

Designing Aggregation Mechanisms for Reputation Systems in Online Marketplaces

CHRISTINA APERJIS

Hewlett Packard Labs

and

RAMESH JOHARI

Stanford University

A seller in an online marketplace with an effective reputation mechanism should expect that dishonest behavior results in higher payments now, while honest behavior results in higher reputation—and thus higher payments—in the future. We briefly survey recent results on the Window Aggregation Mechanism, a widely used class of mechanisms that shows the average value of the seller's ratings within some fixed window of past transactions. We suggest approaches for choosing the window size that maximizes the range of parameters for which it is optimal for the seller to be truthful. We show that mechanisms that use information from a larger number of past transactions tend to provide incentives for patient sellers to be more truthful, but for higher quality sellers to be less truthful. We then discuss a broader class of aggregation mechanisms that weight recent ratings more heavily and show that the same insight applies.

Categories and Subject Descriptors: K.4.4 [Computers and Society]: Electronic Commerce

General Terms: Economics; Theory

Additional Key Words and Phrases: Aggregation mechanisms, reputation systems, electronic marketplaces

1. INTRODUCTION

In online trading communities, sellers have a temptation to dishonesty, because potential buyers have to decide how much to pay for an item without being able to observe it firsthand. In particular, the buyer typically cannot know in advance whether the seller is describing the item honestly, and hence may be afraid that he might be exploited if he trusts the seller. This effect is exacerbated because a buyer will often interact with sellers with whom he has never interacted before and may only seldom interact in the future. The absence of trust created by this information asymmetry may result in market failure [Akerlof 1970].

If there is a *reputation system* in place, dishonesty involves a greater immediate payoff at the expense of a lower long-term payoff. In such a mechanism, after a transaction the buyer can rate the seller. The system typically relies on an *aggregation mechanism* that aggregates past ratings into the seller's score. For example, the system might maintain the empirical average of feedback across *all* the seller's past transactions; the empirical average over only a fixed window of past transactions; or, more generally, a weighted average of past transactions.

When a seller posts a description for an item, potential buyers also observe the seller's score. Using this information, we postulate that buyers employ simple

Authors' addresses: christina.aperjis@hp.com, ramesh.johari@stanford.edu

heuristics (e.g., [Tversky and Kahneman 1974]), Bayesian techniques, or some combination of the two, in order to decide how much to bid [Resnick et al. 2006]. In our model the expected payment to the seller is thus a function of his description and his score, and may also depend on the aggregation mechanism. Empirical studies have shown that sellers with high scores enjoy a *price premium*: on average they sell at higher prices than less reputable sellers; see [Resnick et al. 2006] for a survey. Motivated by these studies, we assume that the expected payment to the seller is increasing in his score.

A natural goal for a reputation system is to encourage sellers to describe their items truthfully. While incentivizing truthfulness is intuitively appealing, it is also connected to *efficient trade*. Trade is efficient if the item is allocated to the agent that values it most; this also corresponds to aggregate surplus maximization. The buyers' valuations depend on the item's characteristics, which only the seller knows before an online sale. Thus, in order to achieve efficient trade, it is important that the seller gives an accurate description to potential buyers.

In [Aperjis and Johari 2010a] we show that when ratings are averaged over the entire lifetime of the seller (the Unweighted Aggregation Mechanism), then for a large class of payment functions, the seller may be incentivized to falsely advertise after some rating histories. In particular, the temptation for the seller to become dishonest increases as the total number of ratings increases, because each additional rating has a smaller and smaller effect on the seller's score. On the other hand, if ratings from recent transactions affect the seller's score more than older ratings, then it is possible to incentivize the seller to be truthful. In this note, we briefly describe recent results on such aggregation mechanisms reported in [Aperjis and Johari 2010a] and [Aperjis and Johari 2010b].

2. WINDOW AGGREGATION MECHANISM

In [Aperjis and Johari 2010b] we study the *Window Aggregation Mechanism*, a widely used mechanism in which the seller's score is the average value of his T most recent ratings. With the goal of incentivizing the seller to describe the item truthfully, we address the *design question* of choosing the right window size T . We define an optimal window size as one which maximizes the range of parameters for which it is optimal for the seller to be always truthful, and we study the dependence of the optimal window size on the parameters of the model.

Our main result is an interesting qualitative tradeoff: informally, *increasing the window size is more likely to make patient sellers truthful, while it is less likely to make high quality sellers truthful*. First, increasing the window size is more likely to make patient sellers truthful. This is an intuitive result, since *patient sellers value future payments more, and thus an aggregation mechanism with longer memory can successfully couple current behavior with distant future payoffs*. On the other hand, a larger window is less likely to make high quality sellers truthful. When the seller has high value items for sale most of the time and the window is large, the seller is likely to have a high score regardless of what actions he takes when he receives a low value item, because most items have high value. This makes a smaller window more desirable, because it *magnifies the impact of the seller's actions in those periods where he has a low value item for sale*.

