CESifo WORKING PAPERS

11171
2024

June 2024

Learning from Online Ratings

Xiang Hui, Tobias J. Klein, Konrad O. Stahl

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 11171*

Learning from Online Ratings

Abstract

Online ratings play an important role in many markets. However, how fast they can reveal seller types remains unclear. To study this question, we propose a new model in which a buyer learns about the seller's type from previous ratings and her own experience and rates the seller if she learns enough. We derive two testable implications and verify them using administrative data from eBay. We also show that alternative explanations are unlikely to explain the observed patterns. After having validated the model in that way, we calibrate it to eBay data to quantify the speed of learning. We find that ratings can be very informative. After 25 transactions, the likelihood of correctly predicting the seller type is above 95 percent.

JEL-Codes: D830, L120, L130, L810.

Keywords: online markets, rating, reputation, Bayesian learning.

Xiang Hui
Washington University at St. Louis
St. Louis / MO / USA
hui@wustl.edu

Tobias J. Klein
Tilburg University
Tilburg / The Netherlands
t.j.klein@tilburguniversity.edu

Konrad O. Stahl
University of Mannheim / Germany
konrad.stahl@uni-mannheim.de

May 6, 2024
We are grateful for comments by Liran Einav, Andrey Fradkin, Germain Gaudin, Ali Hortaçsu, Ginger Jin, Fiona Scott Morton, André Stenzel, Jonas von Wangenheim, Christine Zulehner, as well as seminar participants at Chicago Booth, Johns Hopkins, Rochester, Tilburg University, the European Commission/DG Comp, Bocconi, Washington University, APIOC, the Bass FORMS Conference, the CESifo ICT conference, the ISMS Marketing Science conference, the ZEW ICT conference, and the German Economic Association IO Meeting. We would like to thank in particular Christoph Carnehl for constructive discussions and help in the conceptualization of our model. We are grateful to eBay for providing access to the data. Konrad Stahl gratefully acknowledges support from the Deutsche Forschungsgemeinschaft through CRC TR 224 (project B04). An earlier version of this paper has been shared under the title "When and Why Do Buyers Rate in Online Markets?"

1 Introduction

Online ratings play an important role in many markets. Consumers regularly consult ratings before they book a hotel, choose a restaurant, hire a plumber, choose a doctor or a lawyer, or make an online purchase.[1] This can lead to better choices and sizable welfare effects (Reimers and Waldfogel, 2021). However, how many transactions it takes until rating records reveal seller types remains unclear. This question is relevant for buyers when choosing a seller, for the platform when ranking search results, and for competition economists who are interested in quantifying the entry cost associated with building an online reputation.

The question of how informative buyer ratings are is generally not easy to answer. Administrative data sets generally contain a wealth of information related to transactions. However, seller types, transaction experiences, and buyers' beliefs are unobserved. Beyond this lack of data, it is not straightforward to estimate even average buyer experiences from ratings because there is rating bias—buyers are less willing to leave a rating if they had a negative experience as compared to a positive one (Nosko and Tadelis, 2015). Moreover, there is no canonical model that relates rating decisions to buyer beliefs and learning about the seller type.[2] Such a model could be calibrated to data to quantify the speed of learning.

We propose such a model and calibrate it to administrative data from eBay. We formalize the idea that consumers are more likely to share their experience when they learn in the transaction about seller quality. We also allow for the aforementioned rating bias. We first validate the model by showing that it predicts two stylized facts that we document in administrative data from eBay. We then use the model to quantify how fast buyers can learn from online reviews about seller quality. For this, we ran a survey to elicit some of the model parameters and calibrate others to match the patterns in eBay data. We find that after 25 transactions, the likelihood of correctly predicting the seller type is above 95 percent. Thus, learning is fast even though ratings are biased.

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

[2] The rating decision is of course modeled in many papers. However, the focus is often not on how it depends on both publicly available and private information. For instance, Vellodi (2020) theoretically studies entry barriers associated with building an online reputation and assumes that all transactions are rated. Acemoglu et al. (2022) theoretically study learning from online reviews when early buyers differ from late buyers. They assume that only the intensity of the experience with the seller determines whether or not a rating is left.

In more detail, the typical seller is of either high or low quality, and the buyer can have a good or a bad experience in a transaction. Her experience is imperfectly informative about seller quality: she can have a good experience with both types of sellers, but it is more likely that she has a good experience with a high-quality seller than with a low-quality one. Before that experience, she forms a belief summarized in the probability that the seller is of high quality, which incorporates all publicly available information about that seller, e.g., information about the environment in which he operates inasmuch it is indicative of seller quality, and in particular his public rating record. Based on her transaction experience, the buyer updates that belief in a Bayesian sense and decides whether to share that update in a rating.

Towards that, she compares the benefit of leaving a rating to a stochastic cost. The benefit represents a general willingness to share information. Its size depends on how much the posterior probability that a seller is of high quality differs from the prior probability, and on whether the difference is positive or negative. The difference in beliefs captures her learning intensity: The more she has learned, the likelier she will rate because the more information is conveyed in the rating. However, she may be more reluctant to share her experience if the difference is negative.[3] This leads to rating bias because positive experiences are more likely to be shared than negative ones.

We first analyze theoretical properties of our model. We derive two empirical predictions. These predictions are related to the *effects* of changes in beliefs: later buyers are better informed about seller quality because they can interpret online ratings the seller has received and a negative rating has a negative effect on the belief that a seller is high-quality. The first prediction from our model is that the likelihood that sellers receive a rating for a transaction will be lower for later transactions. The intuition behind this result is that buyers tend to learn more about seller quality in earlier transactions than they will in later ones because prior beliefs in later transactions are either closer to zero or one; hence they are also more likely to leave a rating for earlier transactions. For the second empirical prediction, we focus on sellers with a good rating record. For them, a negative shock to prior beliefs, for instance, due to a negative rating that is given by another buyer, has a positive effect on the likelihood that a rating is left. We show that

[3] She may be afraid of retaliation by the seller, or she may generally feel bad when talking negatively about others.

the likelihood of leaving a rating increases more when the buyer's own experience is negative rather than positive. Both predictions illustrate that the main driving force of the model is that through learning, signals become less and less informative. The advantage of a simple model like ours, with two types of sellers and two types of experiences, lies in its clarity and ease of understanding. We illustrate in an appendix that the same driving force also naturally arises in a model with a continuum of types and a continuum of possible transaction experiences.

For our empirical analysis we leverage access to administrative data from eBay, one of the most established and successful online marketplaces.[4] The classic eBay feedback system is characterized by positive, neutral, and negative ratings. It embodies core elements, such as voluntary user-generated feedback on transactions to indicate a seller's or buyer's reliability. With its introduction in 1995, it was one of the first to exist. Most online marketplaces have since used it as a model to design their reputation or feedback system.[5] We construct a sample of newly entering sellers who had their first listing in March 2011. We follow these sellers for one year. Our data are at the level of the individual transaction and contain transaction details, information on ratings, as well as buyer and seller characteristics.

We first validate our model by testing our two empirical predictions. To test the first, we construct a balanced panel of sellers, with at least 86 transactions in the first year. We use a balanced sample for this because later transactions could be rated only for sellers who perform better and therefore have more transactions than others (that is, only better-performing sellers have these later transactions, which is a form of dynamic selection). We use the balanced panel to construct a plot of the likelihood that a transaction is rated, against the index for the transaction number that ranges from 1 to 86. It shows that earlier transactions are indeed more likely to be rated than later transactions. We also carry out regressions to control for differences in rating behavior across buyers, seller fixed effects, month fixed effects, and goods category fixed effects. Together with using a balanced panel, this allows us to attribute the observed decrease in the likelihood that a transaction is rated to buyers' learning about seller types.

The challenge when testing our second prediction on the effects of a negative shock to

[4]In 2023, eBay was the second-biggest worldwide in terms of visits, after Amazon. See https://www.webretailer.com/b/online-marketplaces/ (October 2023).

[5]Tadelis (2016) points out that "practically every online marketplace has adopted some form of a reputation or feedback system that is closely related to the one that eBay introduced in 1995."

beliefs is that subsequent ratings could be confounded by unobserved transitory shocks. In particular, a negative rating following a negative one could be caused by correlated seller behavior rather than by a change in rating behavior as predicted by our model. We make use of quasi-random variation in the timing of ratings and an institutional detail to overcome this challenge, by using ratings that were given by mistake and removed at a later point in time. We then use two empirical strategies to quantify the effect of the negative rating that was left by mistake. Our main approach is akin to a difference-in-differences analysis. Our second approach is akin to an event study. Findings from both are in line with our second prediction: A mistakenly given negative rating increases the probability that a rating is left; and increases the probability that a rating that is given is negative.

After having tested the two *qualitative* empirical predictions, we turn to calibrating our model to eBay data to *quantify* the speed of learning and study the formation of (unobserved) beliefs.[6] The data are not directly informative about some of the parameters of the model. Therefore, we conducted an online survey. We targeted participants who lived in the U.S. and shopped online at least once a month. In line with our model, we asked them to think of sellers as high vs. low quality, and of transaction experiences as either good or bad. Then, we asked them to quantify three key parameters of our model: the likelihood that a newly entering seller who has his first listing is of high quality; the likelihood that a transaction experience with a high-quality seller is good; and the likelihood that a transaction experience with a low-quality seller is good.

We then calibrate the remaining parameters to fit our model to the eBay data. We form moments and minimize an objective reflecting the difference between model simulations and key patterns found in the eBay data, such as the percentage of negative ratings, or numbers related to our empirical predictions. We find that buyers are more inclined to leave a rating when their experience was positive and that they are more likely to leave a rating when the transaction experience had a bigger effect on their beliefs about seller quality.

Then, we conduct simulations. We find that after 25 transactions, the likelihood of correctly predicting the seller type is above 95 percent. By that time, sellers have received about 17

[6] One can think of the model as helping us to impute beliefs.

ratings. We also simulate the effect of changes to the rating system. Learning would be even faster if the platform would be able to mandate that all transactions are rated. This would also remove rating bias. Another way to remove rating bias would be to encourage buyers to share negative experiences. The latter would lead to learning almost as fast as if all transactions were rated.

Related literature We relate and contribute to three branches of the literature. The first branch focuses on rating bias. Rating bias is present, e.g., when negative experiences are underrepresented in rating averages (Schoenmüller et al., 2019). One reason may be that buyers fear retaliation when they share a negative experience (see, e.g., Resnick and Zeckhauser (2002), Klein et al. (2009), Dellarocas and Wood (2008), Bolton et al. (2013), Hui et al. (2018), and Fradkin et al. (2021)). Retaliation by sellers was accounted for by eBay, by abandoning the option to rate buyers negatively.[7] Rating bias may also be present because ratings may be manipulated or bought (see, e.g., Mayzlin et al., 2014; Luca and Zervas, 2016; He et al., 2022). However, manipulating ratings at eBay is costly because ratings are linked to transactions. Nosko and Tadelis (2015) document that nonetheless, buyers are reluctant to share negative experiences on eBay. Our contribution to this branch of the literature is twofold. First, we add evidence from a survey among a population of people shopping regularly online. In the eyes of our survey participants, the likelihood of having a good experience with a high-quality seller is 10 percentage points lower than average ratings suggest when taken at face value. Our main contribution is to use the results from the survey to calibrate our theoretical model to eBay data and simulate how fast online ratings can reveal seller types in the presence of rating bias.

A second branch of the literature focuses on the dynamics of rating and purchase decisions. Moe and Schweidel (2012) study how previous ratings affect whether and what buyers rate, and Moe and Trusov (2011) estimate the value of this in terms of extra sales. They look at product ratings. We look at seller ratings. They analyze observational data using Bayesian estimation, while we base our analysis on a balanced panel of starting sellers and exploit quasi-experimental variation. Dellarocas et al. (2006) and Dai et al. (2018) focus on the construction of rating aggregates that are more informative than raw averages. Cabral and Hortaçsu (2010),

[7] That event was used by Klein et al. (2016) to show non-marginal consequences.

Hu et al. (2017), and Li and Hitt (2008) show empirically that product ratings provided by early buyers are different from ratings provided by late buyers. Ishihara and Liu (2017) show that buyer selection, meaning that early buyers are different from late buyers, is important in this context. Carnehl et al. (2023) study how this affects pricing decisions. Acemoglu et al. (2022) also study buyer selection by developing a theoretical model of Bayesian learning from online reviews. They assume that early buyers have a stronger taste for the product that is sold. The intensity of the experience with the seller determines whether a rating is left. They derive theoretical results on the rate of learning and compare that rate between different rating systems. We contribute to this branch of the literature by proposing a new model of rating behavior in which the rating decision depends on how much the typical buyer learns about the seller's quality in the transaction, relative to a belief based on previous ratings. We derive two new and specific empirical predictions related to the dynamics of ratings. We show that these empirical predictions hold in eBay data. We do not explicitly study buyer selection, but control for it in our analysis. We also calibrate our model to the eBay data and use it to quantify how fast buyers learn from online reviews.

A third branch of the literature focuses on different motives behind the rating decision. Chakraborty et al. (2022) propose a game involving a monopolist, early adopters, and future buyers involving a difficult-to-test hypothesis, by which ratings are left when they influence the purchase decision of future buyers. This aligns with our model's central idea that buyers are more inclined to share their transaction experiences if they have gained more insights from those experiences on their own. We provide empirical support for our model and thereby indirectly for the model by Chakraborty et al. (2022). Luca and Reshef (2021) show that price increases lead to lower average ratings. Anderson and Sullivan (1993) and Ho et al. (2017) hypothesize that consumers are more likely to leave a rating when their experience in a transaction is not in line with what they expected, which is sometimes referred to as the "surprise hypothesis". While this bears some similarities with the driving forces in our model, it is at odds with our second empirical prediction: we predict that the likelihood that a negative experience is shared *increases* when there is a negative shock to the prior; by the surprise hypothesis, however, it *decreases* with such a shock, because then a negative experience is less of a surprise. Our empirical results are in line with our empirical prediction. Thus our evidence speaks against the "surprise hypothesis". We contribute more generally to this branch of the literature by relating the rating decision to Bayesian learning about seller quality and quantifying the speed of learning from online reviews.

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

2 Model

Our analysis is guided by a simple model. In this section, we present key elements of the model and informally discuss the two empirical predictions we take to the data to validate the model. These predictions are related to the effects of changes in beliefs. Full details on the model and related derivations can be found in the Theoretical Appendix. In Section 7 we make additional assumptions that allow us to simulate the evolution of beliefs and calibrate the model to the data.

2.1 Setup

Consider an online marketplace with two types of sellers, low- and high-quality, and two types of experiences, good and bad.[8] The likelihood of having a good experience with a high-quality seller is $\rho_h > 1/2$ and the likelihood of having a good experience with a low-quality seller is $\rho_l < \rho_h$.

Potential buyers do not observe the seller type. Before the transaction, the buyer forms a prior belief $\mu(\lambda, y)$ about seller quality. μ increases in her perception that the sellers on the platform provide good service on average, summarized in the probability λ, and in the particular seller's public reputation record, summarized in some index y. When performing the transaction, she observes the transaction experience $s \in \{g, b\}$, where s stands for signal. When

[8] In Online Appendix A, we present a generalization accommodating a continuum of transaction experiences and more than two types of ratings.

the transaction experience was good, then her posterior belief that the seller is of high quality is

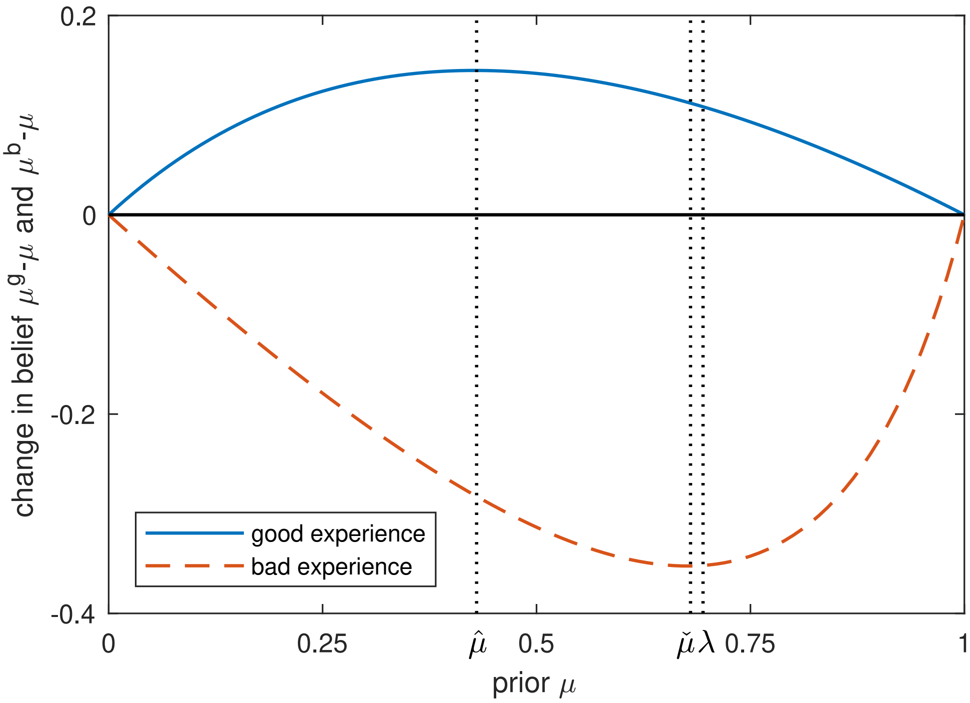
$$\mu^g = \frac{\mu \rho_h}{\mu \rho_h + (1-\mu)\rho_l}. \tag{1}$$

When the transaction experience was good, then her posterior belief is

$$\mu^b = \frac{\mu \cdot (1-\rho_h)}{\mu \cdot (1-\rho_h) + (1-\mu)\cdot(1-\rho_l)}. \tag{2}$$

We assume that ratings, when left, are truthful in the sense that they are equal to transaction experiences. This is in line with instructions on eBay and many other online platforms. The central element of our model is that the more the buyer updates her beliefs after conducting a transaction with a seller, the more she expects others to learn from her rating, and therefore the more likely she will be to leave a rating. This is natural given that we assume that she shares her signal: when other buyers are like her, then others will learn more from the signal if she learned more from it. Formally, the likelihood that she rates the seller increases in the absolute difference between her prior belief μ and her posterior belief μ^s about seller quality. The absolute difference is meant to indicate that the intensity of learning drives the rating decision. We also allow for a baseline inclination to leave a rating even if there is no learning. This baseline inclination is allowed to depend on the transaction experience itself. This part of the model captures that some buyers feel more comfortable sharing positive experiences as compared to negative ones.

The aim of this paper is to study learning from online ratings when the rating decision depends on learning. In Section 7, we use a specification for which the likelihood that a rating is left is equal to $b^s + b \cdot |\mu^s - \mu|$, where $s \in \{g, b\}$ and $b > 0$. The more buyers learn in a transaction about seller quality the higher $b \cdot |\mu^s - \mu|$, and therefore, the higher the likelihood that they leave a rating. Note that our specification can also capture other motivations for leaving a rating than the one we focus on. Consider for instance reciprocity as a behavioral concept discussed broadly in the literature and surveyed in detail by Rabin (1998). Positive reciprocity (having a good experience motivates leaving a positive rating) can be captured by $b^g > 0$ and negative reciprocity (having a negative experience motivates retaliating by a negative rating)



Notes: This figure shows the change in beliefs as a function of the prior and the type of experience. See Section 2 for a summary of the model and the Theoretical Appendix for details. Plotted for values of ρ_h and ρ_l that we elicited using a survey. See Section 6 for details on the survey.

