GOVERNANCE AND THE EFFICIENCY
OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 426

Market Transparency, Adverse Selection, and Moral Hazard

Tobias J. Klein *
Christian Lambertz **
Konrad O. Stahl ***

* CentER, TILEC, Tilburg University
** University of Mannheim
*** University of Mannheim, C.E.P.R., CESifo, and ZEW

September 2013

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Klaus M. Schmidt · Department of Economics · University of Munich · D-80539 Munich,
Phone: +49(89)2180 2250 · Fax: +49(89)2180 3510

Market Transparency, Adverse Selection, and Moral Hazard*

Tobias J. Klein

Christian Lambertz

Konrad O. Stahl†

September 2013

Abstract

We study the effects of improvements in market transparency on eBay on seller exit and continuing sellers' behavior. An improvement in market transparency by reducing strategic bias in buyer ratings led to a significant increase in buyer valuation especially of sellers rated poorly prior to the change, but not to an increase in seller exit. When sellers had the choice between exiting—a reduction in adverse selection—and improved behavior—a reduction in moral hazard—, they preferred the latter because of lower cost. Increasing market transparency improves on market outcomes.

JEL classification: D47, D83, L15.
Keywords: Anonymous markets, adverse selection, moral hazard, reputation building mechanisms, market transparency, market design.

*We would like to thank the participants of the 2013 ICT Workshop in Evora, the 2nd MaCCI day in Mannheim, the 11th ZEW ICT Conference, the 2013 EARIE conference in Evora, seminar participants in Tilburg, and in particular Xiaojing Dong, Benjamin Engelstätter, Ying Fan, John List, Volker Nocke, Martin Peitz, David Reiley, Martin Salm, and Stefan Weiergräber for helpful comments. Special thanks go to Michael Anderson and Yossi Spiegel for unusually insightful reactions to an earlier version of this paper. Financial support from the Deutsche Forschungsgemeinschaft through SFB TR-15 and Microsoft through a grant to TILEC is gratefully acknowledged. The latter was provided in accordance with the KNAW Declaration of Scientific Independence. The views expressed here are not necessarily the ones of the Microsoft Corporation. Previous versions of this paper were circulated under the title "Adverse Selection and Moral Hazard in Anonymous Markets".

†T. J. Klein: CentER, TILEC, Tilburg University; C. Lambertz: University of Mannheim; K. O. Stahl: University of Mannheim, C.E.P.R., CESifo, and ZEW. E-Mail for correspondence: T.J.Klein@uvt.nl.

1 Introduction

Informational asymmetries abound in anonymous markets, such as those opening in the internet almost every day. In particular, before trading takes place, the typical buyer does not know whether her anonymous counterpart, the seller she is confronted with, appropriately describes and prices the trading item, and whether he conscientiously conducts the transaction, so she receives the item in time and in good condition.

Without remedies, these informational asymmetries invite adverse selection and moral hazard. *Adverse selection* may arise along Akerlof's (1970) classical argument. Conscientious sellers leave—or may not even enter—the market, as long as their behavioral trait, and their effort, are *ex ante* unobservable to the buyers, and thus the buyers' willingness to pay or even to trade is hampered. For complementary reasons, opportunistically exploitative and careless sellers tend to self-select into such a market, because they can cheat on buyers by incorrectly claiming to offer high quality products and good delivery service at high price. *Moral hazard* may arise because effort on both sides of the market is costly: therefore, sellers may package goods badly; or delay, or default on delivery. Likewise, buyers may delay, or default on payments.

Whereas the consequences of adverse selection and moral hazard are well understood conceptually, empirical tests on the direction of the effects as predicted by theory, and evidence on their magnitude are still scarce and centered around insurance markets. Theory predicts, however, that these effects should be pervasive. In this paper, we use data on seller behavior in classical product markets, and show that a change in the market design led to a reduction in moral hazard, but did at the same time not trigger exit from the market. Thereby, we complement earlier findings that were obtained in the context of insurance markets using a novel identification strategy in a very different context. In view of the fact that the effects of adverse selection and moral hazard cannot be easily separated in theoretical models accounting jointly for them,[1] it is particularly interesting that we could obtain evidence by which we are able to disentangle them empirically.Internet transactions provide a useful environment for collecting such evidence and conducting tests. Faced with adverse selection and moral hazard in these markets, the market organizers designed remedies early on. In particular, they constructed mechanisms under which buyers and sellers mutually evaluate their performance; and documented them, so that agents

[1] See, for instance, Laffont and Martimort (2002), Ch.7.

on both sides of the market could build reputation capital. These reporting mechanisms were adjusted over time in reaction to opportunistic reporting behavior on one, or both sides of the market, that was not anticipated by their designers. The changes in the reporting mechanism are typically unexpected by market participants, and thus can be perceived as natural experiments.

We collected reporting data before, and after such changes to address a question we feel to be central in evaluating the performance of anonymous markets: *Does a reduction in seller/buyer informational asymmetry improve on the performance of markets, and if so: via a reduction in adverse selection, or a reduction in moral hazard?*

It would be optimal to study this question by observing the agents' factual behavior in the individual transactions. Yet in internet markets, that behavior is typically not documented by the intermediaries, but rather its evaluation by the counterparty, which needs interpretation. At the same time, that evaluation provides more information than typically available in traditional offline markets, where data on seller behavior other than assortment and prices are often not available.[2]

eBay's classic reputation mechanism allows buyers and sellers to mutually evaluate their performance in just completed transactions. In May 2007, eBay added a new, second rating system, called Detailed Seller Ratings (DSR), that allows buyers to rate seller performance in detail—but not *vice versa*. The DSRs are reported as moving averages over the last 12 months, so unlike under the classic feedback scheme, the buyer's individual evaluation cannot be identified by the seller under the DSR scheme. One year later, in May 2008, eBay also changed the symmetry between buyer and seller rating in its classic feedback scheme, by forbidding negative seller rating and with it, removing buyer fear of seller retaliation to a bad rating by the buyer, that was likely to have had an influence on buyer ratings before the change was enacted.

Lacking documentation about the agents' behavior in the transaction, which is not recorded by eBay, we exploit these changes in eBay's feedback mechanism as follows. Since under the DSR scheme, the buyer's evaluation of a particular transaction is not identifiable by the seller, the DSR can be considered an unbiased reflection of buyers' satisfaction of that seller. We use this unbiased measure to identify how buyer satisfaction was affected by the change in eBay's classic feedback mechanism one year later.

We show that this second change leads to a *significant and quantitatively important increase*

[2] See, however, the emerging trend towards consumer evaluations of offline transactions in the internet.

in buyer satisfaction with the incumbent sellers. By contrast, we also show that this change did *not lead to an increase in the exit rate of poorly performing sellers.* We should emphasize that adverse selection is a relevant issue in the world considered here. The reason is that we find sellers to be naturally and consistently heterogeneous in their behavior, as reflected in the unbiased buyer evaluations. In view of this, the exit of badly performing sellers constitutes a relevant option. Yet, rather than exiting from the market, these sellers improve on all DSR dimensions, and this substantively and significantly more than the average incumbent sellers.

Towards our interpretation of these results, we develop a toy model from which we predict effects on seller adverse selection and moral hazard. Sellers differ by the dis-utility they suffer from engaging in effort towards satisfying a consumer at a given (competitive) price. Within this toy model, removing buyer fear of adverse retaliation by the seller incentivizes the buyer to report truthfully rather than opportunistically; and in particular to report bad experiences. That, in turn, leads sellers to change their behavior, in alternatively two ways: first, sellers with high dis-utility of effort may leave the market, thus ameliorating adverse selection; second, the sellers remaining in the market may engage in more effort towards improving on buyer satisfaction, thus ameliorating moral hazard.

We assess in detail the robustness of our empirical results against alternative interpretations and exclude the possibility that other developments have led to the observed significant increase in buyer satisfaction. This is important because our key identifying assumption is that average ratings would not have changed without the change to the feedback system. In particular, we establish that there was no time trend in average ratings before the change, that effects of another change—the introduction of Best Match—can be isolated and did not affect our results, and that changes in other factors such as the decreased popularity of the classical auction format have all not led to the increase in ratings after the change to the feedback system. Moreover, we document empirical patterns supporting the assumption that ratings are related to transaction outcomes; and that buyers are concerned about their reputation and therefore, the change to the feedback system had an effect.

This then allows us to conclude that the interpretation based on our toy model fits best, that is: *the reduction of the informational asymmetry due to a reduction of buyers' reporting bias disciplines sellers, and results in a reduction of seller moral hazard rather than a reduction of seller adverse selection.* The reason is that the additional costs sellers incur when they change their behavior are actually small relative to the benefits that are associated with being active, even for the badly performing sellers. Otherwise, we should have observed an increased rate of exit after the increase in transparency.

An increase of transparency removes buyer regret and thereby leads to higher quality outcomes. Given the small cost in implementing the observed change in the reporting mechanism, the significant increase in buyer satisfaction generated from that; and given that the sellers' material costs of changing their behavior are arguably small, we claim this increase in market transparency to have a beneficial welfare effect.[3]

As sellers and buyers who are active on eBay are very much alike users of other online platforms, we expect our result, namely that increasing transparency will lead to higher quality outcomes and will likely be welfare-enhancing, to also apply to other contexts in which reporting mechanisms are used to discipline user behavior in online markets. Beyond that, it could provide guideline for changes in the numerous reporting mechanisms on seller performance in offline markets, for instance for restaurant and hotel services.

Towards detailing procedure and results, we proceed as follows. In the next Section 2, we report on the literature pertinent to what we do. In Section 3, we describe the eBay Feedback Mechanism, and in particular the two changes we focus on. Section 4 contains the description of our data. In Section 5, we present our central results. In the ensuing Section 6 we develop our toy model, from which we derive our preferred explanation and interpretation of the results. In Section 7 we assess in much detail the robustness of our interpretation and provide additional supportive results. We conclude with Section 8.

2 Literature

Adverse selection and moral hazard are related to asymmetric information between two contracting parties.[4] Take a car insurance company and an individual insurance taker. Adverse selection arises if the individual self selects into buying high coverage because she knows to be an unsafe driver. Moral hazard is present if she pursues lesser accident-preventing effort in the

[3] Unfortunately, a more rigorous welfare analysis is beyond the scope of this paper, as it would require us to observe or infer seller cost as well as buyer preferences. Only then we could compare the increase in consumer surplus to the decrease in seller rents.

[4] Sometimes, a distinction is made between *ex ante* moral hazard and *ex post* moral hazard. The terminology adverse selection refers to the former in this paper, and the terminology moral hazard to the latter.

face of the better coverage, and is therefore more likely to have an accident.

It is difficult to disentangle adverse selection and moral hazard by just observing the incidence of accidents conditional on coverage. Therefore, authors have first focused on showing that asymmetric information affects economic outcomes. In general, this is done by relating individual choices to *ex post* outcomes (see Chiappori, 2000, for an early review). For example, Chiappori and Salanié (2000) use data on contracts and accidents in the French market for automobile insurance to test whether insurance contracts with more comprehensive coverage are chosen by individuals who then have higher claim probabilities. If this is the case, then this can either be explained by moral hazard, or adverse selection, or both, without further discrimination.[5]

From the theorist's point of view, the inability to disentangle adverse selection and moral hazard effects does not come as surprise: the analyst typically cannot observe self selection *ex ante* by type because the type is largely private information. In addition, with an endogenous change of effort, that type can modify the outcome. In the interpretation of our results preferred by us, we can strictly separate the two effects *ex post*, by observing poorly performing sellers' exit from the market as reflecting a reduction in adverse selection; and ongoing sellers' effort taking, as reflecting a reduction in moral hazard.

At any rate, following up on Chiappori and Salanié (2000), Abbring, Chiappori, and Pinquet (2003) and Abbring, Heckman, Chiappori, and Pinquet (2003) argue that dynamic insurance data allow researchers to isolate moral hazard effects, by looking at insurance contracts in which the financial loss associated with a second claim in a year is bigger, so that exercising moral hazard becomes more costly, and therefore the incentive to do so decreases. One can isolate moral hazard effects in this context because one naturally follows an individual over time, and therefore the factors influencing adverse selection stay the same, while incentives to exert moral hazard change. In the context of deductibles in health insurance Aron-Dine, Einav, Finkelstein, and Cullen (2012) follow-up on this idea and investigate whether individuals exhibit forward looking behavior, and reject the hypothesis of myopic behavior.

Focusing on adverse selection, Einav, Finkelstein, Ryan, Schrimpf, and Cullen (2011) show that some individuals select insurance coverage in part based on their anticipated behavioral response to the insurance contract, and term it "selection on moral hazard". For this, they

[5] See also Finkelstein and Poterba (2004) for a similar approach in the context of annuitization and mortality and Fang, Keane, and Silverman (2006) in the context of health insurance.

exploit variation in the health insurance options, choices and subsequent medical utilization across different groups of workers at different points in time. Also Bajari, Hong, and Khwaja (2006) study individual selection of insurance contracts. They provide, as we do in our very different context, evidence of moral hazard, but not of adverse selection. Their result is based on a structural model of demand for health insurance, in which, in order to isolate selectivity *ex ante* and lacking exogenous variation, they need to control in an elaborate way for individual risk and risk preference.

