

Goodwill Hunting: An Economically Efficient Online Feedback Mechanism for Environments with Variable Product Quality

Chrysanthos Dellarocas

Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139
dell@mit.edu

Abstract. Most online feedback mechanisms publish unbiased statistics (usually averages) of past ratings. Such mechanisms fail in environments where the same seller sells products of many different qualities, such as marketplaces of used cars and collectibles. This paper presents a novel feedback management mechanism that succeeds in facilitating efficient transactions in such settings. One particularly interesting aspect of this mechanism is that it uses the threat of biased future reporting of quality in order to induce sellers to truthfully declare the quality of their items.

1 Introduction

In many marketplaces sellers have more information about the quality of the goods they sell than buyers do. Self-interested sellers are expected to try to take advantage of such information asymmetries by misrepresenting (inflating) the true quality of their goods and charging higher prices. Knowing this, however, rational buyers will ignore the sellers' quality signals and will only be willing to pay prices corresponding to the average expected quality of all goods in the marketplace. Left to itself, this set of behaviors will gradually drive high cost/high quality sellers out of the market and might even lead to the total dissolution of the marketplace [1].

Since, in electronic markets, face-to-face interaction between the buyer and the seller and direct inspection of goods is usually not possible, the problem of quality uncertainty is exacerbated [7]. In such settings, the marketplace operator, acting as a trusted third party, can play a vital role in reducing uncertainty and facilitating efficient transactions. One obvious role would be to inspect and certify the quality of all goods offered for sale. However, such pre-sale inspection is not always practical and, even when it is practical, it would be prohibitively expensive and certainly not scalable to the size of some of today's electronic markets.

Another, more plausible, role for the marketplace is to operate a feedback or reputation mechanism [4], [9]. Feedback mechanisms encourage buyers to rate the quality of goods received (as perceived by them). Submitted feedback is analyzed, aggregated with feedback received from other members and made publicly available

to the community in the form of seller reputation profiles. Several examples of such mechanisms are already being used in a number of well-known online communities, such as eBay, Amazon, ePinions, Slashdot, etc. A growing body of empirical evidence seems to indicate that these systems have managed to provide remarkable stability in otherwise very risky trading environments [3], [5], [6], [8], [10].

Almost all commercial feedback mechanisms in use today act as impartial rating aggregators: they accumulate individual ratings and publish an unbiased aggregate measure such as the sum or average of those ratings per product or per seller. In this paper I show that such mechanisms fail to induce cooperation in settings where the same seller may sell products of several different qualities, such as marketplaces of used cars, collectibles, etc. Furthermore, I propose a novel feedback management algorithm called “Goodwill Hunting” or GWH (for reasons that will become obvious in Section 3) that works in such settings. The most interesting feature of GWH is that it uses the threat of biased reporting of quality in order to induce sellers to truthfully reveal the qualities of their items.

The rest of the paper is structured as follows. Section 2 explains why simple feedback aggregation mechanisms fail in settings where the same seller may sell products of various qualities. Section 3 presents and analyzes the Goodwill Hunting mechanism under the assumption that buyers submit truthful feedback. Section 4 presents some extensions to the algorithm that provide incentives to buyers to submit truthful feedback. Section 5 discusses additional considerations that arise if we assume that players eventually exit the market. Finally, Section 6 concludes.

2 Shortcomings of Simple Feedback Aggregation Mechanisms

The majority of today’s feedback mechanisms operate as unbiased aggregators of consumer ratings: they solicit feedback regarding the buyers’ perception of quality following transactions and publish a statistic (most often the average) of such feedback received so far for each seller. The purpose of this section is to show that such simple mechanisms fail in settings where sellers sell products of various qualities.

More specifically, the setting of this paper is a marketplace where a large number of long-lived sellers sell a wide variety of goods for which the demand is independent. Therefore, each seller can be considered a monopolist in his class of goods. The marketplace also contains a large number of long-lived buyers¹. The quality of each good can be described by a single scalar parameter $q \in \mathbf{R}$. Each seller k sells an infinite sequence of goods whose qualities are distributed around a mean μ_k with variance π_k^2 . These parameters are specific to each seller and are initially unknown to buyers. There is agreement among buyers and the seller as to what constitutes a product of a given quality. Examples of settings that meet these assumptions include

¹ In the rest of the paper, I will use the subscript i to refer to products, j to refer to buyers and k to refer to sellers.

marketplaces of used cars, collectibles (rare stamps, rare coins, comic books, antiques, etc.) and many professional services.

Sellers sell their items sequentially using Vickrey auctions. Following each transaction, the winning bidder is encouraged to rate the seller by reporting the quality q_i^b of the current item as perceived by her. The system collects these ratings and publishes the sample mean m_k and variance s_k^2 of ratings for each seller.