Figure 1: Change in beliefs

can be captured by $b^b > 0$. Alternatively, b^g and b^b can respectively capture the intensity of the experience net of the perceived (psychological) cost of leaving a rating (see, e.g., Acemoglu et al., 2022).[9] Irrespective of the exact motivation behind leaving a rating, our specification can accommodate rating bias. One source of rating bias is that the probability of leaving a rating depends on $|\mu^s - \mu|$. In addition, there is rating bias whenever $b^g \neq b^b$. In particular, average ratings are more favorable than average experiences if $b^g > b^b$.[10]

2.2 Implications

Figure 1 shows the change in beliefs that result from applying Bayes' rule for a given prior μ. We used the parameter values we elicited from the online survey discussed below in Section 6. A direct consequence of applying Bayes' rule is that there is updating if, and only if, $0 <$

[9]As discussed in the introduction, Luca and Reshef (2021) show that price increases can lead to lower average ratings. This can be seen as a different experience, and in that sense, it is captured by our model.

[10]As explained in the Introduction, our data do not support the surprise hypothesis (Anderson and Sullivan, 1993; Ho et al., 2017) The surprise hypothesis states that the likelihood to rate depends positively on $|1\{s = g\} - \mu|$, where $1\{s = g\}$ is an indicator for having a good transaction experience.

$\mu < 1$ and that a positive signal always implies that beliefs increase, while a negative signal always implies that they decrease. This follows directly from equations (1) and (2) above. In the Theoretical Appendix, we show that the maximum of the solid line summarizing updates resulting from good experiences is attained at a value $\hat{\mu}$, that is below the value $\check{\mu}$ at which the minimum of the dashed line is attained, that summarizes updates from bad experiences: Buyers learn the most from a positive experience for a value of μ that is lower than the value for which they learn the most from a negative signal. Moreover, we show that these extrema are unique and that the one for a negative signal is always obtained at a higher value of μ. For the specific parameter values for which the figure was drawn, positive signals are less informative about seller quality than negative signals are: for any prior μ the buyer learns more from a negative experience, and most when priors are between 0.5 and 0.8.

Next, we sketch the aforementioned two empirical predictions related to the dynamics of ratings, which we later take to the data.[11] The reasons for resorting to just these predictions are practical ones. Both are about the dynamics in observed ratings. For the second prediction, we consider negative feedback as a shock to beliefs, as the most common event is that positive feedback is given; we look at the special case of μ sufficiently close to 1 because we follow a set of sellers in our empirical analysis who are most likely high-quality sellers.

Empirical prediction 1. *The likelihood that a rating is given may first increase in the transaction index, will then reach its maximum, and will ultimately be a decreasing function in the transaction index.*

The underlying intuition can best be explained using Figure 1. Think of a new seller on eBay that has not conducted any transactions. Buyers believe that he is of high quality with probability λ, elicited in our survey to amount to 0.7 for eBay, and indicated on the horizontal axis in Figure 1.

For a high-quality seller, beliefs μ will move on average toward 1: no matter whether the experience is good or bad, the absolute value of the change in beliefs will decrease over time, and with this, the likelihood that a rating is left. Note that while the likelihood could have first increased, it does not happen for the average high-quality seller since the initial belief λ is large

[11]More formal statements and proofs are provided in the Theoretical Appendix.

enough. For low-quality sellers, we also start with beliefs λ, but then move toward 0 for the average seller. When the transaction experience is positive, then the difference in beliefs first increases and then decreases. When the transaction experience is negative, the difference in beliefs decreases, except in the very beginning and very slightly. Therefore, for low-quality sellers, the likelihood that a rating is left first increases and then decreases.

The observed pattern in the data is a mixture between the pattern for high-quality (always decreasing) and low-quality (first increasing, then decreasing) sellers. When there are enough high-quality sellers, then the likelihood that a rating is given is monotonically decreasing in the transaction index. Otherwise, it will first be an increasing function and then a decreasing one.

Empirical prediction 2. *When initial beliefs are sufficiently close to 1, then a negative shock to the prior leads to an increase in the likelihood that a rating is left; when a rating is left, then it is more likely to be negative.*

When initial beliefs are sufficiently close to 1, a negative shock to beliefs will move μ away from 1 in Figure 1.[12] As a consequence, a buyer learns more from a transaction. Therefore, she is more likely to leave a rating both when her experience in the transaction was good or bad. It turns out that the effect is bigger in relative terms if the experience is bad; therefore the likelihood increases that a rating that is given is negative.

3 Empirical setting, data, and descriptives

3.1 Empirical setting

In Sections 4 and 5, we take our two empirical predictions from Section 2 to the data. We also quantify the magnitudes involved. We use some of those in Section 7 when we calibrate the model. Here we describe our empirical setting and our dataset.

[12] We focus on the case in which beliefs are close to 1, because our sample is a sample of sellers that are likely of high quality.

3.2 Sample

We use rich administrative data from eBay for the U.S.[13] Our starting point is the set of sellers who had their first listing ever in March 2011 (sample 0).[14] From this set of sellers we construct two subsamples. The first subsample consists of the set of sellers who have at least 86 transactions in the first year (sample 86).[15] The second subsample consists of the set of sellers who have at least 338 transactions in the first year (sample 338). 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 respective samples of sellers, we construct an unbalanced panel of all transactions for the sellers in sample 0, and balanced panels with the first 86 transactions for the sellers in sample 86, and the first 338 transactions for sellers in sample 338. The transactions we use are for the so-called core products on eBay. For instance, real estate and cars are not in our data.

3.3 Variable definitions and summary statistics

We report summary statistics for all three samples in Table 1. There are three panels. Panel A contains seller characteristics. For each variable, we first create one observation per seller. Panel A then reports the average of those observations across sellers.[16] For the 141,138 sellers in sample 0 who had their first transaction in March 2011, average sales in the first year are $1,218. In that year they have on average 24 transactions and sell products in 5 so-called leaf categories.[17] The percentage positive is one of eBay's two rating aggregates that is displayed next to a seller's name on the platform.[18] It is calculated as the number of positive feedback the seller has received, relative to the number of feedback that was either positive or negative. Thus, neutral feedback is discarded. We report the eBay percentage positive at the end of the

[13] On eBay, online ratings are called feedback. We use the two terms interchangeably.

[14] 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.

[15] To be precise, here and in the following these are always transactions until the end of February 2012, irrespective of the exact date of the first listing.

[16] 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.

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

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

Table 1: Summary statistics

	(1) sample 0 (unbalanced)	(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 leaves 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 the same seller before	0.150	0.142	0.179
share of transactions with any feedback	0.622	0.670	0.655
share of transactions with neutral or negative feedback	0.016	0.011	0.007
share of transactions with low DSR	0.022	0.019	0.013
share of transactions with a claim	0.020	0.012	0.006
days between transaction and feedback	13.2	12.5	12.9
observations	3,413,354	609,310	477,256

Notes: Averages for three samples. Sample 0: all sellers who had their first listing ever in March 2011. Sample 86: top 5% of those sellers in terms of transactions. Sample 338: top 1%. In Panel A, one observation is one seller. We report statistics on all transactions they respectively conducted in the first year. 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.

first year.

It is not surprising that the top 5% (sample 86) and top 1% (sample 338) sellers in terms of the number of transactions within the first year have a higher sales volume and more transactions in more leaf categories compared to all sellers (sample 0), and that their percentage positive feedback is higher by 6 and 7 percentage points, respectively. Some of the sample 0 sellers who started in March 2011 will stop being active on eBay. Those sellers are a negative selection, and more likely to receive negative feedback (see for instance Cabral and Hortaçsu, 2010).

The buyers who bought from these sellers are characterized in Panel B. Our starting point here is all transactions in the first year for the sellers in 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 respective set of buyers, and for those, we calculated three measures. For sample 0, for instance, 1,792,076 distinct buyers bought from the 141,138 sellers who had their first transaction in March 2011.

We calculated two measures of buyer experience. The first measure is the number of transactions buyers conducted in the year before March 2011 (the month in which our sellers had their first listing). For these transactions, we can also calculate how often they left feedback and call this the inclination to leave feedback, and we calculate the share of negative feedback, which we refer to as buyer criticalness. The second measure of buyer experience, which we will use most of the time, is whether they have registered before March 2009.[19] This is the case for 71.3% of the buyers in sample 0.

Recall that sellers in sample 0 are rated worse on average, as compared to those in sample 86 and sample 338. Interestingly, in Panel B both measures of buyer experience for the sellers rated in sample 0 are lower than for those in sample 86 and sample 338. At the same time, buyer inclination to leave feedback and buyer criticalness are remarkably similar across samples. Averages of both correspond well to the eBay percentage positive between 98% and 99% that we report in Panel A for samples 86 and 338.

In Panel C, we report summary statistics for transactions in the first year. We use all transactions for sample 0, the first 86 transactions for sample 86, and the first 338 transactions for sample 338. The first statistic shows that buyers and sellers do not interact repeatedly in the vast majority of transactions. Less than 20% of the transactions are with a seller from which the buyer has bought before. The remaining statistics relate to feedback. For the transactions in our sample, feedback is left for about two out of three transactions. This corresponds well to the numbers reported in Panel B. The share of feedback that is neutral or negative is 2.6% in sample 0. Putting this side-by-side with the percentage of 92% positive feedback in Panel A, we can see that sellers with a higher number of transactions have a higher percentage of positive feedback. This is also in line with the lower shares of neutral or negative feedback in samples 86 and 338, which are 1.1% and .07%, respectively.

The next measure is the share of the transactions with low detailed seller ratings (DSR).

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

Buyers on eBay are asked to provide a DSR after providing a classical rating. DSRs are ratings in four dimensions, item description, communication, shipping time, and shipping and handling charges. Ratings are left on a five-point scale. We define a DSR as low if at least one of the 4 DSR dimensions has 1 or 2. It is interesting to see that the share of anonymous low DSRs is higher than the share of non-anonymous neutral or negative feedback, which indicates that buyers avoid public negative statements about sellers. Lastly, the gap between the date on which a transaction took place and the date on which the transaction is rated (if at all) is around 13 days on average and is similar across the samples.

4 Inclination to leave a rating

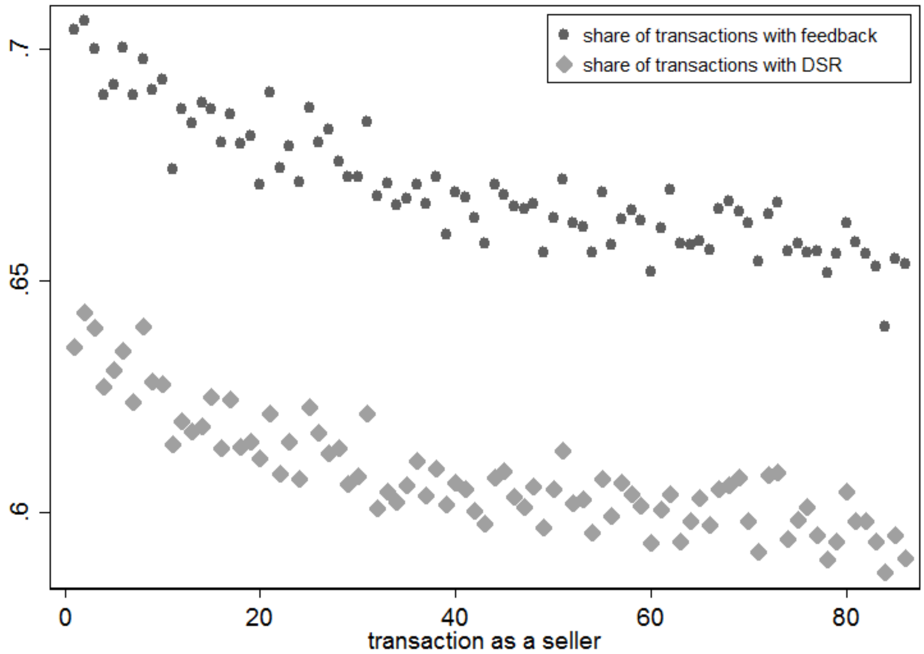
The first empirical prediction from our model is that the likelihood that a given transaction will be rated decreases with an increasing number of transactions performed by the average seller. By our explanation, buyers learn on average less and less from a transaction and are therefore less and less inclined to share their experience by leaving a rating.

4.1 Results from our balanced panel

We use our balanced panel of transactions to test this prediction. This allows us to plot the likelihood of receiving a rating against the transaction number and to interpret the results as if we followed sellers over time.[20] In Figure 2, we plot the share of transactions with feedback and DSR, respectively, against the number of transactions performed by the sellers. Both the share of transactions with feedback and the share of transactions with DSRs are lower for later transactions.

When constructing this figure, we do not control for differences across sellers, calendar time effects, differences across products, and buyer characteristics. Therefore, the patterns in Figure 2 could be driven by seller heterogeneity, time effects, differences across products, or selective buying. For instance, products that are more likely to be rated could also more likely to be sold

[20] By using a balanced panel we circumvent the econometric problem of dynamic selection. Dynamic selection happens when sellers who have more transactions are more likely to be a high type. This would lead to selection bias.



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

Figure 2: Probability to receive feedback

by sellers starting on eBay. Or buyers who are more likely to rate in general could wish to buy from a starting seller.

We address these concerns in the regressions reported in Table 2.[21] 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 of receiving a rating for the 10th transaction is lower by 0.491 percentage points, as compared to the first transaction.

We successively add controls in the ensuing columns. In column (2), we control for seller fixed effects and calendar month fixed effects. In column (3), we control for product type using information on the leaf category. This addresses the concern that the likelihood of leaving a rating depends on the type of product. Results suggest that the likelihood of leaving a rating is lower by 0.389 percentage points for every 10 additional transactions.

We control for buyer experience in columns (4) and (5). The estimated coefficient on the interaction term between buyer experience and the transaction index in column (5) suggests

[21] We present results for DSRs in Table B.1 in the appendix. They are similar.

that the dependence of the likelihood to receive a rating does moderately depend on buyer experience: the effect of 10 additional transactions is different by -0.178, from a baseline of -0.211. In column (6) we control for the buyers' respective inclination to leave feedback. We find that the effect is weaker for buyers who are more likely to leave a rating in the first place. But overall, we still find a negative overall effect: earlier transactions are more likely to receive a rating. We present a host of additional results and robustness checks in Appendix B.

4.2 Alternative explanations

4.2.1 Buyer selection

An important alternative explanation for our finding could be that earlier consumers select themselves into transactions because they expect to receive a higher surplus, and are therefore more likely to leave a rating. This mechanism plays an important role in the theoretical contribution by Acemoglu et al. (2022). In columns (4) through (6) of Table 2 we have controlled for buyer experience and in columns (5) and (6) also for the inclination to leave a negative rating. To some extent, this already controls for buyer selection. To further investigate whether it could be an alternative explanation for our findings, we assess the dependence of two measures of consumer surplus on the transaction index. For the first measure, we use the difference between the winning bid and the second-highest bid (plus a small, fixed increment) of the second-price auctions conducted on eBay, which is a good proxy for consumer surplus.[22] Furthermore, we look at transactions in posted price format. Specifically, the second measure is the price consumers paid for the same item when it is new. For this, we use the product IDs eBay assigned to some of the offered products. Our second measure should be inversely related to consumer surplus. For neither of the two measures, consumer surplus decreased in the transaction index, as can be seen from Figures B.2a and B.2b in Online Appendix B.1.2. Taken together, our findings suggest that selective buying cannot explain our findings.

[22] Auctions account for around 15% of sales on eBay

Table 2: Probability to receive feedback

	(1)	(2)	(3)	(4)	(5)	(6)
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 × buyer exp.					-0.178**	-0.184**
					(0.0762)	(0.0753)
buyer inclination to leave feedback						26.90***
						(0.275)
trans. num/10						0.141**
× buyer inc. to leave fdbk						(0.0570)
seller FE	No	Yes	Yes	Yes	Yes	Yes
month FE	No	Yes	Yes	Yes	Yes	Yes
leaf category	No	No	Yes	Yes	Yes	Yes
adj R-squared	0.000671	0.0621	0.0712	0.0656	0.0656	0.131
observations	609310	609310	607135	515978	515978	515978
number of clusters	7085	7085	7085	7083	7083	7083

Notes: Results of regressions of 100 times an indicator for receiving feedback on the transaction number divided by 10, as well as other controls and interaction terms. One observation is a transaction. Standard errors are clustered at the seller level and account for heteroskedasticity. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$. Based on sample 86.

4.2.2 Friendly reviews

Another alternative explanation for the observed pattern in Figure 2 could be that starting sellers ask friends to buy from them and leave a favorable rating. The cost of doing so would be the listing fee and the effort it takes to coordinate with friends. Alternatively, sellers could buy these ratings from a third party (see, e.g., Mayzlin et al., 2014; Luca and Zervas, 2016; He et al., 2022). Then, the cost would be the fee the third party charges plus the transaction fee on eBay, as reviews are linked to transactions.

While we cannot directly exclude this alternative explanation, we believe that three pieces of evidence speak against it generating the pattern we observe. First, Figure B.3(a) in the Online Appendix suggests that the cost of having no reputation record is a 5-10% lower price sellers can charge for the first 10 or so transactions, at least in our sample of sellers who turn out to become very successful. Arguably, this is not too much and therefore, the incentive to ask friends is relatively low. Second, Figure B.3 suggests that the likelihood of selling is lower in the beginning when the number of feedback that sellers have received to date is low. If it was the case that many sellers ask friends to buy from them early on or when they buy reviews from a third party, then it should be higher, not lower. Third, when said friends or third parties do this systematically, then they would appear in our data as a buyer who rates many transactions. We control for this in one of our specifications in Table 2 and the results don't change.

5 Effect of a negative rating on subsequent ratings

Our second empirical prediction is that a negative shock to beliefs will increase the likelihood that a future transaction receives a rating and the likelihood that this rating is negative as opposed to positive. To test this prediction, we cannot simply relate the likelihood of receiving a rating to the reputation record the seller has at that point in time, as transitory shocks could confound ratings that are left shortly after one another.