We instead develop our results from a natural experiment, involving, in our interpretation, self selection and adjustment of moral hazard *ex post.* We follow sellers over time, which allows us to control for unobserved differences across sellers by means of fixed effects when studying moral hazard. We then study whether an improvement of the mechanism led to increased exit from the market on the one hand, and a reduction of moral hazard on the other hand.

The institutional change the effects of which we study led to an increase in market transparency. Market transparency also plays an important role in several other, distinct literatures. In the context of restaurants, Jin and Leslie (2003) show that quality disclosure for restaurants, by means of requiring them to display quality grade cards in their windows, causes them to make hygiene quality improvements. Anderson and Magruder (2012) relate online ratings of restaurants to restaurant reservation availability and find that an extra half-star on the popular platform Yelp.com causes restaurants to sell out 19 percentage points more frequently. Also in finance, there is a literature on the effects of mandatory disclosure. For instance, Greenstone, Oyer, and Vissing-Jorgensen (2006) show that financial investors valued an extension of disclosure requirements by documenting abnormal returns for firms that were most affected by this. In the context of competition policy, Henze, Schuett, and Sluijs (2013) conduct an experiment in which they vary the extent to which consumers are informed about quality. They find effects of this on the quality firms provide in equilibrium, and conclude that information disclosure is a more effective tool to raise welfare and consumer surplus than theory would lead one to expect.

There is also a literature on quality disclosure in electronic markets which in turn is related to Avery, Resnick, and Zeckhauser's (1999) somewhat sweeping general hypothesis that the internet has greatly reduced the cost of distributing information and that there is an efficient provision of evaluations by users. Dranove and Jin (2010), Bajari and Hortasçsu (2004) and Cabral (2012) provide reviews of the theoretical and empirical literature on quality disclosure on the internet. For eBay, the general finding is that better ratings benefit sellers by an increase in the probability to sell a product, and in its selling price. See, e.g., Melnik and Alm (2002), Lucking-Reiley and Reeves (2007) and Jin and Kato (2008) for evidence using field data, and Resnick, Zeckhauser, Swanson, and Lockwood (2006) for experimental evidence.

These results show that ratings on eBay convey information, but it is unclear how much. The reason is that, due to the design of the reputation mechanism, ratings were biased before the implementation of DSR, and the removal of symmetric classic feedback. Resnick and Zeckhauser (2002) provide reduced-form evidence that points towards underreporting of negative experiences and Klein, Lambertz, Spagnolo, and Stahl (2006) complement this by showing that the probability to leave a negative rating increases substantially towards the end of the period in which feedback can be left.

Klein, Lambertz, Spagnolo, and Stahl (2009) provide detailed information on the actual structure of the feedback mechanism and provide first descriptive evidence on the newly introduced DSRs. Bolton, Greiner, and Ockenfels (forthcoming) also provide such evidence and complement it with an experimental study. Focusing on why classic ratings are left at all, Dellarocas and Wood (2008) estimate a model of rating behavior, assuming that ratings, once given, are truthful, and estimate the true underlying distribution of satisfaction. This can be seen as controlling for the selection bias that comes from traders being much more likely to leave a rating when satisfied.

Cabral and Hortaçsu (2010) provide indirect evidence for the presence of moral hazard on eBay. They find that, when a seller first receives negative feedback, his sales rate drops and he is more likely to receive negative feedback and to exit. Moreover, they find that just before exiting, sellers receive more negative feedback than their lifetime average. With our paper we complement the aforementioned studies by providing direct evidence on one of the most policy-relevant questions, namely the relationship between the design of the feedback mechanism and the presence of moral hazard or adverse selection.

3 eBay's Feedback Mechanism

eBay's feedback mechanism by which sellers and buyers could evaluate the performance of their trading partners was introduced in February 1996, just a few months after the first auction had taken place on its website.[6] In its earliest form, the system allowed any eBay user to leave feedback on the performance of any other user, i.e., a "positive," "neutral," or "negative" rating accompanied by a textual comment. This feedback was immediately observable on his or her "Feedback Profile" page, together with all ratings and comments that a user had ever received by other users.

In February 2000, four years after its institution, the mechanism was changed to transaction-specific feedback. Since then, all new ratings must relate to a particular transaction, i.e. only the seller and the buyer in a particular transaction can rate each other regarding their performance in that transaction.

From early on, the feedback mechanism has led to conflicts and heated discussions about unfairly biased reporting. As a consequence, eBay repeatedly modified the system. The two major changes we focus on here were made in May 2007 and May 2008, respectively. In May 2007, eBay introduced a new form of unilateral rating by buyers: Detailed Seller Ratings (DSR). In addition to the original bilateral rating available heretofore, buyers could now separately rate the transaction items *accuracy of the item description*, *communication*, *shipping speed*, and *shipping charges* with one to five stars. These detailed ratings are left anonymously and only unilaterally by the buyer; they are anonymized by being published in summarized form only, so that the individual rating cannot be identified by the seller.

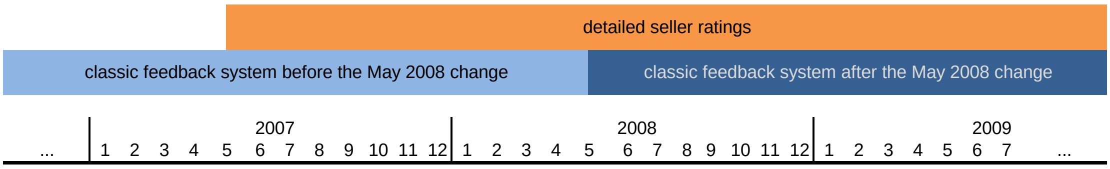
This change tackles what was felt to be a substantial flaw in eBay's original bilateral feedback mechanism, namely the buyer's fear of retaliation when leaving a negative rating *before* the seller—a problem well known to many eBay users and well discussed among scholars for some time.[7] An important detail is that DSRs can only be left when a classic rating is left. The two ratings need not be consistent, however. That is, for the very same transaction, a buyer could leave a positive classic rating identifiable by the seller—and a negative, truthful set of DSRs not identifiable by him. Not that the two ratings are only imperfect substitutes. This is because DSRs reflect averages, whereas classic ratings are individually observable. The classical ratings thus show in addition how the seller behaved at the margin, i.e. in the most recent transactions.

In May 2008, the classic bilateral feedback mechanism was transformed to effectively a unilateral one as well: sellers could only leave positive ratings on buyers—or none at all. Thereby,

[6] An early description of the basic mechanism and an analysis of rating behavior are given in Resnick and Zeckhauser (2002).

[7] Klein, Lambertz, Spagnolo, and Stahl (2006) gave an early account of this.

Figure 1: Changes to the feedback mechanism



eBay removed the possibility that the seller would strategically postpone his rating, in order to implicitly or explicitly threaten the buyer with retaliation to a negative rating.[8] The two changes are summarized in Figure 1.[9]

Since buyers can leave DSR without threat of retaliatory feedback by the seller, we take it that a buyer's DSR constitutes a possibly subjective, but strategically unbiased evaluation of seller performance.[10] We use this as the basis for investigating how individual seller performance reacts to the May 2008 change, when all ratings were effectively made unilateral. As the May 2008 change enacted by eBay addresses the classic feedback system that was actually better established amongst users than DSR, we do expect a significant impact of that second change, even though the May 2007 change to DSR already had made room for candid ratings that could have affected seller behavior.

4 Data

Our data are monthly information on feedback received for about 15,000 eBay users over a period of three years, between July 2006 and July 2009. The data were collected from eBay's

[8] In fact, eBay stated the reasons for this step in a public announcement in January 2008: *Today, the biggest issue with the system is that buyers are more afraid than ever to leave honest, accurate feedback because of the threat of retaliation. In fact, when buyers have a bad experience on eBay, the final straw for many of them is getting a negative feedback, especially of a retaliatory nature...Now, we realize that feedback has been a two-way street, but our data shows a disturbing trend, which is that sellers leave retaliatory feedback eight times more frequently than buyers do... and this figure is up dramatically from only a few years ago. So we have to put a stop to this and put trust back into the system...here's the biggest change, starting in May: Sellers may only leave positive feedback for buyers (at the seller's option).* (Taken from http://announcements.ebay.com/2008/01/a-message-from-bill-cobb-new-pricing-and-other-news/, last accessed in June 2013.) Note that while the change in the rating system was announced, its timing was not, which allows us to treat the time as exogenous at which the change was enacted.

[9] Additional changes aiming at alleviating sellers' concerns about strategic feedback extortion by the buyers but not of interest in this study were at several points in time. For instance, buyers could threaten to leave a negative rating if not given a discount, without any fear of retaliation by the seller. To remedy this, eBay abandoned any options to mutually withdraw feedback. Another change was the introduction of a new way of sorting auction listings. We discuss the latter change in more detail in the form of a robustness check in Section **??**.

[10] As eBay does not follow up the transactions process, there was no unbiased information about eBay transactions available at all before the DSR mechanism was implemented.

U.S. website using automated download routines and scripts to parse the retrieved web pages for the details in focus. In May 2007, we drew a random sample of, respectively, 3,000 users who offered an item in one of five different categories. The categories were (1) Laptops & Notebooks, (2) Apple iPods & Other MP3 Players, (3) Model Railroads & Trains, (4) Trading Cards, and (5) Food & Wine.[11] We chose these categories because they were popular enough to provide us with a large list of active sellers, and because they appeared reasonably heterogeneous to us not only across categories, but also within categories as none of them was dominated by the listings of a few sellers. From June 1, 2007 onwards we downloaded these users' "Feedback Profile" pages on 18 occasions, always on the first day of the month. The last data collection took place on July 1, 2009. The information dating back from May 2007 to July 2006 was inferred from the data drawn in June, 2007.[12]

Towards capturing changes in adverse selection, we needed to specify *seller exits* in the observation period. We identified as the date of exit the date after which a user did not receive any new DSRs during our observation window. This is a proxy, as it may also apply to users not completing any transaction for a period of time beyond our observation window, but being active thereafter. However, when we follow individuals who become inactive in the very beginning of our sampling period, then we see that it is very unlikely that such users will become active again within a couple of months.[13,14]

Out of the 15,000 user names we drew in May 2007, we were able to download feedback profiles for 14,937 unique users in our first data collection on June 1, 2007.[15] One year later, we could still download data for 14,684 users, and two years later for 14,360 users.[16]

[11]See Table 5 in Appendix A for the exact categories.

[12]See Figure 8 in Appendix A for a graphical representation of the times at which we collected data. This data collection design implied that we were unable to recover seller entry in the relevant time interval.

[13]We are more likely to mis-classify sellers as inactive towards the very end of our observational period. Yet this is unlikely to affect our results, because we have collected data for more than a year after the feedback change, and we mostly use information around the May 2008 change to the system. See also footnote 27.

[14]We also re-did our analysis using a different definition of exit, namely the time after which a user did not receive any classic ratings anymore. This criterion captures the activity of users when active as a buyer or a seller, as classic ratings can be received when acting in either role. If users are equally likely to stop being active as a buyer before and after the change to the classic feedback mechanism then finding an increase in the probability that became inactive according to this more strict criterion also indicates that adverse selection was affected by the feedback change. The empirical patterns are very similar for both definitions of exit and therefore we do not report the results below.

[15]There were download errors for 11 users and we decided to drop three users from our panel for which eBay apparently reported wrong statistics. Moreover, there are 48 users in our sample who had listings in two categories (and therefore are not unique), and two users who had listings in three of our five categories. We dropped the duplicate observations.

