Buying Reputation as a Signal of Quality: Evidence from an Online Marketplace*

Lingfang (Ivy) Li† Steven Tadelis‡ Xiaolan Zhou§

June 12, 2020

Abstract

Seller reputation, generated by buyer feedback, is critical to fostering trust in online marketplaces. Marketplaces or sellers may choose to compensate buyers for providing feedback. Signaling theory predicts that only sellers of high-quality products will reward buyers for truthful feedback, especially when a product lacks any feedback and when the seller is not established. We confirm these hypotheses using Taobao's "reward-for-feedback" mechanism. High-quality products, especially without established feedback, are chosen for feedback rewards, which cause sales to increase by 36%. Marketplaces and consumers can therefore benefit from allowing sellers to buy feedback and signal their high-quality products in the process. *JEL* Classifications: D47, D82, L15, L86.

Keywords: reputation, feedback, ratings, signaling, rebate mechanism, Taobao.

*We thank Alibaba for providing us access to the data. We are grateful to Tom Blake, Luis Cabral, Dimitriy Masterov, Mo Xiao, and Hanghui Zhang for helpful discussions and to participants at several seminars for their valuable feedback. Lingfang Li is grateful for the support from the National Natural Science Foundation of China (grants #71522007 and #71490721), Shanghai Education Development Foundation and Shanghai Municipal Education Commission (Shuguang Program, No. 17SG05), and Alibaba Huoshui Research Plan. Xiaolan Zhou is grateful for the support from the National Natural Science Foundation of China (grant #71603159). This paper was previously presented under the title "Does Money Really Talk? Evidence from Reward-for-Feedback Mechanism in Alibaba's C2C Platform." The corresponding author's email is xiaolan.z.zhou@gmail.com.

†School of Management, Fudan University. e-mail: lingfangivy.li@gmail.com

‡UC Berkeley, NBER, CEPR, and Cesifo. e-mail: stadelis@berkeley.edu

§Faculty of Economics and Management, East China Normal University, e-mail: xiaolan.z.zhou@gmail.com

1 Introduction

The tremendous growth of trade in online marketplaces such as eBay, Taobao, and Amazon Marketplaces, to name a few, is remarkable because buyers are purchasing items that they cannot inspect from anonymous and distant sellers. Reputation and feedback systems are often credited as mechanisms that foster trust and make buyers feel comfortable transacting in these large anonymous marketplaces (Dellarocas, 2003). Many studies provide evidence that buyers indeed respond to a seller's reputation in ways that confirm the role played by reputation mechanisms.[1]

However, recent evidence suggests that online reputation mechanisms suffer from several shortcomings. First, feedback is a public good that may be underprovided (Bolton et al., 2004; Chen et al., 2010; Lafky, 2014). Second, user-generated feedback is often biased, with extreme levels of "grade inflation" across several platforms (Nosko and Tadelis, 2015; Horton and Golden, 2015; Zervas et al., 2015). Third, and less explored, is the notion that established reputations of existing products may become a barrier to entry for new products, thus creating a "cold-start" problem that may stifle market expansion. Several market design questions naturally arise: How can feedback become more informative, especially for newly introduced products? Should marketplaces offer rewards for informative feedback or should they encourage sellers to do this? If sellers are allowed to pay for feedback, how will it affect market outcomes?

In this paper, we argue that established theory can shed light on these questions, and we exploit a unique dataset to test whether the theoretical implications have merit. Namely, we argue that sellers will offer to pay for feedback only if they expect feedback to be positive, implying that an offer to pay for feedback acts as a reliable signal of product quality. This simple argument echoes the ideas put forth in Nelson (1974) (and later formalized by Kihlstrom and Riordan (1984) and Milgrom and Roberts (1996)) who argued that sellers of high-quality products will pay to advertise their products, implying that advertising acts as a signal of quality. Intuitively, high-quality sellers will be willing to spend on advertising because they will benefit from repeat purchases by happy buyers, a benefit not enjoyed by low-quality sellers. We argue that, in a similar way, only high-quality sellers who list high-quality products will pay buyers to leave feedback on these products because they expect to receive positive feedback, which boosts future sales. Instead, sellers of low-quality products, or those who will deliver a poor quality service, will receive negative feedback, which

[1] For recent surveys of this literature see Cabral (2012) and Tadelis (2016).

stifles future sales. Hence, in equilibrium, a market for feedback will allow sellers of high-quality products to buy their reputation by rewarding buyers for feedback, therefore distinguishing their products early in the sales process and thereby solving the cold start problem and increasing future sales.

This separating equilibrium can be sustained only if sellers can commit to paying buyers for information regardless of the sentiment of this information. Sellers themselves, however, may not be able to commit to paying for feedback if this feedback might be negative because they would prefer to receive positive reviews. Thus, payment for feedback must be unconditional on sentiment and instead reward buyers for providing meaningful information regardless of whether the sentiment is positive or negative. Though sellers may have trouble committing to such a mechanism, a market designer—particularly for an online marketplace—would benefit from taking on this centralized role. By designing such a market for feedback, an online marketplace can help reduce informational frictions and increase sales to the benefit of both buyers and sellers in the market.

Alibaba Group's Taobao—the world's largest online marketplace—launched such a mediated market for feedback, which allows us to shed light on several questions that follow from our theoretical framework. First, do sellers signal the high quality of their products by offering to reward buyers for leaving informative feedback? Second, do buyers respond to these signals by buying more of the signaled products, and if so, what are the returns to sellers who reward buyers for feedback in terms of sales and future feedback? Third, does sellers' behavior suggest that the returns to buying feedback are highest for items that have no established feedback? Finally, are the returns to buying feedback lower for sellers who themselves have an established strong reputation?

Using proprietary data from Alibaba, we exploit Taobao's "Rebate-for-Feedback" mechanism (RFF), which allows sellers to set a rebate amount for any product they list; this rebate is awarded to buyers who leave feedback after they purchase that product. Taobao guarantees that the rebate is transferred from the seller's account to a buyer who leaves what Taobao determines to be informative feedback. Importantly, the informativeness of feedback does not depend on whether it is favorable but instead is determined by a machine learning (Natural Language Processing) algorithm that examines the content and length of the buyer's detailed feedback and whether key features of the product are mentioned.

Our panel dataset consists of all transactions that were made from 13,018 randomly selected sellers who sold at least one product between September 2012 and February 2013 on Taobao.com in at least one of four distinct categories: cellphones, memory cards, cosmetic masks, and jeans. A key feature of the data is that products are given a unique identifier (item ID) for each and every seller. In other words, an item ID in our data is a product-seller pair, which allows us to use a fixed-effects panel model that controls for seller and product attributes simultaneously and identifies variation within items to establish how the adoption of RFF impacts an item's sales and the feedback left by buyers. We use past effective feedback (defined in Section 2) as a measure of a product's quality and use the absence of feedback to classify "cold-start" products.

Our main findings confirm that RFF is a potent signaling mechanism that is strategically used by sellers. First, sellers are more likely to adopt RFF for "cold start" products and for products with a high measure of quality. Second, as sellers become more established, they are less likely to use the RFF signal, consistent with their own stronger reputation being a substitute for product-level reputation. Third, buyers respond strongly to the RFF signal—sales of an item are approximately 36% higher when the seller chooses the rebate option. Although a concern may be that buyers are responding to the price discount that results from receiving a rebate, we use some unique features of Taobao's feedback system to establish that at least 27% of the effect is due to signaling alone. Finally, RFF adoption induces buyers to write more detailed and informative feedback but does not bias buyers towards positive feedback.

Our paper offers several contributions to the literature. First, we believe that this paper is the first to empirically analyze a novel feedback-enhancing mechanism in a large online marketplace and to show that it provides significant signaling and public goods provision benefits. Second, we use our rich data to provide compelling empirical evidence on the role of signaling in online markets, showing that sellers indeed send credible signals to which buyers respond rationally. This approach complements two recent studies by Backus et al. (2019) and Kawai et al. (2013) that use rich data from online marketplaces to provide evidence consistent with signaling equilibria. Third, we expand on the signaling narrative of Nelson (1974), which was directly tested in a recent paper by Sahni and Nair (2016). Finally, we contribute to the growing market design literature with respect to managing asymmetric information in online markets. Unlike Hui et al. (2016), Nosko and Tadelis (2015) and Masterov et al. (2015), who emphasize how marketplaces can manage the asymmetric information problem by regulating seller quality, we show that marketplaces can improve market outcomes by allowing sellers to self-select using RFF signaling mechanisms.

More broadly, several papers have focused on the public goods nature of feedback and proposed two ways to incentivize buyers to leave more feedback: either online marketplaces can provide incentives to buyers to leave feedback (Miller et al., 2005; Fradkin et al., 2015) or sellers can provide these incentives (Li, 2010). Li (2010) suggests an RFF mechanism and shows that rebate mechanisms play a dual role: first by incentivizing buyers to leave feedback and, second, by providing a device for sellers to exert effort to provide a high-quality transaction. Incorporating both adverse selection and moral hazard, Li and Xiao (2014) extend this idea to listed-price online markets and show that bad sellers can exert effort to provide high-quality transactions and that if a seller chooses RFF for an item, the seller will exert effort to provide high-quality transactions. As a consequence, buyers will avoid sellers that do not choose the RFF option. Li and Xiao (2014) also test a variant of the RFF mechanism in lab experiments and find evidence consistent with the theoretical predictions in Li (2010). Cabral and Li (2015) run a series of controlled field experiments on eBay where sellers propose monetary rewards for providing (any) feedback and find that buyers grant these sellers more frequent feedback and more favorable feedback when transaction quality is high, but when transaction quality is low, offering a rebate significantly decreases the likelihood of negative feedback. The mixed results of Li and Xiao (2014) and Cabral and Li (2015) most likely emerge because the implementation of the mechanism is fundamentally different in the two studies. In Li and Xiao (2014), once a seller chooses RFF, the experimenter guarantees that the reward will be transferred, but in Cabral and Li (2015), the sellers themselves promise a reward. This scenario may cause buyers to believe that they will not be paid for negative feedback. Importantly, when Taobao implemented RFF, the platform guaranteed that the reward would be paid by the seller to the buyer based on the informativeness of the feedback content and not on whether the review was positive or negative. Our findings are consistent with those in Li and Xiao (2014), showing the importance of a commitment to reward feedback for its informativeness and implying that platforms must play an active role in effectively designing and mediating the market for feedback.