We instead estimate the effects of negative ratings that are documented to be given by mistake. We can identify these ratings in the raw data because buyers went through a procedure to have them removed at a later point in time.[23] We look at later buyers' ratings before the mistakenly given negative rating was removed. These ratings should generate the intended effect, namely changing buyers' prior beliefs about seller quality. At the same time, the problem that two ratings are confounded by a negative shock is not present, because the feedback was not meant to be negative.[24]

5.1 Difference-in-differences analysis

Our data is at the level of transactions t for each seller i. We code $rating_{it} = 1$ if a rating is left for seller i on transaction t, and $rating_{it} = 0$ if no rating is left. We also define similar indicators for the event that a positive and a negative rating is left, respectively. We distinguish between sellers i who at some point receive a negative rating by mistake and those who don't. We call those who receive at some point a negative rating by mistake the treated sellers and indicate this using the variable $treated_i = 1$. The control group coded $treated_i = 0$ consists of sellers who never receive a negative by mistake. Finally, for the treated sellers we code $post_{it} = 1$ for all transactions that take place after the wrong negative feedback was received.[25]

We estimate the linear probability model

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

Our parameter of interest is γ, which is the effect of a negative rating that was given by mistake on the likelihood that a rating is given for transaction t by seller i. We control for transaction fixed effects α_t. They capture the dependence of the inclination to leave a rating on the length

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

[24] One may be worried that a rating was negative due to a negative quality shock retracted later because the seller bullied the buyer or bought her out. This should be infrequent, as a seller can only request the revision of feedback received in the last 30 days, and can do it only once per transaction: if the buyer rejects it, he cannot request it again. Beyond that, the seller can only request up to five revisions for every 1,000 feedback he receives. After all, if the seller bought out the buyer, then one should expect that the seller chooses reasons 1 or 3 in order not to unnecessarily suggest that the buyer is at fault. For that reason, we selected only ratings that were removed with reason 2.

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

of the feedback record. β captures differences between sellers who receive at some point a negative rating by mistake. This could be, for instance, because they sell in different categories.

This is a difference-in-differences model. The identifying assumption, often termed "parallel trends", is that the likelihood that a rating is left changes with the transaction number in the same way for sellers who at some point receive a negative rating by mistake, as it does for sellers who never receive such a negative feedback by mistake.[26]

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

In columns (4) - (6), we repeat the analyses with sample 86. Because the sample size is much smaller, we report robust standard errors but do not cluster at the seller level.[27] The results are similar—except that the effect on the likelihood of receiving a negative rating is higher.

These results imply that the likelihood that a rating is negative (if given) increases after a negative rating has been given by mistake. In the pre-period, it is[28]

$$\frac{.011 + .00724}{(.670 - .0666) - (.011 + .00724)} = .03023.$$

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

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

[28] Based on the summary statistics for sample 86 in Table 1, Panel C, column (2), and the results presented in Table 3, columns (4) and (6). The probabilities are all calculated for the treatment group that received a negative feedback signal by mistake. For example, for the post-period 0.011 is the share of transactions with neutral or negative feedback from the total sample. Added to this is 0.00724, the correction of that share for the treated group. Further added to this is 0.00966, the additional share of negatives left by the treated group after the shock. The probability of positive feedback involves all feedback minus negative feedback received.

Table 3: Effect of a negative rating on subsequent ratings

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

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

and in the post-period, it is

$$\frac{.011 + .00724 + .00966}{(.670 - 0.0666 + .0468) - (.011 + .00724 + .00966)} = .04483.$$

In line with our prediction, the relative likelihood that the rating is negative rather than positive increases by almost 50%.

5.2 Event study

We also directly estimated the effect of a mistakenly given negative rating on the likelihood that a subsequent rating is negative, using again sample 0. This can be seen as a generalized version of an event study. The event is a negative rating that was given by mistake; we exploit the randomness in the time it takes until a transaction is rated and use it to identify the effect of that negative rating. We see this event study as a complement to our difference-in-differences analysis above, as it allows us to control for transaction quality (as explained below) and to obtain additional results on the impact of a mistakenly given negative rating on transaction quality.

In principle, a seller can receive several mistakenly given negative ratings. We denote the time at which such a negative rating was received for the first time 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.[29]

Then, we classified these transactions as follows:

- class 1: transaction and feedback no later than t_0
- class 2: transaction no later than t_0 and feedback after t_0
- class 3: transaction and feedback after t_0.

Thus, class 2 contains transactions that were conducted before the mistakenly given negative rating. For these transactions, the seller could not have reacted to the mistakenly given negative rating by, for instance, providing a better transaction experience.

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

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

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

[29] Table 1 shows that for our sample the average number of days between the transaction and the feedback is 12.5. We chose 30 days because we also wanted to allow for possibly longer reaction times. 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.

Table 4: 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 $\times$ buyer experience		-0.0156**			-0.0164**
		(0.00689)			(0.00691)
class 2 $\times$ new product with ID			-0.0100		-0.0111
			(0.0103)		(0.0104)
class 2 $\times$ >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 $\times$ buyer experience		-0.00322			-0.00377
		(0.00367)			(0.00367)
class 3 $\times$ new product with ID			-0.00652		-0.00595
			(0.00516)		(0.00484)
class 3 $\times$ >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: Sample 0 restricted to all transactions in a time window of 30 days around the time of a negative feedback that was later retracted. Class 1 (omitted) involves transactions and feedback before the later retracted feedback was given. Class 2 involves transactions that took place before the later retracted feedback was given, wich were rated after the later retracted feedback was given. Class 3 involves feedback and transaction after the retracted negative. 73 is the 75th percentile of the number of positive feedbacks that all sellers in the event study had accumulated, i.e. sellers with first negative feedback that was given by mistake. Standard errors are clustered at the seller level and account for heteroskedasticity. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

We report the results in Table 4. Column (1) shows that indeed, the likelihood that a transaction in class 2 receives a negative rating is one percentage point higher. This is a big effect when we compare it with the likelihood of 2.6% that a rating is neutral or negative (0.016 / 0.622 from Table 1). But it is also surprisingly consistent with the estimates we presented in Table 3: calculating the baseline probability that a given rating is negative on the basis of sample 0 analogously to that for sample 86 in calculation in footnote 28, we obtain $0.0144/0.5822$ based on Table 1 and Table 3, and changes to $(0.0144+0.00335)/(0.588+0.0419)$ after the negative feedback signal. The difference between these two numbers is 1.36 percentage points, not much larger than the 1.05 percentage point increase reported in column (1) of Table 4. Lastly, the fact that the coefficient estimate for class 3 is similar to that for class 2 suggests that changes in seller behavior after observing a negative rating are small.

The results in the following columns show that this effect is driven by inexperienced buyers who registered after 1 March 2009 (column (2)) and that it is not important whether a product is new and standardized (column (3)). Furthermore, the effect becomes smaller when the seller has already more than 73 positive feedback on his record (column (4)).[30]

5.3 Alternative explanation and robustness

5.3.1 Confirmation bias

An alternative explanation for our findings could be that users who had a negative experience themselves feel reassured when they observe that another user leaves a negative feedback. As a result, they are more likely to share their experience as well. This could be called confirmation bias in our context and could explain the positive effect that a negative feedback that has been given by mistake has on the likelihood that a negative rating is left (columns (3) and (6) in Table 3, Table 4). However, it cannot explain the positive effect on the likelihood that a positive rating is left (columns (2) and (5) in Table 3) and also not the size of the overall probability that a rating is left (columns (1) and (4) in Table 3). In fact, one would expect that a negative rating has a negative effect on the likelihood that a positive experience is shared. Therefore, confirmation

[30] 73 is the 75th percentile of the number of positive feedback that all sellers in the event study, i.e., sellers with first negative feedback that was given by mistake, had accumulated.

bias is unlikely to explain our findings. In contrast, our findings naturally arise in our model, where a negative rating that is given by mistake has a positive effect on the likelihood that a rating is left for either experience.

5.3.2 Robustness

Section B.2 in the Online Appendix provides supporting evidence for the parallel trends assumption we make to obtain the results in Table 3. It also contains the results of three robustness checks for the results from the event study in Table 4.

6 Survey evidence on beliefs

We ran an online survey to elicit the beliefs of market participants about three key model parameters: the likelihood ρ_h of having a good transaction experience with a high-quality seller on eBay, the likelihood ρ_l of having a good transaction experience with a low-quality seller, and the fraction λ of newly starting sellers that are of high quality. In practice, eBay is very good at blocking sellers that are offering truly bad service to buyers. Therefore, we think of *low quality* as *not-so-good quality*, and have formulated our question accordingly.

Appendix E contains details on the implementation, summary statistics of characteristics of the participants, and histograms of the survey answers. In brief, we paid and asked questions to 1,000 participants who were living in the U.S. and shopped online at least once a month. The survey was conducted on October 3, 2022. 98.5 percent of the participants indicated that they had at least some experience shopping on eBay, and 65 percent of the participants had bought at least 20 items on eBay over the last 10 years. On average, they were 39 years old. Half of them were female.

Table 5 contains mean responses and corresponding standard errors. In the eyes of the respondents, the likelihood of having a good transaction experience with a high-quality seller is about 88 percent. With a not-so-good quality seller, this drops to 49 percent. Furthermore, the respondents believe that the likelihood that a newly entering seller is of high quality is about 69 percent. We used these values already in Figure 1, and they will serve as the input into our

Table 5: Survey responses

parameter	mean	standard error	explanation
ρ_h	0.8838	0.0037	$\Pr(s = g \mid q = q_h)$
ρ_l	0.4931	0.0075	$\Pr(s = g \mid q = q_l)$
λ	0.6946	0.0063	prior new seller is high quality

Notes: Mean responses and corresponding standard errors. 1000 participants. See appendix E for details.

simulation study we report on in the following section.

These results are remarkable by themselves, as we explained to the survey participants that they should think of eBay sellers as either being of high or not-so-high quality. The results here suggest that on average, buyers believe that even if all sellers on eBay would be of high quality, then the likelihood to have a positive transaction experience would be about 88 percent. This is substantially lower than the percentage of the ratings on eBay that are positive, which is about 98.6 percent (Table 1, Panel C, column 1, 3rd number). This suggests that buyers are aware of the rating bias. The survey results are closer to the finding by Dellarocas and Wood (2008) that buyers are satisfied in 81.5 percent of the transactions and mildly dissatisfied 17.4 percent of the time. They are also in line with the evidence presented by Nosko and Tadelis (2015) suggesting that average ratings on eBay are biased.

7 The speed of learning

We now turn to the question of how fast consumers can learn about the seller type. For this, we calibrate the model to moments of the data and perform appropriate simulations.

7.1 Setup

For our model of the typical buyer's rating decision, we posited that our buyer formed a prior belief about a specific seller's quality before conducting a transaction with that seller, based on the initial belief λ that a randomly drawn seller currently on the platform is of good quality, and on the scalar y that summarized all publicly available information about our particular seller.

To simulate dynamics, we assume that our buyer has rational expectations, and knows the entire history of ratings, including which transactions have not been rated.[31] In the present context, rational expectations mean that a buyer uses all the available information to form his beliefs. Assuming rational expectations is useful for our purposes because it allows us to solve for the implied belief at any point in time.

Turning to specifics, index a seller's transactions by t and let z_t be the state variable containing information about the seller's rating record that is available to the buyer before she conducts transaction t. The state variable z_t is empty at $t = 1$, as the seller does not have prior transactions. It contains one element at $t = 2$, and so on. The elements of z_t are either 0 when no rating was left, or contain the rating r_t that was left on transaction t. Prior beliefs for the buyer in transaction t are denoted by μ_t. In transaction t, the buyer receives signal $s_t \in \{g, b\}$ and then updates her beliefs to μ_t^s.

We use a linear specification for the net benefit derived from leaving a rating,

$$u_t^s \equiv b^s + b \cdot |\mu_t^s - \mu_t| - c_t.$$

There is a baseline benefit b^s of leaving a rating. It depends on the signal s, allowing for an additional, e.g., psychological cost of leaving a negative rating. Naturally, $b > 0$. $\mu_t \equiv \mu(\lambda, z_t)$ is determined by Bayes' rule (see below). We assume that c_t is uniformly distributed (on the unit interval).[32] This implies that the probability that a rating is left is equal to $\bar{u}_t^s \equiv b^s + b \cdot |\mu_t^s - \mu_t|$.[33]

7.2 Simulating paths of ratings

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

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

[32] We also experimented with a logit specification, where c_t is logistic. The results were similar. The advantage of assuming a uniform distribution is that the parameters can be interpreted directly.

[33] This requires $0 \leq \bar{u}_t^s \leq 1$. For the parameter values we use (Table 5 above and Table 6 below), this condition holds, as $|\mu_t^s - \mu_t|$ is less than 0.4 (Figure 1).

Now, if a rating was left, the one element of z_2 is called r_1. The buyer in the second transaction knows that. In addition, $\mu_1 = \lambda$ is common information, so the buyer in the second transaction has all the information to calculate μ_1^s; hence $\mu_2 = mu_1^s$. If no feedback was left in the first transaction, the buyer in the second transaction knows—by solving the model—the probability π_1^s that the buyer in the first transaction leaves a rating for a given signal s. Using this, she obtains by Bayes' rule

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

where

$$A = \mu_1 \cdot \left\{ \rho_h \cdot (1-\pi_t^g) + (1-\rho_h) \cdot (1-\pi_t^b) \right\}$$
$$B = (1-\mu_1) \cdot \left\{ \rho_l \cdot (1-\pi_t^g) + (1-\rho_l) \cdot (1-\pi_t^b) \right\}.$$

In words, A is the joint probability that a seller is high quality and no rating was left.[34] $A+B$ is the probability that no rating was left.[35]

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

7.3 Calibration and parameter values

We used the survey responses for ρ_h, ρ_l, and λ reported in Table 5, and calibrated the three utility parameters b_0^g, b_0^b, and b to our eBay data. For this, we specified a set of moments. The moments are related to the two empirical predictions we have discussed above. We used the L2 norm of those moments as a criterion function to find the parameter values reported in Table 6. Appendix C contains details on the calibration procedure.

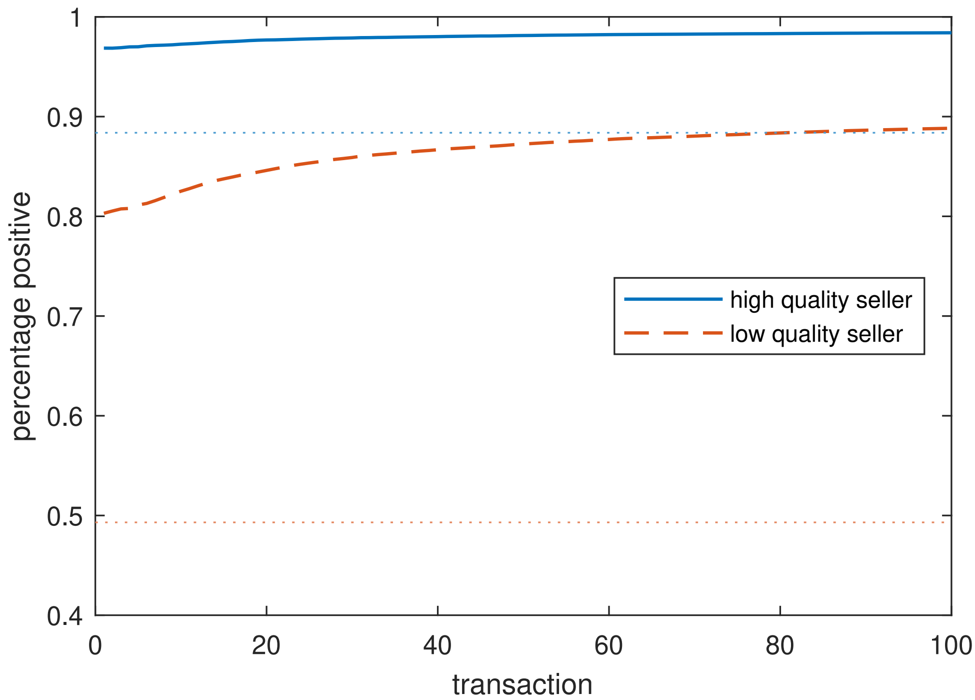
[34]This is equal to the probability that the seller in the first transaction was high quality (μ_1) times the probability that no rating was left provided that the seller is high quality (in curly brackets). The expression in curly brackets is the likelihood that a good signal was received from a high-quality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

[35]Figure D.6 in Appendix D shows the change in beliefs associated with a missing rating when $\rho_h = 0.8838$ and $\rho_l = 0.4931$, the parameter values we elicited in our survey (details in Section 7.3).

Table 6: Parameter values

parameter	value	explanation
b^g	0.7359	baseline probability to rate when experience was good
b^b	0.0747	baseline probability to rate when experience was bad
b	0.3371	additional benefit from leaving a rating when beliefs change

Notes: This table shows the values of the calibrated parameters. See Appendix C for details.



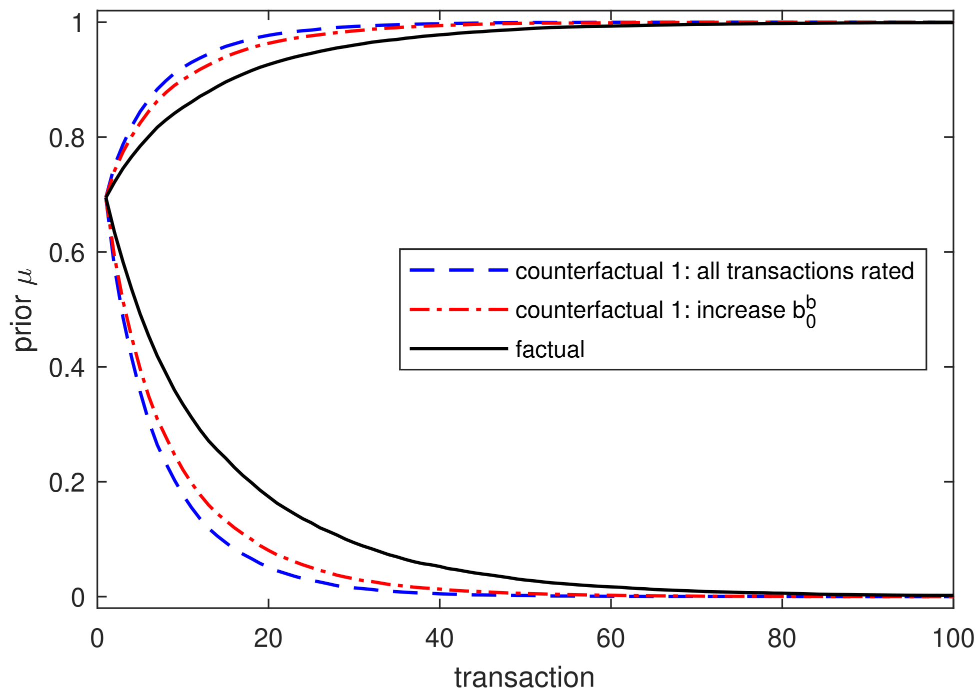
Notes: Evolution of percentage positive feedback to date for a low- and high-quality seller, respectively when we average across 10,000 simulation runs. The upper and lower dotted lines are the respective likelihood of having a good experience with a high- and low-quality seller. Based on setup and procedure described in Section 7.1 and parameter values in Section 7.3.

Figure 3: Rating bias and evolution of percentage positive ratings

7.4 Rating bias

In line with the survey evidence presented in Section 6, we predict with our model that the percentage of positive ratings is systematically lower than the percentage of positive experiences.

In Figure 3, we show how the percentage of positive ratings to date evolves for good and bad sellers, and compare it to the respective likelihood that a transaction experience is positive. The figure shows that ratings are biased and the bias becomes stronger over time. Even though the likelihood of having a good transaction experience does not depend on time, the percentage of positive ratings to date increases for both types.