Alternatively, we can interpret this result by directly looking at the effect of the window size on the seller's future scores. By choosing a window size T , the system designer determines *how much* and for *how long* the seller's future scores decrease if he does not describe his current item truthfully. We identify the following tradeoff between the intensity and the duration of this score reduction: the intensity is decreasing in T , while the duration is increasing in T . The duration effect is more important for patient sellers, so the optimal window size increases. On the other hand, the intensity effect becomes more important for high quality sellers, so the optimal window size decreases.

Our analysis proceeds through three different models of buyer feedback. First, we consider the case of *perfect monitoring*, where the buyer rates the seller accurately after every transaction. Then, we consider two types of *imperfect monitoring*: the seller may not receive ratings after some transactions (missing feedback), and the rating that the seller receives may not accurately reflect his action (incorrect feedback). In all these settings, under reasonable assumptions, we find the same qualitative tradeoff described above. The analysis suggests that the tradeoff between incentivizing high quality sellers and patient sellers is quite robust to noisiness in the feedback from buyers.

We note that the Window Aggregation Mechanism is widely used in online marketplaces, such as eBay and the Amazon Marketplace. In particular, in both marketplaces the average value of the seller's ratings in the last 12 months is shown in a prominent position: next to the description of the item on eBay and below the seller's name on Amazon. Finally, the Window Aggregation Mechanism is simple, since it is characterized by a single parameter (the window size) and is easily understood by buyers and sellers.

3. WEIGHTED AGGREGATION MECHANISM

In [Aperjis and Johari 2010a] we study the *Weighted Aggregation Mechanism*, a mechanism in which the seller's score is a weighted average of his past ratings. Let w_i be the weight for the i -th most recent rating. The Window Aggregation Mechanism with window size T is a special case of the Weighted Aggregation Mechanism, where $w_i = 1/T$ for the T most recent ratings, and $w_i = 0$ for older ratings.

We address the design question of choosing the vector of weights \mathbf{w} that best incentivizes the seller to be truthful. We find that it is optimal to weight recent transactions more, and include only a finite number of past ratings in the seller's score. We then establish (under a related but different set of assumptions) the same tradeoff identified above for the Window Aggregation Mechanism: averaging over a longer past history of ratings is *more likely* to incentivize patient sellers to be truthful, but *less likely* to incentivize high quality sellers to be truthful. Despite these properties, such mechanisms are not characterized by the simplicity or ubiquity of Window Aggregation Mechanisms.

4. NONEQUILIBRIUM VS. EQUILIBRIUM APPROACHES

In our model, we consider the best response of the seller to a *fixed* behavior of the buyers; that is, a fixed payment function which only depends on the information available to buyers (i.e., the aggregation mechanism, the seller's score, and the de-

scription of the item). We believe this nonequilibrium approach may be reasonable in practice because the large and dynamic set of participants in the major online markets makes the rationality, knowledge, and coordination required for the usual notion of perfect Bayesian equilibrium difficult to ensure. Indeed, an alternate approach to justify our analysis is to view the payment function seen by the seller as the equilibrium outcome in a *large system limit* of an electronic marketplace, where the number of buyers and sellers approaches infinity. In ongoing work, we are applying the notion of *mean field equilibrium* for dynamic games to establish that the design insights provided here are valid in a scaling limit where the number of market participants grows large. This technique may be of value for other mechanism design problems in dynamic settings; in such problems characterizing equilibria is often intractable, eliminating any hope of developing engineering insight with respect to dynamic equilibria. Instead, large system limits simplify the characterization of equilibria, and thus optimization over classes of dynamic mechanisms as well.

REFERENCES

- ADLAKHA, S., JOHARI, R., WEINTRAUB, G., AND GOLDSMITH, A. 2010. Mean field equilibrium in concave stochastic games. Working paper.
- AKERLOF, G. A. 1970. The market for lemons: Quality uncertainty and the market mechanism. *Quart. J. of Econom.* 84, 3, 488–500.
- APERJIS, C. AND JOHARI, R. 2010a. Designing reputation mechanisms for efficient trade. Tech. rep., Stanford University.
- APERJIS, C. AND JOHARI, R. 2010b. Optimal windows for aggregating ratings in electronic marketplaces. *Management Sci.* 56, 5 (May), 864–880.
- RESNICK, P., ZECKHAUSER, R., SWANSON, J., AND LOCKWOOD, K. 2006. The value of reputation on eBay: A controlled experiment. *Experimental Econom.* 9, 2 (June), 79–101.
- TVERSKY, A. AND KAHNEMAN, D. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185, 1124–1131.