Eliciting Honest Feedback

Empirical Results for Quadratic and Logarithmic Scoring Rules

Alexander Sheh and Kenny Fong
Division of Engineering and Applied Sciences
Harvard University, Cambridge, MA 02138
{sheh,kcfong}@fas.harvard.edu

May 14, 2003

Abstract

Electronic markets face the challenge of motivating honest feedback from strangers who have limited interaction with each other. Miller, Resnick and Zeckhauser [3] propose a mechanism for eliciting truthful evaluations by providing monetary incentives. However, their paper does not present valuable empirical data that would help one better understand their system. For example, knowing how noise and deception affect a buyer's feedback payment gives insight into the mechanism's robustness to poor evaluators and prevaricators. Payment distribution data also allow us to assess the strengths and weaknesses of various scoring rules. In this experimental study, we simulate auctions with sequential interaction involving 500 buyers, 5 discrete signal types, and 5 discrete seller types. We generate payment distribution data across 10,000 auctions using the quadratic and logarithmic scoring rules. Additionally we explore the relationship between payment distributions and the degree of buyer noise and dishonesty.

1 Introduction

In the real world, people interact with each other over a long period of time and establish a history of interactions. This history provides information about the ability and character of individuals; i.e., past interactions establish your reputation. In expectation of reciprocity and retaliation, it is in an individual's interest to establish a good reputation. This incentive for good behavior is referred to as the "shadow of the future" [1]. If one is known for providing accurate and trustworthy information, then it is more likely that he will receive similarly authentic information. On the other hand, if an individual has a reputation for lying or poor judgement, then it is unlikely that others will be willing to provide him with valuable information. There is even the possibility of deliberate sabotage. In other words, the information that one provides is commensurate with the information that one receives. More simply put, if you scratch my back, then I will scratch yours.

It is relatively easy to build trust between individuals in long-term relationships, and the above discussion assumes this. How exactly do we build trust among strangers? This is an important question for electronic markets, since it can be difficult to know an individual's past history, prospects of future interactions may be small, and interactions may be cloaked in anonymity. Reputation systems strive to establish the shadow of the future in this context [4].

To be effective, reputation systems require three properties. First, individuals must have longevity in order to establish expectation of future interactions. Secondly, feedback of current interaction must be captured and distributed. Finally, feedback must be used to guide trust decisions. Miller, Resnick and Zeckhauser [3] propose a method for motivating truthful feedback. The paper describes a mechanism that provides monetary incentives for buyers to accurately evaluate sellers. The center gathers and distributes feedback information, and issues associated feedback payments. Higher payments are awarded for "better" evaluations. The quality of an evaluation is defined as the degree to which it agrees with those of others. More specifically, a buyer is paid according to how well his seller evaluation predicts the subsequent buyer's evaluation.

Three strictly proper scoring rules are proposed for calculating feedback payments: quadratic, spherical, and logarithmic. Under these rules, an agent maximizes his expected wealth when truthfully revealing his private information. It is similarly in his interest to put sufficient effort into the evaluation process, as his payment will otherwise be diminished. In the paper, Miller et al. present an effective system for eliciting honest feedback via feedback payments. However, there are still unanswered questions: What is the distribution of these payments - on average how large are the payments and how much do they vary? How exactly does noise affect a buyer's feedback payments? How exactly does lying affect a buyer's feedback payments? How is a buyer's payment distribution affected by the degree to which his information is noisy or false? These are important questions. Knowing how noise and deception affect a buyer's feedback payment can help us determine how robust the mechanism is to poor evaluators and prevaricators. Payment distribution data also allow us to assess the strengths and weaknesses of various scoring rules.

In this paper, we present and discuss our experimental results obtained from simulating auctions with sequential interaction involving 500 buyers, 5 discrete buyer signal types, and 5 discrete seller types. We generate payment distribution data across 10,000 auctions using the quadratic and logarithmic scoring rules. Furthermore, we explore the relationship between payment distributions and the degree of buyer noise and dishonesty. Our experimental results show that both quadratic and logarithmic payments can induce evaluation effort and motivate honest reporting. Both rules are equally effective for the former problem, while logarithmic payments are slightly better at promoting truth-telling in the case of extreme lying strategies. Logarithmic payments result in faster simulations runs, while quadratic payments are more consistent for all buyer types.