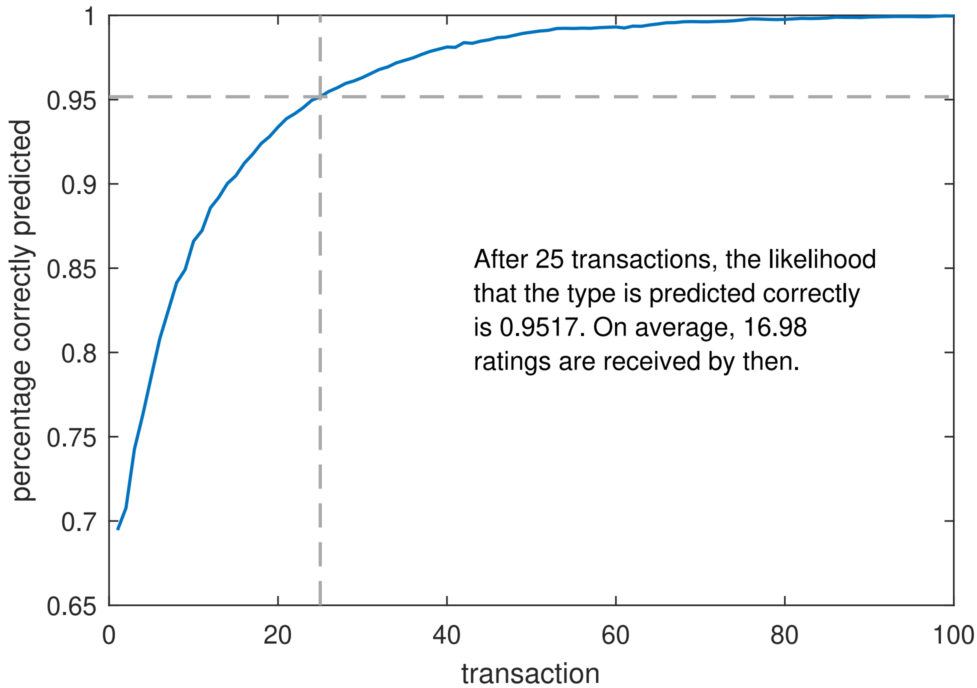
Notes: Evolution of the prior for a high-and a low-quality seller in two counterfactual scenarios. Average across 10,000 simulation runs. In scenario 1, all transactions are rated. In scenario 2, the baseline probability to rate is the same for good and bad experiences. Based on setup and procedure described in Section 7.1 and parameter values in Section 7.3.

Figure 4: Learning from ratings

7.5 Evolution of beliefs

Figure 4 contains the simulated evolution of the separate priors for the average high- and low-quality seller (solid lines). This can be seen as a measure of the speed of learning. Before the first transaction takes place, the prior for both high-quality and low-quality sellers is $\lambda = 0.6946$. The prior increases for the average high-quality seller and decreases for the average low-quality seller. Learning is already fast in the baseline situation. After 50 transactions, buyers essentially know whether a seller is of low or high quality.

We conducted two counterfactual experiments. In the first counterfactual experiment, we assume that all buyers rate their transactions. This shows how fast buyers can learn in principle. In the second more realistic counterfactual experiment, buyers are encouraged to rate. For this, we change the baseline probability to rate when their experience with the seller was bad from $b_0^b = 0.0747$ to equal that when their experience was good, $0.7359 = b_0^g$. We see that the speed of learning is almost as fast as when all transactions are rated. Overall, our simulations show that buyers can learn fast about seller quality even if ratings are given selectively—and more



Notes: The accuracy is defined as the likelihood that a seller is correctly classified. A seller was classified as high type when the prior μ was above a threshold value that depended on the number of transactions. We solve for the optimal thresholds under the assumption that the fraction of high-quality sellers is λ. Calculated for 10,000 simulated sellers. Based on setup and procedure described in Section 7.1 and parameter values in Section 7.3.

Figure 5: Accuracy of seller classification

importantly even if aggregate ratings are biased, as long as the typical buyer does not remain ignorant about that rating bias.

7.6 Accuracy of seller classification

Another way to quantify the speed of learning is to assess the accuracy of seller classification. Accuracy is defined as the percentage of seller types that are predicted correctly. We assumed that the fraction of high-quality sellers in the population is given by λ. Then, we classified sellers as high-quality when the prior is above a threshold value. We allowed the threshold to depend on the number of transactions at the moment of the classification. We numerically solved for the threshold value that maximized the accuracy. This is possible because we use simulated data and we observe the seller type in our simulations. Then, we calculated the percentage of sellers that was correctly predicted.

Figure 5 shows the result. In the first transaction, beliefs are the same for all sellers and given by λ. Therefore, all sellers are classified as high-type, which means that a fraction λ is classified correctly. Over time, the prior for high types increases, and the prior for low types decreases (see Figure 4 above). Therefore, more and more buyers are classified correctly. After 25 transactions, the likelihood that the type is predicted correctly is 0.9517. On average, 16.98 ratings are received by then.

8 Summary and policy recommendation

In this paper, we study how fast online ratings reveal seller types. For this, we develop a new model of rating behavior. We posit that the buyer is more likely to leave a rating, the more she learns in a transaction about seller quality. We show that two central empirical predictions of the model are in line with patterns in administrative data from eBay, elicit beliefs of market participants in a survey, and calibrate the model to the eBay data to quantify the speed of learning.

A central element of the model is the typical buyer's belief about the quality of the seller. This belief is informed by the seller's public rating record. The buyer's experience from the transaction generates an additional informative private signal. The buyer is more inclined to share this signal, the more that experience changes her belief about seller quality. This implies that ratings are informative, as they are more likely to be given if a buyer learns more from the transaction, but they are also selective, as the likelihood to rate depends on the prior and the signal. The two empirical predictions are that first earlier transactions are more likely to be rated, and second, in contrast to predictions generated elsewhere, a negative shock to beliefs about seller quality leads to an increase in the likelihood that the seller is rated. These predictions are born out of the data.

Selective ratings lead to a bias: the average rating for a seller is higher than the average transaction experience. We generate new evidence using an online survey suggesting that buyers on eBay are aware of this. We calibrate our model to study how much can still be learned from online ratings about seller quality. Our simulation results show that even though the bias is quantitatively important, online ratings are very informative and learning is fast. After only 25 transactions, the likelihood of correctly predicting the seller type is above 95 percent.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2022). Learning from reviews: The selection effect and the speed of learning. *Econometrica 90(6)*, 2857–2899.

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

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

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

Cabral, L. (2012). Reputation on the internet. *The Oxford Handbook of the Digital Economy*, 343–354.

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.

Carnehl, C., A. Stenzel, and P. Schmidt (2023). Pricing for the stars: Dynamic pricing in the presence of rating systems. *Management Science*.

Chakraborty, I., J. Deb, and A. Öry (2022). When do consumers talk? Cowles Foundation Discussion Paper No. 2254.

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

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

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

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

Fradkin, A., E. Grewal, and D. Holtz (2021). Reciprocity and unveiling in two-sided reputation systems: Evidence from an experiment on airbnb. *Marketing Science 40*(6), 1013–1029.

He, S., B. Hollenbeck, and D. Proserpio (2022). The market for fake reviews. *Marketing Science 41*(5), 896–921.

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

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

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

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

Ishihara, M. and Y. Liu (2017). A dynamic structural model of endogenous consumer reviews in durable goods markets. *Available at SSRN 2728524*.

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

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

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

Luca, M. and O. Reshef (2021). The effect of price on firm reputation. *Management Science 67*(7), 4408–4419.

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.

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

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

Nosko, C. and S. Tadelis (2015). The limits of reputation in platform markets: An empirical analysis and field experiment. NBER Working Paper 20830, National Bureau of Economic Research, Cambridge, MA, U.S.A.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature 36*, 11–46.

Reimers, I. and J. Waldfogel (2021). Digitization and pre-purchase information: the causal and welfare impacts of reviews and crowd ratings. *American Economic Review 111*(6), 1944–71.

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.

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).

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. Working Paper, Paris School of Economics, Paris, France.

Theoretical Appendix

Model setup

The buyer's belief λ that a new seller entering the platform is of good quality is updated to form the prior belief $\mu(\lambda, y)$ using the seller's public reputation record summarized by the index $y \in \mathbb{R}$. That index could be any function of the information available to date on the seller—in particular a function of previous ratings. It could either be computed by the platform or the buyer herself. It could be thought of as a known function that is increasing in the number of positive, and decreasing in the number of negative ratings. Examples are the percentage positive feedback, an estimate of the likelihood of having a positive experience with the seller, or the fraction of transactions with a positive rating.[36]

To update her initial belief λ in a Bayesian sense, the buyer specifies probability $\sigma_h(y) = \Pr(y|q = q_h)$ that our seller's reputation score is y, given that he is of high, and $\sigma_l(y) = \Pr(y|q = q_l)$, given that he is of low quality. The Bayesian belief that the seller is of high quality is then given by

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

This shows that $\mu(\lambda, y)$ is increasing in λ. That a better rating record increases the belief $\mu(\lambda, y)$ is specified by the monotone likelihood ratio property (MLRP) commonly assumed in models of Bayesian updating,

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

Intuitively, a rating record is the more likely to reflect a high-quality seller the better it is. To simplify notation, we now keep the dependence on λ and y implicit unless needed for the argument, and simply write μ for the buyer's prior belief about seller quality before conducting a transaction. The binary signal s about the seller's quality received when conducting the transaction is considered informative: the probabilities $\rho_h \equiv Pr(s = g|q = q_h)$ and $\rho_l \equiv Pr(s = g|q = q_l)$ satisfy

[36] In Section 7 we assume that buyers have rational expectations and are fully informed about the rating record in all past transactions, including information on transactions that were not rated.

Assumption 2 (Signal). *(i) $\rho_h > \rho_l$ and (ii) $\rho_h > 1/2$.*

Hence the probability that the signal is good is assumed to be higher for a high quality than a low quality seller, and larger than the probability that the signal is bad.

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

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

$$u^s(d,c) \equiv b^s(d) - c \geq 0. \tag{3}$$

The symmetry assumption behind the absolute difference reflects the idea that the intensity of learning drives the rating decision. The larger d, the more the buyer has learned from the transaction, and the more she may expect others to learn from her rating. This rationalizes the positive dependence of $b(d)$ on d. In addition, $b^s(0)$ can capture a baseline inclination to leave a rating, even if there is no learning, and that this baseline inclination varies with the signal s. In all, the likelihood that a buyer leaves a rating is given by $b^s(d)$.[37]

Updating of beliefs

By Bayes' rule, the buyer forms a *posterior belief* μ^g that the seller is of high quality when her experience was good,

$$\mu^g \equiv Pr(q = q_h | s = g) = \frac{\mu\rho_h}{\mu\rho_h + (1-\mu)\rho_l} \tag{4}$$

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

and

$$\mu^b \equiv Pr(q = q_h | s = b) = \frac{\mu(1-\rho_h)}{\mu(1-\rho_h) + (1-\mu)(1-\rho_l)} \tag{5}$$

when her experience was bad.

Figure 1 in Section 2 illustrates this. Starting with $\mu(\lambda, 0)$, the figure shows patterns that are specific to the parameter values we have chosen, and other patterns that hold more generally. First to the more general patterns. It follows directly from (4) and (5), that there is no updating when the prior is either $\mu = 0$ or $\mu = 1$, and there is updating if $0 < \mu < 1$. A positive signal always implies that beliefs increase, and a negative one that they decrease. Moreover, (4) and (5) are continuous in μ. For the specific parameter values chosen for the figure, positive signals are less informative about seller quality than negative signals. By this, for any prior μ the buyer learns more from a negative experience, and most when priors are between 0.5 and 0.8.

Change in beliefs and the rating record

We now study how the amount of learning that influences the buyer's rating decision depends on the seller's rating record through its effect on prior beliefs μ. Recall that by Assumptions 1 and 2 prior beliefs μ are strictly increasing in the rating record y. Therefore, we can focus on the dependence of learning on the level of prior beliefs μ.

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

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

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

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

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

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

$$\mu^g - \mu = \frac{\mu \rho_h}{\mu \rho_h + (1-\mu)\rho_l} - \mu$$

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

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

is equivalent to

$$\rho_h \rho_l \gtreqqless \mu^2 (\rho_h)^2 + 2\mu(1-\mu)\rho_h\rho_l + (1-\mu)^2(\rho_l)^2.$$

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

$$\frac{\rho_l}{\rho_h} \gtreqqless \frac{\mu^2}{(1-\mu)^2}. \tag{6}$$

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

Furthermore, by Assumption 2,

$$\frac{\rho_l}{\rho_h} < 1.$$

Hence equality in (6) holds for prior beliefs

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

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

$$\mu - \mu^b = \mu - \frac{\mu(1-\rho_h)}{\mu(1-\rho_h) + (1-\mu)(1-\rho_l)}.$$

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

$$\frac{1-\rho_l}{1-\rho_h} \gtreqless \frac{\mu^2}{(1-\mu)^2}.$$

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

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

For an interpretation, recall our central posit that a buyer tends to rate a transaction when she has learned a lot from it relative to the quality she has expected beforehand from interpreting the seller's performance score. Formally, the probability to rate increases with the absolute difference between prior and posterior beliefs. By Proposition 1, if that performance score was low, a favorable rating shifts our buyer's posterior belief sufficiently much to eventually contribute another rating. But if the performance score was already high, another favorable rating does not shift the belief enough to induce a rating.

This holds for all transactions conducted by our buyer, no matter her experience. However, the difference between positive and negative experiences is one of detail contained in part (iii) of the proposition: The cutoff in terms of the rating record y between when the rating probabilities increase and decrease is lower for positive experiences than for negative ones—with the difference between the two cutoffs increasing in the difference between ρ^h and ρ^l. Rather intuitively, the probability that our buyer rates negatively continues to increase with further positive ratings, when that of rating positively already decreases.[38]

Rating dynamics

Here we study how the inclination of the typical buyer to rate good and bad transaction experiences changes over the sequence of transactions conducted by the typical seller, and over sellers by type. For a given buyer belief μ, denote the probability that she leaves a rating if the experience was good by $\hat{p}^g(\mu)$, and by $\hat{p}^b(\mu)$ if it was bad. Since $\mu(y)$ is monotonically increasing, Proposition 1 implies that $\hat{p}^b(\mu)$ increases for $\mu < \hat{\mu} \equiv \mu(\hat{y})$ from $\mu = 0$ and decreases thereafter towards $\mu = 1$, and $\hat{p}^g(\mu)$ increases for $\mu < \check{\mu} \equiv \mu(\check{y})$, and decreases thereafter. In particular, it establishes that both functions are increasing in μ for μ sufficiently close to 0, and decreasing in μ for μ sufficiently close to 1, with $0 < \hat{\mu} < 0.5 < \check{\mu} < 1$.

We now translate this into predictions by seller type that are not conditional on whether the buyer's actual experience is positive or negative. Let $p_i(\mu), i \in \{h, l\}$ denote the likelihood that a rating is given to a high-quality vs. a low-quality seller as a function of μ. Then

$$p_h(\mu) = \rho_h \cdot \hat{p}^g(\mu) + (1 - \rho_h) \cdot \hat{p}^b(\mu) \tag{7}$$

and

$$p_l(\mu) = \rho_l \cdot \hat{p}^g(\mu) + (1 - \rho_l) \cdot \hat{p}^b(\mu). \tag{8}$$

Proposition 1 implies that both $p_h(\mu)$ and $p_\ell(\mu)$ increase in μ for $\mu < \hat{\mu}$, and decrease in μ for $\mu > \check{\mu}$. Furthermore, $p_h(\mu)$ is maximal for some $\hat{\mu}_h \in (\hat{\mu}, \check{\mu})$, and similarly $p_l(\mu)$ for some $\check{\mu}_l \in (\hat{\mu}, \check{\mu})$. Consider for the moment a population of sellers $i, i \in N$. Denote the typical buyer's beliefs on the quality of seller i before transaction t by μ_{it}. Denote the c.d.f. of beliefs

[38] 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.

in transaction t for sellers i by

$$F_{\mu_{it}|h,t}(\mu) \equiv \Pr(\mu_{it} \le \mu | q = q_{h,t})$$

if they are of high quality and correspondingly, $F_{\mu_{it}|\ell,t}$ if of low quality. Consider two values of the transaction index t_1 and t_2, with $t_2 > t_1$.

Let us define learning about seller quality in the following way: (a) if i is a high-quality seller, then $F_{\mu_{it}|h,t_1}(\mu) > F_{\mu_{it}|h,t_2}(\mu)$ for all μ; (b) if i is a low-quality seller, then $F_{\mu_{it}|l,t_1}(\mu) < F_{\mu_{it}|l,t_2}(\mu)$ for all μ. By (a), the c.d.f. at t_2 is to the right of the one at t_1, which implies that beliefs are converging to 1. By (b) the c.d.f. at t_2 is to the left of the one at t_1, which implies that beliefs are converging to 0.

Consider the high-quality sellers. The likelihood that a rating is left for the average seller for transaction t is

$$\tilde{p}_{h,t} \equiv \int p_h(\mu)\, dF_{\mu_{it}|h,t}(\mu).$$

As there is learning about the seller type, the higher t, the more sellers will have a μ such that $p_h(\cdot)$ is decreasing. In the limit, all high-quality sellers will have a μ that is equal to 1. Overall, the likelihood of receiving a rating for the average seller may first increase. But will decrease from some point onward.[39] By a similar argument, the likelihood of low-quality sellers may first increase but will eventually decrease. For them, learning means that $\mu_{it} = 0$ in the limit.[40] We summarize in

Corollary 1 (Empirical prediction 1)**.** *Assume that there is learning about seller quality in the sense that the distribution of beliefs of high-quality sellers, $F_{\mu_{it}|h,t}(\mu)$, is increasing in t for all μ and the distribution of low-quality sellers, $F_{\mu_{it}|l,t}(\mu)$, is decreasing in t for all μ. Then, the likelihood that a rating is given may first increase in t, will then reach its maximum, and will ultimately be a decreasing function in t.*

[39] It may well be that it decreases from the very beginning, if the buyers' initial belief λ is big enough, and enough sellers have a μ_{it} big enough even for low t, so that from one transaction to the next the likelihood to receive a rating decreases for them.

[40] Note that buyers will stop buying from them earlier, implying that we may not observe them long enough. They will very likely never be in the sample on which we focus our analysis.

Interim ratings and their effect

Our model obviously allows buyers to have bad experiences with a high-quality seller. We now study how within our modelling framework the likelihood that buyers share positive or negative experiences is conditioned on immediate previous ratings rather than the rating aggregate only. In line with empirical observations, we assume a negative rating arises with with substantially lower probability than a positive one. By focusing on a window in which the seller cannot react strategically to such a negative rating, we can isolate the ensuing buyers' reactions to this event.

Rating decisions are almost always taken with a delay after the purchase. In the interim period between purchase and rating decisions, new ratings may arrive. To study this, index beliefs and rating records by time t and consider the following observation and decision taken by the buyer:

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

By Proposition 1 the negative rating shock increases the probability to rate when $y > \check{y}$. We wish to determine whether upon observing a new negative rating, the buyer rates a negative experience with a higher or lower probability than a positive one, both relative to the baseline probabilities. Let $f_1(y) \equiv \mu^g(y) - \mu(y), f_2(y) \equiv \mu(y) - \mu^b(y)$, and $z(y) \equiv f_1(y) - f_2(y)$. $z(y)$ is the difference in the probability of rating after a positive and a negative experience, respectively.