[16]We devoted a large amount of effort to following users when they changed their user names. Still, we were not successful in doing so for all users. One reason for this is that after an account is closed, the associated feedback

Table 1: Summary statistics

	obs.	mean	std.	percentile 5	25	50	75	95
June 1, 2007								
duration membership in years	14,937	3.83	2.76	0.09	1.33	3.54	6.12	8.46
feedback score	14,937	563.66	2704.53	0.00	18.00	88.00	339.00	2099.00
percentage positive classic ratings	14,189[a]	99.09	5.67	97.10	99.70	100.00	100.00	100.00
member is PowerSeller	14,937	0.07	-	-	-	-	-	-
number classic ratings previous 12 months	14,937	273.10	1351.22	0.00	10.00	43.00	161.00	975.00
percentage positive classic ratings previous 12 months	13,943[b]	98.95	6.51	96.49	100.00	100.00	100.00	100.00
June 1, 2008								
number classic ratings previous 12 months	14,684	282.16	1247.45	0.00	10.00	45.00	164.00	1042.00
percentage positive classic ratings previous 12 months	13,812[c]	97.95	9.95	93.10	99.53	100.00	100.00	100.00
number DSR previous 12 months	4,429[d]	378.78	1240.91	12.00	28.00	78.25	265.50	1378.25
DSR score	4,429	4.71	0.19	4.35	4.65	4.75	4.82	4.90
number DSR relative to number classic feedbacks	4,429	0.42	0.19	0.10	0.27	0.44	0.59	0.70
June 1, 2009								
number classic ratings previous 12 months	14,360	200.45	1039.00	0.00	2.00	20.00	97.00	761.00
percentage positive classic ratings previous 12 months	11,524[e]	99.47	4.19	98.18	100.00	100.00	100.00	100.00
number DSR previous 12 months	3,272[f]	376.41	1249.90	12.00	26.38	72.00	255.75	1378.00
DSR score	3,272	4.78	0.16	4.53	4.73	4.82	4.88	4.95
number DSR relative to number classic feedbacks	3,272	0.46	0.20	0.11	0.29	0.48	0.63	0.74

Notes: The table shows summary statistics for our sample of sellers. There are three panels, each one for a different point in time for which we report summary statistics. These points in time are the first point in time at which we collected data, as well as one and two years after that. DSRs were introduced in May 2007, so the first point in time is the beginning of the first month after this. The change in the classic feedback mechanism whose effect we analyze occurred in May 2008, i.e. in the month prior to the second point in time for which we report summary statistics. The third point in time is one year after that. The feedback score is the number of users who have mostly left positive feedback in the classic system, minus the number of users who have mostly left negative feedback. The PowerSeller status is awarded by eBay if a seller has a particularly high transaction volume and generally a good track record. DSR is the average DSR, per user, across the four rating dimensions. [a]Calculated for those 14,189 users whose feedback score is positive. [b]Calculated for those 13,943 users who received classic feedbacks in the previous 12 months. [c]Calculated for those 13,812 users who received classic feedbacks in the previous 12 months. This is also inferred from other points in time, hence the number of observations is higher than it is for other statistics in the same panel. [d]Calculated for those 4,429 users who received enough DSRs so that the score was displayed. The statistics in the following two rows are calculated for the same users. [e]Calculated for those 11,524 users who received classic feedbacks in the previous 12 months. [f]Calculated for those 3,272 users who received enough DSRs so that the score was displayed. The statistics in the following two rows are calculated for the same users.

Table 1 gives summary statistics. As described above, the first data collection took place on June 1, 2007. On that day, the average user in our sample was active on eBay for almost four years. Proxying user experience by the length of time a user has registered, the most experienced user in our sample had registered with eBay more than eleven years before we collected our first data, and the least experienced user just a few days before our observation window opened. About 2,000 of our users had registered their accounts before the turn of the millennium, and about 3,000 users only within two years before the May 2008 changes.

On eBay, the feedback score is given by the number of distinct users who have left more positive classic ratings than negative ones, minus the number of users who have left more negative ratings than positive ones. At the time the observation window opened, the mean feedback score of our users was 564, the median score was 88, and 769 users had a feedback score of zero. The average share of positive feedback users have ever received was 99.09 per cent, which corresponds well to findings in other studies. The median number of feedbacks received during the year before that was 43. In the following year, users received roughly as many classic ratings as in the year before, and also the percentage positive ratings was very similar. On June 1, 2008, statistics for the DSRs are available for the 4,429 users who received more than 10 DSRs. The reason is that otherwise, anonymity of the reporting agent would not be guaranteed, as a seller could infer the rating from the change in the DSR. DSR scores are available for about 15 per cent of the users one month after their introduction in May 2007, and for about 30 per cent of users one year later. The DSR score we report on here is the average reported score across the four rating dimensions. Yet another year later, the picture looks again similar, except for the number of classic ratings received, which has decreased.

At this point, it is useful to recall the objective of our analysis: it is to study sellers' reactions to the May 2008 system change, on the basis of unbiased ratings by their buyers effective with the introduction of DSR one year before. Users may sometimes act as sellers, and sometimes as buyers. With our sampling rule, we ensure, however, that they were sellers in one of the five specified categories in May 2007. Moreover, DSRs can only be received by users when they act as sellers. Hence, the average DSR score will reflect only how a user behaved in that very role.[17]

profile is no longer available, and therefore, we could not find it. Over time, it became more difficult to download data, as eBay put measures in place that prevented us from downloading enough pages within a short amount of time.

[17] One may still wonder how often the users in our sample acted as buyers. On June 20th, 2008, eBay reveals in a statement that buyers leave DSR 76 per cent of the time when leaving "classic" feedback. In our data collection

Still, it is also important to keep in mind that we will not be able to observe the reaction of sellers who receive less than 10 DSRs per year.[18] However, looked at it in a different way, we capture behavior that is associated with most of the transactions on eBay, as those sellers who receive less than 10 DSRs per year are not involved in most of the sales on eBay.

5 Results

5.1 Incumbent Sellers' Reactions

The May 2008 change provided additional means to buyers to voice negative experiences without fear of negative seller reaction. This should have incentivized continuing sellers to prevent negative buyer ratings by significantly reducing shirking, i.e. not describing and pricing goods as of higher than the true quality; and increasing their effort in the other dimensions towards satisfying buyers. Therefore, we expect a significant increase in buyer satisfaction, as measured by the DSR scores.

This expectation is verified in Figure 2 in which we summarize the evolution of DSRs in the relevant two 12-month intervals before and after the May 2008 change. Recall that at any point in time, DSR indices are published in four categories, for every seller that has received more than 10 DSRs up to that point, with ratings aggregated over the respective preceding 12 months. In the figure, each dot reflects the overall average across sellers and categories. The patterns by category resemble one another closely.[19]

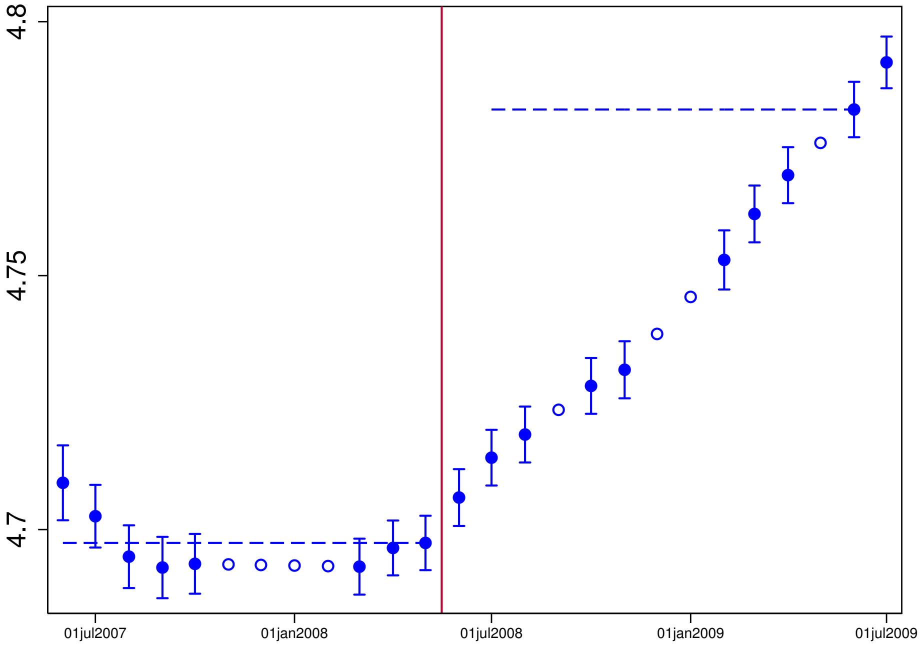
When interpreting Figure 2 it is important to once again keep in mind that DSR scores show the average of all DSR ratings given in the previous 12 months. Therefore, if all ratings received after the change were higher, one would observe unchanged ratings reflected in a flat curve before the change, then an increasing function in the 12 months after the change, and again

just before this statement, the mean overall "DSR to classic" ratio of users for whom a DSR is displayed is about 43 per cent. The difference between those 76 per cent, where users acted as sellers, and the 43 per cent, where they acted as buyers or sellers comes about because they may also have acted as buyers. Looked at it in a different way, the 43 per cent in our sample is a lower bound on the probability that a user has acted as a seller in a given transaction, because DSRs can only be left when a classic rating is left at the same time. It is a rather conservative lower bound because it assumes that a user receives a DSR every time he acts as a seller and it takes a bit more effort for the buyer to leave a DSR, as compared to leaving only a classic rating. In any case, DSR scores reflect how users have behaved when acting in their role as sellers.

[18] Throughout, we will control for seller fixed effects. This is important because sellers for whom DSRs are available may be different from those for whom DSRs are not available; and because sellers who exit at some point may be different from those who will not exit. See also the discussion below.

[19] In Figure 10 in Appendix B, they are reported in detail for the sellers with below median performance before the change, on which we later concentrate our analysis.

Figure 2: Evolution of Detailed Seller Ratings



Notes: The figure shows how DSRs changed over time. The vertical line denotes the time of the May 2008 change to the classic feedback mechanism. The dots are averages across users for whom DSRs are displayed and the error bars depict corresponding 95 per cent confidence intervals. The circles are linearly interpolated values for the periods in which we did not collect data. Before averaging DSRs across users we calculate the average DSR per user, across the four categories. The horizontal dashed lines visualize that the dots are averages over the 12 months prior to the point in time at which the DSRs are displayed.

a flat curve (at a higher level) after that. The full effect of the change equals the difference between the DSR rating one year after the change and the DSR rating right before the change, and is approximately proportional to the slope of the line after the change.

The horizontal lines in Figure 2 depict the relevant average scores that were accumulated at the beginning of May 2008 and June 2009, respectively.[20] The difference in the horizontal lines equals the difference between buyer ratings one year after the change, i.e. after the change in buyer ratings is completely reflected in the 12 months moving average, relative to the score just before the change. The figure clearly shows that the DSRs have increased after the May 2008 change.[21]

[20]The change occurred in mid-May 2008. Hence, the DSR score at the beginning of June, 2008 contains no DSRs left before the change because it is calculated from the ratings received in the preceding 12 months. Conversely, the DSR score at the beginning of May, 2008 contains no ratings received after the change. Figure 8 in Appendix A shows at which points in time data were collected and depicts over which periods, respectively, the DSR scores were calculated.

[21]Unfortunately, we were not able to collect data for more than one year after the change, because eBay started to ask users to manually enter words that were hidden in pictures when more than a small number of pages were

We performed regressions to quantify the effect controlling for fixed effects. Towards their specification, denote by DSR_{it} the average score across the four DSR rating dimensions reported for seller i in period t. Recall that our data is always drawn on the first day of the month, and that DSR_{it} is the average of all ratings seller i has received over the previous 12 months. Let w_{it}^{τ} be the implicit weight that is put, in the construction of the index, on $dsr_{i\tau}$, the average of all ratings given in month τ. Clearly, this weight is zero for $\tau < t-12$ and $\tau \geq t$. Otherwise, it is given by the number of ratings received in τ divided by the total number of ratings received between period $t-12$ and $t-1$. Hence $\sum_{\tau=t-12}^{t-1} w_i^{\tau} = 1$ and

$$DSR_{it} = \sum_{\tau=t-12}^{t-1} w_{it}^{\tau} \cdot dsr_{i\tau}. \tag{1}$$

We want to estimate how $dsr_{i\tau}$ changed after May 2008. That is, we are interested in estimating the parameter β in

$$dsr_{i\tau} = \alpha + \beta POST_{i\tau} + \alpha_i + \varepsilon_{i\tau},$$

where $POST_{i\tau}$ takes on the value 1 after the change, and zero otherwise. The change occurred between the 1st of May and the 1st of June, 2008, and therefore we code $POST_{i\tau} = 1$ if τ is equal to July 2008, or later, and $POST_{i\tau} = 0.5$ if τ is equal to June 2008. This means that we assume that half of the ratings received in May 2008 correspond to transactions taking place after the change.[22] α_i is an individual fixed effect with mean zero and $\varepsilon_{i\tau}$ is an individual- and time-specific error term. We cannot estimate β directly by regressing $dsr_{i\tau}$ on $POST_{i\tau}$ because $dsr_{i\tau}$ is not observed. However, by (1), the reported DSR score is the weighted average rating received in the preceding 12 months, so that

$$DSR_{it} = \alpha + \beta \left(\sum_{\tau=t-12}^{t-1} w_i^{\tau} \cdot POST_{i\tau} \right) + \alpha_i + \left(\sum_{\tau=t-12}^{t-1} w_i^{\tau} \cdot \varepsilon_{i\tau} \right).$$

$\sum_{\tau=t-12}^{t-1} w_i^{\tau} \cdot POST_{i\tau}$ is the fraction of DSRs received after the 2008 change of the system. Hence,

downloaded from their server. Otherwise, we would be able to assess whether the curve indeed flattens out one year after the change. The remarkable fact, however, is that the scores start increasing rapidly and immediately after the change.