I assume that quality perception by buyers is noisy². More specifically, if the quality of item i , as perceived by the seller is q_i^s , then the buyer perceives $q_i^b = q_i^s + \varepsilon_i$, where ε_i is a noise term with mean zero and variance σ^2 . To simplify the analysis, all distributions in this paper are assumed to be normal. Under these assumptions, if the number of accumulated ratings for a seller k is large enough, then $m_k \approx \mu_k$ and $s_k^2 \approx \pi_k^2 + \sigma^2$.

Since items are sold through Vickrey auctions, buyers bid amounts equal to their expected valuations. Suppose that the winning bidder has a valuation function $W(q^b)$, where q^b is an item's quality as perceived by the *buyer*. Since the buyer only knows a seller's average quality μ_k and variance $s_k^2 \approx \pi_k^2 + \sigma^2$ then she will bid an amount equal to $G(\mu_k, s_k^2) = \int u_{s_k^2}(\varepsilon) \cdot W(\mu_k + \varepsilon) \cdot d\varepsilon$, where $u_{s_k^2}(\cdot)$ is the density function of a normal distribution with mean zero and variance s_k^2 .

According to the preceding reasoning, the revenue of a single auction round will be constant for a given seller and independent of the actual quality of the item being auctioned. Let $c(q^s)$ be the cost of producing an item of quality q^s . I assume that cost increases with quality. Following an argument similar to that of Akerlof ([1]), if the seller's quality variability is high enough (in other words, if π_k^2 is high enough) and the profit margin is low enough then there will be some quality \bar{q}^s , with the property that if $q^s > \bar{q}^s$ then $c(q^s) > G(\mu_k, s_k^2)$. But then it will not be profitable for the seller to sell items of qualities $q^s > \bar{q}^s$. This, in turn, will lower the seller's mean quality, which will lower the revenue of a single auction round, which will further lower the quality threshold of profitability, and so on, until the seller ends up selling only the lowest possible quality items in his inventory.

One might think that the situation can be improved by giving sellers the opportunity to declare the quality of each of their individual items. More specifically, suppose that the mechanism asks each seller to declare the quality q_i^d of each item

² Our model of noise is meant to capture both genuine differences in perceptions between buyer and seller as well as non-deterministic fluctuations of product quality despite the seller's best intentions.

and then publishes the triplet (q_i^d, m_k, s_k^2) . Since, in this simple mechanism, buyers are asked to report the qualities they perceive, a seller's future m_k, s_k^2 do not depend on what quality he declares on the current round. Therefore, it suffices to study the seller's decision problem in a single round.

The seller's problem is to select a declared quality q_i^d that maximizes his single-round auction revenue. The buyer's problem is to use the information available to her in order to make the best possible estimate of her expected valuation. Suppose that the buyer assigns probability $\alpha > 0$ to the event the seller truthfully declares his quality. Then, the buyer's expected valuation is given by $G = \alpha \cdot H(q_i^d) + (1 - \alpha) \cdot G(m_k, s_k^2)$,

where $H(q^s) = \int u_{\sigma^2}(\varepsilon) \cdot W(q^s + \varepsilon) \cdot d\varepsilon$ is the expected utility to the buyer of an item

whose quality as perceived by the seller is q^s ³. Knowing this, the seller will always declare the value that maximizes $H(\cdot)$, which is the maximum possible quality. But this contradicts the belief that the seller sometimes tells the truth. The only case where the beliefs of the buyer are consistent with the optimal response of the seller is when $\alpha = 0$, that is, the seller always lies and the buyer ignores his declared qualities and bids his average valuation $G(m_k, s_k^2)$.

To conclude, this section has shown that simple feedback aggregation mechanisms fail to facilitate efficient transactions in settings where each seller sells products of various different qualities. Even when such mechanisms ask sellers to truthfully declare the quality of their individual items, self-interested sellers will never find it optimal to do so.

3 The “Goodwill Hunting” Algorithm

The main argument of this paper is that feedback mechanisms can sometimes increase efficiency in the marketplace by taking a more active role than that of unbiased feedback aggregators. More specifically, I present a concrete mechanism that succeeds in facilitating efficient transactions in environments where sellers sell items of various qualities. The mechanism uses the threat of biased future reporting of quality in order to induce sellers to truthfully declare the individual qualities of their items.