Our paper also contributes to a growing literature that empirically studies the workings of reputation systems in online markets with asymmetric information. A series of recent studies have shown that because feedback is user generated, reputation scores are often biased (Dellarocas and

Wood, 2008; Nosko and Tadelis, 2015), inflated (Horton and Golden, 2015; Zervas et al., 2015), and possibly manipulated by market players (Mayzlin et al., 2014; Luca and Zervas, 2016; Xu et al., 2015).[2] We assert that RFF mechanisms can help promote honest and informative feedback while offering the added benefit of a signaling mechanism that promotes high-quality products. Furthermore, RFF mechanisms help solve the cold-start problem, thus reducing inefficiencies in online markets with asymmetric information.

The paper proceeds as follows. In Section 2, we describe Taobao's RFF mechanism, and in Section 3, we lay out our theoretical arguments and testable hypotheses. Section 4 describes the data, while Section 5 presents our analyses and results. Section 6 concludes and discusses some implications of our analyses.

2 The Taobao Reward-for-Feedback Mechanism

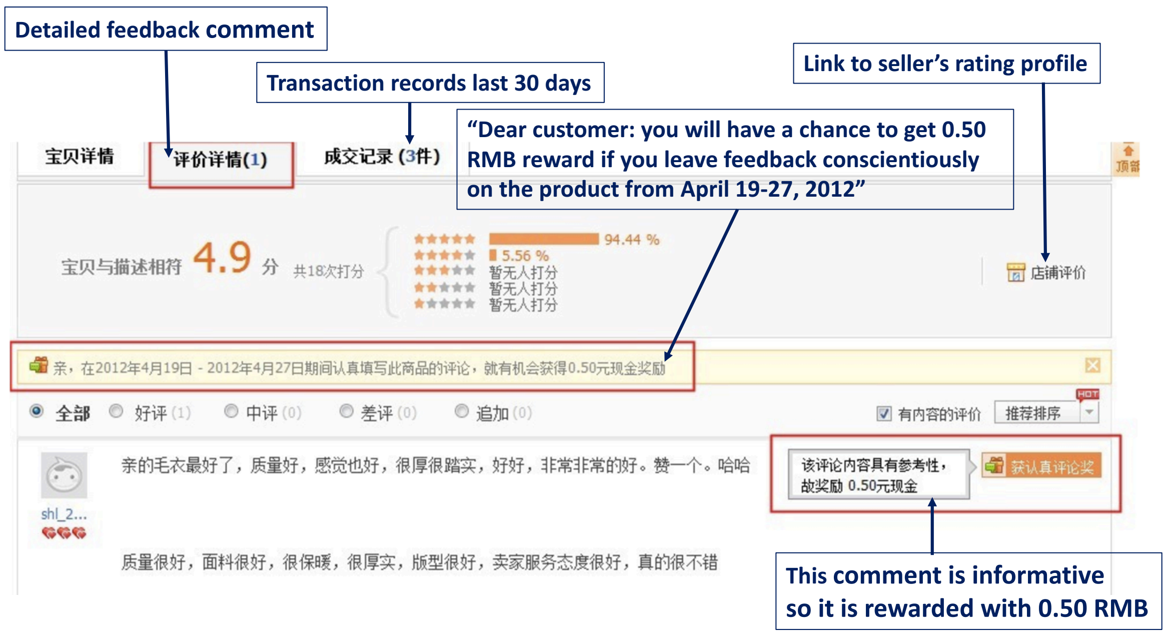
Launched in 2003, Taobao Marketplace (www.taobao.com) has become the most popular C2C online marketplace in China, with close to 500 million registered users. On an average day, more than 60 million visitors access more than 800 million product listings, and an average of 50,000 products are traded every minute.[3] Like eBay, Taobao intermediates between buyers and sellers, but unlike eBay, Taobao earns revenues from advertising and other services, not from listing or commission fees from sellers. The majority of the products on Taobao are new merchandise sold at fixed prices. Customers pay for their purchases using Alipay, which is comparable to eBay's use of PayPal. However, unlike Paypal, Alipay takes money up front, puts it in an escrow account and verifies that the customer is satisfied with the product before payment is released.

An "item" on Taobao refers to a product-seller combination, and any product sold by a seller is assigned a unique item ID. If another seller sells the same product, it is assigned a different item

[2] Some sellers begin selling cheap items to climb up the reputation ladder, as noted in Brown and Morgan (2006) and Fan et al. (2016). Proserpio and Zervas (2017) show that when hotel management responds to online reviews, they then receive fewer but longer negative reviews. Online reviews also help consumers learn, as explored in the context of restaurant reviews by Wu et al. (2015).

[3] BBC India. See https://www.facebook.com/bbcindia/posts/741334802577552 (accessed on February 16, 2017).

Figure 1: Taobao.com page with feedback reward scheme



ID.[4] This definition of an item differs from that in other marketplaces such as eBay, where an item refers to a product and not a product-seller pair.

Similar to eBay, after a transaction is completed on Taobao, buyers and sellers can leave each other positive, neutral, or negative feedback, as well as detailed comments about the transaction. However, Taobao's feedback system differs from eBay's in three important ways. First, Taobao

[4] Suppose A and B are two sellers on Taobao that sell the same two products x and y. Each product sold by each seller has an unique item ID, for example, A's product x has ID $\#A01$, and her product y has ID $\#A02$, while B's products x and y have IDs $\#B01$ and $\#B02$, respectively. If a buyer purchases product x from seller A and leaves a rating for the transaction, the rating information (positive/neutral/negative and comments) will be recorded to the "item rating" page for A's product x, ID $\#A01$, and the positive/neutral/negative point will be counted towards A's seller rating grade, which is shown as a heart, diamond, crown, or gold crown on the item's page as well as on the seller's profile page.

separately reports a user's rating score as a seller and as a buyer, whereas on eBay a user's total rating score is aggregated for sales and purchases.[5] Second, Taobao reports buyers' feedback on an item on both the item's rating page and the seller's rating page. An item's rating page refers to the "detailed feedback comment" page as shown in Figure 1. Buyers can see all ratings for the item when browsing an item's page. A seller's rating page displays ratings for all items sold by a seller on the seller's profile page. By contrast, eBay only displays seller rating profile pages and does not provide item rating pages. Third, if a buyer does not leave any feedback within 15 days after a seller leaves feedback for the buyer, then Taobao's system leaves automatic positive feedback for the seller. In the "comments" area for the transaction, it displays "Feedback provider didn't leave feedback on time; the system offered an automatic positive rating!" as a message containing 18 Chinese characters.[6] In contrast, on eBay, if no feedback is left after a transaction, future buyers will not even know that the transaction has occurred. On the one hand, Taobao's automatic feedback feature makes it possible to track all sales of an item, but on the other hand, it may bias the ratio of positive ratings to be higher than it should, thus affecting the informativeness of the feedback system.

The special features of Taobao's feedback system mean that reviews can take one of three forms: Sentiment only (positive, negative, or neutral) without any description, resulting in feedback with zero Chinese characters; sentiment together with some description, resulting in feedback with a positive number of Chinese characters; and an automatic positive review with 18 Chinese characters. We refer to the second form, sentiment with some description, as an "effective" review because the buyer clearly wrote something about the product or the experience. We refer to the other forms of reviews as ineffective, which play an important role in building our hypotheses and in executing some of our empirical analyses.

On March 1, 2012, Taobao launched a RFF feature for sellers.[7] This feature offers sellers the option of selecting items for which they set a rebate value, in the form of cash back or a store

[5] Hereafter we use the term "rating" to include both positive/neutral/negative sentiment as well as detailed comments. An item's rating therefore reflects a buyer's opinion about the merchandise and the seller.

[6] Sellers almost always leave feedback for a buyer in order to obtain an automatic positive feedback in case the buyer leaves none. Fan et al. (2016) also provide an introduction to Taobao's feedback system. Figure 2 in the online Appendix provides an example of automatic rating, zero-word rating, and effective rating.

[7] The RFF feature implemented by Taobao is similar to the mechanism proposed in Li (2010) and Li and Xiao (2014). In fact, Li suggested the RFF mechanism to Alibaba Research towards the end of 2011, and several months later, Taobao launched the RFF mechanism. See http://www.aliresearch.com/blog/article/detail/id/20486.html (accessed on June 15, 2015). In the Appendix we provide a translation of Taobao's announcement of the new online service. One of the announced goals is to "increase the ratio of non-automatic ratings for sellers." Another is to

(seller-specific) coupon, that is awarded to buyers who leave informative feedback. Sellers choose which of their items will adopt the RFF, for how long it will be offered, and the form and monetary value of rebates. If a seller chooses the RFF option for one of their items, Taobao guarantees that the rebate will be transferred from the seller's account to a buyer who leaves informative feedback. Taobao measures the informativeness of feedback using a natural language processing algorithm that examines the comment's content and length and verifies whether key features of the item are mentioned. Importantly, informativeness does not depend on the sentiment, that is, on whether the feedback is positive or negative. For the execution of payment, the seller deposits a certain amount in advance for a chosen period to adopt the RFF, and Taobao freezes the deposit until the end of the rebate period or until all the funds have been depleted by buyers who obtained a rebate. Hence, funds are guaranteed for buyers who meet the rebate criterion, and sellers cannot choose which buyers will or will not receive the rebate. Any rating that earned its buyer a rebate will be identified on the item's rating page as one for which an RFF was granted so that future buyers can know that this rating was awarded a rebate.

According to a Taobao survey (published in March 2012), 64.8% of buyers believe that they will be more willing to buy items that have the RFF feature, and 84.2% of buyers believe that the RFF option will make them more likely to write detailed comments.[8] Figure 1 shows a Taobao.com page with the RFF feature. The box just below the 4.9 score includes a feedback reminder that reads: "Dear customer, you will have a chance to obtain a 0.50 RMB reward if you leave conscientious feedback on the product from April 19–27, 2012." The box on the lower right corner, in turn, includes a notice "this comment is informative, so it was rewarded with 0.50 RMB."