[22] This is conservative in the sense that, if anything, it would bias our results downwards because we would partly attribute a positive effect to the time prior to the change. Then, we would (slightly) underestimate the effect of the change. See also the discussion in Section **??** and the robustness check in Section 7.3.

we can estimate α and β by performing a fixed effects regression of the reported DSR score on a constant term and that fraction.[23] We can control for time trends in a similar way.[24]

It is important to control for fixed effects in this context because the DSR score is only observable for a selected sample of sellers, namely those who were involved in enough transactions so that the DSR score was displayed. Otherwise, results may be biased; for example because the DSR score of worse sellers with lower α_i's would be less likely to be observed before the change because by then they would not have received enough ratings. Thereby, we also control for seller exit when studying effects on staying sellers' behavior. Controlling for fixed effects is akin to following sellers over time and seeing how the DSR score changed, knowing the fraction of the ratings that were received after the feedback change. This is generally important because we are interested in the change in the flow of DSRs that is due to the change of the May 2008 change of the feedback mechanism.

Our results will turn out to be robust to controlling for a time trend, however. In light of Figure 2 this is not surprising, as it already shows that there was no time trend in the reported DSR scores before May 2008. After that, DSR scores increase slowly over time, but we will show that this is mainly driven by the fact that DSR scores are averages over DSRs received in the previous year, and the fraction of DSRs received after May 2008 increased gradually over time. Consequently, the DSR scores will also only increase gradually, even if the flow of DSRs jumps up and remains unchanged at a higher level after the change.

Table 2 shows the regression results using DSR scores averaged over the four detailed scores of all sellers and using all 18 waves of data. We first look at the first three columns and will discuss the last two below. In specification (1), we use the whole sample and find an effect of 0.0581. In specification (2), we restrict the data set to the time from March 1 to October 1, 2008; hence there are only 30,488 observations. We do so to estimate the effect locally, because this

[23]One may be tempted to object that the weights enter both the regressor and the error term and therefore, the estimates will be biased. What is important here, however, is that $POST_{i\tau}$ is independent of $\varepsilon_{i\tau}$ because the change to the system was exogenous. Then, the fact that the weights are known will allow us to estimate the effect of the change. To see this suppose that there are two observations for each individual, consisting of the DSR score and the fraction of DSR received after the change, respectively. Then, one can regress the change in the DSR score on the change of that fraction, constraining the intercept to be zero. This will estimate the change in the mean of received DSR before vs. after the change, which is our object of interest.

[24]For two separate time trends, the regressors are weighted average times before and after the change. When we subtract the time of the change from those, respectively, then the coefficient on the indicator for the time after the change is still the immediate effect of the change. The change in the trend can be seen as part of the effect. We will also make a distinction between a short-run effect and a long-run effect. For this, the regressors will be the fraction of ratings received until the end of September 2008, and thereafter.

Table 2: Effect of the May 2008 change

	(1) full sample	(2) small window	(3) time trend	(4) DSR<4.75	(5) DSR≥ 4.75
average DSR before change	4.7061***	4.7030***	4.7149***	4.5912***	4.8138***
	(0.0007)	(0.0005)	(0.0034)	(0.0011)	(0.0006)
effect of feedback change	0.0581***	0.0414***		0.0904***	0.0316***
	(0.0024)	(0.0047)		(0.0044)	(0.0021)
effect of feedback change until September 2008			0.0169**		
			(0.0083)		
effect of feedback change after September 2008			0.0589***		
			(0.0184)		
linear time trend before change			0.0009**		
			(0.0004)		
linear time trend after change			0.0007		
			(0.0019)		
fixed effects	yes	yes	yes	yes	yes
R^2	0.0580	0.0131	0.0603	0.0809	0.0466
number sellers	5,225	4,919	5,225	2,337	2,337
number observations	67,373	30,488	67,373	31,260	33,508

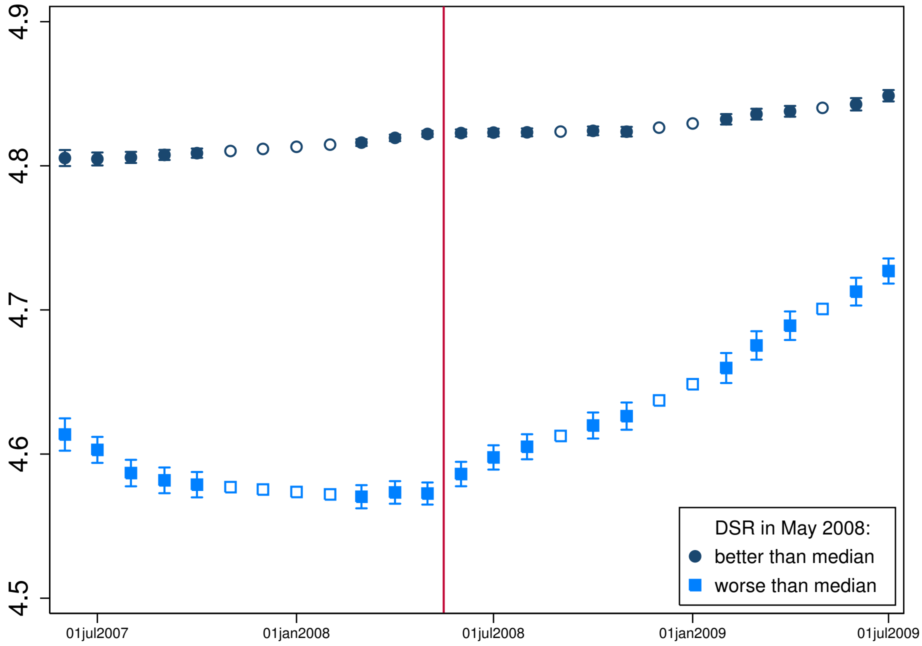
Notes: This table shows the results of regressions of the average DSR score, averaged over the four categories, on a constant term and the fraction of feedbacks received after May 2008. We assume that half of the feedbacks in May 2008 were received before the change and the other half after the change. In specification (2), we do not include observations after October 2008 and before March 2008. In specification (3) we distinguish between the effect until September 2008 and after that date, and also account for a piecewise linear time trend. Specification (4) includes only those sellers who had a DSR score below the median of 4.75 in May 2008 and (5) only those above the median. One observation is a seller-wave combination. Throughout, we control for fixed effects. The R^2 is the within-R^2. Standard errors are cluster-robust at the seller level and significance at the 5 and 1 per cent level is indicated by ** and ***, respectively.

allows us to see how much of this global effect is due to an immediate response by sellers. The estimated effect is equal to 0.0414, which suggests that most of the effect occurs from mid-May to October 1, 2008. In specification (3), we instead allow for a piecewise linear time trend over the entire observation window. We find that the time trend before the change is very small and not significantly different from zero after the change. The effect of the change is estimated to be a short-run effect of 0.0169, until the end of September 2008, and a bigger effect of 0.0589 after that.[25]

To assess the magnitude of the effect, it is useful to express the numbers in terms of quantiles of the distribution of DSR scores among sellers prior to the May 2008 change. According to the results in the first column, the average DSR before the change is 4.7061, and after the change,

[25] Without the piecewise linear time trend the short run effect is estimated to be equal to 0.0325 and the long run effect is estimated to be 0.0711, with standard errors 0.0057 and 0.0028, respectively. Then, the magnitude of the short run effect is comparable to the one of the effect using the smaller sample that is reported in column (2).

Figure 3: Evolution for two different groups



Notes: The figure shows how the average DSR score changed over time, with sellers split into those who had a DSR score above the median of 4.75 prior to the May 2008 change, and those who had a score below that. The circles and squares are linearly interpolated values for the periods in which we did not collect data. The error bars depict pointwise 95 per cent confidence intervals.

it is $4.7061 + 0.0581 = 4.7642$. This corresponds to roughly the 40 and 60 per cent quantiles of the distribution of ratings prior to the change, respectively. Hence, the May 2008 change has led to a *significant and sizeable increase in seller effort.*

We now look at how this increase is differentiated between sellers with low, and high DSR before the change. Towards that, we split our sample at the median DSR of 4.75 just before the change between high and low ranked sellers before the May 2008 change. Figure 3 gives the picture. The increase in DSR score is stronger for sellers with below-median score *ex ante*. Furthermore, the difference in the increase between the above-median and the below-median sellers is significantly different from zero.[26] The last two columns of Table 2 report the corresponding estimates, again controlling for seller fixed effects. The difference between the effect for above- and below-median sellers is significantly different from zero. We obtain similar results

[26]Of concern may be that the increase for the sellers with low DSR before the change may be driven by mean reversion. Indeed, we have divided sellers based on their score. To check whether mean reversion has to be accounted for we instead divided sellers according to the median score on August 1, 2007. With this, scores for the bad sellers also only increase after the change. This shows that mean reversion is not of concern here.

when we perform regressions for those two different groups only for a smaller time window, as in specification (2), or control for time trends, as in specification (3). In the second part of Table 6 in Appendix B discussed later within robustness checks, we show the effects of the feedback change by decile of sellers' DSR rating. Note the decline in the magnitude and significance of the effect, with increasing decile.

Recall again that the system change was not with respect to DSR, but with respect to the classic reporting mechanism. The anonymous and unilateral DSR were established one year before the May 2008 change whose consequences we consider here, and they remained anonymous and unilateral thereafter. Already with the DSR introduced in May 2007, buyers had been able to express their true valuation of seller performance without fear of retaliation by that seller. By looking at the effect of changing the non-anonymous established reporting mechanism, we pick up only an additional effect. It is remarkable that this effect shows up as clearly as documented above.

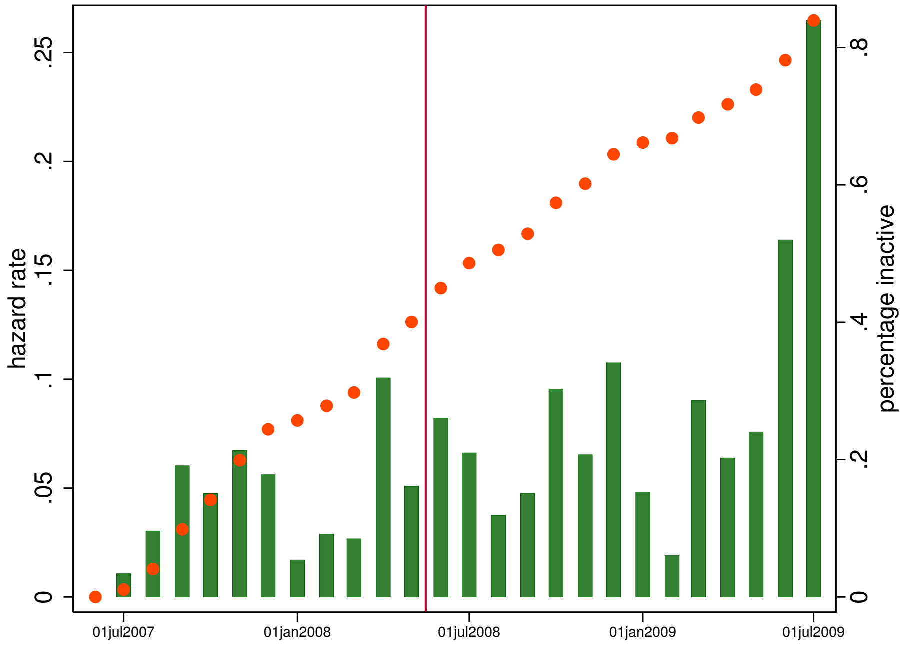
In all, the empirical evidence provides support of our hypothesis that abandoning negative buyer rating by sellers—and thereby reducing impediments against negative seller rating by buyers—has led to substantive seller reactions. Buyer ratings—especially of *ex ante* poorly rated sellers—improved significantly, which, as we will argue in the ensuing section, results from an improvement in the behavior especially of poorly performing sellers, and with this, a reduction in moral hazard.

5.2 Seller Exit

Recall that the May 2008 change provided additional means to buyers to voice negative experiences without fear of negative seller reaction. This should also have incentivized sellers performing poorly before the change to leave the market. To empirically assess this, we looked at the exit rate of the 10 per cent worst sellers, as measured by their DSR rating on May 1, 2008. Figure 4 shows how the hazard rate into inactivity and the fraction of individuals who have become inactive changed over time.[27] We see that overall, many sellers leave over time, both before and after the change. By June 1, 2009, almost 80 per cent of those sellers have become inactive. It is interesting to contrast this to the pattern for the 10 per cent best sellers,

[27] The stronger increase in that fraction towards the end of that window is due to the truncation bias naturally reflected in our observations, by which we incorrectly sort infrequent sellers into the set of exiting firms towards the end of the window.