The setting is identical to that of the previous section. Monopolist sellers sell infinite sequences of goods $\{q_i^s\}$, where q_i^s denotes the quality of good i as perceived by the seller. Sellers sell their items sequentially using Vickrey auctions. The mechanism asks the seller to declare the quality q_i^d of each good and publishes a

³ $u_{\sigma^2}(\cdot)$ is the density function of the error term $\varepsilon = q^b - q^s$.

quality q_i^p , which is a function of both q_i^d and past feedback received for this seller. Sellers cannot contact buyers directly. If $W_j(q_i^b)$ denotes buyer j 's valuation function then her bids are equal to $G_j(q_i^p) = E(W_j(q^b) | q_i^p)$ and the expected auction closing price is given by $G(q_i^p) = E(W(q^b) | q_i^p)$ where $W(q_i^b)$ is the second higher bidder's valuation function for this type of goods. We assume that the distribution of buyer valuations for a given class of goods remains constant over time or changes relatively slowly to reflect changes in tastes, standard or living, etc. at the population level. Therefore, at equilibrium, the auction revenue function $G(q^p)$ for a given seller is expected to be constant or slowly changing with time. The mechanism can therefore "learn" that function by keeping track of past auction results (each auction result provides a noisy sample of the function $G(q^p)$). In the following analysis, I assume that the mechanism knows $G(\cdot)$

I assume that sellers have no control over the qualities and cost of individual goods in their sequences. This assumption is reasonable in settings where the act of accumulating inventory is independent from that of (re)selling it. For example, consider a user car dealership where one department is in charge of buying used cars (accumulating inventory) and another department is in charge of reselling them. I further assume that sellers cannot control the exact sequencing of goods they sell. This assumption simplifies the analysis because it eliminates the possibility of strategic sequencing: if sellers could decide on the sequencing of items, they could try to build a reputation by honestly selling low value items and then, once they have built a good reputation, "make a killing" by cheating on a few high value items. This assumption is more difficult to justify. It is reasonable in environments where, as above, accumulation and reselling of inventory are independent and concurrent activities and where inventory levels are kept low. In the used car dealership example, this would correspond to a situation where resellers sell cars at roughly the same pace as buyers accumulate them, so that at any given time, the number of cars on inventory is very low (in order for the assumption to strictly hold, this number must be equal to one)⁴.

With these two constraints in place, given a sequence of goods of qualities $\{q_i^s\}$ the seller's optimization problem is to select a sequence of declared qualities $\{q_i^d\}$, such that the expected value of the seller's discounted lifetime revenues

$$V = \sum_{i=0}^{\infty} \delta^i \cdot G(q_i^p) \text{ is maximized}^5.$$

⁴ The implications of removing this second assumption are the topic of ongoing research.

⁵ Production cost does not enter into V because, as a consequence of the above assumptions, it is already *sunk* by the time an item is sold.

Initialization	
Set initial goodwill	$g_0 = 0$
Iteration ($i = 0 \dots \infty$)	
1. Seller finds out quality of next item	q_i^s
2. Seller decides what to declare and declares quality to mechanism	$q_i^d = f(q_i^s)$
3. Mechanism publishes declared quality adjusted to compensate/punish seller for previously accumulated/owed goodwill	$q_i^p = G^{-1}(G(q_i^d) + (1 - \delta) \cdot g_i)$
4. Buyer pays seller	$G_i = G(q_i^p) = G(q_i^d) + (1 - \delta) \cdot g_i$
5. Buyer perceives quality of item	$q_i^b = q_i^s + \varepsilon_i$
6. Buyer decides what to report to mechanism and submits rating	$q_i^r = h(q_i^b)$
7. Mechanism adjusts seller's goodwill	$g_{i+1} = g_i + \frac{G(q_i^r) - G(q_i^d)}{\delta} = g_i + \frac{\Delta G_i}{\delta}$

Fig. 1. The Goodwill Hunting Mechanism

q^s	Item quality as perceived by seller
q^d	Item quality declared by seller to mechanism
q^p	Item quality published by mechanism to buyers
q^b	Item quality as perceived by buyer
q^r	Item quality reported by buyer to mechanism
g_i	Seller's goodwill at the beginning of round i
$W(q^b)$	Winning bidder's valuation function
$H(q^s)$	Winning bidder's expected utility given seller's truthful quality declaration
$G(q^p)$	Seller's revenue from a single transaction
$\{q_i^s\}$	Infinite sequence of items for sale
$V(\{q_i^s\})$	Seller's expected discounted lifetime payoff from selling item sequence $\{q_i^s\}$
δ	Seller's discount factor
ε_i	Difference between seller's and buyer's perception of an item's quality (noise term)

Fig. 2. Notation used in the Goodwill Hunting mechanism

The goal of the mechanism is to facilitate efficient transactions by providing incentives to sellers to truthfully declare the qualities of their goods. To that end, the mechanism keeps track of a quantity that we will call the seller’s *goodwill*. This quantity, at any given time is equal to the discounted sum of “unfair” monetary gains or losses that the seller has incurred during his entire stay in the marketplace relative to the reports (ratings) of the buyers that transacted with him. All sellers enter the market with zero goodwill. Following each transaction, if a buyer reports the product to have quality higher than what the seller declared, then the seller has incurred an “unfair” loss equal to $G(q_{reported}) - G(q_{declared})$ and the seller’s goodwill is increased by that amount. If, on the other hand the buyer reports lower quality, the seller has realized an “unfair” gain and the seller’s goodwill is reduced accordingly.