3 Theory and Hypotheses

Rather than laying out a formal model of RFF signaling, we adapt the seminal theory of advertising (Nelson, 1974; Kihlstrom and Riordan, 1984; Milgrom and Roberts, 1996), which naturally translates to our setting. As a precursor to this literature, Nelson (1970) introduced the concepts of *search goods* and *experience goods*: an experience good is a product or service for which product characteristics are difficult to observe in advance, but these characteristics can be ascertained upon consumption; a

"increase the quality of buyers' comments," where Taobao's machine learning algorithm is used to judge feedback quality.

[8] http://bbs.taobao.com/catalog/thread/513886-256229600.htm, accessed June 24, 2012.

search good is a product or service with features and characteristics that are easily evaluated before purchase. Hence, adverse selection problems are typically more severe for experience goods, and some mechanism is needed to help buyers infer which products are high quality and which are not.

The advertising-as-a-signal literature established that only sellers of high-quality goods will spend money on ads to promote their experience goods because only they can be confident that they will receive positive returns from their expenditures. According to the theory, advertising–a form of "burning money"–acts as a signal that attracts buyers who, in equilibrium, correctly believe that only high-quality goods will be advertised. Incentive compatibility is achieved through repeat purchases: buyers who are attracted to sellers that advertise will buy the good, experience it and will return in the future only if the good sold is of a high enough quality. Advertising has to be costly enough to deter sellers of low-quality items from being willing to spend the money and then sell only once to each customer because they will not attract repeat purchases. Hence, ads act as signals that separate sellers of high-quality products from low quality ones and, in turn, convince buyers to purchase them.

We now argue that the RFF mechanism plays a similar signaling role as ads do. A customer's satisfaction with an online transaction is determined both by the quality of the product itself, as well as the service that the seller provides (adequate wrapping, shipping, correct description, etc.) Hence, a satisfactory experience must be a consequence of both a high-quality product and a high quality of the seller's service. Assuming that consumers express their experiences truthfully in written feedback, any consumer who buys a product and is given incentives to leave feedback will leave positive feedback only if the buying experience was satisfactory. This scenario naturally implies that a seller will offer RFF incentives to buyers only if the seller expects to receive positive feedback, which occurs only when the seller provides a high-quality product together with a high quality of service. If a seller knows that its goods or services are unsatisfactory, then paying for feedback will generate a negative review that will harm the seller's future sales.[9] It therefore follows from equilibrium behavior that RFF acts as a credible signal of high quality, separating good experiences from bad ones, which in turn will attract more buyers and result in more sales. Note that the monetary value of the RFF is not what supports the separation of "good" versus "bad" types, but

[9] Li (2010) and Li and Xiao (2014) use moral hazard to demonstrate that RFF can act as a commitment to exert effort and provide a high quality transaction. The RFF acts like a bond; if the seller exerts effort, a positive review will follow and will encourage future purchases, while if she does not then a negative review will follow, which in turn will suppress future sales.

instead it is the quality of feedback that a buyer will leave. These arguments generate our main signaling hypotheses,

S1: *RFF is more likely to be adopted for high- rather than low-quality items. (Signaling Hypothesis)*

B1: *RFF generates more item sales. (Buyer Belief Hypothesis)*

There are two alternative narratives that would result in RFF increasing sales. First, a buyer who leaves informative feedback and receives a rebate effectively receives a discount on the item's price. If it is costless to leave feedback, then higher sales can simply be a consequence of the discount effect of a rebate and not a consequence of signaling. If it is costless to leave informative feedback, then all buyers should leave informative feedback and receive the discount. This pattern, however, is not borne out in the data – many buyers leave feedback with no description or even leave no feedback for RFF items, suggesting that leaving feedback imposes some costs. A second alternative explanation consistent with the notion that leaving feedback is costly is that RFF can be a price-discriminating mechanism. Buyers with a higher opportunity cost of time pay the full price, and those with a lower opportunity cost of time leave feedback and receive a discount.

To explore this idea further, imagine that buyers have some idiosyncratic cost of leaving feedback that is distributed over an interval $c \in [0, \bar{c}]$, and assume that the RFF value offered by sellers is set at r, where $0 < r < \bar{c}$. If RFF contains no signaling information then there should be an increase in demand from people with low costs of leaving feedback, $c < r$, and no change in the demand of those with cost $c > r$. As a consequence, there should be more effective feedback left by the increase in the number of consumers with $c < r$, no additional sales to those who choose to not leave reviews and, in turn, no increase in the number of *ineffective* ratings (either 0 or 18 Chinese characters as explained in Section 2.)

If, instead, RFF *does* contain signaling value, then buyers should infer this value, regardless of the costs of leaving feedback. Hence, only those consumers with a low cost of leaving feedback, $c < r$, will leave effective feedback, and the increased sales of these consumers is driven both by the signaling effect and the discount effect. In contrast, those with costs above $c > r$ will not leave effective feedback but will be motivated to buy more because of the signaling effect, a behavior not induced by the price discrimination narrative. As a consequence, the signaling narrative implies that buyers will leave more ineffective ratings for RFF items:

B2: *RFF items receive more ineffective ratings (Ineffective Rating Hypothesis)*

The above predictions are generated by a static (one-time) signal with dynamically generated incentive compatibility (the effect on future sales). Notice, however, that because RFF generates more feedback by design, it has dynamic consequences for the item's (and seller's) reputation. In other words, RFF creates a virtuous cycle: the increased number of sales and effective positive feedback will attract more future buyers. This ties in with the theoretical literature on seller reputation, which shows that building a reputation is more valuable at the beginning of a seller's career (Bar-Isaac and Tadelis, 2008). Similarly, the advertising literature suggests that a firm will "burn money" to promote brand awareness in its early stages (Milgrom and Roberts, 1996; Bagwell, 2007), after which its reputation will be established. Therefore, if a seller has a high-quality product, then the incentive to choose RFF as a signal to attract buyers is strongest for a new product that has not yet received ratings, and once the product starts developing a good reputation through past ratings, then RFF is less necessary to boost sales. That is,

S2: *RFF is more likely to be adopted when an item has no ratings. (Reputation Building Hypothesis)*

There is a question of how RFF adoption may change feedback. The machine learning algorithm rewards buyers for what is deemed to be informative feedback, which is highly correlated with length. Hence, any buyer motivated to obtain a rebate must leave longer feedback. We therefore have

B3: *RFF induces longer feedback relative to non-RFF items. (Long Ratings Hypothesis).*

Last, there is the question of whether RFF adoption may bias feedback. On the one hand, RFF is designed to offer rebates based on the informativeness of feedback rather than its positive or negative sentiment. Hence, there is arguably no reason why feedback induced by RFF should be biased compared to items without RFF. A common belief, however, is that buyers with extreme opinions are more likely to leave feedback than those who have moderate opinions. It is possible that RFF will change how extreme or moderate opinions will impact feedback propensity, in which case, we will observe some bias in feedback for RFF items. Hence, this possibility remains a theoretically open question that we will explore in our empirical analysis.

4 Data Description

Our data consist of all transactions sold by 13,018 randomly selected sellers who sold at least one unit in four chosen categories between September 2012 and February 2013.[10]. There are 114,090 items in the four categories with positive sales in the sample period; all of these items are new merchandise offered at fixed prices. Among the 13,018 sellers, 60.82% used a rebate at least once during the six months for some product (not necessarily in the four chosen categories). The categories we chose represent both search goods and experience goods in low and high price ranges. Cellphones and memory cards are search goods because their quality is generally known prior to purchase (as long as the products are authentic). The service quality, such as the shipping speed and return policy, may vary across sellers. Cosmetic masks and jeans are experience goods because their true quality can be evaluated only by actually trying them after they are shipped.[11]

Because an *item* on Taobao refers to a product-seller combination, the same product sold by different sellers will be assigned different item IDs. A *transaction* is defined as the sale of an item (one or multiple units). For each transaction, our data contain the transaction ID, item ID, category ID to which the item belongs, buyer ID, seller ID, quantity sold, total transaction price (including shipping fees), a time stamp (e.g., 2013-01-08 23:15:49), and the corresponding rating information if it has been rated. The *rating information* includes whether the rating is positive, neutral, or negative, its length (in Chinese characters), a time stamp, and whether the rating is a first- or second-time rating.[12] Table 1 provides summary statistics for item attributes at the item-month level.

Following Section 2, we define an "effective rating" as one that includes information left by a buyer, which excludes all ratings with zero Chinese characters or positive ratings with exactly 18 Chinese characters. Because we do not have the actual text content, by excluding positive ratings

[10]Unfortunately, we do not have access to Taobao data pre-RFF, which prevents us from exploring a difference-in-differences analysis to uncover causal impacts; therefore, we resort to other methods as described in section 5.

[11]Nelson (1970) uses clothes as an example of search goods because buyers can try on clothes in the shop before purchasing, while he considers a TV as an experience good because a buyer cannot know how well a TV works. Almost 50 years later, in online markets, there is less information asymmetry about TVs than about clothes. Most cellphones and memory cards are well-known branded products with many quantifiable product characteristics on which buyers can read online reviews. In contrast, there are many more small and generic brands of jeans in China—in our dataset, the revenue share of the top 10 jeans brands is only 1.97%, and most jeans and facial masks sold on Taobao cannot be found or tried in offline stores.

[12]A buyer is allowed to leave feedback more than once because for some products quality cannot be determined until it is used for some time.