Figure 4: Exit from the market



Notes: The bars show the fraction of users that have become inactive in the previous month, among the 10 per cent poorest performing sellers as measured by their DSR rating on May 1, 2008, with the percentage on the right axis. The dots denote the cumulative fraction of users who have become inactive since June 1, 2007, with the hazard rate on the left axis.

again measured by their DSR rating on May 1, 2008. For them, we find that only about 45 per cent have become inactive by that time. This shows on the one hand that eBay is a very dynamic environment, and suggests on the other hand that better performing sellers are more likely to stay in the market, as compared to poorly performing sellers.[28]

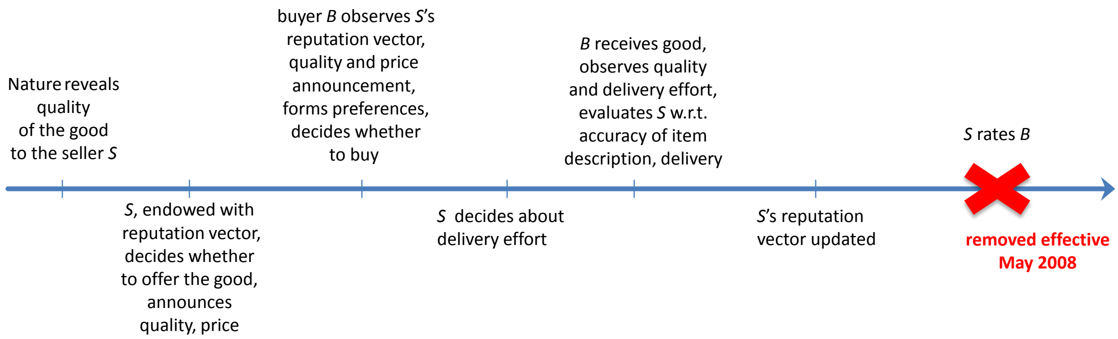
The most important finding, however, is that *the May 2008 change did not trigger any significant increase in the exit rate of sellers.* We also formally tested whether the hazard rate was different before and after the change. Towards this, we excluded the month of June because the change occurred Mid-Mai, and tested whether the hazard rate in July and August was equal to the hazard rate in April and May. The difference between the two is estimated to be 0.003 with a standard error of 0.002, so it is not significantly different from zero. If we focus on the 10 per cent worst sellers, as measured by their DSR rating on May 1, 2008, we find that the difference is estimated to be -.0247 with a standard error of 0.0148, so also for those sellers the

[28]Recall that we follow a sample of sellers over time, and hence we cannot study entry.

estimated change in the hazard rate is not significantly different from zero.[29]

6 A Simple Explanatory Paradigm

Figure 5: Sequence of decisions in a typical eBay transaction



In this section, we develop an explanation of these results, and in the ensuing section, we defend it against alternative explanations. Our explanation is based on a toy model involving one stage in an infinitely repeated seller-buyer game. Rather than concentrating on auctions that have substantively lost market share in internet markets, our toy model is based on a fixed price offer by a typical seller, and one transaction with an interested buyer. The explanation concentrates on the effects of removing the hold-up on the typical buyer's evaluation, that before May 2008 was generated in the classic feedback system by the fact that the seller could retaliate any negative rating by the buyer.

The sequence of decisions in such a typical eBay transaction is condensed in the time line contained in Figure 5. We focus on a rating sequence involving the seller's rating of the buyer *after* the buyer's rating of the seller, with the following justification. In Klein, Lambertz, Spagnolo, and Stahl (2006), we found that the seller rated his counterpart before the buyer did so in only 37 per cent of all cases in which the rating was mutual[30] ; and that in this sequence, a positive rating by the seller was followed by a negative buyer rating in less than one per cent, indicating that the hold up situation we consider to be at the root of the phenomenon analyzed here is not prevalent in this case.

Assume that the good to be traded can take on one of two qualities, q_h and q_l with $q_h > q_l$,

[29] We have also checked whether exit rates increased for other groups after the feedback change, but found no evidence for that.

[30] See also Bolton, Greiner, and Ockenfels (forthcoming) for a similar finding.

selected by nature and revealed to sellers at the beginning of the stage game. The good can be offered at competitive price $p_h > p_l > 0$ and delivered at some dis-utility, or cost of effort. Sellers are differentiated by that effort cost. When engaging in high effort, seller type i faces effort cost c_i, $i \in \{L, H\}$, with $0 < c_L < c_H$. When engaging in low effort, that effort cost is normalized to zero for both types of sellers. The typical seller is endowed with publicly known reputation capital consisting of two DSRs, namely *accuracy of item description* denoted by k^d, and *quality of shipping* denoted by k^s, both taking on values on some closed interval on the positive real line. That reputation capital is built from buyer reactions to his behavior in previous transactions.

When offering the good (at production cost normalized to zero), the typical seller S decides whether to announce it at its true quality q_j and the appropriate price p_j, $j \in \{l, h\}$, which he always does if $j = h$; or to shirk if $j = l$, by announcing the low quality good as of high quality, q_h, at high price p_h.[31] If the buyer orders the good, the seller decides whether to spend effort on its delivery. That seller, however, is considered an opportunistic neoclassical: unless punished via a reduction in his reputation capital, he exploits on the anonymity in the market, by shirking on the description of the quality of the good to be traded, and by providing low effort.

A randomly arriving buyer B, not knowing the true quality of the good, observes the quality-price tuple announced by the seller denoted by $[\hat{q}_j, \hat{p_j}], j \in \{l, h\}$, as well as the seller's reputation vector $[k^d, k^s]$. On their basis she forms an expected utility $\mathbb{E}u[\hat{q}_j, \hat{p}_j, k^d, k^s]$, $j \in \{l, h\}$. Natural assumptions on this utility are that it increases in the first, the third and the fourth argument, i.e. the (announced) quality and the seller's reputation; and decreases in the second argument, i.e. the announced price. She decides to buy the item if $\mathbb{E}u[\hat{q}_j, \hat{p}_j, k^d, k^s] \geq \widetilde{u}$, where $\widetilde{u}$ is her exogenously specified outside option.

In case the buyer decides to buy, seller type i decides whether to engage in effort towards delivery. If providing low effort, seller type i's pay off is $\hat{p}_j$, and thus always positive. If providing high effort, seller type i's pay off is $\hat{p}_j - c_i$, which is always positive if $j = h$ no matter i, positive if $j = l$ and $i = L$, but negative if $j = l$ and $i = H$. Hence, if the good is of low quality and announced this way by the high effort cost seller, his zero profit participation constraint in the stage game assumed here is violated when he intends to provide effort. Finally, buyer B receives

[31] Recently, this format in which prices are announced, has become increasingly popular on eBay, as compared to the classic auction format (Einav, Farronato, Levin, and Sundaresan, 2013). In our toy model, we could replace it by an auction. In that case, the winning bid would reflect the quality of the good as (more or less incorrectly) described by the seller..

the good, observes the accuracy of the item description and the shipping quality, and rates S in terms of these two dimensions. This results in a natural upwards, or downwards revision of both, k^d and k^s, that enters next period as the seller's reputation capital.

Before May 2008, the sequence of decisions involving such a transaction was typically concluded by the additional step indicated in Figure 5, in which the seller rated the buyer along the classic scale, and this rating was publicly observed. We assume that the buyer derives value from being rated well, e.g. because she intends to use such ratings as reputation capital when entering as a seller in the future.[32] Decisions are supposed to be taken rationally, that is, with backward induction in that simple stage game, that is repeated infinitely.

Towards results from this toy model, consider first the sequence of decisions before the May 2008 change. Since the buyer derives value from being rated well, the typical seller S can opportunistically condition his rating on the buyer's rating observed by him, by giving a negative mark if the buyer does so. Recall in this context that the anonymous DSR cannot be given without a non-anonymous classical rating. Hence retaliation by the seller implies that a buyer with strong reputational concerns is captive to the seller's rating, and thus forced to rate S positively in the classic feedback necessarily preceding the DSR, no matter the seller's decisions taken before—and observed by the buyer after the transaction has taken place. In this case, if nature selects q_l, the seller, being opportunistic, shirks with probability 1 on the buyer, by announcing the low quality good at high price p_h, and by not taking any effort to deliver the good—yet still receives a positive contribution to his reputation capital.

By contrast, with eBay's change in the rating mechanism effective May 2008, even the buyer with reputational concerns can give an unbiased negative rating without fearing retaliation. There is abundant evidence that a seller, who intends to stay in the market, must be very concerned with his reputation because he can sell more rapidly, and at higher price. The May 2008 change then implies that such a seller must accurately describe the item even if of low quality, and quote an appropriately low price. Towards obtaining a mark by the buyer that does not negatively influence his reputation capital, he must also take effort in delivering the item. With the assumptions made above, this implies a negative stage payoff$\hat{p}_l - c_H$ for the high effort

[32]More generally, while there is little theoretical work on why exactly buyers have such reputational concerns, empirical findings stingily suggest that this is the case (Resnick and Zeckhauser, 2002; Klein, Lambertz, Spagnolo, and Stahl, 2006; Bolton, Greiner, and Ockenfels, forthcoming). Buyers may have reputational concerns because they simply don't like to be rated negatively, or because it could harm them in a future transaction, when acting as a buyer or as a seller.

cost seller.

Consider now the special case of a seller endowed with low reputation capital before the May 2008 change. He faces two alternatives: either to exit the market—but before then profitably depleting his reputation capital, by shirking, i.e. selling the low quality good at high price and by not providing costly effort towards delivery, resulting in stage payoff $p_h > 0$; or alternatively to forgo that short run rent and to continue operating in the market—but then to provide goods in a way that his reputation capital increases very strongly even if his stage pay off is negative, because this allows him to sell high quality goods at high price later on.

In all, on the basis of this toy model, the May 2008 change disciplines sellers, and thus results in two main effects: *a reduction in moral hazard exercised by the sellers intending to stay in the market;* and/or, *a reduction in adverse selection exercised by the exit of poorly rated sellers.* Moral hazard is reduced via an increasingly accurate item description and increased delivery effort of the sellers remaining in the market and results in an improved buyer evaluation; adverse selection is improved via an increase in the exit rate of sellers poorly rated before the May 2008 change. Alternatively, if the poorly rated sellers continue to stay in the market, we should see an above average contribution to the reduction in moral hazard, towards an improvement in their reputation capital.[33]

Rather than observing both effects, we only observe a significant improvement in buyer ratings, which we interpret as a factual improvement of seller behavior; and do not observe any increase in the exit rate of poorly rated sellers—yet a particularly strong increase in their ratings after the May 2008 change.

Why does the poorly performing sellers' reaction to the May 2008 change appear to be so asymmetric, against exit, and for improved performance in place? Along the lines of our toy model, the share of high opportunity cost sellers appears to be small, so *improving on the performance*—as reflected in buyer evaluations—*is still profitable for sellers* in the long run, *even if nature selects a low quality/low price item for them.* Clearly, giving up on shirking, by correctly describing and selling a low quality good at low price, involves the opportunity cost of foregoing a possibly large rent of selling that good at a high price. Yet that rent must be held

[33] As to detail, if buyers' feedbacks are delayed, then we predict from this simple paradigm a downward jump in the feedback score right after the May 2008 change, resulting from the fact that before that change, sellers exercised moral hazard in transactions rated negatively by the buyers right after the change, whereas in transactions after the change, sellers would strategically anticipate unbiased buyer rating.

against the depletion of the reputation record.

7 Additional Empirical Support, Competing Explanations, and Robustness

In Section 5, we have shown that removing negative seller ratings of buyers in eBay's classic feedback system, and with it potential retaliation to negative buyer ratings, has resulted in a significant improvement in DSRs especially for sellers that previously were rated poorly; and no change in sellers' exit behavior, especially that of the poorly rated ones. In Section 6, we gave an explanation that is consistent with these results. In this section, we present additional evidence that supports the underlying assumptions. We then go through a list of competing explanations and show that these are likely not to hold. We finally discuss the robustness of our results to assumptions on the timing of the change, and on delays between transaction and rating that we have made in order to estimate the effect of the change.