If a seller accumulates nonzero goodwill, then the mechanism adjusts the published quality q^p of his subsequent goods relative to what the seller declares, so as to compensate/punish the seller for the amount of goodwill he has accumulated/owes. More specifically, if a seller has negative goodwill, (which occurs if the seller has *overstated* the quality of some of his goods in the past relative to the buyers perceptions), the mechanism reduces the published quality of all subsequent goods sold by the seller (relative to the declared quality) by a small amount, calculated to, over time, take away from the seller whatever unfair gains he previously realized by overstating the quality of his goods. Similarly, if the seller accumulates positive goodwill (which occurs if the seller has *understated* the quality of some of his goods in the past relative to the buyers perceptions), the mechanism increases the published quality of subsequent goods by a small amount, calculated to compensate the seller for whatever unfair losses he previously incurred. That way, as I will formally show below, the mechanism makes it impossible for the seller to realize any long-term gains by misrepresenting the quality of his products.

Figure 1 summarizes the simplest version of the GWH mechanism. Figure 2 summarizes the notation used in this and the next section. The rest of the section explains why the mechanism indeed induces self-interested sellers to truthfully declare the qualities of their items. To simplify the analysis, this section makes the assumption that buyers always provide truthful (albeit noisy) feedback. Section 4 presents a few simple extensions of the GWH mechanism that provide incentives to buyers to indeed behave in this way.

Analysis of the GWH mechanism

In the setting described in this paper, given a sequence of items $\{q_i^s\}$, the seller’s objective is to select a sequence of declared qualities $\{q_i^d\}$ so as to maximize his discounted lifetime payoff:

$$V(\{q_i^s\}) = \sum_{i=0}^{\infty} \delta^i \cdot G(q^p(q_i^d, g_i)) \tag{1}$$

In the following analysis, we will use the notation \mathbf{q}_i^s to refer to the sequence $\{q_i^s, q_{i+1}^s, \dots\}$, i.e. the sequence of item qualities that remains after the first i members

are removed. The seller’s optimization problem can equivalently be expressed using a Bellman equation formulation as follows:

$$V(\mathbf{q}_i^s, g_i) = \max_{q_i^d} \left\{ G(q^p(q_i^d, g_i)) + \delta \cdot E \left[V(\mathbf{q}_{i+1}^s, g + \frac{G(q^r(q^b(q_i^s))) - G(q_i^d)}{\delta}) \right] \right\} \quad (2)$$

Before we proceed, it will be useful to prove the following lemma:

Lemma 1: For any sequence of item qualities $\{q_i^s\}$, $i = 0, \dots, \infty$ and any $g \in \mathbf{R}$,

$$V(\{q_i^s\}, g) = V(\{q_i^s\}, 0) + g$$

Proof: See Appendix.

We are now ready to prove the main result of this section:

Proposition 1: If buyers submit truthful reports of their perceived quality to the mechanism (i.e. if $q^r = q^b$) then truthful declaration of qualities (that is, $q^d = q^s$) is a subgame perfect equilibrium strategy for the seller.

Proof: By the Optimality Principle of dynamic programming it suffices to show that there is no point at which a seller can make a one-time play different from the equilibrium play that raises his total payoff.

Making use of Lemma 1, (2) can be rewritten as:

$$\begin{aligned} V(\mathbf{q}_i^s, g_i) &= V(\mathbf{q}_i^s, 0) + g_i \\ &= \max_{q_i^d} \left\{ G(q^p(q_i^d, 0)) + \delta \cdot E \left[V(\mathbf{q}_{i+1}^s, \frac{G(q^r(q^b(q_i^s))) - G(q_i^d)}{\delta}) \right] \right\} + g_i \\ &= \max_{q_i^d} \left\{ G(q^p(q_i^d, 0)) + \delta \cdot E \left[V(\mathbf{q}_{i+1}^s, 0) + \frac{G(q^r(q^b(q_i^s))) - G(q_i^d)}{\delta} \right] \right\} + g_i \\ &= \max_{q_i^d} \{ G(q^p(q_i^d, 0)) - G(q_i^d) + G(q^r(q^b(q_i^s))) + \delta \cdot V(\mathbf{q}_{i+1}^s, 0) \} + g_i \end{aligned} \quad (3)$$

From Figure 1, $G(q^p(q_i^d), 0) = G(q_i^d)$. Substituting into (3) we get:

$$V(\mathbf{q}_i^s, g_i) = \max_{q_i^d} \{ G(q^r(q^b(q_i^s))) + \delta \cdot V(\mathbf{q}_{i+1}^s, 0) \} + g_i \quad (4)$$

which shows that the seller’s payoff does not directly depend on the declared qualities $\{q_i^d\}$ and therefore cannot be increased by deviating from truth-telling. ■

Proposition 1 shows that the seller is indifferent between truthfully representing and misrepresenting the quality of his items. Under the assumption of risk-averse buyers, a small extension of the mechanism will strengthen the result even further and

will make it strictly preferable for the seller to truthfully declare the qualities of his items.