Table 1: **Summary Statistics (item-month)**

	Obs.	Mean	Std.	25%	Median	75%
Rebate	284,263	0.181	0.385	0	0	0
Coupon rebate	284,263	0.136	0.343	0	0	0
Cash rebate	284,263	0.0454	0.208	0	0	0
Amount of cash rebate (RMB) conditional on a cash rebate	6,373	1.409	2.235	0.500	1	1
- Cellphone	1,216	2.809	4.479	1	1.100	3
- TF card	42	2.008	2.114	0.500	1	3
- Mask	2,188	0.851	0.798	0.500	0.500	1
- Jeans	2,927	1.236	0.962	0.500	1	1.047
Average item transaction price (RMB)	284,263	172.8	892.4	29	68.09	117.6
- Cellphone	30,068	956.4	2216	223.8	505.3	1138
- TF card	7,263	134.2	1842	28	40.77	80
- Mask	111,515	49.10	501.6	5.610	17.50	56.02
- Jeans	135,417	102.8	196.2	60.36	80	114.5
Ratio of cash rebate to average item price	6,373	0.0420	0.136	0.00638	0.0117	0.0269
- Cellphone	1,216	0.00627	0.0106	0.00130	0.00299	0.00674
- TF card	42	0.0206	0.0128	0.0121	0.0158	0.0269
- Mask	2,188	0.0967	0.222	0.0107	0.0251	0.0875
- Jeans	2,927	0.0163	0.0176	0.00749	0.0112	0.0199
Never received ratings (cold start)	234,925	0.319	0.466	0	0	1
Ratio of item pstv-effective ratings (cumulative)	159,406	0.459	0.303	0.25	0.444	0.647
Ratio of item neg/neu-effective ratings (cumulative)	159,406	0.00778	0.0463	0	0	0
Monthly item sales (monthly)	284,263	39.29	692.7	1	3	11
Number of item ineffective ratings (monthly)	284,263	12.09	307.9	0	1	4
Number of item ratings (monthly)	284,263	21.96	545.1	1	2	6
Ratio of item effective ratings (monthly)	240,308	0.477	0.374	0.111	0.475	0.889
Ratio of item long ratings (monthly)	240,308	0.178	0.287	0	0	0.250
Ratio of item pstv ratings (monthly)	240,308	0.990	0.0705	1	1	1
Average number of days before leaving ratings (monthly)	238,452	10.01	6.332	5	8.625	14
Diamond grade	284,263	0.425	0.494	0	0	1
Crown grade	284,263	0.390	0.488	0	0	1
Gold crown grade	284,263	0.0441	0.205	0	0	0
Ratio of seller positive ratings (cumulative)	283,922	0.991	0.0137	0.987	0.995	0.999

Notes: The observations are at the item-month level. An item in a month is excluded if item sales in that month were zero. Effective ratings exclude those with zero or 18 Chinese characters. Long ratings have at least 24 Chinese characters. Cumulative ratings refer to ratings from the the beginning of the sample to the end of month $t-1$.

with 18 characters, we are discarding some ratings that are genuine together with the automated ratings. Figure 2b shows that less than one percent of ratings have 17 or 19 characters, implying that authentic consumer comments that contain 18 characters are likely to be less than one percent as well. We define a “long rating” to be a rating with at least 24 Chinese characters, which is the 75th percentile of all the ratings excluding the 18-character ratings.

For each seller, the data contain information about their location (province), service promises (e.g., whether they accept any returns within seven days, etc.) and daily reputation, including the seller’s rating score, the seller’s rating grade, and the ratio of positive ratings. As described in the Appendix, there are a total of 21 seller grades, where a higher grade is identified by a higher interval of rating scores, calculated as the number of positive minus negative ratings. Grades are highly correlated with the number of sales because of Taobao’s automatic positive ratings and the low number of negative and neutral ratings overall. Hence, a seller’s grade is a useful measure of how experienced the seller is on Taobao. We bunch sellers into four broad grades, which are “gold crown” (the highest set of grades), “crown” (the next highest set), “diamond” (the third highest set), and “heart” and “no-grade” for all others (the lowest set).

We define a *period* as one month because information on Taobao is displayed for the previous 30 days. For example, when a potential buyer searches for a product and sorts the search outcomes by items’ sales, the results are ranked based on sales in the previous 30 days. A seller’s ratio of positive ratings is reported for the last month and last six months on Taobao. Hence, for each item, we summarize its characteristics as the monthly sales, monthly average price, and monthly number of positive, neutral, negative, and effective ratings. For each seller, we use the seller’s rating grade at the beginning of each month and the ratio of the seller’s positive ratings at the monthly level. As a robustness test in the online Appendix, we used other time frames for period duration, and the main findings continue to hold.

There are two types of rebates a seller can choose: cash or coupon. Approximately 26.33% of rebates are cash, with an average value of RMB 1.41. Unfortunately, we have the monetary value of only cash rebates, while for coupon rebates, we only know if they were adopted or not. Coupons can only be used against a future sale from the same seller, whereas a cash rebate is paid out regardless of a buyer’s future purchases.[13] The data contain a rebate’s starting date but not its ending date. We therefore define a rebate-month dummy that takes the value of 1 if either a seller initiated at least one rebate for the item in that month or if a buyer receives a rebate from a transaction that

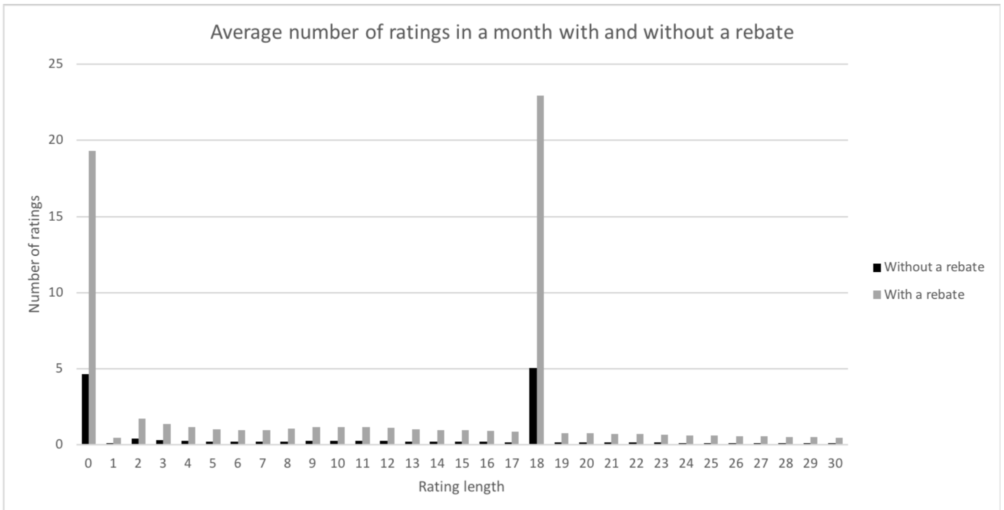
[13]In the dataset, the values of cash rebates are reported only if they are granted to at least one buyer. The average value of cash rebates for cellphones, memory cards, masks, and jeans are RMB 2.81, RMB 2.01, RMB 0.85, and RMB 1.24 respectively. The average ratio of cash rebates to transaction prices for the four categories—cellphones, memory cards, masks, and jeans—are 0.00627, 0.0206, 0.0967, and 0.0163, respectively.

includes the item in that month. Since some sellers changed an item's rebate form (cash or coupon) within a month, we define the rebate form as a cash rebate if cash was chosen at least 50% of the time in the month; otherwise, it is defined as a coupon rebate.

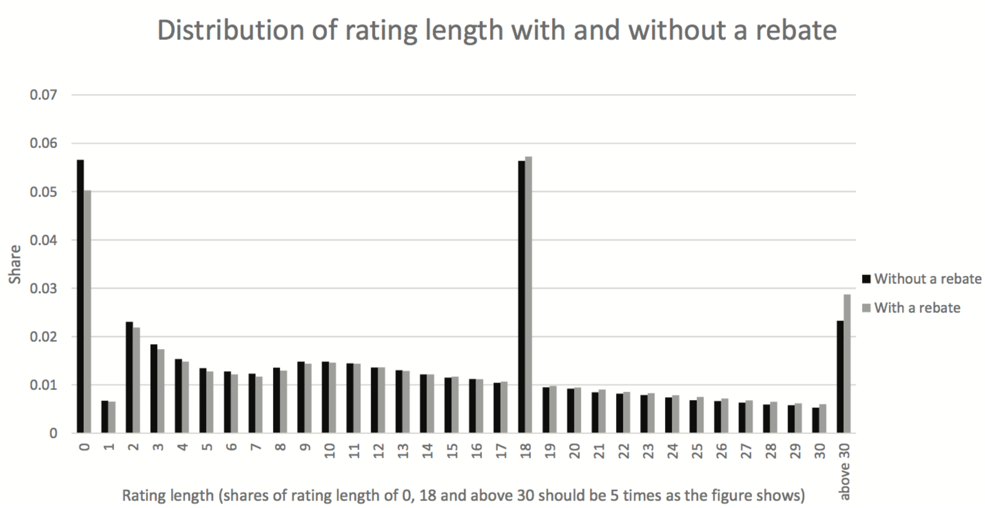
5 Empirical analysis

Before turning to the regression analyses, it is illustrative to look at the raw data that describe how buyers respond to items for which RFF were sometimes available and sometimes not available during our sample period. First, as Figure 2a shows, there are more ratings of any character length with rebates than without rebates. Second, as Figure 2b shows, the distribution of the length of ratings in Chinese characters with and without a rebate differs as well. The fraction of ratings with less than 14 characters is larger for items without rebates compared to items with rebates, while the opposite is true for ratings with at least 15 characters. In fact, the distribution of the number of characters for items with rebates first-order stochastically dominates that of items without rebates, which suggests that rebates play a role in motivating people on the extensive margin to write more ratings overall and those on the intensive margin to write longer comments.

Figure 2: (a) Distribution of Rating Length with and without Rebates. (b) Number of Ratings of Each Length with and without Rebates.