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

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

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

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

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

Then,

$$\operatorname{sgn}\left\{\frac{d}{d\mu}z(\mu(y))\right\} = \operatorname{sgn}\left\{\frac{\rho^h\rho^l}{[\mu\rho^h+(1-\mu)\rho^l]^2}+\frac{(1-\rho^h)(1-\rho^l)}{[\mu(1-\rho^h)+(1-\mu)(1-\rho^l]^2}-2\right\}. \quad (9)$$

From the discussion ensuing Assumption 2, there is a $\bar{y}(\varepsilon) > \check{y}$ such that $\mu(y) > 1-\varepsilon$ for all $y > \bar{y}(\varepsilon)$, implying that μ approaches 1 as y becomes sufficiently large. For simplicity, evaluating (9) at $\mu = 1$, we obtain $\frac{\rho^l}{\rho^h} + \frac{1-\rho^l}{1-\rho^h} \gtreqless 2$. Expanding and simplifying, we get $\rho^h > \frac{1}{2}$. Hence $\frac{d}{dy}z(\mu(y)) > 0$ as y gets large, and thus $\mu(y)$ approaches 1. □

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

This leads to our second empirical prediction:

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

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

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

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

Online Appendix

A A more general version of our model

We formulated our model for the simple two-by-two-by-two case in which seller quality is either high or low, buyer experiences can either be good or bad, and ratings—if given at all—are positive or negative. This allowed us to highlight the idea that the rating decision depends on the amount of learning. The model also fits well to the eBay context in which we conduct our empirical analysis. Because of its simplicity, we were able to calibrate the model to eBay data after eliciting some of the model parameters in a survey.

In this appendix, we generalize our model to more than two types of transaction experiences, more than two types of ratings, and more than two types of sellers. We show that we can still simulate the evolution of beliefs in a similar way as in Section 7 and that the main driving force of the much more general model is the same as in our simple model: buyers learn from online reviews even if they are biased; and therefore, signals become less and less informative.

A.1 Setup

We assume that there is a continuum of seller types θ_i. Seller types are drawn from a beta distribution with parameters $\alpha = 10$ and $\beta = 5$.[1] Transaction experiences for seller i are drawn from a normal distribution with mean θ_i and variance $\sigma^2 = 0.04$. Hence sellers differ with respect to the mean of the distribution of experiences they produce, but not with respect to the variance.

Buyers' prior beliefs are given by the same beta distribution that the sellers are drawn from.[2] They observe transaction experiences and can leave a rating on a five-point scale. Experiences are mapped into ratings, if they are given, in the following way: if the experience is below 0.5, then a rating of 1 is given, if it is between 0.5 and 0.55, then a 2 is given, if it is between 0.55 and 0.6, then a 3 is given, if it is between 0.6 and 0.7, then a 4 is given, and if it is above 0.7, then a 5 is given. The likelihood that a rating is given is equal to $\max\{0, \min\{1, b_j + b \cdot m\}\}$ for $j = 1, \ldots, 5$.[3] b_j is the baseline inclination to leave a rating, m is the amount of learning in

[1]This choice and other distributional choices are just for illustration. The model can be solved for other choices in the same way.

[2]Results are similar when we use other distributions, e.g. a uniform distribution.

[3]This can be micro-founded in a similar way as in Section 7.

the transaction, and b is a parameter translating the amount of learning into a probability. As in the main part of the paper, the amount of learning m is specified as the Euclidean distance between the posterior and the prior density. The specific parameter values are given by $b_1 = 0.5$, $b_2 = 0.4$, $b_3 = 0.3$, $b_4 = 0.5$, $b_5 = 0.6$.[4] Ratings are biased in the sense that not all b_j are equal to one another.

A.2 Learning from experiences

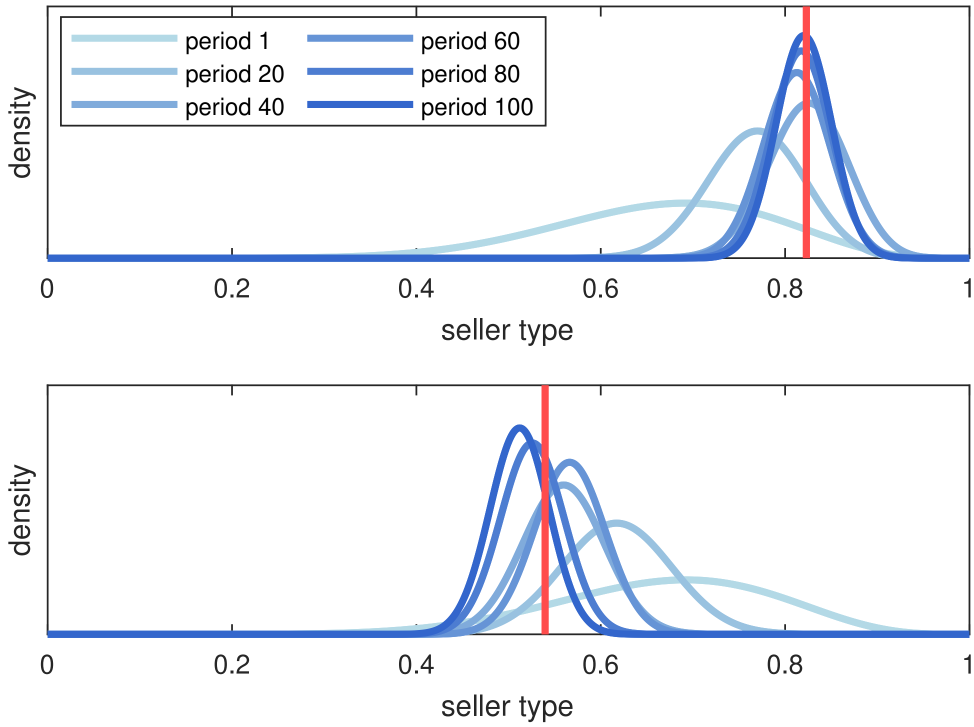
In the simple model in the main part of the paper, buyers share experiences when they leave a rating. In the more general model considered here, they share a coarse version of the signal on a five-point scale. Therefore, the buyer learns more from a transaction than another market participant who only observes the rating. Therefore, we calculate two posterior distributions on a grid for the seller type, one where the updating is based on the signal and one where the updating is based on the rating. We use 1,000 grid points between 0 and 1 (including boundaries). As in Section 7, the buyer is considered to observe the full history of ratings (hence also whether no rating was left). Calculating the posterior for buyers (who learn from experiences) and other buyers (who learn from ratings and their absence, respectively) is technically more difficult and involves more steps and computations. However, the main lines are similar to what we do in Section 7.

A.3 Findings and discussion

Figure A.1 illustrates our general findings. It shows the distribution of prior beliefs for 2 simulated sellers at 6 points during the learning process. The vertical red line is respectively the seller type, θ_i. Before any transaction and rating, no learning has taken place (so this is the beta distribution we assume). The successive concentration of beliefs demonstrates learning of potential buyers from ratings rather than their own experience with the seller.

We have shown in Section 7, that in our simple model learning takes place over time even if there is rating bias, and that beliefs μ converge to either 0 or 1. Mechanically, also the variance

[4] The model can generally produce the well-known J-distribution for ratings, by which on a scale between very positive and very negative ratings, many are very positive and a few are very negative, with very few in-between. For example, Hu et al. (2009) find that pattern for reviews provided on Amazon.



Notes: Figure shows the evolution of beliefs for two simulated sellers. They differ in the mean of the distribution of experiences they generate (θ_i, respectively the vertical red line), but not in the variance (σ^2). The prior (distribution in period 1) is respectively given by the beta distribution with parameters $\alpha = 10$ and $\beta = 5$. See Online Appendix A for details.

Figure A.1: Evolution of beliefs for two simulated sellers of beliefs (which is given by $\mu \cdot (1-\mu)$) converges to zero. We have shown in Section 2 that as a consequence less and less is learned from experiences and therefore, it is less and less likely that experiences are shared (our first empirical prediction).

In the more general model, the mean and the variance of the beliefs are not mechanically linked to one another. But we generally find, at least for the parameter values we chose and for others we tried, that the mean of the beliefs converges to the true value and the variance shrinks over time. As for the simple model, this is the case even though ratings are biased.

Overall, this illustrates that it is conceptually straightforward to generalize our model and that it is technically possible to solve it. The key advantages we see in our simple model are that it captures the main driving forces and that we can calibrate it to our data because we can easily elicit three of its parameters in a survey and find the remaining three through calibration.

B Additional results and robustness

B.1 Empirical prediction 1

B.1.1 Alternative measures

Table B.1 reproduces Table 2 for DSRs as the outcome. In Tables B.2 and B.3 we use alternative definitions of buyer experience indicated in the Notes. Table B.4 reproduces column (1) in Table B.3 for separate DSR categories. None of these invites a reinterpretation of our results.

B.1.2 Changes in consumer surplus over time

In Acemoglu et al. (2022)'s model, the likelihood to rate is directly related to the surplus consumers get from a transaction. A lower surplus for lower transactions could therefore be an alternative explanation for the pattern in Figure 2. In Figures B.2, we study the evolution of consumer surplus over time. In Figure B.2a, we study auctions and use the winning bid minus the sales price in an auction as a proxy for consumer surplus. Auctions on eBay share many features of second-price auctions and therefore the winning bid is a good approximation of the winner's maximum willingness to pay. In Figure B.2b, we use the sales price as inversely related to consumer surplus. For this, we use transactions with posted price where the product had a product ID and was in new condition. We then regress the logarithm of price on dummy variables for the transaction index, controlling for seller fixed effects, transaction month fixed effects, and product ID fixed effects, to account for potential changes in the product portfolio as a seller grows. In both figures, we do not observe dramatic changes in consumer surplus; if anything, Figure B.2b suggests a price decrease as the transaction index increases, which by Acemoglu et al. (2022)'s model should lead to an increase in the likelihood of rating. Hence, the results suggest that changes in consumer surplus over time cannot explain the empirical patterns in rating behavior in Figure 2.

Table B.1: Probability to receive DSR

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

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

Table B.2: Probability to receive feedback with alternative definitions buyer experience

	(1) feedback	(2) feedback	(3) feedback	(4) feedback	(5) feedback	(6) 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), buyer experience is indicator for registration before 01feb2005, buyer inclination to leave feedback is indicator for above 0.857143 (median). In columns (4)-(6), buyer experience is indicator for at least 78 transactions (75th percentile), buyer inclination to leave feedback is indicator for at least 0.981982 (also 75th percentile). Standard errors clustered at seller level, account for heteroskedasticity. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

Table B.3: Event study with alternative definition of experience

	(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 B.4: Event study by DSR category

	(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. Low means 1 or 2 out of 5. *** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

B.1.3 Effect of ratings on price

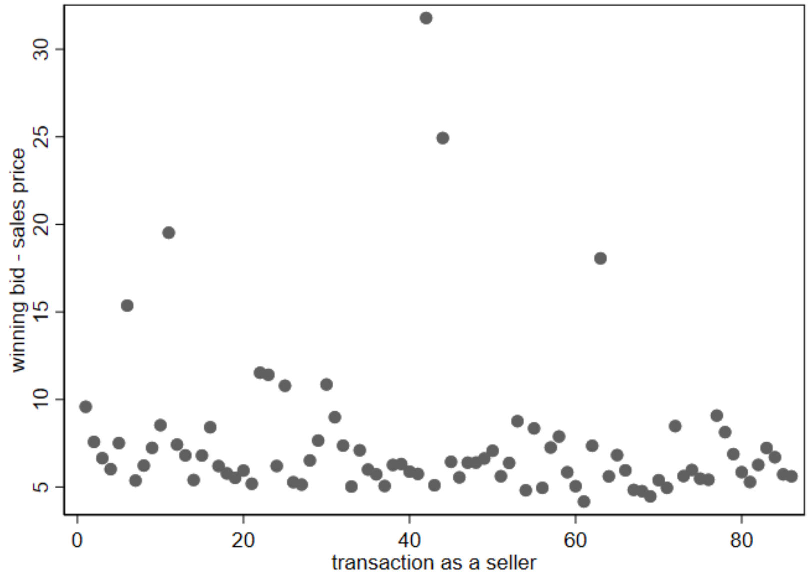
The central idea behind our model is that the likelihood of leaving a rating is higher when the buyer herself has learned more from the transaction. Our samples 86 and 338 consist of sellers who have established themselves on eBay, so their rating record must have improved over time. According to our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

However, not only the potential buyers receive the information communicated through ratings, but also the seller does. He could use the improvement in his reputation to increase the price of the products sold. And if, as we claim, 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.

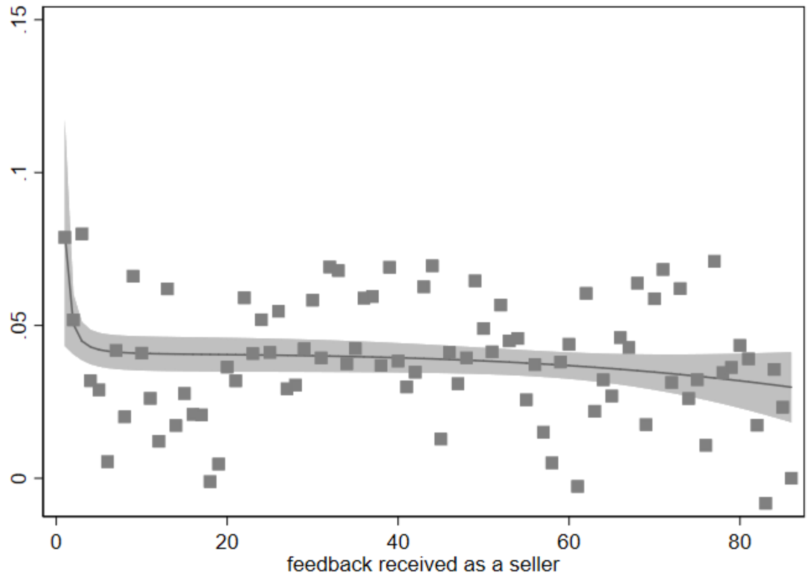
Thus we relate the effect of the length of the feedback record on the price and the likelihood of selling. By controlling for the percentage negative feedback, we measure the effect of the amount of information. Figures B.3 show the result. Both the price and the likelihood to sell depend positively on the number of feedback received. The relationship is concave. Figure B.3a shows that prices are about 7% lower for the first transactions.[5] Figure B.3b shows that

[5] Recall that we include a full set of transaction index dummies and drop the one for the last, 156th transaction.

(a) Winning bid minus sales price



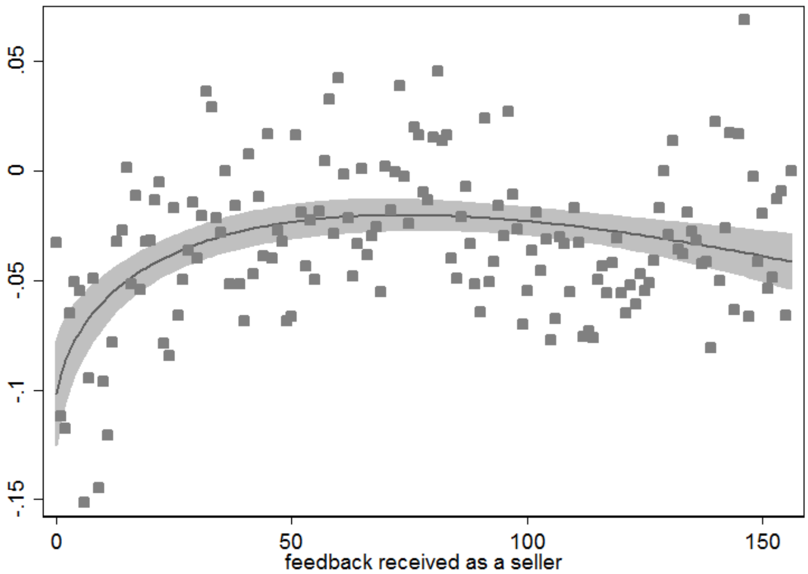
(b) log(sales price)



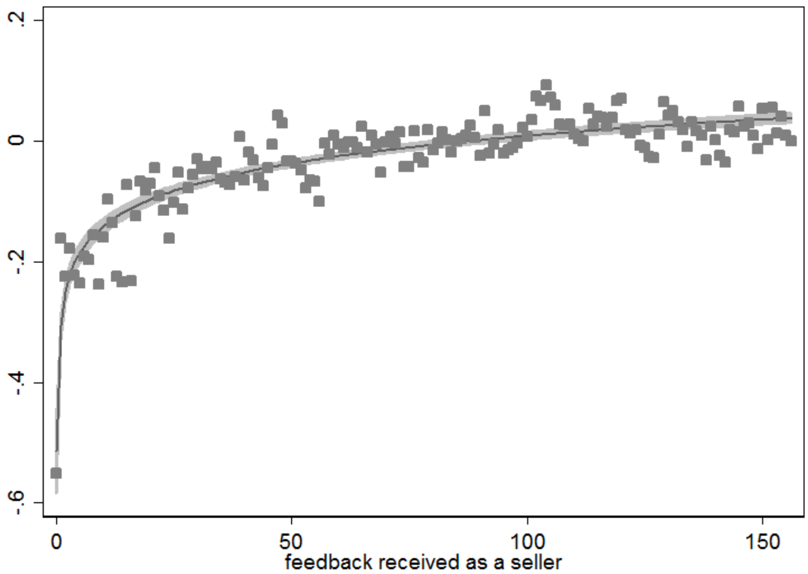
Notes: Figures based on sample 86 show dependence of two measures related to consumer surplus on the transaction index. Figure B.2a: All auction sales, average winning bid minus the sales price against transaction index. Figure B.2b: Transactions with posted price, where the product had a product ID and was in new condition. Regression of logarithm of sales price on dummy variables for the transaction index, controlling for seller fixed effects, transaction month fixed effects, and product ID fixed effects. Dummy for transaction index = 86 dropped as the benchmark. Figure plots coefficients on the transaction index dummies and a local polynomial fit.

Figure B.2: Dependence of consumer surplus on transaction index

(a) Effect of positive feedback on price



(b) Effect of positive feedback on probability of selling



Notes: Sample 338 restricted to transactions with a product ID. First 156 feedback from sellers who had at least 156 feedback (90th percentile of feedback index) in the restricted sample. Figure B.3a: Regression of 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. Sample further restricted to new items in the posted price format. Dummy for feedback index = 156 dropped as the benchmark. Figure B.3b constructed by getting all listings with product ID from sellers in sample 338 in their first year, and regressing the dummy variable for whether a listing sells (1 item or more) on dummy variables of feedback indices, logarithm of listed price, percentage negative to date, controlling for seller fixed effects and product ID fixed effects. Sample further restricted to items in the posted price format.

