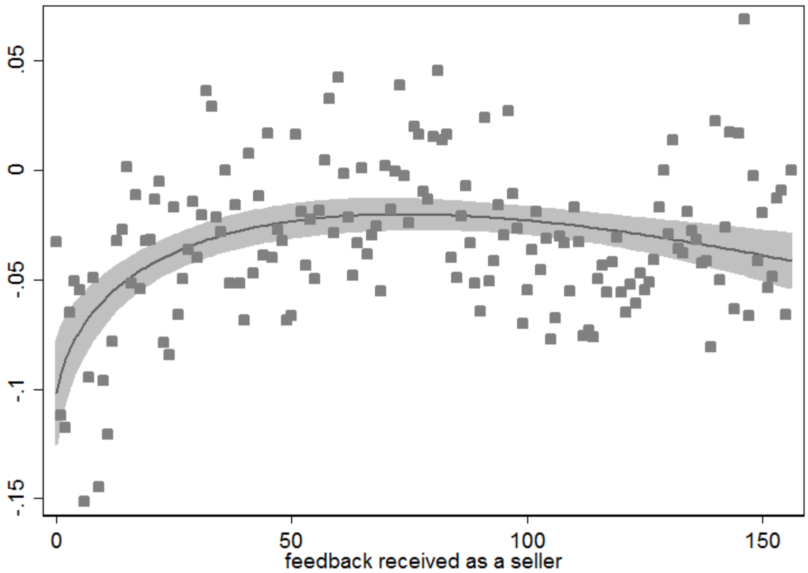
4.3 Learning effects

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.

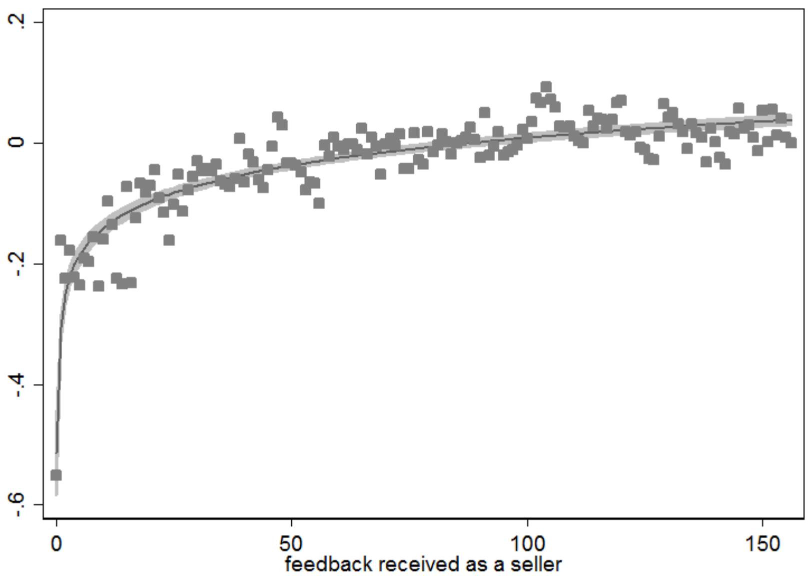
One may wonder whether the same empirical pattern could be driven by seller learning. Suppose first that a sellers learns over time and therefore transaction quality is higher for later transactions. Consequently, buyers would be positively surprised, implying that they would be more inclined to leave a rating. The pattern documented in Figure 3 and Table 3 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. If true, then this could generate the pattern we observe. We should then

(a) Effect of positive feedback on price



(b) Effect of positive feedback on probability of selling



Notes: We use the Sample 338 and restrict the sample to transactions with a product ID. We then look at the first 156 feedback from sellers who had at least 156 feedback in the restricted sample. The number 156 was chosen because it's the 90th percentile of feedback index in the restricted sample. Figure 4a 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 4b 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 4: Effects of a positive feedback

Table 4: Transaction quality

	(1) claim	(2) neutral or negative	(3) low DSR
transaction number/10	0.107*** (0.0146)	0.0718*** (0.0119)	0.0501*** (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: This 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 $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

also 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 regressed 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 used the same specification as in column (3) of Table 3. As before, the dependent variable is respectively an indicator for the event, times 100. This means that the units of coefficient estimates are percentages. Table 4 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 2.