In order to understand how the mechanism must be extended, it is important to understand the properties of the auction revenue function $G(q^p)$.

Proposition 2: The auction revenue function $G(q^p)$ is a decreasing function of the variance of the error term $\varepsilon = q^b - q^s$.

Proof Sketch: The basic idea of the proof is the following: Suppose that the winning bidder has a valuation function $W(q^b)$, where q^b is an item’s quality as perceived by the buyer. In that case, $H(q^s) = \int_{\varepsilon} u_{\sigma^2}(\varepsilon) \cdot W(q^s + \varepsilon) \cdot d\varepsilon$ represents that bidder’s expected utility from purchasing an item whose quality as perceived by the seller is q^s . If the bidder is risk-averse (in other words, if $W(q^b)$ is concave) and ε is a normally distributed error term, then the classic results of Rothschild and Stiglitz ([11]) imply that $H(q^s)$ is a decreasing function of the variance of the error term ε . The proof derives $G(q^p)$ as a function of $H(q^p)$ and uses the above property of $H(\cdot)$ to show that $G(q^p)$ decreases as the variance grows.

Proposition 2 shows that $G(\cdot)$, and therefore the seller’s revenues increase as the (buyers perception of the) variance of ε decreases. Therefore, it is to the seller’s advantage to keep the buyers’ perception of the variance as low as possible. The mechanism can induce the seller to always tell the truth by publishing an estimate of that variance, equal to the sample variance of the random variable:

$$z = (q^r - q^d) \cdot y$$

where y is a random variable that takes the value +1 with probability 0.5 and -1 with probability 0.5. It is easy to see that $E(y) = 0$ and $E(y^2) = Var(y) = 1$ ⁶.

Suppose now that the seller attempts to misrepresent the quality of some of his items by declaring $q^d = q^s + \xi$, where ξ is a random variable. Given that $q^r = q^b = q^s + \varepsilon$, $z = (q^r - q^d) \cdot y = (\varepsilon - \xi) \cdot y$. Then:

$$E(z) = E[(\varepsilon - \xi) \cdot y] = E[\varepsilon - \xi] \cdot E[y] = 0$$

⁶ The reason why we multiply $q^r - q^d$ by the random variable y is to avoid the situation where the seller always inflates his quality by a constant factor ξ_o . In that case, $Var(q^r - q^d) = Var(\varepsilon - \xi_o) = Var(\varepsilon)$, therefore, if it were not for the factor y , the seller would have been able to get away without any increase in variance.

$$E(z^2) = E[(\epsilon - \xi)^2 \cdot y^2] = E[(\epsilon - \xi)^2] \cdot E[y^2] = E[\epsilon^2] + E[\xi^2] = \text{Var}(\epsilon) + E[\xi^2]$$

$$\text{Var}(z) = E(z^2) - E(z) = E(z^2) = \text{Var}(\epsilon) + E[\xi^2] > \text{Var}(\epsilon)$$

We see that any attempt of the seller to deviate from truth telling increases $\text{Var}(z)$, which is the system’s estimate of $\text{Var}(\epsilon)$ and therefore only serves to decrease the seller’s revenues. If the mechanism publishes z , the seller will therefore find it optimal to always truthfully declare his perception of his products’ qualities.

4 Incentives for Truthful Buyer Participation in the Feedback Mechanism

The previous section demonstrated that if buyers rate according to their truthful perception of quality, sellers find it optimal to truthfully declare the qualities of their items. This section presents some ideas on how the mechanism can be extended in order to induce buyers to indeed behave in this manner.

The problem of inducing participation (let alone truthful participation) to feedback mechanisms has been identified by several authors, such as Avery, Resnick and Zeckhauser ([2]). The principal issue is that feedback information is a public good: its production incurs a (small) cost for the rater but, once produced, it can be freely consumed by everybody in the community. Therefore, community members have an incentive to free ride, consuming everybody else’s feedback while submitting none of their own.