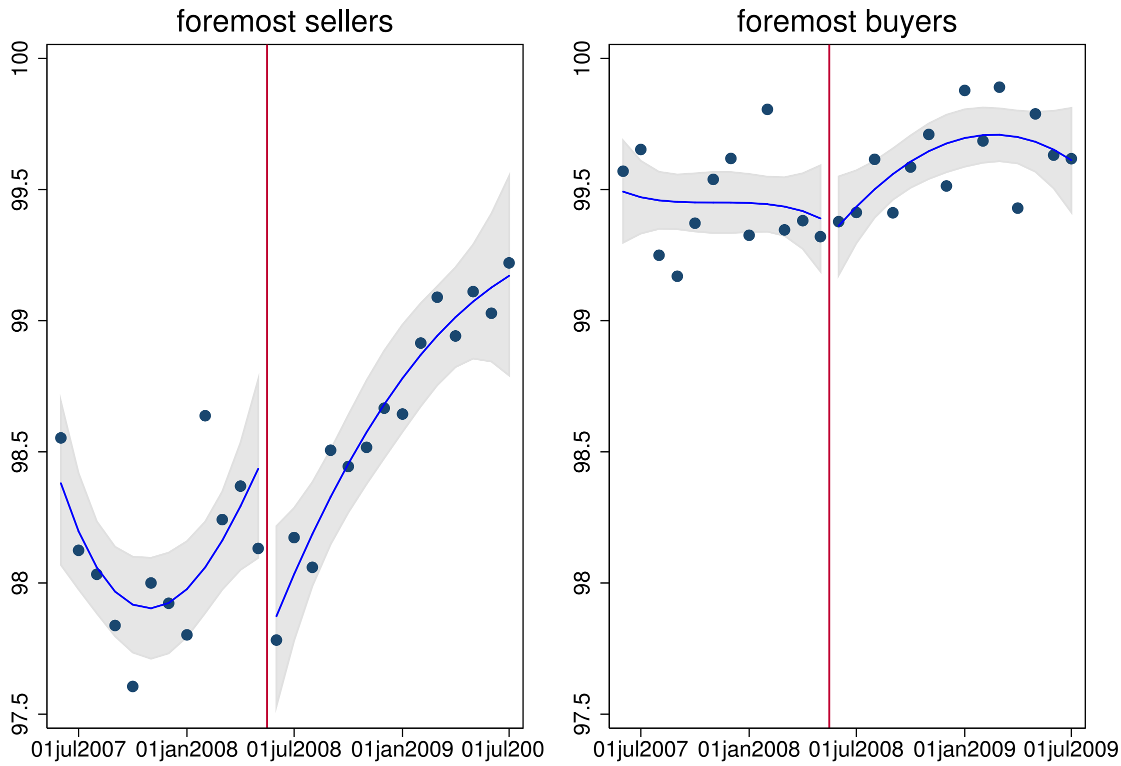
7.1 Empirical Support for Assumptions Underlying Our Explanation

One might first want to question whether there actually are different seller types so that adverse selection can arise at all—otherwise, only moral hazard would play a role. Towards an answer, recall that we have included two parameters in the specification of our key regression, namely a seller fixed effect α_i and a seller-specific time varying effect ε_{it}. The fraction of the variance of $\alpha_i + \varepsilon_{it}$, at a given point in time and across sellers, that is due to variation in α_i gives us an indication of the relative strength of the seller fixed effect.[34] In the five specifications reported in Table 2, this fraction (x100) amounts to 84, 94, 84, 77 and 54 per cent, respectively. Only the last fraction is low. But that reports on the above-median sellers. This demonstrates that most of the heterogeneity across sellers is time-invariant, so that that differences across sellers are much more important than seller specific differences in outcomes over time. This is in line with our view that sellers differ in their (opportunity) cost.

Next, one might wonder whether buyers have reputational concerns. If this was not the case, then buyers would have no incentive to modify their rating behavior after the change. Recall that the classic feedback given to a specific transaction was (and is) identifiable by the seller

[34] We calculate it in the usual way after performing fixed effects estimation.

Figure 6: Effect on classic feedbacks



Notes: The left figure shows how the percentage positive feedbacks changed over time. The lines are fitted values of local quadratic regressions and the shaded area shows pointwise asymptotic 95 per cent confidence intervals, respectively. We used the Epanetchnikov kernel with a bandwidth of 200. The dots are averages per wave. The solid vertical line depicts the change to the classic feedback mechanism. The left figure is for those eBay users who were foremost sellers, and the right figure for those who were foremost buyers. To classify users, we used the ratio between DSR and classic feedbacks received on May 1, 2008. In particular, we classify those 25 per cent users with the highest ratio as foremost sellers and the 25 per cent with the lowest ratio as foremost buyers.

(and the observer). Recall also that our analysis so far was concentrating on the subsample of the users with enough transactions as sellers so they obtained DSR records. We study the effect of the May 2008 change on the classic feedback ratings received by all users sampled, buyers and sellers. Towards conducting the analysis separately for these two groups, we first classify all users sampled as being foremost sellers or buyers on eBay. We based our classification on the ratio between DSR and (cumulative) classic feedbacks received by May 1, 2008. Specifically, we classify those 25 per cent users with the highest ratio as foremost sellers and the 25 per cent with the lowest ratio as foremost buyers.

In Figure 6 we compare the percentage of positive feedbacks obtained for the two subpopu-

Table 3: Effect on classic feedbacks

	foremost sellers				foremost buyers			
bandwidth	50	100	200	300	50	100	200	300
local linear	-.369	-.542*	-.328	-.256	.052	.053	-.002	.030
	(.470)	(.296)	(.227)	(.204)	(.219)	(.177)	(.122)	(.125)
local quadratic	-.490	-.408	-.727***	-.762***	-.085	.098	-.049	-.081
	(.638)	(.497)	(.296)	(.349)	(.378)	(.230)	(.187)	(.158)

Notes: This table shows estimated effects of the feedback change on the percentage positive classic ratings received by users who were either foremost sellers or buyers. These were obtained by performing kernel regressions. We used an Epanetchnikov kernel. The cells contain estimates for local linear and local quadratic regressions and the respective standard errors in parentheses. Each column corresponds to a different bandwidth. To classify users, we used the ratio between DSR and classic feedbacks received on May 1, 2008. In particular, we classify those 25 per cent users with the highest ratio as foremost sellers and the 25 per cent with the lowest ratio as foremost buyers. This leads to 22,762 observations for the first group and 26,262 for the second group, coming from 1,168 and 1,169 users, respectively. Bootstrapped standard errors are cluster-robust at the seller level. Significance at the 10 and 1 per cent level is indicated by * and ***, respectively.

lations in the observation window. It shows very clearly that effective May 2008, the percentage of positive feedbacks dropped for users identified as foremost sellers, but remained unchanged for those identified as foremost buyers. Our explanation is as follows: Typically there is a delay between any transaction and its evaluation by the buyer. As the date of the change enacted in May 2008 could not be anticipated by the sellers, they did, on average, exhibit more shirking and carelessness in delivery right before, rather than right after the May 2008 change. Effective this very date, however, buyers could leave negative classic ratings without the risk of retaliation by the seller. In the left figure, we therefore observe a downward jump in buyer ratings right after the May 2008 change, i.e. before sellers could react to that change. The left figure suggests also that, as ratings increase, they do react as time goes on.[35]

Table 3 contains the corresponding formal tests. The four columns on the left contain results for foremost sellers, and the four columns on the right results for foremost buyers, following our classification. Each column corresponds to a different bandwidth for the kernel regressions, and in the rows we show results for a local linear regression and a local quadratic regression. Figure 6 suggests that a bandwidth of 200 together fits the data well when we use a local quadratic specification. The corresponding estimate for sellers is -.727. It is significant at the 1 per cent

[35]Keep in mind, however, that some of these ratings could still have been received by those users in their role as buyers.

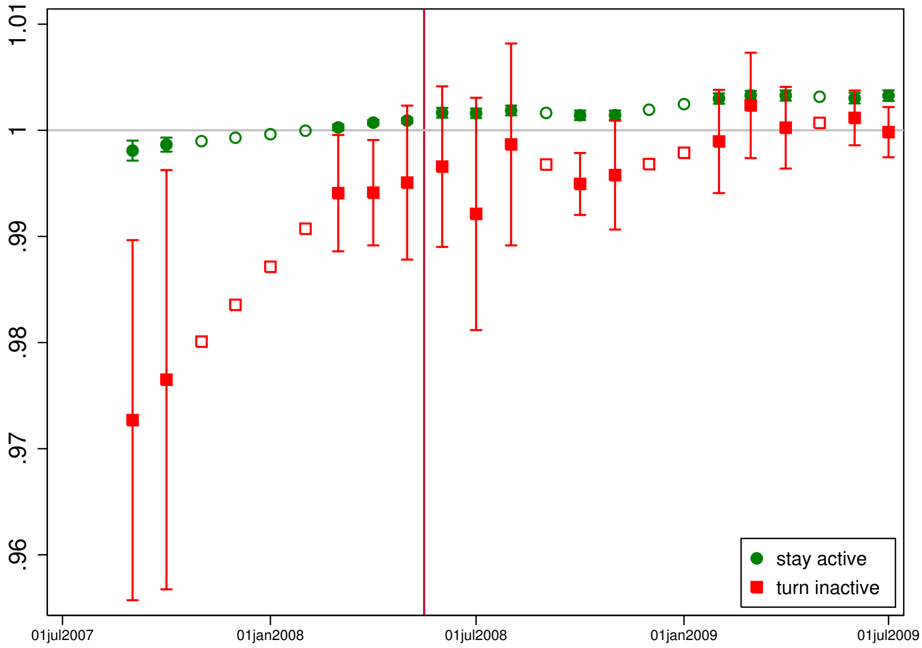
level.[36] Observe finally that the estimated effect for buyers is never significantly different from zero. Taken together, this provides strong evidence for buyers having reputational concerns.

Finally, in an anonymous market such as the one under discussion, we expect rational sellersto change their behavior just before leaving the market. In particular, we expect them to exert more moral hazard. In light of this, we consider it as an indication supporting our claim that buyers correctly value the transaction via the DSR if we see buyer ratings on exiting sellers to degenerate. In Figure 7, we compare the continuing and the exiting sellers' DSR scores, relative to those obtained three months earlier. Whereas the relative DSR scores of the continuing sellers remain essentially unchanged in the time window considered, the exiting sellers' DSR relative scores go down on average. The error bars seem to suggest that this difference is not significantly different from zero. As before, however, the confidence intervals are point-wise. The ratio is significantly different between those individuals exiting and those staying when we pool over the time periods. The corresponding regression with standard errors clustered at the seller level shows that the point estimate of the intercept, which is the average over the dots for the stayers in the figure is 1.001, with a standard error of 0.0001. Towards appreciating this interpretation, keep in mind that exits concentrate on the poorest sellers, and that we report here on the relative degeneration of their DSR scores. This suggests the causality indicated above that DSRs decline because sellers plan to exit. We cannot exclude the opposite argument, that sellers exit because DSRs are declining. Yet it would not violate our claim that DSRs are reasonable measure of seller performance if that would be the correct causality.

More importantly, the coefficient on an indicator for becoming inactive is -0.005 with a standard error of 0.0008. This means that the ratio is significantly lower for individuals who retire from the market, indicating that performance trends downwards before retirement. The standard deviation of the ratio in a given wave, e.g. May 2008, is 0.0077, so the effect is equal to 65 per cent of this, which arguably is non-negligible. In all, this reinforces the results derived by Cabral and Hortaşçsu (2010) we have referred to in the literature section 2, and supports our view that on the one hand, sellers have reputational concerns, and on the other hand DSRs are

[36]The pointwise confidence interval in Figure 6 would suggest that the effect of the change is not significantly different from zero at the 1 per cent level. Any conclusion based on this would be mis-leading, however. The proper test is based on the estimated size of the jump and the standard error for that size: the variance of the difference between the two curves is not the sum of the two variances, but smaller, because the covariance between the two is positive.

Figure 7: Behavior prior to inactivity



Notes: In this figure we compare the ratio between the DSR and the DSR three months before for exiting users (depicted by the squares) to that of the stayers (depicted by the dots). We used linear interpolation in case we did not collect data for the latter DSR score when calculating the ratio. Circles and squares are linearly interpolated values for the periods in which we did not collect data. The error bars depict pointwise 95 per cent confidence intervals.

related to the outcome of the transaction.[37]

7.2 Competing Explanations

The empirical evidence in this paper consists of a comparison of DSRs received after the change to DSRs received prior to the change. Our interpretation that the increase in DSRs after the change hinges on the assumption that this was not caused by other contemporaneous changes to the rules of eBay, or changes in buyer or seller behavior that are unrelated to the change to the system. Ideally, we would assess this using a "control group" from a market in which comparable sellers and buyers interact exactly in the same way as they did on eBay, except that there was no change to the feedback mechanism. Unfortunately, such a market does not exist. In the following, however, we go through an extensive list of competing explanations, and

[37] One might wonder the strong attenuation in the effect after May 2008 on buyers turning inactive. One should be careful, however, in the interpretation for the effect in the first two months reported in the figure (the following months indicate imputed averages), as they suffer from a leftward truncation bias.

conclude that none of those is likely to have have caused the increase in DSRs.

First, the results could have been generated simply by grade inflation rather than sellers' effort. Figure 2 strongly speaks against that, as there is no grade increase before, but a significant one after the May 2008 change.

The second competing explanation is that before the change, some buyers who wanted to leave a negative rating without retaliation could do so only by leaving a negative DSR. After the change, they could safely leave a negative classic rating, and therefore abstain from leaving that negative DSR. By this, DSR ratings would just increase because negative DSRs would not be left anymore by those users. One way to test this is to check, for each seller, whether the number of DSRs relative to the number of classic ratings has decreased after the change. The numbers in Table 1 already suggest that this was not the case. Towards a formal test we ran a regression, controlling for fixed effects, to estimate the change between that ratio on May 1, 2008 and July 1, 2009. In both cases, the ratio is for the preceding 12 months. This regression uses only sellers for which DSR ratings were available at both points in time. The ratio on May 1, 2008 is 0.4211 and the estimated change in the ratio is 0.0384, with a standard error of 0.0025. This shows that, if anything, the number of DSRs per classic ratings has increased, invalidating the above mentioned concern.

An alternative, third explanation of our results could be another change in eBay's allocation mechanism implemented within our observation window. Three months prior to the change whose effects are discussed here, eBay changed the order in which listings were displayed when buyers on eBay searched for an item. Before that, offers were simply ranked by the time remaining until the offer was closed. Under the new ranking scheme, called "Best Match" (BM), a number of factors determined which listings appeared first. One of these factors was the DSR score, and therefore the introduction of BM provided an incentive for sellers to improve their performance.[38] The ranking scheme was modified several times since. The exact way of ranking listings is a trade secret highly guarded by eBay, as is e.g. Google's search algorithm. We now assess whether the introduction of BM could have geared our results.