These effects point in the direction that for our balanced sample of 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 3 that the likelihood to leave a feedback decreases by about 3.8 percentage points. Taken together, our interpretation is thus that part of this can probably be attributed to lower transaction quality, but not all of it.

5 Inclination to Share a Negative Experience

The second stylized fact is related to Section 2's Proposition 2, which establishes that a reduction in the index y induced by a negative rating will increase the likelihood that another buyer leaves a negative rating right thereafter, when the transaction quality was perceived to be negative.

We first briefly review the argument. 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 leaves 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 difference to her prior, because the updated prior attaches a lower probability to the seller being a good type. This increases the probability that she will rate.

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 quality shock.

Our approach is instead to use first negative feedback that were 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.[20] Unlike us, a later buyer does not know that such a negative feedback was given by mistake. Therefore, these feedback should generate the effect we want to quantify. At the same time, the aforementioned problem that two feedback are confounded by a negative shock is not there, because the first feedback was not meant to be negative.[21]

To operationalize this, 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 2 shows that the average number of days between the transaction and the feedback is 12.5 for our sample.[22]

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

[20] 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 also https://www.techjunkie.com/retract-feedback-ebay/.

[21] One may be worried that there was a negative feedback that was due to a negative quality shock, and that was then retracted 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.

[22] 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 1: transaction and feedback no later than t_0
- class 2: transaction time no later than t_0 and feedback after t_0
- class 3: transaction and feedback after t_0.

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.

It is useful to relate this specification to our theoretical model before we discuss the results. 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. The same will be true for class 3. If it is the case that sellers adjust their behavior so that transactions that took place after a negative feedback was left (even if it was later retracted) had a higher transaction quality, then the coefficient on the class 3 indicator will capture both effects.

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 1.1% (Table 2) that a transaction results in a neutral or negative feedback.

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

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.

Coming back to the concern that there may still be confounding by quality shocks, 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. Here, the coefficient on the class 2 indicator is estimated from transactions that took place in a day that contains the day of the transaction for which negative feedback was given and later retracted, and for which feedback was given after the first negative feedback that was later retracted.

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

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**	0.0198***	0.0109**	0.0410***	0.0508***
	(0.00437)	(0.00726)	(0.00457)	(0.0147)	(0.0170)
class 2 × buyer experience		-0.0156**			-0.0164**
		(0.00689)			(0.00691)
class 2 × new product with ID			-0.0100		-0.0111
			(0.0103)		(0.0104)
class 2 × >73 previous positive feedback				-0.0353**	-0.0349**
				(0.0154)	(0.0154)
class 3	0.0103**	0.0122**	0.0106**	0.0368***	0.0390***
	(0.00420)	(0.00501)	(0.00435)	(0.00981)	(0.0107)
class 3 × buyer experience		-0.00322			-0.00377
		(0.00367)			(0.00367)
class 3 × new product with ID			-0.00652		-0.00595
			(0.00516)		(0.00484)
class 3 × >73 previous positive feedback				-0.0314***	-0.0312***
				(0.0108)	(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: We restrict the sample to all transactions that are in a time window of 30 days around the time of a negative feedback that was later retracted. Class 1 (omitted) is the set of transactions where transaction and feedback time are before the feedback time of the retracted negative feedback. Class 2 has the transaction time before that and the feedback time after it. Class 3 has feedback and transaction time after the retracted negative. 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 × 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: We use the 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 $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

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 negatives 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. We select all transactions between the 30th and the 60th day for all sellers, 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 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

Table 7: Inclination to leave a negative feedback: 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: The sample consists of all transactions between the 30th and the 60th. Classes are defined using the first non-positive feedback. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

because it is observable to future buyers.[23] 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) 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).

6 Summary and Policy Recommendation

We study when and why buyers rate. In order to structure the typical buyer's decision we provide a simple model in which a buyer tends to rate when, relative to her expectations, she has learned a lot from her transaction with a particular seller. Formally, this corresponds to the magnitude of the Bayesian update of her prior beliefs as influenced by the observed aggregate of the previous ratings of the seller. With this simple model we can explain seemingly unrelated empirical results we derive from eBay's raw data, ranging from feedback concentrating on extreme experiences that is well documented for most online markets, to a probability of ratings decreasing in the number of transactions conducted by a sellers, to an increase in the probability of a negative rating when the rating record diminishes due to a negative rating. While the latter can be easily explained within a standard social learning framework, the decline in the probability of rating cannot –nor, of, course, the distribution of ratings on the scale from (very) negative to (very) positive. Indeed, our contribution to this framework consists of the endogeneization of decision to communicate the private learning result.