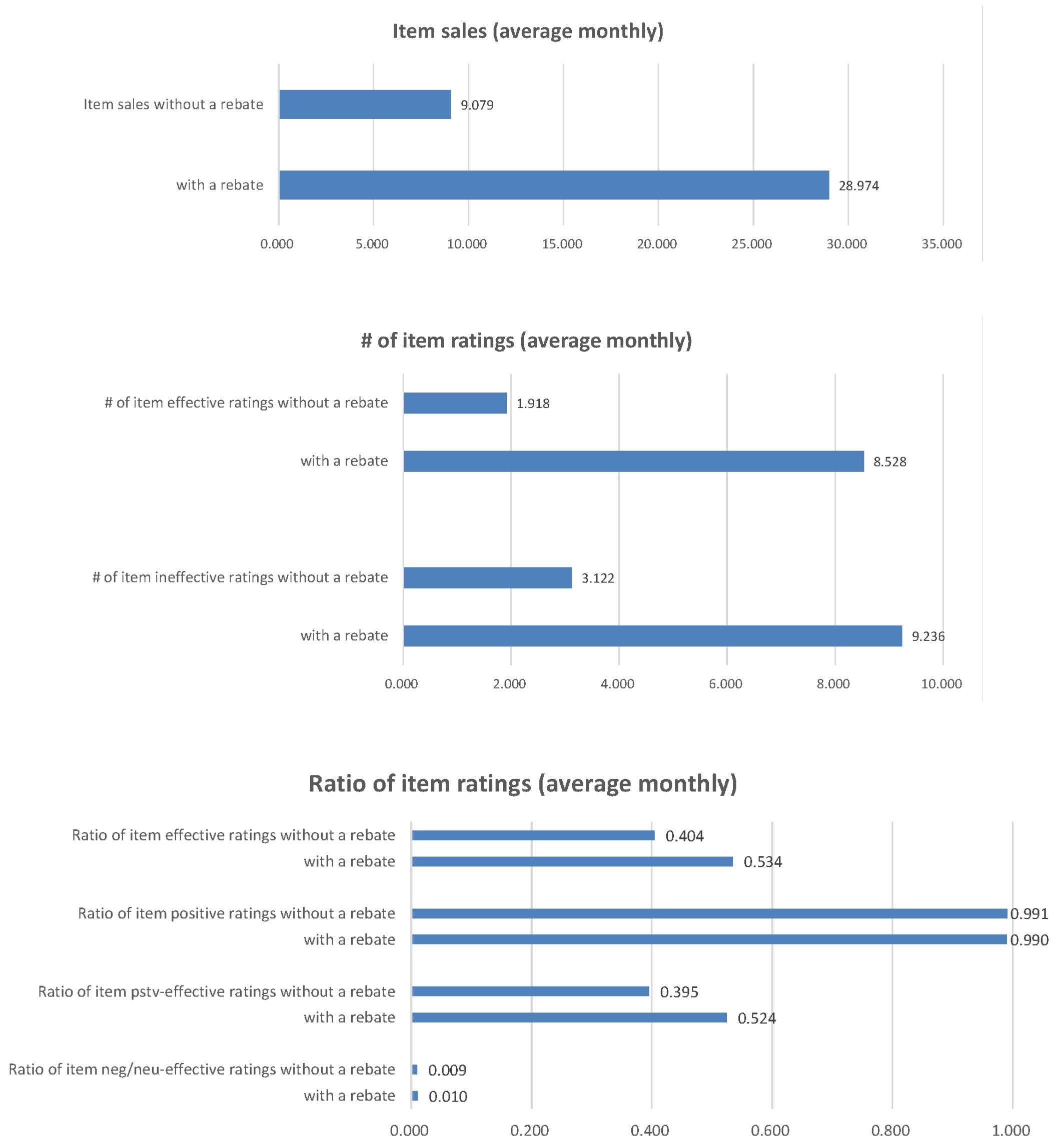
(a)



(b)

Figure 3 shows the relationship between rebates and sales and that between rebates and ratings. The average monthly sales of an item with a rebate is much higher than that without a rebate, which is indicative of the buyer belief hypothesis. The same applies to the average number of monthly effective and long ratings. Interestingly, there are more ineffective ratings, indicative of the ineffective ratings hypothesis, which confirms the signaling content of RFF. We now turn to a more careful analysis that teases out all of the hypotheses.

Figure 3: Sales and Ratings with and without a Rebate



5.1 Rebate Adoption

In this subsection, we empirically examine the hypotheses related to sellers' strategic behavior, namely, that sellers use RFF to signal high-quality items (S1) and that they use it more when for items with no feedback (S2). We estimate the following panel regression:

$$y_{i,s,t} = \alpha + \pi \cdot Item_{i,s,t-1} + \gamma \cdot Seller_{s,t-1} + \mu_t + \delta_s + \varepsilon_{i,s,t}, \tag{1}$$

where $y_{i,s,t} \in \{0, 1\}$ is an indicator equal to 1 if item i of seller s in period t offers a rebate; $Item_{i,s,t-1}$ is a vector of item characteristics in period $t-1$, the month before the adoption choice, including the following variables: the logarithm of item sales (plus 1); an indicator equal to 1 if the item has not received any ratings (cold start); the ratio of positive effective ratings for the item (conditional on the cumulative number of item ratings being greater than 0); and the ratio of negative and neutral effective ratings for the item. Item sales takes the value in month $t-1$ rather than the accumulated value because for sales, buyers only observe those in the previous 30 days. All other variables for ratings take the cumulative value from the beginning of period 1 to the end of period $t-1$ because only the lagged values are observed by buyers. Monthly fixed effects μ_t are included to control for time trends, and we run specifications with and without seller fixed effects δ_s, as described below.

The explanatory variables we choose are those that can be observed by buyers at the end of period $t-1$ and that influence a seller's strategic choice of whether or not to offer RFF for an item. Note that we do not include price in the analysis because price is determined during period t and is endogenously determined by the seller. Indicators for a seller's grade, which is also observed by buyers, control for seller experience.

Recall that we consider quality to be the buyer's satisfaction with the transaction, which is determined both by the item's quality as well as the quality of the seller's service. It is possible that a seller may choose to put in more effort in the delivery of items for which the seller chooses to use RFF. Because we take the view that the item's inherent quality is a key input into the choice of using RFF as a signal, we use the ratio of an item's positive effective ratings up to and including period $t-1$ as a measure of quality rather than this ratio for period t. The indicator of an item never receiving feedback before period t identifies items for which the cold-start problem exists and for which the signaling incentives are high.

Table 2 shows the results of the regression in equation (1). The basic specification in column (1) is a logit model that identifies the decision to adopt RFF across items and across sellers, and column (1') presents the marginal effects. Columns (2)-(3) show the results from a linear probability model that we use to include seller fixed effects while avoiding the incidental parameters problem. All specifications show that RFF is more likely to be chosen for an item if it has high quality, measured by the cumulative ratio of an item's positive effective ratings (the row "Ratio of item pstv-effective ratings"). Namely, sellers seem to choose their *best* items to further guarantee a higher likelihood of

Table 2: Adoption of a rebate for an item

Dependent variable	Indicator = 1 if a rebate is adopted for an item in t			
	Logit	Logit Marginal effects	Linear	Linear
	(1)	(1')	(2)	(3)
Seller characteristics in $t-1$				
Diamond grade	0.3247***	0.0424***	0.0369***	-0.0641***
	(0.0208)	(0.0027)	(0.0025)	(0.0046)
Crown grade	0.4597***	0.0600***	0.0560***	-0.0958***
	(0.0214)	(0.0028)	(0.0027)	(0.0074)
Gold crown grade	0.3917***	0.0511***	0.0463***	-0.2831***
	(0.0328)	(0.0043)	(0.0044)	(0.0170)
Ratio of seller positive ratings	5.5000***	0.7182***	0.4343***	0.2255**
	(0.5129)	(0.0669)	(0.0490)	(0.0969)
Item characteristics in $t-1$				
ln(sales + 1)	0.3488***	0.0456***	0.0531***	0.0425***
	(0.0041)	(0.0005)	(0.0006)	(0.0005)
Never received ratings (cold start)	0.5363***	0.0700***	0.0668***	0.0482***
	(0.0207)	(0.0027)	(0.0026)	(0.0023)
Ratio of item pstv-effective ratings	0.4941***	0.0645***	0.0617***	0.0184***
	(0.0245)	(0.0032)	(0.0031)	(0.0028)
Ratio of item neg/neu-effective ratings	-0.3286*	-0.0429*	-0.0414**	-0.0375**
	(0.1806)	(0.0236)	(0.0202)	(0.0172)
Item category				
Cellphone	-0.6022***	-0.0786***	-0.0800***	-0.0708
	(0.0212)	(0.0028)	(0.0026)	(0.0531)
Memory card	-1.4616***	-0.1909***	-0.1348***	-0.1721***
	(0.0592)	(0.0076)	(0.0049)	(0.0536)
Mask	-0.5370***	-0.0701***	-0.0720***	-0.0771***
	(0.0133)	(0.0017)	(0.0017)	(0.0280)
Constant	-8.7366***		-0.3939***	-0.0977
	(0.5102)		(0.0486)	(0.0976)
Seller fixed effect	No	No	No	Yes
Item fixed effect	No	No	No	No
Month fixed effect	Yes	Yes	Yes	Yes
Observations (item-month)	230,365	230,365	230,365	230,365
R2			0.091	0.086

Notes: Regressions are at the item-month level and standard errors are in parentheses. The dependent variable is an indicator of whether RFF was adopted for an item in month t. Seller and item characteristics are measured in month $t-1$, where ratios of ratings are calculated for cumulative values of ratings up until month $t-1$ inclusive. Asterisks indicate significance at 10% (*), 5% (**) and 1% (***).

making their customers happy so that she leaves a positive review. Column (1') shows that the marginal effects of the logit model imply that when the ratio of an item's positive effective ratings is 10% higher, the average probability of choosing RFF increases by 0.65% (approximately 0.18% with seller fixed effects), consistent with the Seller Signaling Hypothesis (S1). Note also that in all specifications, RFF is more likely to be chosen if the seller has a higher ratio of positive ratings, which lends support to our approach that a transaction's quality, and the implied likelihood of a buyer having a satisfactory experience, is determined by both the quality of the item and the seller's service.

Turning to the Reputation Building Hypothesis (S2), all specifications show that an item is more likely to be chosen for RFF if it had not yet received any ratings. In periods after an item receives its first rating, the probability of the seller adopting RFF for the item decreases by approximately 7% (approximately 5% with seller fixed effects). Interestingly, column (3) shows that when seller fixed effects are included, then as sellers move up the grade segments (from heart to diamond, to crown, and eventually to golden crown), they are incrementally less likely to use RFF, which suggests that a seller who obtains a higher grade is more confident that buyers will purchase any item from them, consistent with reputation building at the seller level. Nevertheless, within each grade bucket, sellers are more likely to use RFF if they have a higher ratio of positive ratings as discussed above.[14]

Turning to item categories, columns (1') and (2) show that the default category of jeans is approximately 7-8% more likely to be chosen for RFF than masks or cellphones, and approximately 13-19% more likely to be chosen than memory cards. If our judgment of search versus experience goods is convincing, then this suggests that a seller is more likely to choose *experience goods* and *expensive goods* to participate in RFF, further supporting the signaling hypothesis advocated by Nelson (1974).