We begin by arguing that, if everybody rates truthfully, buyers are better off than if everybody rates randomly (or does not rate at all). If all buyers rate randomly then a seller’s expected value function

$$V = \sum_{i=0}^{\infty} \delta^i \cdot G(q_i^p) = \sum_{i=0}^{\infty} \delta^i \cdot \{G(q_i^d) + \frac{1-\delta}{\delta} \sum_{j=0}^{i-1} [G(q_j^r) - G(q_j^d)]\}$$

$$= \sum_{i=0}^{\infty} \delta^i \cdot \{G(q_i^d) - \frac{1-\delta}{\delta} \sum_{j=0}^{i-1} G(q_j^d)\} + \sum_{i=0}^{\infty} \delta^i \cdot \frac{1-\delta}{\delta} \sum_{j=0}^{i-1} G(q_j^r)$$

$$= \sum_{i=0}^{\infty} \delta^i \cdot \{G(q_i^d) - \frac{1-\delta}{\delta} \sum_{j=0}^{i-1} G(q_j^d)\} + \{\text{Random value}\}$$

becomes independent of the seller’s true item qualities and depends only on what the seller declares. Knowing this, rational buyers will ignore the signal q^p published by the system and will always bid an amount equal to their expected utility averaged

over all product qualities of all sellers in the marketplace. $\bar{G} = \int_q p(q) \cdot H(q) \cdot dq$.

Using arguments similar to those of Section 2, it can be shown that this situation will gradually drive sellers with higher average product qualities (and therefore higher average cost) out of the market and eventually will make everybody except sellers of

lowest quality/cost leave the market. Buyers will therefore lose all surplus associated with purchasing from those vendors.

Having established that truthful ratings benefit the community as a whole, I will now propose a simple mechanism that is designed to prevent individual buyer deviations from the community rating norms. I am making the assumption that, because the number of buyers and sellers in the marketplace is very large, each buyer transacts with a given seller at most once in her lifetime. Therefore, buyers gain nothing from strategically manipulating the feedback profile of a seller (with whom they will never transact)⁷. The objective of the mechanism is to simply make sure that (a) buyers take the “pain” of visiting the feedback submission screen and (b) buyers submit ratings, such that the mean and variance between the buyer’s and the seller’s perception of quality are consistent across the entire buyer community.

The principal idea is to levy a periodic membership fee F_b from buyers and then offer buyers periodic rebates whose amount is contingent on the buyer’s rating behavior.

The following are the rules of rebate calculation:

1. Buyers receive no rebate if they have failed to submit feedback regarding any transactions they have completed in the last period. This rule provides incentives for buyer participation in the feedback mechanism.
2. Buyers receive no rebate if the sample variance of the quantity $q^r - q^p$ is equal to zero in the past period. This rule is designed to prevent buyers from simply reporting perceived qualities equal to published qualities (plus a constant).
3. Let $d = q^r - q^d$ be the difference between an item’s reported and declared quality, \bar{d}_j be the sample mean of differences corresponding to buyer j ’s ratings (of multiple sellers) in the last period, D_j be the sample variance of the same and D be the sample variance of d across all buyers. Buyer j then receives a rebate equal to:

$$\max\{0, F_b \cdot (\min\{\frac{D}{D_j}, \frac{D_j}{D}\} - K \cdot |\bar{d}_j|)\}$$

where K is a constant. The above formula is maximized when $D = D_j$ and $\bar{d}_j = 0$ and penalizes buyers either for deviating from the community-wide

⁷ This assumption is somewhat difficult to reconcile with my previous assumption that each seller is a monopolist (i.e., sells goods for which the demand is independent from that for the goods of other sellers). One way to interpret the assumption of one-shot buyers in such a setting is by assuming that a buyer only buys a good of a given type (e.g. a used car) from a given online marketplace (e.g. eBay) once in her life. Clearly, this is not a very satisfactory interpretation. If we assume that repeat purchase from the same seller is a possibility, then buyers have an incentive to intentionally under-report quality in order to secure lower prices in subsequent rounds. Left to itself, this behavior will eventually drive sellers out of the market. Further research is needed to develop mechanisms that prevent such behavior.

quality perception error variance, or for submitting ratings that, on the average, are biased relative to the sellers' declaration of quality.

We do not claim that the above procedure is completely foolproof. For example, one way in which a buyer can earn high rebates without disclosing her true perception of quality is by submitting ratings equal to $q^r = q^p + \eta$, where η is a normally distributed random variable with mean zero and variance equal to the published variance for that seller. However, given that buyers do not benefit from strategically gaming the feedback mechanism, the author believes that developing a truly foolproof feedback elicitation mechanism is not a priority in the setting of this paper. What is needed is a mechanism that induces "lazy" buyers to spend the decidedly small effort required to navigate to an online marketplace's feedback submission screen and submit sensible (as opposed to completely random) feedback. The simple set of rules presented in this section satisfies those requirements.

5 Endgame Considerations

This section briefly discusses the implications of relaxing the assumption of infinitely lived buyers and sellers and proposes extensions to the mechanism that address the resulting considerations. Because of space limitations, the detailed proofs are omitted.