Figure B.3: Effects of a positive feedback the likelihood of selling increases steeply as a result of the first feedback that are received, and less and less so for later feedback. The figures suggest that the first 50 feedback are particularly informative about the seller type.[6]

B.2 Empirical prediction 2

B.2.1 Difference-in-differences analysis

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

B.2.2 Event study

Table B.6 contains the results of three robustness checks. In column (1) we control for seller-day fixed effects. This means that the coefficient on the class 2 indicator is estimated from transactions in a day in which a transaction with later retracted negative feedback was conducted, and for which feedback was given after the later retracted negative feedback. The coefficient on the

This means that the figure shows prices relative to prices in the 156th transaction. The figure shows that for the first transaction, prices are about 10% lower and from the 50th transaction onward, they are about 3% lower than prices in the 156th transaction.

[6]Figure 2 shows that for the first 86 transactions the likelihood of leaving a rating is about 62 percent on average. This means that the 86 transactions in Figure 2 are comparable to the 50 feedback mentioned here.

class 3 indicator is absorbed in the fixed effect. The effect of main interest is the coefficient on the class 2 indicator. It is very similar to our estimate in column (1) in Table 4. The two placebo tests for which we report results in columns (2) and (3) define classes based on events that are not observable to others, and which should therefore not have an effect. Placebo 1 defines the classes based on a wrong claim without a negative or neutral feedback. A wrong claim is a claim that was later removed because eBay decided that the underlying issue was either no one's fault or the fault of the buyer. Placebo 2 defines it as a positive feedback and a low DSR at the same time, where the low DSR has later been revised to a high DSR. In both cases considered in the Placebos, a buyer was not satisfied with the transaction, but this was not due to the behavior of the seller, similar to the negative feedback that was later changed and that we use in our main analysis. Since confounding by quality shocks is not likely and since claims and low DSRs 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 difference-in-differences approach for the event study (not to be confused with the more classical difference-in-differences specification underlying Table 3). The underlying idea is that negative and neutral ratings both convey that something did not go well in the transaction. Here, we look at the differential effect of the two, that is the difference between a wrongly left negative and a wrongly left neutral rating. 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 as opposed to neutral and create interaction terms between the class indicators and this indicator for a first negative feedback.

Results are presented in Table B.7. The first outcome is whether a negative feedback is left for a given transaction. For this outcome, the coefficients on the class indicators will be the respective likelihood that a negative feedback is left in the three classes. The likelihood is higher in class 2 than in class 1, meaning that there is some evidence for time-varying quality that confounds a first neutral feedback and a subsequent negative one. The coefficient on the class 3 indicator is smaller again, suggesting that quality improves after a first neutral feedback.

Importantly, we are not interested *per se* in this pattern, but in the interaction between the indicators for class 2 and negative. This is the additional effect a first negative feedback has because it is observable to future buyers.[7] 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 it is random whether a wrong feedback is negative or neutral, 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. Even though we use a different sample and a different empirical approach, the size of the effect we estimate is similar to the estimate in Table 4.

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

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

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

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

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

Table B.6: Inclination to leave a negative feedback: robustness checks

	(1) fixed eff.	(2) placebo 1	(3) placebo 2
class 2	0.00869*	0.00408	0.000553
	(0.00455)	(0.00451)	(0.00159)
class 3		0.0176**	0.00367**
		(0.00805)	(0.00170)
seller FE	No	Yes	Yes
transaction index	Yes	Yes	Yes
seller × transaction date FE	Yes	No	No
adj R-squared	0.0926	0.0565	0.0420
observations	19526	12764	73904
number of clusters	162	145	714

Notes: Same specification as column (1) in Table 4. Fixed eff. controls for seller × transaction date fixed effects. 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 $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

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

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

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

Table C.8: Moments

quantity	target value	scale factor	re-scaled difference
probability rating is left for $t = 1$	0.7000	0.01	0.6843
probability rating is left for $t = 86$	0.6689	0.01	-0.6843
percentage negative ratings for $t = 1, \ldots, 86$	0.0164	0.001	0.0014
increase in probability negative rating after negative	0.00335	0.001	-2.4467
increase in probability any rating after negative	0.0419	0.01	-3.9025

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

C Details on the calibration procedure

Here we describe the procedure we used to calibrate our model. Recall that we use the survey responses reported in Table 5 for the parameters ρ^h, ρ^l, and λ. The parameters we intend to find with our calibration procedure are the preference parameters b^g, b^b, and b.

We use 5 moments. The first two columns of Table C.8 show the quantities we target and respective target values. The first quantity is the probability that a rating is left for the first transaction. The target value of 0.7 is taken from Figure 2. The second quantity is the probability that a rating is left for the 86th transaction. The target value is given by the value 0.7 from before minus 0.0311. This is the estimate of -0.362 in the second column of Table 2 times 86 and divided by 1,000. The third target quantity is the percentage of negative feedback for the first 86 transactions. The number 0.0164 is calculated by dividing the share of transactions with neutral or negative feedback reported in Table 1 for sample 86, which is 0.016, by the share of transactions with any feedback, which is 0.670. The fourth and fifth target quantities are the effect of a negative rating on the probability that a negative rating is given and on the probability that any feedback is given, respectively. The target values are our estimates in Table 3 for sample 0.

To find our calibrated parameter values, we simulate sequences of buyer experiences and ratings from our model. We do so for 10,000 simulated sellers and respectively 200 transactions. We then calculate averages of the first three target quantities over all simulated paths. For the last two target quantities, we add an additional negative rating in the 40th transaction for each seller and then calculate the average effect this has on the likelihood that a negative rating is given and on the likelihood that any rating is given.

Next, we construct the goal function that we numerically minimize to find the values of the parameters.[8] The goal function is the sum of the squared re-scaled differences between the respective values of the simulated quantities (in the first column of Table C.8) and the target values (in the second column), with the scale factors specified in the third column.

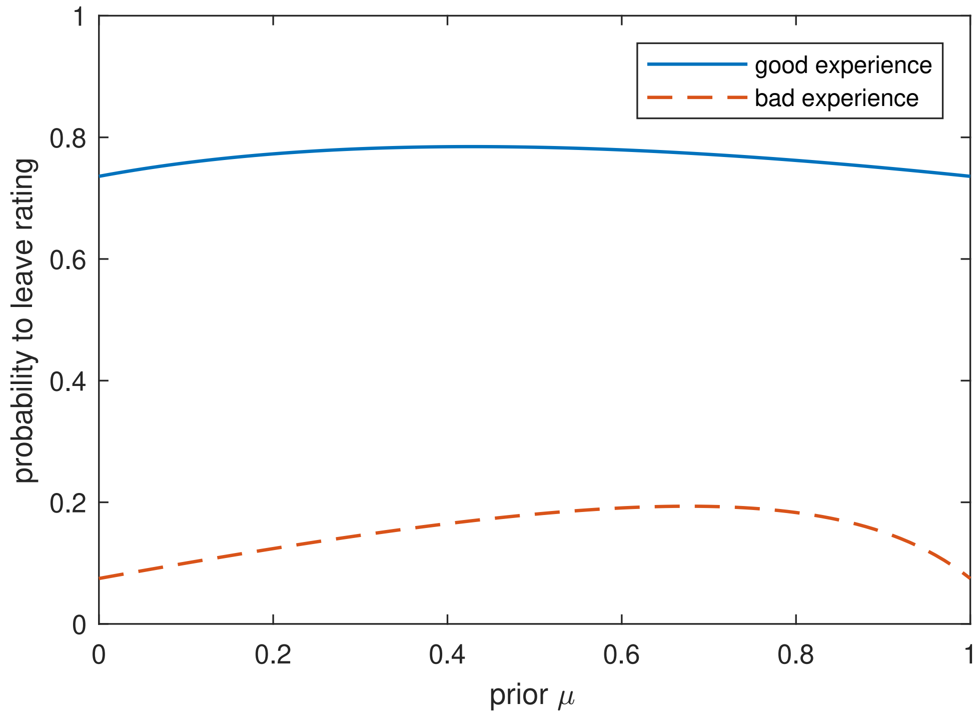
Table 6 in the main text reports the values of the calibrated parameters. The last column in Table C.8 contains the values of the re-scaled difference at the calibrated parameters. Our model fits the first three target quantities particularly well. This is the case because the three parameters b^g, b^b, and b are directly related to the probability of leaving a positive and a negative rating, respectively, and how that probability changes over time.

D Additional figures for the simulation study

Figure D.4 shows how the probability that a rating is left depends on the prior and the experience. From this, we can calculate for a given prior the likelihood that a rating is negative. It is the product of the likelihood that an experience is negative $(1-\rho^l)$ and that a rating is given when the experience is negative (see Figure D.4), divided by the likelihood that a rating is given. The latter likelihood is computed as the likelihood that the experience was positive (ρ^l) times the probability that a rating is given when the experience is positive (see Figure D.4) plus the likelihood that the experience was negative $(1-\rho^l)$ times the probability that a rating is given when the experience is negative (see Figure D.4). The result is plotted in Figure D.5. In Figure D.6, we extend Figure 1 by buyer changes in beliefs based on rational expectations, but not observing any ratings, to show how future buyers can learn from past ratings.

We now turn to simulated paths of experiences and ratings. Figure D.7 shows the evolution of the prior and the percentage of positive feedback for one simulation run for a high-quality seller. One negative rating is left, at the time at which the dashed red line jumps down. The thick blue line shows the evolution of the prior. We start off at λ. The prior increases with every

[8]Simulating 10,000 paths that are each 200 transactions long involves taking random draws for the experiences and for the rating cost. We take $2 \cdot 10,000 \cdot 200$ random draws. As usual, we take the random draws for all 10,000 paths before optimizing over the parameters. That is, we do take the random draws only once and not every time anew when we evaluate the goal function at another trial value of the parameters. We use the same random draws to simulate the effects of an additional negative rating in the 40th transaction.



Notes: Probability that a rating is left, by experience. Based on setup and procedure described in Section 7 and parameter values in Tables 5 and 6.

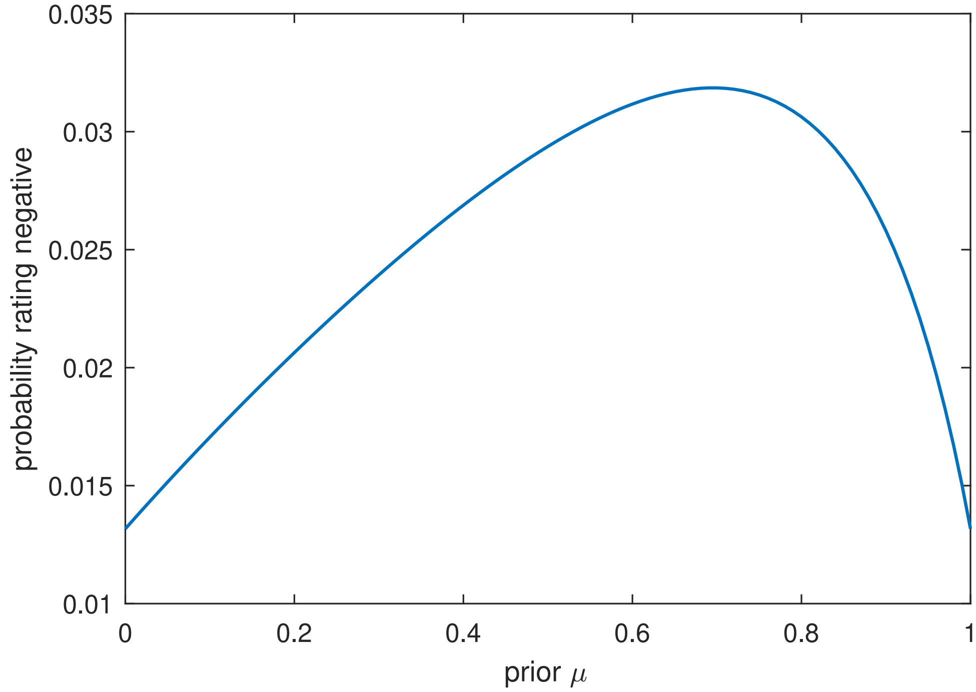
Figure D.4: Probability to leave rating

positive rating. The small downward movement is due to no feedback being left, as no feedback is more likely to be based on a bad signal. The horizontal dotted line shows the likelihood that a positive signal is received, which is equal to ρ_h. Interestingly, after a few transactions, the percentage positive feedback lies above that value.

The remaining figures were obtained by simulating 10,000 paths and taking the average (as we did in our calibration procedure). Figure D.8 shows the distribution of percentage positive feedback after 50 transactions, for high- and low-quality sellers, respectively. Figure D.9 shows how the corresponding distributions of beliefs evolve. Figure D.10 shows the evolution of the prior when buyers ignore the information contained in missing ratings. A missing rating tends to be a negative signal, rendering beliefs too optimistic.

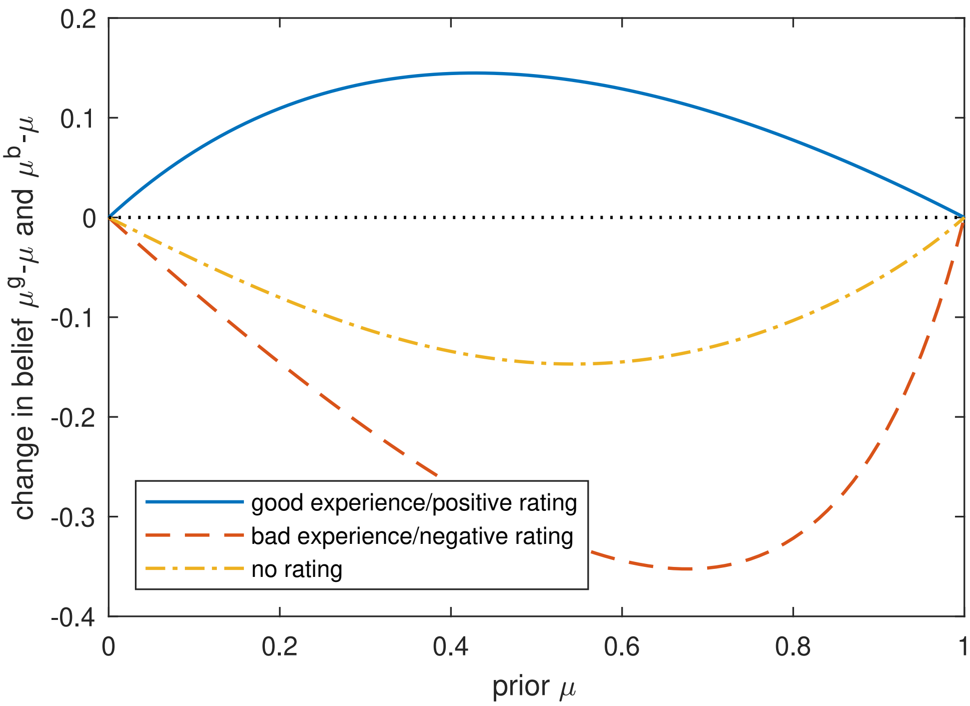
The final two figures show how additional negative ratings affect high-quality sellers. Figure D.11 shows the effect on the likelihood that a transaction is rated and Figure D.12 shows the effect on the likelihood that a rating is negative. Both Figures indicate relatively strong effects of early-on negatives that, however, fade rather quickly from about the 40th transaction.

,



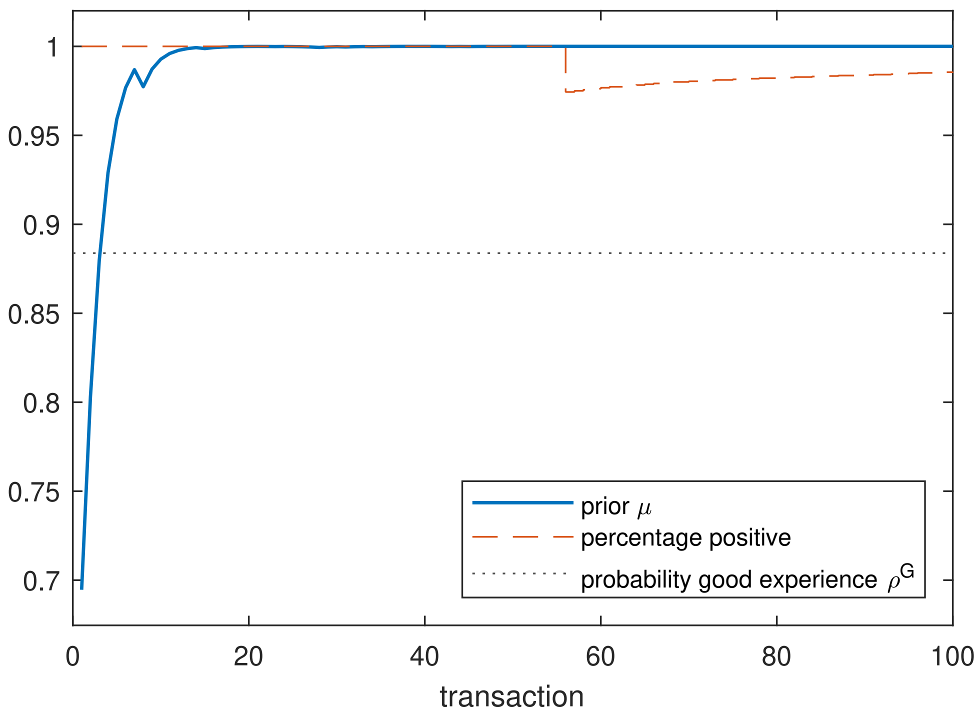
Notes: Probability that a rating is negative provided that it is left. Based on setup and procedure described in Section 7 and parameter values in Tables 5 and 6.

Figure D.5: Probability a rating is negative



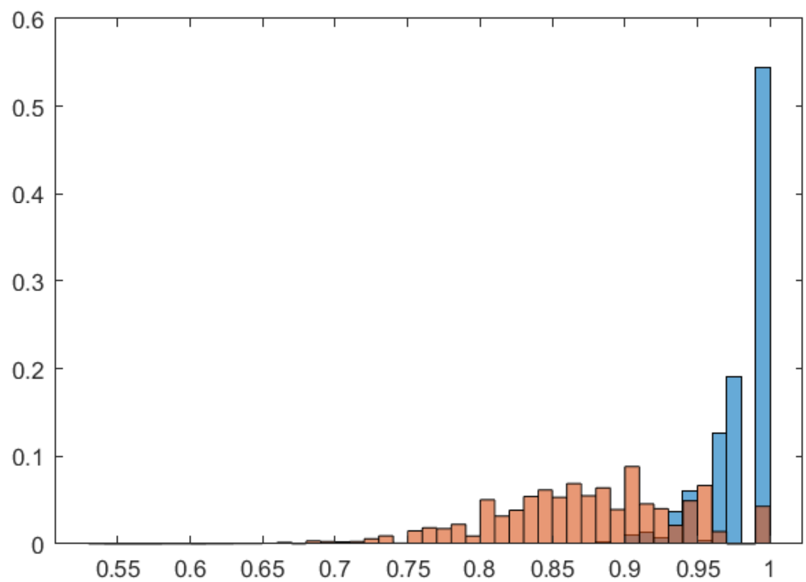
Notes: Extends Figure 1 and shows change in beliefs associated with a positive signal or rating, no rating, and a negative signal or rating. Based on setup and procedure described in Section 7 and parameter values in Tables 5 and 6.