5.2 The Impact of Rebates on Sales and Ratings

We now empirically examine the hypotheses related to how buyers respond to items with RFF adoption along three dimensions. First, they will be more likely to buy them (B1); second, these

[14] Note that without seller fixed effects, sellers who achieved any grade above "heart" are approximately 4-6% more likely to use RFF. This finding seems consistent with the fact that once sellers gain some experience on the site, they are more savvy about using different tools to achieve better outcomes.

purchases are driven by the signaling value and not just the effective price discount (B2); and third, that ratings will be longer (B3). We therefore estimate the following panel regression model:

$$y_{i,s,t} = \alpha + \beta \cdot Rebate_{i,s,t} + \pi \cdot Item_{i,s,t-1} + \gamma \cdot Seller_{s,t-1} + \delta_i + \mu_t + \varepsilon_{i,s,t}, \tag{2}$$

where $y_{i,s,t}$ is the dependent variable of interest (sales or ratings) as shown in in Tables 3 through 6. $Rebate_{i,s,t}$ is a vector that indicates the rebate status of item i sold by seller s in period t, which includes indicators for any rebate and if the rebate is in the form of cash. As before, $Item_{i,s,t-1}$ includes item characteristics, $Seller_{s,t-1}$ includes seller characteristics, and δ_i and μ_t are item and time fixed effects. The variables for item characteristics and seller characteristics in equation (2) include all the corresponding variables in equation (1). In addition, we also include the item price in period t, which is observed by buyers and will affect item sales and item ratings. We use the logarithm value of item sales, the number of item ratings, and the item price. Since price is observed only when a transaction occurs or item sales are greater than zero, the observations with zero sales are dropped out automatically.

We include item fixed effects, which control for item and seller characteristics. The key is that during the six-month period in our data, sellers may vary the adoption of a rebate for any given item, and we want to identify how buyers respond to RFF as it varies for an item. That is, changes in buyer behavior as a response to RFF are identified from variation *within* sellers and *within* products by the definition of the item ID. This is a more refined variant of the "matched listing" approach first used by Elfenbein et al. (2012), who study how sellers on eBay use charity as a substitute for reputation.[15] Hence, if product or seller characteristics drive some results, then our use of item fixed effects within our panel structure should alleviate any such concerns and absorb any product or seller heterogeneity. The coefficient on rebate is therefore identified based on the assumption that the error terms that affect sales within each item are independent of rebate adoption.

It is apparent that we do not have a randomly selected set of products for which RFF is adopted. In fact, the premise is that the choice to adopt RFF is strategically endogenous to product quality and seller reputation. As an attempt to estimate the impact of RFF adoption on sales with an eye towards robustness, we perform a propensity score matching exercise that treats the adoption of a

[15] Unlike on Taobao where an item sold by the same seller has a unique item ID, on eBay an item sold by the same seller several times will be recorded as several "listings". Elfenbein et al. (2012) identify a "matched listing" as a situation where a seller posted multiple items with the same title, subtitle, and starting price that differ in other listing attributes such as committing a fraction of the sale to charity. Einav et al. (2014) rely on variation within matched listings to investigate various sale strategies on eBay.

rebate as a treatment and creates a plausible control set of items for which RFF was not adopted. The results are reported in the online appendix and are consistent with the results described in the following subsections.

5.2.1 The Effect of Rebates on Sales

Table 3 uses $\ln(\text{sales} + 1)$ in period t as the dependent variable in equation (2). Recall that Table 2 relies on sellers offering RFF for some products but not others, whereas Table 3 relies on sellers switching within products. Before turning to our regression results, it is useful to describe some of the magnitudes of this variation. A total of 32,762 sellers offer some but not all products with RFF, and the average share of products offered is approximately 29%, with an interquartile range of 6.7% to 44.4%. Within an item, 31,380 sellers switch from no RFF to RFF, while 37,974 sellers switch in the other direction.

As Table 3 shows, the estimated coefficient on the rebate dummy is large and significant, showing that a rebate increases the quantity sold of an item by approximately 36% on average.[16] This finding supports the *Buyer Belief Hypothesis* (**B1**), which is consistent with the equilibrium behavior of signaling. To put this effect in context, Table 1 shows that the median seller-item is approximately 3 items sold per month; therefore, using RFF will result in approximately one more unit sold. We also estimated the effects of different types of rebates. Column (3) shows that, on average, a coupon rebate increases the quantity sold of an item by 39%, while a cash rebate increases the quantity sold by 28%. We conjecture that this result is probably because a coupon usually has a higher value compared with the average value of a cash rebate.[17]

The other estimated coefficients on item and seller characteristics are as expected and do not vary much across specifications. A low item price in month t, a high number of item sales in month $t-1$, a non-zero cumulative number of item ratings, a high cumulative ratio of positive item ratings, a low cumulative ratio of negative item and neutral long ratings, and a high cumulative ratio of positive seller ratings all attract more sales.[18] In column (4), we follow a robustness exercise similar to that of Elfenbein et al. (2012) and exclude items for which RFF was either never or always

[16] The coefficient is 0.3077 and $exp(0.3077) - 1 = 0.36$. Because several of the coefficient estimates in our log-linear models are substantially distant from zero, we use this conversion for most of the verbal description of the results.

[17] Sellers are probably more generous with coupons because they can only be used against a future sale from the same seller, whereas a cash rebate is paid out regardless of a buyer's future purchases.

[18] It is curious that in all specifications we find that crown grade sellers are associated with fewer sales.

Table 3: Impact of rebate on sales of an item

Dependent variable:	ln(item sales + 1) in month t			
	(1)	(2)	(3)	(4)
Rebate in month t	0.3077***			0.3129***
	(0.0067)			(0.0073)
- Coupon rebate		0.3312***		
		(0.0077)		
- Cash rebate		0.2447***		
		(0.0121)		
- Cellphone rebate			0.3468***	
			(0.0226)	
- Memory card rebate			0.1277**	
			(0.0637)	
- Mask rebate			0.1769***	
			(0.0105)	
- Jeans rebate			0.3994***	
			(0.0090)	
Item characteristics				
ln(price, t)	-0.3811***	-0.3813***	-0.3821***	-0.4695***
	(0.0094)	(0.0094)	(0.0094)	(0.0162)
ln(sales +1)	0.0628***	0.0626***	0.0617***	0.0833***
	(0.0022)	(0.0022)	(0.0022)	(0.0035)
Never received ratings (cold start)	-0.0202**	-0.0202**	-0.0237**	-0.0343**
	(0.0093)	(0.0093)	(0.0093)	(0.0163)
Ratio of item pstv-effective ratings	0.1386***	0.1385***	0.1377***	0.2205***
	(0.0112)	(0.0112)	(0.0112)	(0.0195)
Ratio of item neg/neu-effective ratings	-0.3926***	-0.3927***	-0.3757***	-0.3099**
	(0.0688)	(0.0688)	(0.0687)	(0.1203)
Seller characteristics in $t-1$				
Diamond grade	-0.0016	-0.0037	0.0006	-0.0314
	(0.0137)	(0.0137)	(0.0137)	(0.0214)
Crown grade	-0.0596***	-0.0632***	-0.0541***	-0.0757**
	(0.0210)	(0.0210)	(0.0210)	(0.0329)
Gold crown grade	-0.0150	-0.0115	-0.0075	0.0296
	(0.0448)	(0.0448)	(0.0447)	(0.0624)
Ratio of seller positive ratings	2.3066***	2.3117***	2.2490***	5.0145***
	(0.3091)	(0.3091)	(0.3088)	(0.6265)
Constant	0.6362**	0.6342**	0.6956**	-1.1580*
	(0.3099)	(0.3099)	(0.3096)	(0.6274)
Item fixed effect	Yes	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes	Yes
Observations	230,365	230,365	230,365	83,667
R2	0.099	0.099	0.101	0.152

Notes: Regressions are at the item-month level and standard errors are in parentheses. The dependent variable is an items sales (plus 1) in month t. Ratios of ratings are calculated for cumulative values of ratings up until month $t-1$ inclusive. Column (4) excludes items that either never or always participated in RFF. Asterisks indicate significance at 10% (*), 5% (**) and 1% (***).

adopted during our six month panel. As expected, because the item fixed effects identify variation within items, the results are very similar to those in Column (1).

We further consider the heterogeneous effect of rebates across product categories. Column (3) in Table 3 reports estimates from dummies for each product category. We find that the rebate has the largest effect on cellphones and jeans, and the lowest on memory cards. This finding is consistent with the narrative that signaling is more important when the buyer is more concerned about asymmetric information, which is the case when the product is either more expensive (e.g., cellphones), or when it is an experience good (e.g., jeans). The upshot is that the effect of a rebate on sales is large, and consistent with the Buyer Belief Hypothesis (B1).

As we note in Section 3, an alternative narrative is that the increase in sales is a result of the effective price discount of RFF adoption or, possibly, the use of RFFs to price discriminate across different types of consumers. The next subsection shows that the signaling effect itself is sizeable.

5.2.2 Using Ineffective Ratings to Calculate the Magnitude of the Signaling Effect

We do not try to interpret the coefficient on price in Table 3 because, as in any demand estimation, without exogenous shifts in price, we cannot control for endogeneity. Of course, it is reassuring that the coefficient is negative, but this naturally raises a potential concern regarding the signaling value of RFF. Because a buyer who is able to redeem an RFF will effectively receive the equivalent of a price discount, it is possible that the response to RFF is purely a price-discount effect. With a credible identification of price elasticities, it would be possible to separate the signaling value of RFF from the price-discount effect of RFF. We do not, however, find a credible strategy to identify price elasticities, given the data we have. To address this, we now turn to the Ineffective Ratings Hypothesis (B2).