Sellers with Finite Horizons

In real-life, nobody lives forever. Through the use of the discount factor δ , the preceding analysis does incorporate the possibility that sellers may exit the market. However, it assumes that the decision to exit is based on a Bernoulli trial performed at the end of each cycle. In reality, sellers usually know when they will leave the market well in advance, while buyers ignore that information. Sellers, therefore, will try to take advantage of this information asymmetry by cheating buyers in the last few rounds of their stay on the market, accumulating negative goodwill that will never be repaid back to the community.

One simple, and effective, way of addressing this problem is to extend the mechanism by charging sellers a fee F_0 upon entry in the marketplace. That fee serves as a bond: when a seller exits the marketplace, he is paid back that fee (plus interest) minus the current amount of goodwill owed to the marketplace.

More specifically, the required extensions to the mechanism have as follows:

1. Upon entry to the marketplace, a new seller pays an amount F_0 to the center. The seller can elect to pay any amount F_0 greater than the maximum revenue G_0 of a single transaction.
2. At the end of the i th transaction, the system compounds the bond with interest ($F_{i+1} = F_i / \delta$) and calculates the difference $\Delta_{i+1} = F_{i+1} - g_{i+1}$. If, at any point,

$\Delta_{i+1} < G_0$, the seller is expelled from the marketplace (but may re-enter, by paying a new fee).

3. If a seller voluntarily exits the marketplace at the end of the i th transaction, he is repayed back an amount equal to $\Delta_{i+1} = F_{i+1} - g_{i+1}$.

Buyers with Finite Horizons

Buyers can also exit the market at any time. However, the issues related to buyer “misbehavior” (in the sense of reduced participation in the feedback mechanism or untruthful ratings) right before exit are less severe because the mechanism already charges buyers a periodic fee, part of which is refunded at the end of the period on the basis of buyer rating “performance”. That fee thus acts as a bond that gives buyers incentives to conform to rating norms up until the end of their stay on the market.

6 Concluding Remarks

Online reputation mechanisms are emerging as a promising alternative to more traditional trust building mechanisms in electronic markets. The appeal of reputation mechanisms is that, when they work, they facilitate cooperation without the need for additional costly enforcement institutions. They have, therefore, the potential of providing more economically efficient solutions in a wide range of adverse selection and moral hazard settings.

This paper has studied one such setting, namely marketplaces where sellers sell heterogeneous goods of variable quality. In such settings, simple feedback aggregation mechanisms fail to facilitate efficient transactions between buyers and sellers. This paper presented “Goodwill Hunting” (GWH), a novel feedback management mechanism that succeeds in facilitating efficient transactions in such settings. The GWH mechanism takes a more active stance than most current feedback mechanisms, in that it adjusts the information declared by sellers about the qualities of their products before publishing it to the buyers in ways that provide sellers with incentives to truthfully declare these qualities.

The initial results reported in this paper can be extended in a number of directions. First, this paper assumes that sellers are monopolists. This is not a realistic assumption, especially in large-scale electronic marketplaces, such as eBay. It is important to study the behavior of the GWH mechanism in the presence of multiple competing sellers. Second, I have assumed that sellers do not control the costs and sequencing of the items they sell. Although this assumption can be justified in a number of settings (see discussion in Section 3), it would be interesting to consider the implications of removing it. Third, I have made the simplifying assumption that each buyer buys from a given seller only once. This removes the need to consider strategic misreporting of quality by buyers, but is clearly not very realistic, especially under the assumption of monopolist sellers. It is therefore important to more carefully study to what extent feedback mechanisms are sensitive to strategic feedback manipulation from the part of buyers.

Acknowledgments. Many thanks to Xiaoquan Zhang for his contributions in early phases of this work. This material is based upon work supported by the National Science Foundation under CAREER Grant No. 9984147.

References

1. Akerlof, G. (1970) The market for ‘lemons’: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 84, pp. 488-500.
2. Avery, C., Resnick, P. and Zeckhauser, R. (1999) The Market for Evaluations. *American Economics Review*, 89, 3, pp. 564-584.
3. Bajari, P. and A. Hortascu (2000). Winner’s Curse, Reserve Prices and Endogenous Entry: Empirical Insights From eBay Auctions. Working paper, December 2000.
4. Dellarocas, C. (2002) The Digitization of Word of Mouth: Promise and Challenges of Online Reputation Systems. Working Paper.
5. Dewan, S. and Hsu, V. (2001) Trust in Electronic Markets: Price Discovery in Generalist Versus Specialty Online Auctions. Working Paper. January 31, 2001.
6. Houser, D. and Wooders, J. (2000) Reputation in Auctions: Theory and Evidence from eBay. Working Paper, University of Arizona, 2000.
7. Kollock, P. (1999) The Production of Trust in Online Markets. In *Advances in Group Processes* (Vol. 16), eds. E.J. Lawler, M. Macy, S. Thyne, and H.A. Walker, Greenwich, CT: JAI Press.
8. Lucking-Reiley, D., Bryan, D., Prasad, N. and Reeves, D. (2000). Pennies from eBay: The Determinants of Price in Online Auctions, Working Paper, Vanderbilt University, 2000.
9. Resnick, P., Zeckhauser, R., Friedman, E., Kuwabara, K. (2000) Reputation Systems. *Communications of the ACM*, Vol. 43, (12), December 2000, pp. 45-48.
10. Resnick, P. and Zeckhauser, R. (2001) Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay’s Reputation System. Working Paper for the NBER workshop on empirical studies of electronic commerce. January 2001.
11. Rothschild, M. and Stiglitz, J. E. (1970) Increasing Risk I: A Definition. *Journal of Economic Theory*, 2 (3), September 1970, pp. 225-243.