[38] That change was obviously motivated by the increased attractiveness of the fixed price over the auction format to sellers: *A related reason was, we introduced the fixed price format of listings. They could be 30 days, 60 days, and 90 days. And when you have fixed price listings that can be live on the site for 30 days, 60 days, 90 days, "time ending soonest" which was a sort on eBay, no longer made sense for those types of listings. You have a 30-day listing that might only come up to the top of the results 30 days after it was listed. So we had this problem, lots of fixed price inventory, 30 days and 60 days.* (Taken from http://files.meetup.com/1537023/Best_Match_Transcript.doc, last accessed in June 2013.)

Table 4: Effect of the May 2008 change without 10 per cent worst sellers

	(1) full sample	(2) small window	(3) time trend	(4) DSR<4.75	(5) DSR≥ 4.75
average DSR before change	4.7400***	4.7426***	4.7545***	4.6465***	4.8129***
	(0.0006)	(0.0004)	(0.0030)	(0.0010)	(0.0006)
effect of feedback change	0.0535***	0.0382***		0.0844***	0.0306***
	(0.0022)	(0.0038)		(0.0041)	(0.0021)
effect of feedback change until September 2008			0.0016		
			(0.0074)		
effect of feedback change after September 2008			0.0682***		
			(0.0174)		
linear time trend before change			0.0016***		
			(0.0004)		
linear time trend after change			-0.0013		
			(0.0018)		
fixed effects	yes	yes	yes	yes	yes
R^2	0.0703	0.0165	0.0748	0.1067	0.0437
number sellers	4,047	4,047	4,047	1,794	2,253
number observations	58,004	26,358	58,004	25,390	32,614

Notes: See notes to Table 2. Here, we additionally exclude the 10 per cent worst sellers, as measured by their DSR score on March 1, 2007.

Feiring (2009, 3rd ed, p.16) reports that within the time window of our analysis, the ranking induced under the BM scheme affected only the very poorest sellers, namely those for whom *Item as Described* and *Communication, Shipping Time*, or *Shipping and Handling Charges* were ranked only 1 or 2 (out of 5) stars in more than 3 per cent, and more than 4 per cent of their transactions, respectively. We concentrate our robustness check on these. We will show first, that this is a small group of sellers, and second that excluding them from our analysis does leave our results essentially unaffected.

As the first order effect of introducing BM, we expect the sellers with relatively poor records to realize fewer transactions, and correspondingly obtain significantly fewer DSRs. So we looked at shifts in the number of DSRs received post March 2008 by percentile of sellers distributed by DSR scores. Table 6 in Appendix B shows that the number of DSRs received after the introduction of BM decreased significantly only for the 10 per cent poorest sellers (the effect is -5.76 from a level of 44.44 ratings per month before that, with a standard error of 1.50). Re-doing the regressions that underlie the results in Table 2 and dropping the 10 per cent worst sellers yields Table 4. It is immediate that the results are very similar, thus supporting our claim that our analysis is not affected by the introduction of BM.

A fourth alternative reason why ratings could have increased is that buyer demand has shifted from auctions to fixed-price offerings, as documented by Einav, Farronato, Levin, and Sundaresan (2013), in particular in their Figure 1. As one can see there, however, the decrease was gradual at least until September 2008, while our Figure 2 shows that DSRs increased already before that, right after the change to the feedback mechanism.[39] Moreover, and more importantly, one would not expect that a change in format would have an effect on buyer satisfaction as measured in our paper. The reason is that the DSR score we used is the average over the DSR score in four categories—namely item description, communication, shipping charges and shipping speed—and none of them is arguably related to whether or not the item has been offered in an auction. After all, our interpretation as based on our toy model equally applies to auctions.

7.3 Robustness checks

We don't know the exact date of when eBay enacted the change whose effects we are reporting here. For simplicity of our analysis, we also considered coincident the timing of a transaction and its rating.[40] The effects of moving the date of the change and introducing delays between transaction and rating are clearly confounding—also with delays in the sellers' and the buyers' perceptions of the date of change. In view of the coarseness of our data collection, we will discuss them within one sensitivity analysis.

Suppose for now that both sellers and buyers would know the exact date of the change (around May 15, 2008), and the seller changed his behavior in a transaction right after the change. That transaction was probably completed by May 25 and this was also the time at which the buyer left a DSR for him. Conversely, if a transaction took place before May 15, 2008, then the seller was not able to react to the feedback change, as it was unanticipated. Nevertheless, a feedback could have been left for that transaction in the second half of May 2008. So far, we have assumed that half of the DSRs received in May 2008 corresponded to transactions conducted after the feedback change. Disregarding this reporting delay, we attribute the ratings after the change all to transactions thereafter, and with this tend to underestimate the effect of the feedback change. We do not expect this to have big effects, however, because

[39] See their Section 7 for an explanation why the jump in September 2008 is mechanical.

[40] Net of the analysis of classic feedback discussed in subsection 7.1 by which we further supported our explanation.

the delay is likely to be small relative to the length of our observation period.

We don't have a record on delays between transaction and feedback. Yet Figure 2 in Klein, Lambertz, Spagnolo, and Stahl (2006) shows the distribution of the time between the end of the auction and the moment at which the first feedback was left. The vast majority of feedbacks is positive and for those about 60 per cent are left after 2 weeks and almost 90 per cent are left after 4 weeks.[41] Based on this, we re-did the analysis of Section 7.2 relating to BM, assuming that out of all DSRs received in March 2008, 75 per cent of the transactions took place after the introduction of BM. Moreover, we assumed that out of all DSRs received in May 2008, 25 per cent of the transactions took place after the change to the feedback system. Results were very similar.

Finally, we re-did the analysis underlying Table 2, assuming that 25 per cent of the transactions took place after the change to the feedback system. Table 7 in Appendix B shows the results. They are very similar to the ones reported in Table 2 and 4.

8 Conclusion

In anonymous markets, buyers (and sellers) must rely on reports of each others' performance in order to efficiently execute transactions. In this paper, we use changes in the mechanism by which buyers can report on seller performance, to extract effects, which we interpret as seller adverse selection and seller moral hazard, and separate between the two.

Specifically, in May 2008, eBay changed its established non-anonymous feedback system from bilateral to essentially unilateral ratings, by allowing sellers to evaluate buyer behavior only positively, rather than also neutrally or negatively as before. With this, eBay dismissed with buyer fear of seller strategic retaliation to negative feedback given by buyers which, by eBay's own argument, resulted in buyer opportunism when rating sellers.

One year before the change in focus, eBay had introduced unilateral anonymous Detailed Seller Ratings that already allowed buyers to rate sellers without a bias generated by fear of seller retaliation to a negative rating—but retained the classic rating that, because non-anonymous, could be opportunistically biased. We use the unbiased Detailed Seller Rating as the basis for

[41] For negative feedbacks, the distribution is shifted to the right. Klein, Lambertz, Spagnolo, and Stahl (2006) argue that this may be due to strategic considerations: both parties had an incentive to wait with their first rating if it was negative, because then it was less likely to be retaliated.

checking the effect of removing buyer reporting bias via the May 2008 change in the classic rating mechanism.

We show that this change resulted in improved seller ratings by buyers, but no exit of poorly rated sellers. In fact, the poorly rated sellers' ratings improved more than average. Towards the explanation of these results preferred by us, we develop a toy model that focusses on the effects of this natural experiment, from which we derive first, a reduction in seller moral hazard in preparing and executing the transaction; and second, a reduction in seller adverse selection, as generated by the increased exit of poorly performing sellers from the market. Within this context, we can interpret the absence of an effect on seller adverse selection by the relatively low cost of improving on moral hazard.

We check this explanation against a number of competing ones coming to mind, and develop tests supporting the one given by us. Based on our explanation of the empirical results, we conclude that *improvements in (unbiased) information provided by consumer reports, i.e. increases in market transparency, have a significant disciplining effect on sellers.* Indeed, the *incentives given to them this way should improve on welfare*: were it not in their interest would sellers not increase their effort towards improving on transactions.

Our results are derived within the context of an anonymous internet market—a type of market that obviously increases in importance, and is likely to dominate in transactions volume classical markets in the foreseeable future. From a business policy point of view, our analysis provides for a fine example in which relatively small changes in the design of an information mechanism can have a significant effect. From the point of view of academic research, our result is, to the best or our knowledge, the first in which, at least for classical product markets, the effects of reducing buyer–seller informational asymmetries on adverse selection and moral hazard are clearly separated and directly juxtaposed.

References

Abbring, J., P. Chiappori, and J. Pinquet (2003): "Moral hazard and dynamic insurance data," *Journal of the European Economic Association*, 1(4), 767–820.

Abbring, J., J. Heckman, P. Chiappori, and J. Pinquet (2003): "Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish?," *Journal of the European Economic Association*, 1(2-3), 512–521.

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

Anderson, M., and J. Magruder (2012): "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database," *Economic Journal*, 122(563), 957–989.

Aron-Dine, A., L. Einav, A. Finkelstein, and M. Cullen (2012): "Moral hazard in health insurance: How important is forward looking behavior?," NBER Working Paper No. 17802, National Bureau of Economic Research, Cambridge, MA, U.S.A.

Avery, C., P. Resnick, and R. Zeckhauser (1999): "The market for evaluations," *American Economic Review*, 89(3), 564–584.

Bajari, P., H. Hong, and A. Khwaja (2006): "Moral hazard, adverse selection and health expenditures: A semiparametric analysis," NBER Working Paper 12445, National Bureau of Economic Research, Cambridge, MA, U.S.A.

Bajari, P., and A. Hortaşçsu (2004): "Economic Insights from Internet Auctions," *Journal of Economic Literature*, 42(6), 457–489.

Bolton, G., B. Greiner, and A. Ockenfels (forthcoming): "Engineering Trust: Reciprocity in the Production of Reputation Information," *Management Science*.

Cabral, L. (2012): *Reputation on the Internet*chap. 13, pp. 343–354. Oxford University Press, New York, NY, U.S.A.

Cabral, L. M., and A. Hortaşçsu (2010): "The Dynamics of Seller Reputation: Theory and Evidence from eBay," *Journal of Industrial Economics*, 58(1), 54–78.

Chiappori, P. (2000): "Econometric models of insurance under asymmetric information," in *Handbook of Insurance*, ed. by G. Dionne, Huebner International Series on Risk, Insurance, and Economic Security, pp. 365–393. Kluver, Boston, MA, U.S.A.

Chiappori, P., and B. Salanié (2000): "Testing for asymmetric information in insurance markets," *Journal of Political Economy*, 108(1), 56–78.

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.

Dranove, D., and G. Z. Jin (2010): "Quality Disclosure and Certification: Theory and Practice," *Journal of Economic Literature*, 48(4), 935–963.

Einav, L., C. Farronato, J. Levin, and N. Sundaresan (2013): "Sales Mechanisms in Online Markets: What Happened to Internet Auctions?," Mimeographed.

Einav, L., A. Finkelstein, S. Ryan, P. Schrimpf, and M. Cullen (2011): "Selection on moral hazard in health insurance," NBER Working Paper No. 16969, National Bureau of Economic Research, Cambridge, MA, U.S.A.

Fang, H., M. Keane, and D. Silverman (2006): "Sources of advantageous selection: Evidence from the Medigap insurance market," NBER Working Paper No. 12289, National Bureau of Economic Research, Cambridge, MA, U.S.A.

Feiring, D. (2009): "Best Match Made Simple," available at http://www.auctioninsights.info/bestmatchbook.

Finkelstein, A., and J. Poterba (2004): "Adverse selection in insurance markets: Policyholder evidence from the UK annuity market," *Journal of Political Economy*, 112(1), 183–208.

Greenstone, M., P. Oyer, and A. Vissing-Jorgensen (2006): "Mandated disclosure, stock returns, and the 1964 Securities Acts amendments," *Quarterly Journal of Economics*, 121(2), 399–460.

Henze, B., F. Schuett, and J. P. Sluijs (2013): "Transparency in Markets for Experience Goods: Experimental Evidence," Mimeographed.

Jin, G., and A. Kato (2008): "Price, quality, and reputation: Evidence from an online field experiment," *The RAND Journal of Economics*, 37(4), 983–1005.

Jin, G. Z., and P. Leslie (2003): "The effect of information on product quality: Evidence from restaurant hygiene grade cards," *Quarterly Journal of Economics*, 118(2), 409–451.

Klein, T., C. Lambertz, G. Spagnolo, and K. Stahl (2006): "Last minute feedback," C.E.P.R. Discussion Paper No. 5693, Centre for Economic Policy Research, London, U.K.

Klein, T. J., C. Lambertz, G. Spagnolo, and K. O. Stahl (2009): "The Actual Structure of eBay's Feedback Mechanism and Early Evidence on the Effect of Recent Changes," *International Journal of Electronic Business*, 7(3), 301–320.