In columns (1)-(3) of Table 4 the dependent variable in equation (2) is the log of the number of ineffective ratings. Recall that by the nature of the RFF feature, buyers whose ratings contain no comment will not obtain the rebate according to Taobao's conditions. Hence, as we explain in Section 3, buyers who understand the signaling value of rebates but are uninterested in obtaining the rebate (or have a high cost of leaving feedback) should still flock to items with rebates because of their signaling content. This argument, in turn, implies that any increase in the number of *ineffective* ratings reflects buyers who are not attracted by the price-discount effect of rebates but

Table 4: Impact of rebate on the number of ineffective and total ratings of an item

Dependent variable	ln(no. item ineffective ratings +1) in month t			ln(no. item ratings +1) in month t		
	(1)	(2)	(3)	(4)	(5)	(6)
Rebate in month t	0.1273***			0.2428***		
	(0.0051)			(0.0056)		
- Coupon rebate		0.1370***			0.2562***	
		(0.0058)			(0.0065)	
- Cash rebate		0.1012***			0.2068***	
		(0.0092)			(0.0103)	
- Cellphone rebate			0.1321***			0.2628***
			(0.0172)			(0.0191)
- Memory card rebate			0.0506			0.1031*
			(0.0486)			(0.0539)
- Mask rebate			0.0791***			0.1467***
			(0.0080)			(0.0089)
- Jeans rebate			0.1627***			0.3116***
			(0.0068)			(0.0076)
Item fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	230,365	230,365	230,365	230,365	230,365	230,365
R2	0.395	0.395	0.395	0.337	0.337	0.338

Notes: Regressions are at the item-month level and standard errors are in parentheses. The dependent variable is the log of the number of ratings plus 1 (ineffective ratings in columns 1-3 and effective ratings in columns 4-6) that an item receives in month t. Asterisks indicate significance at 10% (*), 5% (**) and 1% (***).

instead by its signaling value, as stated in the Ineffective Rating Hypothesis (B2). Column (1) of Table 4 shows that rebates increase ineffective ratings by approximately 13.6%, confirming the Ineffective Rating Hypothesis (B2). Columns (2) and (3) show the heterogeneous impact across types of RFF and product categories, which mirror the sales results in Table 3. Columns (4)-(6) of Table 4 repeat the analysis for the total number of ratings.

We can use these results regarding ineffective and total ratings to calculate a lower bound on the signaling effect of RFF. The idea is simple: because Taobao's system will automatically assign an ineffective (18 character) rating to any transaction for which the buyer did not leave feedback but the seller did, then almost every transaction in our data has feedback. Hence, the increase in transactions with ineffective ratings can only be a consequence of the signaling effect and not a result of the RFF's price discount. In fact, even some of the increase in *effective* ratings may be a consequence of signaling alone, but for a lower bound analysis, we assume that *all* the additional effective ratings are driven by the price-discount feature of RFFs.

As column (4) in Table 4 shows, the impact of RFF on the *total* number of ratings is approximately 27.5% (the coefficient is 0.2428),[19] and column (1) in Table 4 shows that rebates increase ineffective ratings by approximately 13.6%. Hence, by proxying for the increase in transactions by the increase in *total* ratings of 27.5%, the proportion increase in ineffective ratings out of the increase in total ratings is $(0.136 \times 0.554)/0.275 = 27.2\%$.[20] That is, more than a quarter of the increase in total ratings, which proxy for total transactions, can be attributed to buyers who had no intention of receiving the discount implied by the RFF. This figure provides a conservative lower bound on the signaling effect of RFF. Proportionally, more than $27.2\% \times 36\% = 9.8\%$ increase in item sales is due to signaling alone. Because many effective ratings have very few characters, making them practically ineligible for a rebate, their signaling effect is likely to be quite large as well.

5.2.3 The Effect of Rebates on the Informativeness and Bias of Ratings

In Tables 5 and 6, the dependent variable in equation (2) is a variety of rating measures. We define the *ratio of item effective ratings* as the number of an item's effective ratings divided by the number of all its ratings. Columns (1) and (2) in Table 5 show that offering a rebate raises an item's ratio of effective ratings by almost 7%, and that coupon rebates have a larger effect than cash rebates. As column (3) shows, the effects are strongest for jeans and cellphones, weaker for masks, and nonexistent for memory cards, similar to previous patterns. Because effective ratings include more information, this confirms that RFF results in more informative feedback.

Columns (4) and (5) in Table 5 show that offering a rebate raises an item's *ratio of item long ratings* (the number of an item's long ratings divided by the number of all its ratings) also by almost 7%, and that coupons have a greater effect than cash rebates. This finding confirms the obvious *Long Ratings Hypothesis* (B3). Column (6) shows that the effects are strongest for jeans and cellphones, weaker for masks, and nonexistent for memory cards.

Using the estimates from Tables 3 and 5 we can provide an estimate of the *dynamic* reputation effect of using RFF. According to Table 3, the impact of an item's ratio of positive effective ratings on sales is 0.1388, which is approximately 15%. According to Table 5, the impact of RFF on the ratio

[19] This is less than the effect of rebates on the quantity of items sold, which is 36% (Table 3), because first, feedback may arrive with a lag and not be part of month t ratings, and second, some transactions include multiple items and ratings are given at the transaction level.

[20] The proportion of ineffective ratings in all ratings for the four categories is 0.554.

Table 5: **Impact of rebate on informativeness of item ratings**

Dependent variable	Ratio of item effective ratings, t			Ratio of item long ratings, t		
	(1)	(2)	(3)	(4)	(5)	(6)
Rebate in month t	0.0689***			0.0675***		
	(0.0028)			(0.0022)		
- Coupon rebate		0.0740***			0.0722***	
		(0.0032)			(0.0026)	
- Cash rebate		0.0548***			0.0546***	
		(0.0051)			(0.0041)	
- Cellphone rebate			0.0592***			0.0710***
			(0.0095)			(0.0077)
- Memory card rebate			0.0462*			0.0327
			(0.0276)			(0.0222)
- Mask rebate			0.0458***			0.0552***
			(0.0044)			(0.0035)
- Jeans rebate			0.0878***			0.0767***
			(0.0038)			(0.0030)
Item fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	194,002	194,002	194,002	194,002	194,002	194,002
R2	0.202	0.202	0.202	0.069	0.070	0.070

Notes: Regressions are at the item-month level and standard errors are in parentheses. The dependent variable is the ratio of an item's effective ratings (columns 1-3) and ratio of an item's long ratings (columns 4-6) that an item receives in month t. Asterisks indicate significance at 10% (*), 5% (**) and 1% (***).

of item effective ratings is approximately 7%.[21] Hence, multiplying these two effects (0.07×0.15) yields the dynamic reputation effect of using RFF, which is an approximately 1% increase in sales in the next period via an increase in positive effective ratings in the current period that are generated from the RFF in the current period. Though much smaller than the direct effect of RFF, which was 36%, this increase is still a significant impact that has cumulative effects over time.

Columns (1)-(6) in Table 6 explore whether rebates change the likelihood of receiving positive feedback, including item fixed effects to compare within items (columns (1)-(3)) and without them and to compare across items (columns (4)-(6)). All the specifications show that offering a rebate is not associated with more positive feedback, suggesting that offering RFF does not bias ratings compared to not offering RFF.

Interestingly, as columns (7)-(9) in Table 6 show, offering a rebate causes the seller to receive ratings earlier, shortening the time to receive a rating by close to 8%. This situation benefits a seller

[21]Specifically, the impact of RFF on the ratio of item positive effective ratings is approximately 7% and the impact of RFF on the ratio of item positive effective ratings is negligible and insignificant. The corresponding table is in the online Appendix.

Table 6: Impact of rebate on bias and time of item ratings

Dependent variable	Ratio of item pstv ratings, t						ln(average number of days between item transaction and rating +1) in month t		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Rebate in month t	-0.0008			-0.0001			-0.0730***		
	(0.0005)			(0.0004)			(0.0045)		
- Coupon rebate		-0.0005			0.0001			-0.0804***	
		(0.0006)			(0.0005)			(0.0052)	
- Cash rebate		-0.0018*			-0.0007			-0.0526***	
		(0.0010)			(0.0008)			(0.0082)	
- Cellphone rebate			-0.0018			-0.0008			-0.0541***
			(0.0018)			(0.0013)			(0.0153)
- Memory card rebate			-0.0010			0.0035			-0.0726
			(0.0052)			(0.0042)			(0.0444)
- Mask rebate			-0.0017**			0.0027***			-0.0440***
			(0.0008)			(0.0007)			(0.0070)
- Jeans rebate			-0.0000			-0.0018***			-0.0973***
			(0.0007)			(0.0005)			(0.0061)
Item fixed effect	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	194,002	194,002	194,002	194,002	194,002	194,002	193,992	193,992	193,992
R2	0.163	0.163	0.163	0.009	0.009	0.010	0.239	0.239	0.240

Notes: Regressions are at the item-month level and standard errors are in parentheses. The dependent variable is the ratio of positive ratings (columns 1-6) and the log of the average number of days between an item's sale in period t and the day the item received feedback. Asterisks indicate significance at 10% (*), 5% (**) and 1% (***).

in two ways: First, the proceeds from the sale will be transferred from Alipay faster, and second, the likelihood of a dispute decreases. As before, the effects are strongest for jeans and cellphones.

In summary, when a seller adopts RFF for an item, monthly sales are approximately 36% higher, and the number of ineffective ratings is nearly 14% higher. We also find that both the ratio of effective ratings and the ratio of long ratings of an item will increase when the item is chosen for RFF, but the ratio of positive ratings for the item does not seem to increase.

5.3 Robustness Checks

The data we obtained from Taobao cover a six-month period after the RFF mechanism was implemented, preventing us from conducting a natural experimental study of RFF adoption using data from before and after the introduction of RFF. Nonetheless, we can perform several robustness tests, which we describe in this section. All the tables that we refer to for these robustness tests appear in the online Appendix.

5.3.1 Measure of Item Quality

We use the ratio of positive effective item ratings as a measure of item quality. We also use the ratio of positive long item ratings as an alternative measure of item quality.[22] Similar to our previous findings, high-quality items are more likely to be adopted for RFF than low-quality items.

5.3.2 Product Categories

In sections 5.1 and 5.2, we report the estimated impact of rebates averaged across our four categories. Because detailed comments of other buyers are more important for goods with more risk involved in the purchase, such as experience goods and more expensive goods, we estimate the average impact of a rebate for each category as a robustness check. We divide our data into four categories—cellphones, memory cards, masks, and jeans—and run the panel regressions within each category.

Similar to our findings with the full sample of goods, we find for each category that rebates increase item sales, the ratio of long item ratings, and the number of long item ratings, whereas they shorten the days between a transaction and a rating. A rebate has almost no effect on an item's ratio of positive ratings in all categories.