Figure D.6: Change in beliefs (extended)



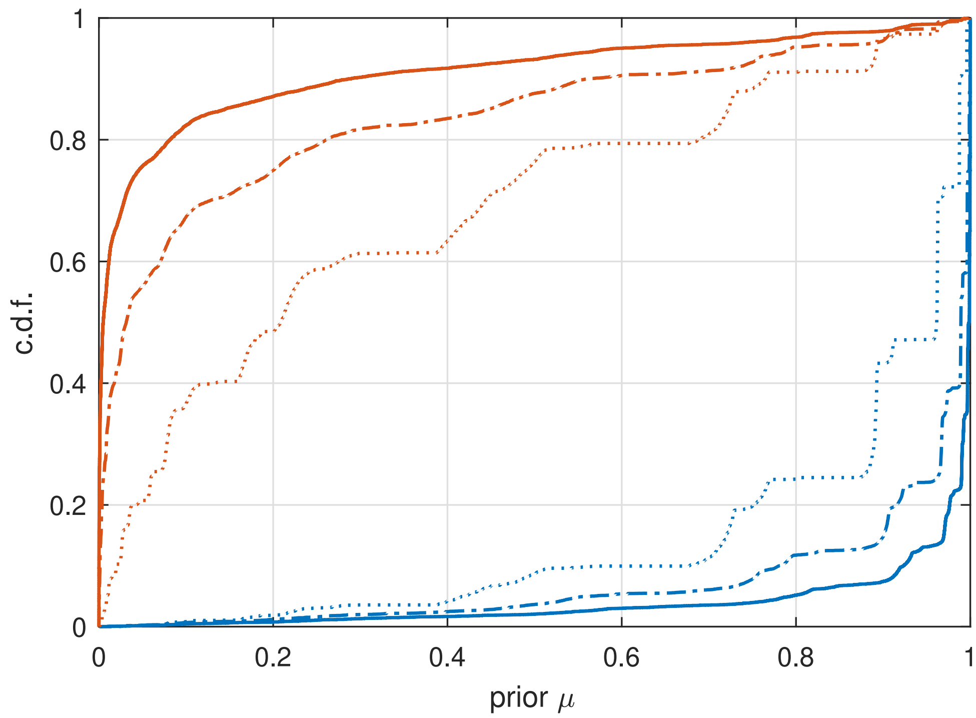
Notes: Evolution of the prior and the percentage positive feedback for one simulation run for a high-quality seller. Based on setup and procedure described in Section 7 and parameter values in Tables 5 and 6.

Figure D.7: Evolution of prior and percentage positive for one simulation run



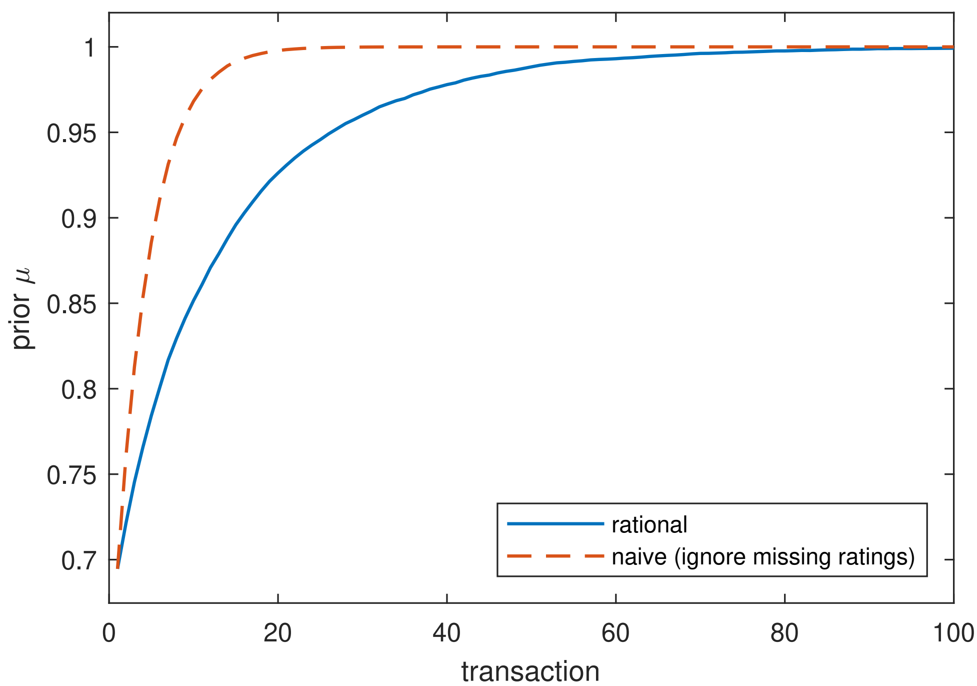
Notes: Distribution of percentage positive feedback after 50 transactions. Plotted for high-quality (blue) and low-quality (red) sellers. Obtained from 10,000 simulated paths. Based on setup and procedure described in Section 7 and parameter values in Tables 5 and 6.

Figure D.8: Percentage positive after 50 transactions



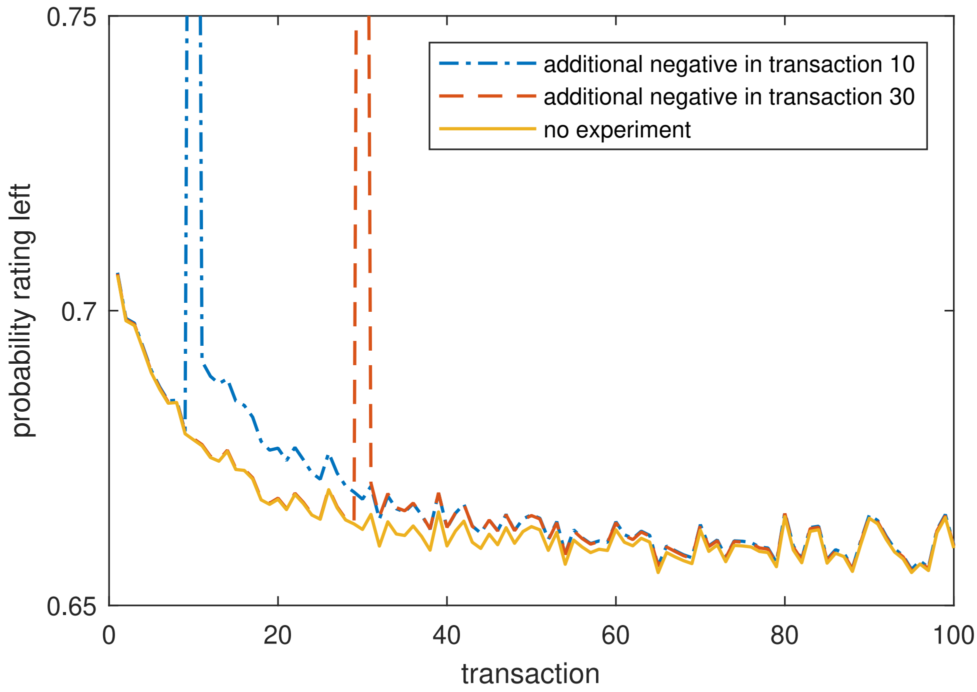
Notes: Distribution of beliefs after 10 (dotted line), 20 (dash-dotted line), and 30 (solid line) transactions. Plotted for high-quality (blue) and low-quality (red) sellers. Obtained from 10,000 simulated paths. Based on setup and procedure described in Section 7 and parameter values in Tables 5 and 6.

Figure D.9: Distribution of beliefs



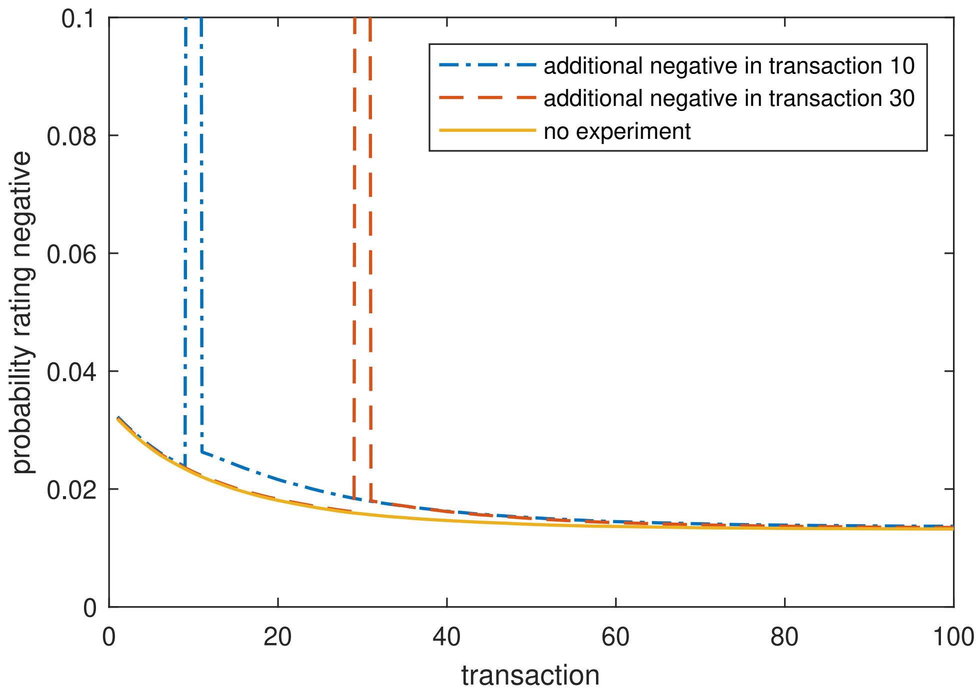
Notes: Comparison evolution of the prior for a high-quality seller when missing ratings are taken into account and when ignored. Average over 10,000 simulated paths. Based on setup and procedure described in Section 7 and parameter values in Tables 5 and 6.

Figure D.10: Evolution of prior with and without taking missing ratings into account



Notes: Evolution of probability that a rating is left. Shown for a high-quality seller. Average of 10,000 simulated paths and for the case that an additional negative rating is left in transactions 10 and 30, respectively. Based on the setup and procedure described in Section 7 and parameter values in Tables 5 and 6.

Figure D.11: Evolution of probability that a rating is left



Notes: Likelihood that given feedback is negative for a high-quality. Average of 10,000 simulated paths and for the case that an additional negative rating is left in transactions 10 and 30, respectively. The vertical axis is truncated at 0.1. Based on the setup and procedure described in Section 7 and parameter values in Tables 5 and 6.

Figure D.12: Evolution of probability that a rating is negative

E Survey to elicit beliefs about model parameters

E.1 Overview

On 3 October 2022, we conducted a survey on Prolific to elicit the beliefs of participants about the parameters ρ^{ℓ}, ρ^{h}, and λ of our model.[9]

We asked Prolific to recruit 1000 participants who have their residence in the U.S. and do online shopping at least once per month. We paid participants $1.50 for filling in the survey. 95% of the participants completed the survey within at most 370 seconds.

Table E.9 shows the characteristics of our respondents. They are about 40 years old on average and about half of them are female. More than 90 percent of them have the U.S. nationality. By design, they live in the U.S. and often shop online. Almost all of them have at least some eBay experience and about half of them have used eBay at least 20 times to buy something.

E.2 Results

Figure E.13 shows the answers to the three main questions. Not everyone answered all the questions. Figure E.13a is based on 993 responses, Figure E.13b is based on 994 responses, and Figure E.13c is based on 1000 responses. Table 6 shows estimates of the respective means and standard errors. The average response for ρ_h is about 0.88, the average response for ρ_l is about 0.49, and for λ the average response is about 0.69.

E.3 Additional details on the implementation and screenshots

The survey was implemented using a Google Form. Figure E.14 contains screenshots. Figure E.14a shows the welcome screen, which was followed by some explanation and the question of whether participants give informed consent in E.14b. If they didn't give consent, the screen in Figure E.14c was shown. If they did give consent, the survey continued by asking them about their Prolific ID. Next, they were asked how often they shopped online, see Figure E.14d. This was used to verify that respondents meet our selection criterion that they shop online at least

[9] Prior to running the survey, IRB approval was given by the IRB at the Tilburg School of Economics and Management and a pilot was run. Small editorial changes were made to the text after running the pilot. The results reported here are very similar to the ones from the pilot (not reported here).

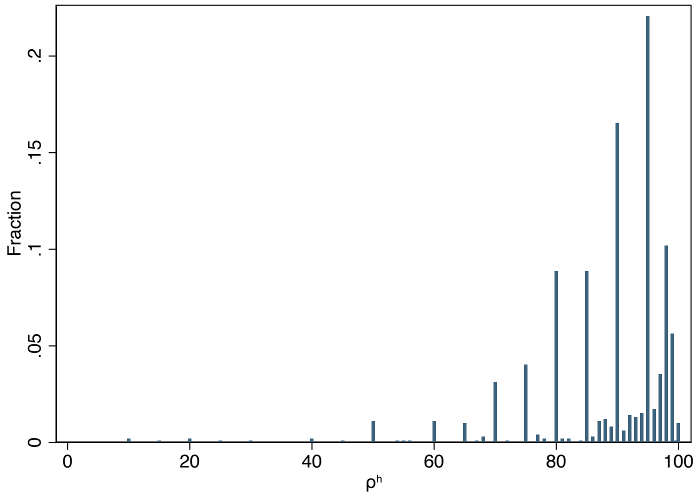
once a month. The wording of the question is the same as the wording of a question potential participants had answered before when Prolific asked them about many potential selection criteria. If participants now answered that they did shopping less often, then their answer was inconsistent with what they had said earlier, and therefore we showed them the screen in Figure E.14f. Otherwise, we continued with our 4 survey questions, which are shown in Figures E.14g-E.14j. Finally, Figure E.14k shows the exit message.

Table E.9: Sample population

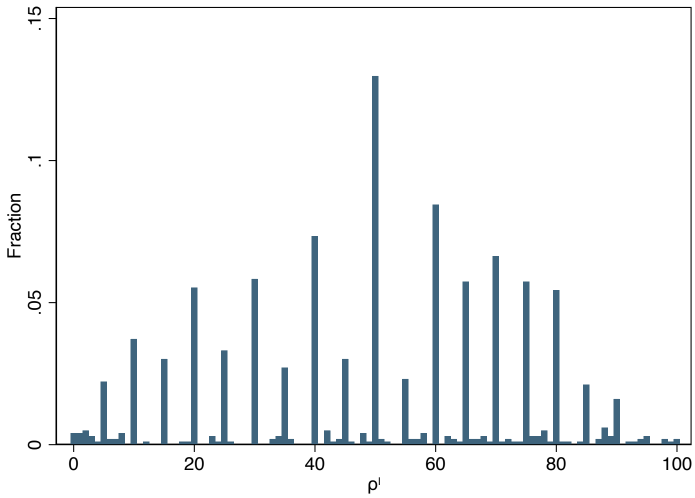
	mean	std.
age	39.03	12.88
female	0.48	-
U.S. nationality	92.00 %	
how often shopping online		
about once a month	9.40%	-
several times a month	30.90%	-
about once per week	27.80 %	-
more than once a week	31.90 %	-
how often bought on eBay over last 10 years		
never	1.50%	-
very rarely, between 1 and 19 times	32.80%	-
quite a few times, between 20 and 99	49.80%	-
many times, more than 100 times	15.90%	-

Notes: Std. shown for the only continuous variable: age. Remaining variables are indicators. Age available for 993, gender for 997, nationality for 998 out of 1000 respondents provided by Prolific. All participants responded to question on frequency of online shopping (see Figure E.14e) and eBay experience (see Figure E.14j).

(a) ρ_h



(b) ρ_l



(c) λ

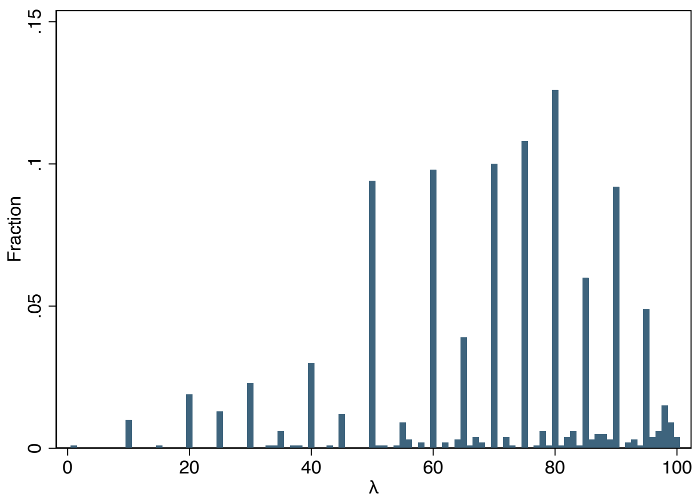
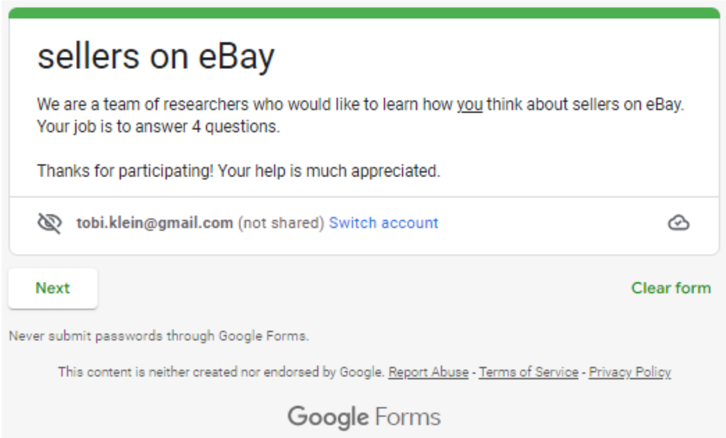