Laffont, J., and D. Martimort (2002): *The Theory of Incentives*. Princeton University Press.

Lucking-Reiley, D., D. B. N. P., and D. Reeves (2007): "Pennies from eBay: The Determinants of Price in Online Auctions," *Journal of Industrial Economics*, 55(2), 223–233.

Melnik, M., and J. Alm (2002): "Does a Seller's Reputation Matter? Evidence from eBay Auctions," *Journal of Industrial Economics*, 50(3), 337–349.

Resnick, P., and R. Zeckhauser (2002): "Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system," in *The Economics of the Internet and E-Commerce*, ed. by M. Baye, vol. 11, pp. 127–157. Elsevier Science, Amsterdam, The Netherlands.

Resnick, P., R. Zeckhauser, J. Swanson, and K. Lockwood (2006): "The Value of Reputation on eBay: A Controlled Experiment," *Experimental Economics*, 9(2), 79–101.

A Further Details on the Data Collection

As described in Section 4, we drew sellers from 5 product categories. They are given in Table 5.

Recall that DSR scores are averages of DSRs received in the previous 12 months. Figure 8 shows the points in time at which we have collected data and the corresponding 12 month periods the DSR scores are calculated for.

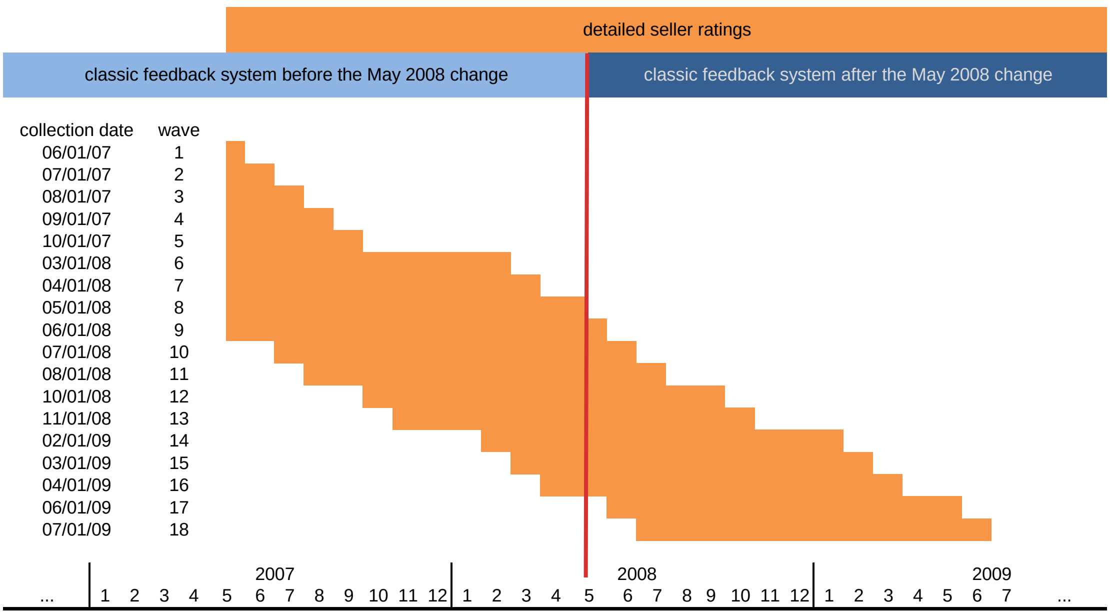
Figure 9 shows how many DSRs and classic feedbacks were received in the preceding year and over time. The number of DSRs received in the previous 12 months increases until May 2008 because DSRs were only introduced in May 2007. Since then, the number is relatively stable. The figure shows that on average (across users), more feedbacks are received than DSRs. The reason for this is that for this figure we have counted DSRs as zero when they were not displayed and DSRs are only displayed if at least 10 DSRs were left in the previous 12 months.

Table 5: Product categories

Home → All Categories → Computers & Networking → Laptops, Notebooks
Home → All Categories → Consumer Electronics → Apple iPod, MP3 Players
Home → All Categories → Toys & Hobbies → Model RR, Trains
Home → All Categories → Collectibles → Trading Cards
Home → All Categories → Home & Garden → Food & Wine

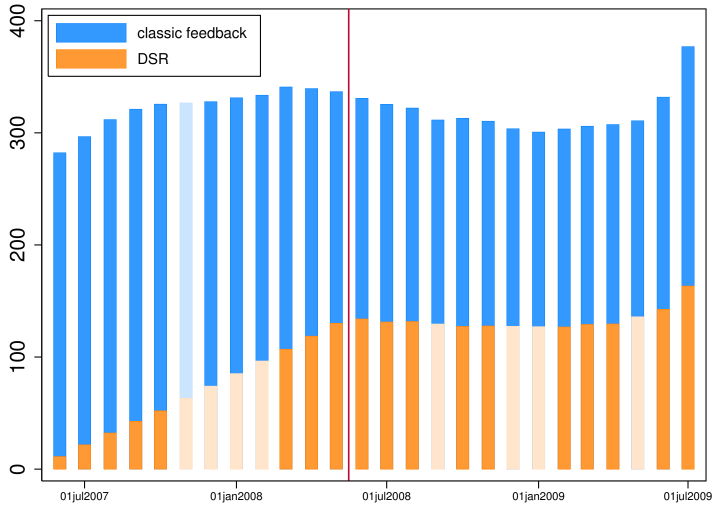
Notes: As of February 2008.

Figure 8: Data collection



Notes: The shaded area depicts the time interval DSR scores are covering. The change to the classic feedback system was implemented during the month of May 2008.

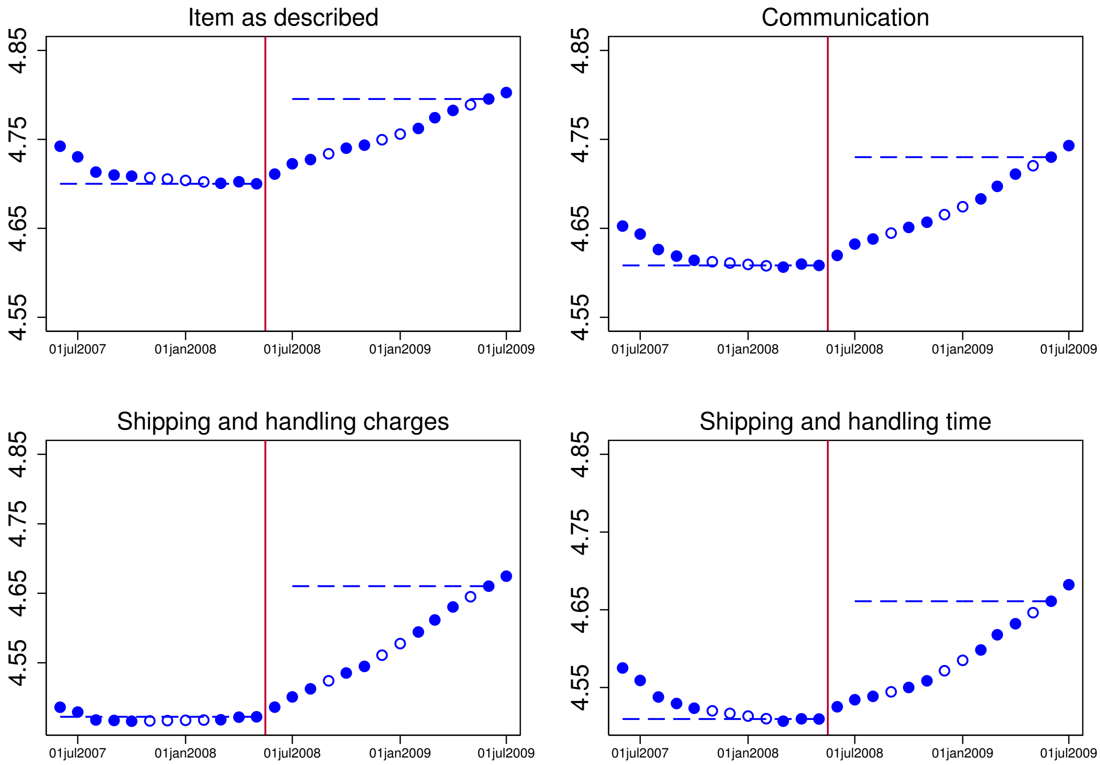
Figure 9: Number of classic feedbacks and DSRs received in the previous 12 months



Notes: The blue bars show the average number of classic feedback received in the previous year and the orange bars show the average number of DSRs received. Both averages are across sellers. The number of DSRs was counted as zero when less than 10 DSRs were received in the previous 12 months, respectively. The vertical line depicts the time of the May 2008 change to the classic feedback system. The parts of the bars that are in lighter color are not directly computed from data but come from linear interpolations of the original data, before averaging across users.

B Additional Tables and Figures

Figure 10: Average DSR score by category for poorly performing sellers



Notes: Figure 3 shows how the average of the four DSRs changed over time, for two different groups of sellers, and also reports error bars for that average. This figure shows how the different DSRs of the sellers performing below the median of 4.75 prior to the May 2008 change performed over time. As in the aforementioned figure, the dots are averages across users. Circles are linearly interpolated values for the periods in which we did not collect data. The vertical line denotes the time of the May 2008 change to the classic feedback mechanism. The horizontal dashed lines visualize that the dots are averages over the 12 months prior to the point in time at which the DSRs are displayed.

Table 6: Effect of Best Match and feedback change by decile of DSR rating

	decile of DSR score on March 1, 2008									
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
number sellers	459	443	450	478	423	451	554	347	451	450
number observations	5,115	6,091	6,177	6,888	6,234	6,700	8,180	5,122	6,554	6,058
effects on the number of DSRs received										
average number DSR before March 2008	44.44***	69.00***	54.17***	41.82***	47.64***	31.97***	34.76***	28.71***	31.53***	18.29***
	(3.48)	(3.16)	(1.50)	(3.93)	(1.41)	(1.15)	(1.00)	(0.94)	(0.85)	(0.71)
effect of Best Match	-5.76***	-0.09	-0.72	-2.02*	-0.24	0.10	0.26	0.02	0.10	0.18
	(1.50)	(1.33)	(0.67)	(1.22)	(0.48)	(0.43)	(0.45)	(0.38)	(0.36)	(0.30)
effect of feedback change	3.09**	-2.13	0.14	0.50	0.15	-0.17	-0.56	-0.11	-0.20	-0.15
	(1.26)	(1.39)	(0.74)	(0.70)	(0.46)	(0.36)	(0.38)	(0.35)	(0.33)	(0.24)
fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.04	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
effects on the DSR ratings received										
average DSR received before March 2008	4.32***	4.55***	4.63***	4.68***	4.72***	4.75***	4.78***	4.81***	4.84***	4.89***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
effect of Best Match	-0.03	-0.03	0.00	0.00	0.04***	0.02	0.00	0.00	-0.01	-0.01
	(0.05)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
effect of feedback change	0.20***	0.14***	0.09***	0.07***	0.03**	0.04***	0.05***	0.03**	0.02*	-0.01
	(0.05)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.08	0.08	0.13	0.11	0.19	0.15	0.11	0.06	0.01	0.02

Notes: The upper part of this table shows the results of regressions of the number of DSR received per month on the fraction of months in which BM and the new feedback system were in place, respectively. The lower part shows regressions of the average DSR score, averaged over the four categories, on a constant term and the fraction of feedbacks received after the introduction of BM and the May 2008 change to the feedback mechanism, respectively. We assumed that 75 per cent of the DSR in March, 2008 were received after the introduction of BM and 25 per cent of the DSR received in May 2008 were received after the change to the classic feedback mechanism. One observation is a seller-wave combination. There are 2,337 sellers. Throughout, we control for fixed effects. The R^2 is the within-R^2. Standard errors are cluster-robust at the seller level and significance at the 1 per cent level is indicated by ***.

Table 7: Effects with time delay

	(1) full sample	(2) small window	(3) time trend	(4) DSR<4.75	(5) DSR≥ 4.75
average DSR before change	4.7067***	4.7034***	4.7150***	4.5919***	4.8143***
	(0.0006)	(0.0004)	(0.0035)	(0.0010)	(0.0006)
effect of feedback change	0.0589***	0.0435***		0.0921***	0.0318***
	(0.0024)	(0.0052)		(0.0044)	(0.0021)
effect of feedback change until September 2008			0.0183**		
			(0.0081)		
effect of feedback change after September 2008			0.0652***		
			(0.0180)		
linear time trend before change			0.0009**		
			(0.0004)		
linear time trend after change			0.0001		
			(0.0019)		
fixed effects	yes	yes	yes	yes	yes
R^2	0.0583	0.0125	0.0606	0.0820	0.0459
number sellers	5,225	4,919	5,225	2,337	2,337
number observations	67,373	30,488	67,373	31,260	33,508

Notes: See notes to Table 2. The difference between the two tables is that here, we assume that only 25 per cent of the feedbacks received in May 2008 correspond to transactions that took place after the change.