[22] Recall that item positive long ratings include ratings with at least 24 Chinese characters, which is the 75^{th} percentile of all the ratings excluding the 18-character ratings.

5.3.3 Alternative Period Windows

Recall that we observe when the RFF offering started but do not observe precisely when it ended. In the analysis in sections 5.2 and 5.3, we use one month as a period window and use item-months as the observation measurement. One concern is that, if an item's rebate period is less than a month, our results may be biased. To address this concern we use two-week blocks instead of one month for the period window as a robustness check.

For each two-week period t, the explanatory variables for $t-1$ are created using indicators for a seller's grade and accumulated ratio of positive seller ratings up to the end of period $t-1$, accumulated item ratings up to the end of period $t-1$, and item sales from the previous 30 days prior to t because when a buyer considers an item for purchase, she can see the item sales for the past 30 days. For example, $\ln(\text{sales}, t-1)$ means the log sales of the item in the 30 days prior to the two-week period t. The results we obtain for each category are very similar to those in Tables 4-6.

Another concern may be that the periods are adjacent, implying that the first days of period t are closer to the last days in period $t-1$ than they are to most days in period t. To address this concern we conduct another set of robustness tests in which a period includes the first 15 days of each month from the 6 months we studied. This roughly two-week gap creates periods in which each day in the period is closer to other days in that same period than to days in any other period. The main results are also robust to this specification.

6 Concluding Remarks

The burgeoning growth of online marketplaces and the increased access they provide to data offer new and exciting opportunities to empirically test how markets work in practice. We exploit a unique dataset from Taobao's online marketplace to examine the effects of allowing sellers to buy a reputation through an RFF mechanism. Our empirical evidence supports the notion that the RFF mechanism creates a "missing market" that allows sellers to signal their high-quality products, especially when these products are new to the market. Namely, sellers with high-quality yet unestablished products use RFF to signal their product's quality, and buyers respond to these signals rationally. This response, in turn, alleviates the adverse selection problem and, more notably, the cold-start problem in anonymous online marketplaces. Our results shed light on the strategic interaction between buyers and sellers in online marketplaces, which in turn offers insights into the design of online markets.

Specifically, we find that high-quality products are more likely to participate in RFF than low-quality products, suggesting the positive selection signaling content of RFF. We also find that an item is more likely to be chosen to participate in RFF before establishing a good item reputation on Taobao and is less likely to be chosen by sellers who have established themselves as reputable. These results suggest that RFF is a substitute for product and seller reputations. We also show that buyers respond to the RFF signal in ways that are consistent with equilibrium behavior: sales of an item are approximately 36% higher when the seller chooses an RFF. It is as effective, on average, as quadrupling the previous month's sales or increasing the number of positive long ratings in the previous month by nearly seven times. Although we cannot estimate price elasticities and completely nail down the actual signaling effect of RFFs, we can use the increase in ineffective ratings to calculate a conservative lower bound of the signaling effect, which accounts for more than 27% of the 36% increase in sales and is likely to be much higher.

It is also notable that using RFFs helps sellers establish a good reputation quickly, creating a "flywheel" effect of sorts. That is, the signaling content of the RFF encourages both more sales and more feedback, with the latter rapidly increasing the seller's and the product's reputation, which in turn attracts more buyers and generates more sales. This situation alleviates the "cold start" problem from which new products typically suffer. Importantly, using RFF does not seem to create bias in feedback, implying that the informational content of the extra feedback is reliable.

Turning to the market design implications of our study, our analysis suggests that marketplaces can help reduce the asymmetric information problem by letting sellers engage in RFF signaling practices. That is, online marketplaces can rely on the strategic sophistication of both sellers and buyers to alleviate some of the asymmetric information problem by leveraging the signaling incentives of high-quality sellers. This last point offers insights into the question of whether marketplaces need to be regulated to improve quality. It is in the interest of marketplaces to reduce the asymmetric information problem, and we show that established market mechanisms such as signaling can be used to enhance marketplace quality.

In light of our findings, a natural question arises: why is RFF not used by other marketplaces that rely on voluntary buyer feedback, such as eBay or Amazon? Indeed, RFF is quite old in "internet time," and one would imagine that its text analysis algorithm for verifying that submitted reviews appear to be authentic is available to other platforms at a low cost. Therefore, why is it not more widely adopted? We can suggest a speculative answer regarding eBay and Amazon, and time will tell if other marketplaces will see the value of adopting some form of RFF mechanism in the future.

Taobao differs from eBay in that eBay has chosen to aggregate a user's ratings, whether for bought or sold items, and does not classify ratings for each product independently. On Taobao, an "item" is a product-seller combination and is therefore different from a "listing" (or product) on eBay. It was to eBay's detriment that it did not change its feedback reporting to create product-level reviews. Amazon's website features a large seller (Amazon itself) and many thousands of smaller third-party sellers who use the website to sell their products. Taobao, instead, is a marketplace with many small sellers and with more differentiated products across sellers. We speculate that this difference makes RFF less effective for Amazon than for Taobao because Amazon uses the "buy box" to feature a product, and if more than one seller sells the same product, then only one seller wins the buy box (see, e.g., https://www.bigcommerce.com/blog/win-amazon- buy-box/#the-changing-buy-box). Hence, because Amazon primarily bunches feedback at the product level, this creates a free-rider problem across sellers because unlike Taobao's, Amazon's feedback is not at the seller-product level.

References

Backus, M., Blake, T., and Tadelis, S. (2019). On the Empirical Content of Cheap-Talk Signaling: An Application to Bargaining. *Journal of Political Economy*, 127(4).

Bagwell, K. (2007). The Economic Analysis of Advertising. In *Handbook of Industrial Organization*, volume 3, pages 1701—1844.

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

Bolton, G. E., Katok, E., and Ockenfels, A. (2004). How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation. *Management Science*, 50(11):1587–1602.

Brown, J. and Morgan, J. (2006). Reputation in Online Auctions: The Market for Trust. *California Management Review*, 49(1):61–81.

Cabral, L. (2012). Reputation on the Internet. In Peitz, M. and Waldfogel, J., editors, *The Oxford Handbook of the Digital Economy*. Oxford University Press.

Cabral, L. and Li, L. I. (2015). A dollar for your thoughts: Feedback-conditional rebates on ebay. *Management Science*, 61(9):2052–2063.

Chen, Y., Harper, M., Konstan, J., and Li, S. X. (2010). Social Comparisons and Contributions to Online Communities: A Field Experiment on Movielens. *American Economic Review*, 100(4):1358–1398.

Dellarocas, C. (2003). The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management science*, 49(10):1407–1424.

Dellarocas, C. and Wood, C. A. (2008). The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias. *Management Science*, 54(3):460–476.

Einav, L., Kuchler, T., Levin, J., and Sundaresan, N. (2014). Assessing Sale Strategies in Online Markets Using Matched Listings. *American Economic Journal: Microeconomics*, 7(2):215–247.

Elfenbein, D., Fisman, R., and McManus, B. (2012). Charity as a Substitute for Reputation: Evidence from an Online Marketplace. *Review of Economic Studies*, 79(4):1441–1468.

Fan, Y., Ju, J., and Xiao, M. (2016). Reputation Premium and Reputation Management: Evidence from China's Largest e-Commerce Platform. *International Journal of Industrial Organization*, 46(1):63–76.

Fradkin, A., Grewal, E., Holtz, D., and M., P. (2015). Bias and Reciprocity in Online Reviews: Evidence from Experiments on Airbnb. *16th ACM Conference on Electronic Commerce, June 2015, (EC 2015)*, page 641.

Horton, J. and Golden, J. (2015). Reputation Inflation: Evidence from an Online Labor Market. *Working paper, New York University*.

Hui, X., Saeedi, M., Shen, Z., and Sundaresan, N. (2016). Reputation and Regulations: Evidence from eBay. *Management Science*, 62(12):3604–3616.

Kawai, K., Onishi, K., and Uetake, K. (2013). Signaling in Online Credit Markets. Working Paper.

Kihlstrom, R. E. and Riordan, M. H. (1984). Advertising as a Signal. *Journal of Political Economy*, 3(92):427–450.

Lafky, J. (2014). Why do People Rate? Theory and Evidence on Online Ratings. *Games and Economic Behavior*, 19(2):554–570.

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

Li, L. I. and Xiao, E. (2014). Money Talks: Rebate Mechanisms in Reputation System Design. *Management Science*, 60(8):2054–2072.

Luca, M. and Zervas, G. (2016). Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud. *Management Science*, 62(12):3412–3427.

Masterov, D. V., Mayer, U. F., and Tadelis, S. (2015). Canary in the e-Commerce Coal Mine: Detecting and Predicting Poor Experiences Using Buyer-to-Seller Messages. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, EC '15, pages 81–93, New York, NY, USA. ACM.

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *American Economic Review*, 104(8):2421–55.

Milgrom, P. and Roberts, J. (1996). Price and Advertising Signals of Product Quality. *Journal of Political Economy*, 4(94):796–821.

Miller, N., Resnick, P., and Zeckhauser, R. (2005). Eliciting Honest Feedback: The Peer Prediction Method. *Management Science*, 51(9):1359–1373.

Nelson, P. (1970). Information and Consumer Behavior. *Journal of Political Economy*, 2(78):729–754.

Nelson, P. (1974). Advertising as Information. *Journal of Political Economy*, 4(82):729–754.

Nosko, C. and Tadelis, S. (2015). The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment. *NBER Working paper*, (No. w20830).

Proserpio, D. and Zervas, G. (2017). Online Reputation Management: Estimating the Impact of Management Responses on Consumer Reviews. *Marketing Science*, 36(5):645–665.

Sahni, N. and Nair, H. (2016). Does Advertising Serve as a Signal? Evidence from Field Experiments in Mobile Search. Stanford University working paper.

Tadelis, S. (2016). Reputation and Feedback Systems in Online Platform Markets. *Annual Review of Economics*, 8:321–340.

Wu, C., Che, H., Chan, T. Y., and Lu, X. (2015). The Economic Value of Online Reviews. *Marketing Science*, 34(5):739–754.

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

Zervas, G., Proserpio, D., and John, B. (2015). A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average. *Boston University Working paper*.