Appendix

Lemma 1: For any sequence of item qualities $\{q_i^s\}$, $i = 0, \dots, \infty$ and any $g \in \mathbf{R}$,

$$V(\{q_i^s\}, g) = V(\{q_i^s\}, 0) + g$$

Proof: Let $\{q_i^s\}$, $i = 0, \dots, \infty$ be an arbitrary sequence of item qualities as perceived by the seller. Furthermore, let $\{q_i^d(q_i^s, g_i)\}$ be the corresponding sequence of declared qualities that maximizes the Bellman equation (2) at round i . In other words, $q_i^d(q_i^s, g_i)$ maximizes $V(q_i^s, g_i)$. Then, the value function

$$V(\{q_i^s\}, g) = \sum_{i=0}^{\infty} \delta^i \cdot G(q_i^p(q_i^d(q_i^s, g_i), g_i)) \text{ where } G(q_i^p(\cdot)) \text{ is the auction revenue}$$

from the i th transaction if the mechanism publishes quality q_i^p . Furthermore, from Figure 1 it is straightforward to show by induction that:

$$g_i = g + \frac{1}{\delta} \sum_{k=0}^{i-1} \Delta G_k \quad \text{where } \Delta G_k = G(q_k^r) - G(q_k^d(q_k^s, g_k))$$

where q_k^r cannot depend on g_k because the current amount of a seller's goodwill is privately known by the mechanism only.

By successive substitutions of quantities from Figure 1, we get:

$$\begin{aligned} V(\{q_i^s\}, g) &= \sum_{i=0}^{\infty} \delta^i \cdot [G(q_i^d(q_i^s, g_i)) + (1 - \delta) \cdot g_i] \\ &= \sum_{i=0}^{\infty} \delta^i \cdot G(q_i^d(q_i^s, g_i)) + (1 - \delta) \cdot \sum_{i=0}^{\infty} \delta^i \cdot g_i \\ &= \sum_{i=0}^{\infty} \delta^i \cdot [G(q_i^r) - \Delta G_i] + (1 - \delta) \cdot \sum_{i=0}^{\infty} \delta^i \cdot \left[\frac{1}{\delta} \sum_{k=0}^{i-1} \Delta G_k + g \right] = \\ &= \sum_{i=0}^{\infty} \delta^i \cdot G(q_i^r) - \sum_{i=0}^{\infty} \delta^i \cdot \Delta G_i + (1 - \delta) \cdot \sum_{i=0}^{\infty} \delta^i \cdot g + (1 - \delta) \cdot \sum_{i=0}^{\infty} \delta^i \cdot \left[\sum_{j=0}^i \Delta G_j \right] = \\ &= \sum_{i=0}^{\infty} \delta^i \cdot G(q_i^r) - \sum_{i=0}^{\infty} \delta^i \cdot \Delta G_i + (1 - \delta) \cdot \frac{g}{1 - \delta} + (1 - \delta) \cdot \sum_{i=0}^{\infty} \delta^i \cdot \left[\sum_{j=0}^{\infty} \delta^j \cdot \Delta G_i \right] = \\ &= \sum_{i=0}^{\infty} \delta^i \cdot G(q_i^r) + g - \sum_{i=0}^{\infty} \delta^i \cdot \Delta G_i + (1 - \delta) \cdot \sum_{i=0}^{\infty} \delta^i \cdot \left[\frac{\Delta G_i}{1 - \delta} \right] = \\ &= \sum_{i=0}^{\infty} \delta^i \cdot G(q_i^r) + g \end{aligned}$$

For $g = 0$ we can similarly get:

$$V(\{q_i^s\}, 0) = \sum_{i=0}^{\infty} \delta^i \cdot G(q_i^r)$$

which gives $V(\{q_i^s\}, g) = V(\{q_i^s\}, 0) + g$ since $\sum_{i=0}^{\infty} \delta^i \cdot G(q_i^r)$ does not depend on g .