Our results have important implications for platforms and policy makers. Specifically, even though the observed rating process is highly non-random, the information sharing process on which our explanation of the rating decisions is based may be rather efficient. The rating average typically reported by sellers or platforms may be a good indicator of seller quality. The reason is simply that based on our theory, buyers tend to discontinue to share information about their private experience when that experience is close to conforming to that average. By contrast, if it is not, and her posterior beliefs that the seller provides good transactions substantially differ from the buyers' prior beliefs, they will tend to share their private learning experience. This will obviously contribute to the desired modification of the seller's rating aggregate.

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

This applies only to the situation when the rating record is small, however; more precisely when, upon observing the seller's rating record, the typical buyer cannot be sure that the seller provides good quality. By contrast, when she is, her incentive to communicate a deviating experience shrinks. Any market design geared to generate an unbiased rating index should account for this.

An attractive possibility could be to use only recent, e.g., the last 30 ratings in the aggregate rating index. By this, the time period within which ratings enter the index would be long for sellers that are infrequently active on the platform, and possibly very short for those frequently active. The concentration on the most recent ratings in the formation of the rating index could be welcome two important reasons. First to prevent, in line with Hoerner and Lambert (2020), sellers from exercising moral hazard, that could even be incentivized by a long (and good) rating record.[24] Second, such a rating index applied to incumbent sellers would reduce the entry barrier emphasized by Vellodi (2020) that is generated the incumbents' established rating record any entrant has to overcome.

[24] Kovbasyuk and Spagnolo (2018) propose short limits for positive records, but long limits for negative ones. While implementable with more difficulties, this would conform with our suggestion.

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*.

Bajari, P. and A. Hortaçsu (2004, June). Economic insights from internet auctions. *Journal of Economic Literature 42*(2), 457–489.

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. 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.

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

Einav, L., C. Farronato, and J. Levin (2016). Peer-to-peer markets. *Annual Review of Economics 8*, 615–635.

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.

Hoerner, J. and N. Lambert (2020). Motivational ratings. *Review of Economic Studies*, forthcoming.

Hopenhayn, H. and M. Saeedi (2019). Optimal quality ratings and market outcomes. 2019.

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.

Kovbasyuk, S. and G. Spagnolo (2018). Memory and markets. *Available at SSRN 2756540*.

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.

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.

Resnick, P. and R. Zeckhauser (2002). 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.

Saeedi, M. and A. Shourideh (2020). Optimal rating design. Technical report, Carnegie Mellon University.

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.

Vellodi, N. (2020). Ratings design and barriers to entry. Technical report, Paris School of Economics.

Online Appendix

A Robustness

This appendix contains additional results and robustness checks. Table A.1 reproduces Table 3 for DSR's as the outcome. Table A.2 and A.3 use alternative defintions of experience. Table A.4 reproduces Table 3 for separate DSR categories.

Table A.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: This 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 A.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 $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Table A.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**	0.0156***	0.0109**	0.0410***	0.0160
	(0.00437)	(0.00456)	(0.00457)	(0.0147)	(0.0105)
class 2 × buyer experience		-0.0107**			-0.00543
		(0.00534)			(0.00605)
class 2 × new product with ID			-0.0100		-0.00681
			(0.0103)		(0.0102)
class 2 × number previous positive feedback				-0.0353**	-0.00764
				(0.0154)	(0.0117)
class 3	0.0103**	0.0152***	0.0106**	0.0368***	
	(0.00420)	(0.00435)	(0.00435)	(0.00981)	
class 3 × buyer experience		-0.0101**			-0.00228
		(0.00434)			(0.00432)
class 3 × new product with ID			-0.00652		-0.000914
			(0.00516)		(0.00517)
class 3 × number previous positive feedback				-0.0314***	0.00632
				(0.0108)	(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 A.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 $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.