Figure E.13: Survey answers

(a) Screen 1: welcome



(b) Screen 2: consent form

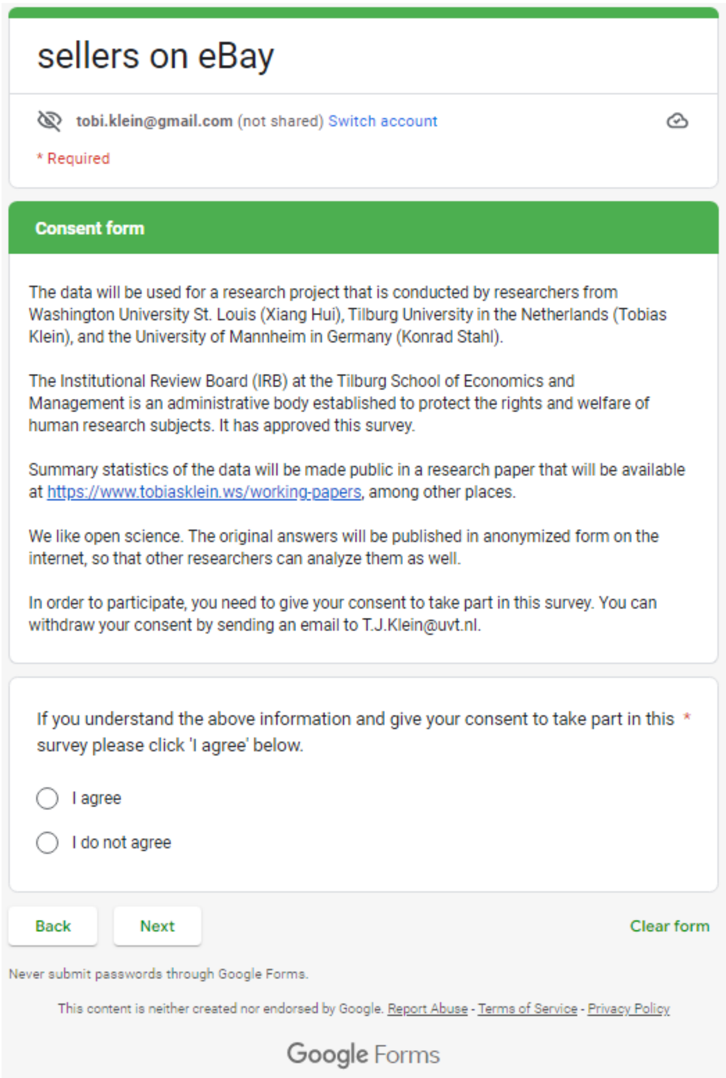
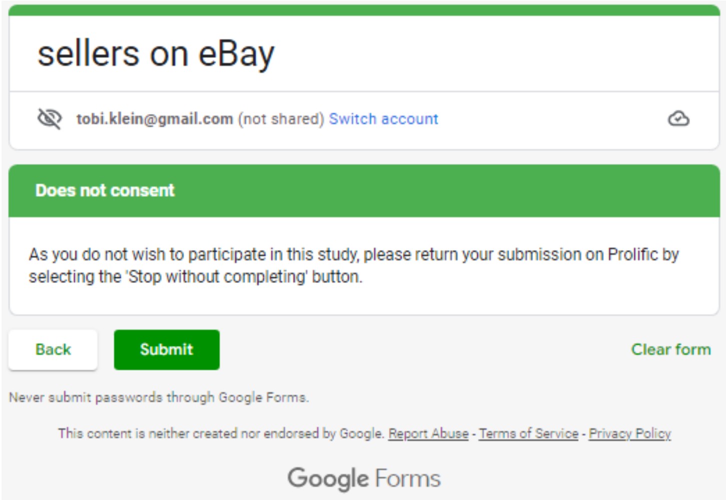


Figure E.14: Screenshots

(c) Follow-up to screen 2: message if consent was not given



(d) Screen 3: ask for Prolific ID

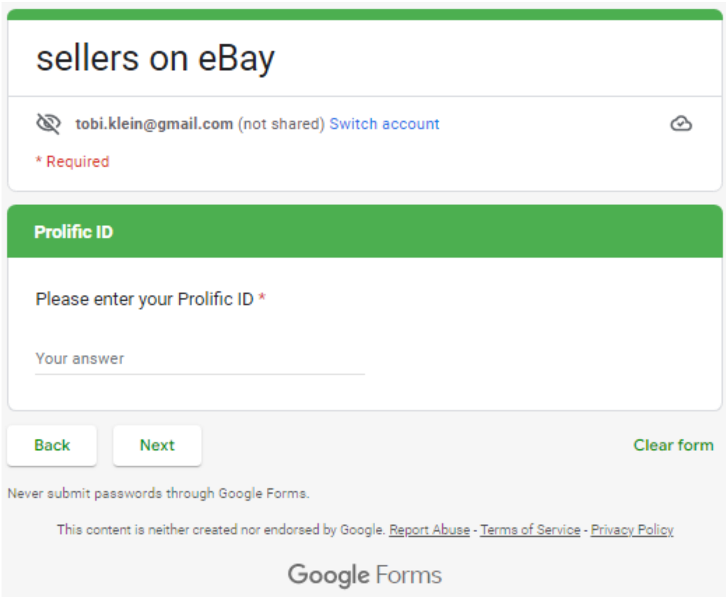
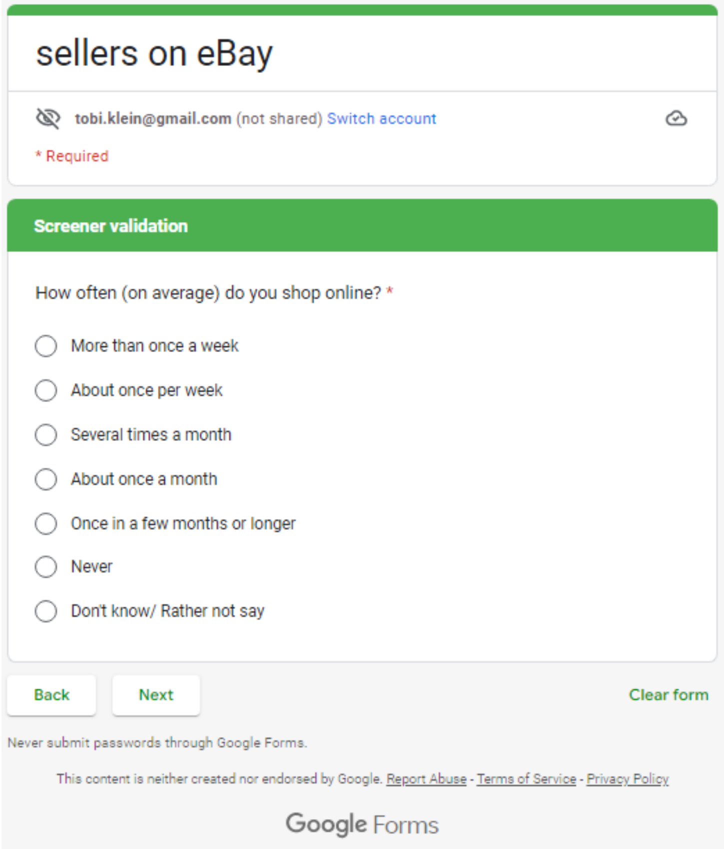


Figure E.14: Screenshots

(e) Screen 4: validation that participant does online shopping at least once per month



sellers on eBay

tobi.klein@gmail.com (not shared) Switch account

* Required

Screener validation

How often (on average) do you shop online? *

- More than once a week
- About once per week
- Several times a month
- About once a month
- Once in a few months or longer
- Never
- Don't know/ Rather not say

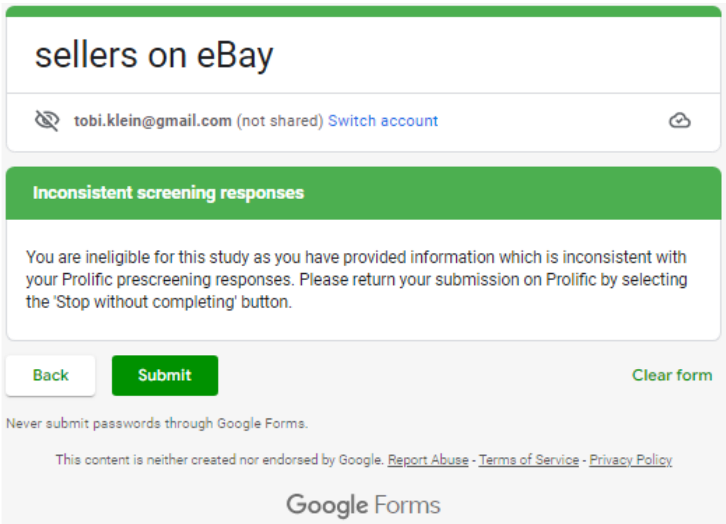
Back Next Clear form

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Privacy Policy

Google Forms

(f) Follow-up to screen 4: message if participant says that he or she does online shopping less than once per month



sellers on eBay

tobi.klein@gmail.com (not shared) Switch account

Inconsistent screening responses

You are ineligible for this study as you have provided information which is inconsistent with your Prolific prescreening responses. Please return your submission on Prolific by selecting the 'Stop without completing' button.

Back Submit Clear form

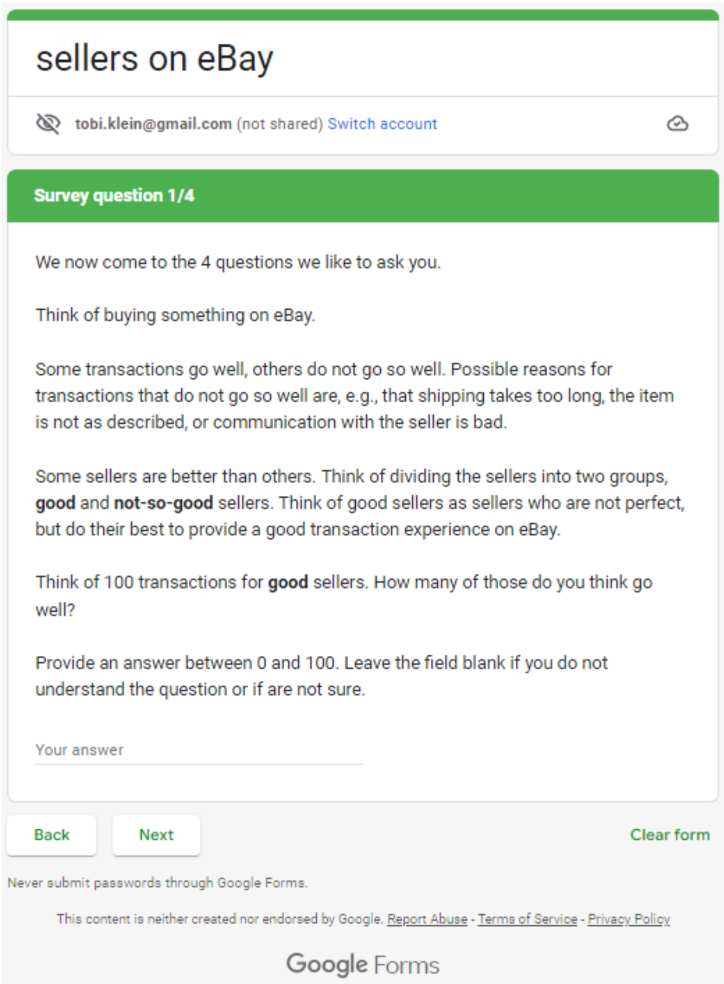
Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Privacy Policy

Google Forms

Figure E.14: Screenshots

(g) Screen 5: survey question 1/4



sellers on eBay

tobi.klein@gmail.com (not shared) Switch account

Survey question 1/4

We now come to the 4 questions we like to ask you.

Think of buying something on eBay.

Some transactions go well, others do not go so well. Possible reasons for transactions that do not go so well are, e.g., that shipping takes too long, the item is not as described, or communication with the seller is bad.

Some sellers are better than others. Think of dividing the sellers into two groups, **good** and **not-so-good** sellers. Think of good sellers as sellers who are not perfect, but do their best to provide a good transaction experience on eBay.

Think of 100 transactions for **good** sellers. How many of those do you think go well?

Provide an answer between 0 and 100. Leave the field blank if you do not understand the question or if are not sure.

Your answer

Back Next Clear form

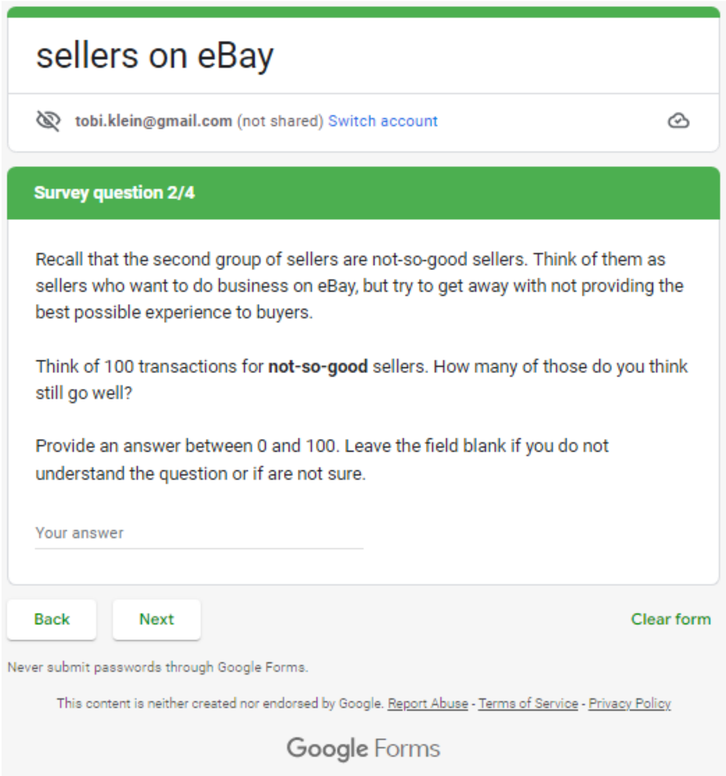
Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Privacy Policy

Google Forms

Figure E.14: Screenshots

(h) Screen 6: survey question 2/4



sellers on eBay

tobi.klein@gmail.com (not shared) Switch account

Survey question 2/4

Recall that the second group of sellers are not-so-good sellers. Think of them as sellers who want to do business on eBay, but try to get away with not providing the best possible experience to buyers.

Think of 100 transactions for **not-so-good** sellers. How many of those do you think still go well?

Provide an answer between 0 and 100. Leave the field blank if you do not understand the question or if are not sure.

Your answer

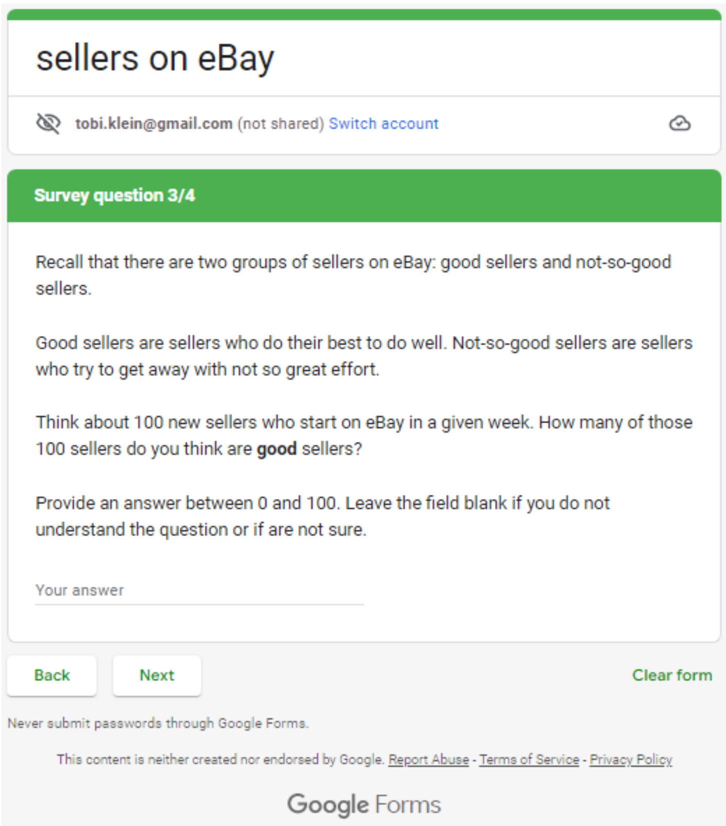
Back Next Clear form

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Privacy Policy

Google Forms

(i) Screen 7: survey question 3/4



sellers on eBay

tobi.klein@gmail.com (not shared) Switch account

Survey question 3/4

Recall that there are two groups of sellers on eBay: good sellers and not-so-good sellers.

Good sellers are sellers who do their best to do well. Not-so-good sellers are sellers who try to get away with not so great effort.

Think about 100 new sellers who start on eBay in a given week. How many of those 100 sellers do you think are **good** sellers?

Provide an answer between 0 and 100. Leave the field blank if you do not understand the question or if are not sure.

Your answer

Back Next Clear form

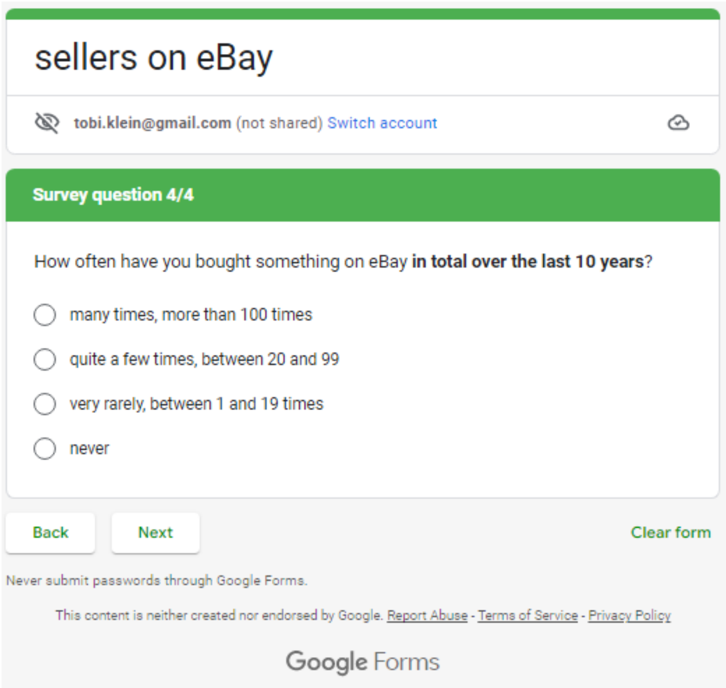
Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Privacy Policy

Google Forms

Figure E.14: Screenshots

(j) Screen 8: survey question 4/4



(k) Screen 9: exit message

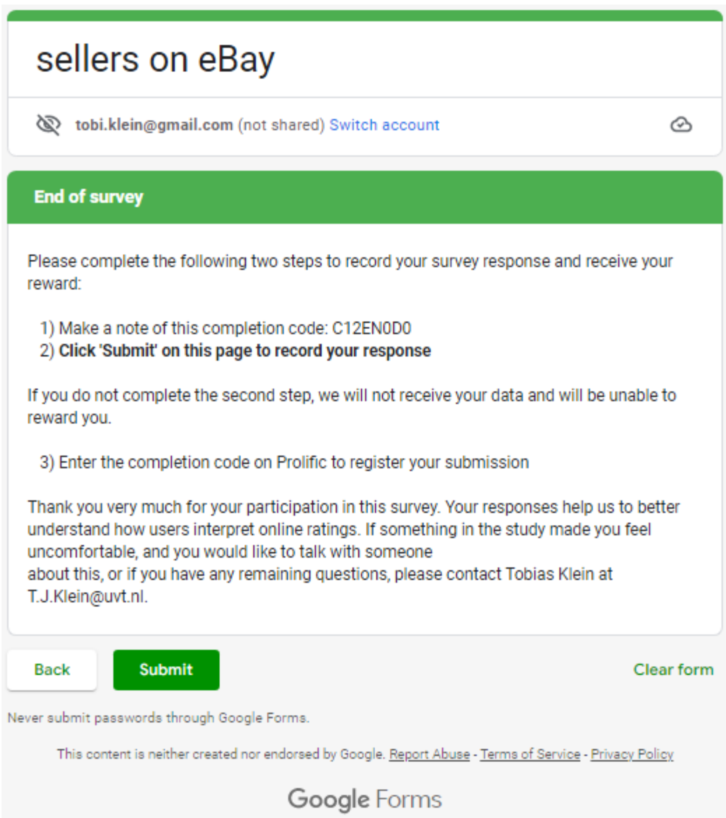


Figure E.14: Screenshots