This paper is organized as follows. Section 2 describes Miller et al.'s mechanism for eliciting honest feedback using money incentives. Section 3 describes our experiment setting and the implementation. We present our experimental results in Section 4 and give a discussion on the data in Section 5. Finally, Section 6 concludes.

2 Mechanism

In this section, we give a concise introduction to the mechanism for eliciting honest feedback proposed by Miller, Resnick and Zeckhauser. Readers are advised to consult [3] for a detailed description of the mechanism.

We consider an environment where a number of buyers engage with a seller and then rate her product for quality. We make the following assumptions in our model:

- The quality of the seller does not vary over time.
- All buyers attach the same value to the seller's quality.
- Quality is observed with some error, so a buyer's perception of the seller's type may be deviated from the seller's type.
- Buyers are risk neutral and seek to maximize expected utility.
- Buyers' perception of the seller's type is private information.

Miller et al. develop a payment system that simultaneously elicits honest buyer feedback as a strict Nash equilibrium and also breaks even. To make sure that the payment system breaks even in the long run, the mechanism needs a common processing facility called the center, whose job is to ask the buyers for their ratings on the seller's quality, then to reward or penalize each buyer on the basis of his rating, and most importantly to play the role of a bank to make transfers to the buyers, ensuring that the mechanism at least breaks even in the long run. Note that the center does not know the buyers' perception of the seller's type, so it rewards or penalizes individual buyers based on the announced ratings of the other buyers only.

Section 2.1 gives a formal description of the mechanism and its underlying assumptions. Section 2.2 defines a strictly proper scoring rule, a payoff structure the mechanism is based on that can induce truthful revelation of ratings from the buyers. In Section 2.3, we show how to construct self-financing transfers that make truthful reporting of ratings a Nash equilibrium in our mechanism. Finally, in Section 2.4, we show how our mechanism and payment rules can be adapted to situations where buyers interact sequentially, upon which our experimental study described in Section 3 is built.

2.1 Formulation

We refer to a seller's quality as her type. We refer to a buyer's perception of a seller's type as his signal. Suppose there are a finite number of seller types, indexed by $t = 1, \ldots, T$. Let $p(t)$ be the commonly held prior probability assigned to the seller being type t. Assume that $p(t) > 0$ for all t.

Let I be the set of buyers, where $|I| \geq 3$. Each buyer, judging from his own transaction, privately observes a signal of the seller's type. Suppose that buyers' signals are independent and identically distributed given the seller's type. Let S^i be the random signal received by buyer i. Let $S = \{s_1, \ldots, s_M\}$ be the set of possible signals, and let $f(s_m|t) = \Pr(S^i = s_m|t)$, where $f(s_m|t) \geq 0$ for all s_m and t. Let $s^i \in S$ denote a generic realization of S^i. We will use s^i_m to denote the event $S^i = s_m$.

In the mechanism proposed by Miller et al., each buyer announces his signal and the center makes a transfer to the buyer that may depend on all buyers' announcements. Under the formulation described above, truthful reporting is a strict Nash equilibrium in the mechanism. That is, for each buyer i, conditional on receiving signal s_m, announcing the real signal s_m uniquely maximizes the buyer's expected transfer, where the expectation is taken with respect to the distribution of all other buyers' signals conditional on $S^i = s_m$. The analysis begins by noting that although S^i and S^j are independent conditional on the seller's type t, they are generally dependent because they are drawn from the same distribution f with parameter t. One way to conceptualize this dependence is to apply the notion of stochastic relevance.

Definition 2.1 *Let S^i and S^j be random variables. S^i is stochastically relevant for S^j if and only if the distribution $g(S^j|S^i)$ of S^j conditional on S^i is different for different realizations of S^i.*

Denote $g(s^j|s^i) = \Pr(S^j = s^j|S^i = s^i)$. Miller et al. [3] show that for any two distinct buyers i and j, S^i is stochastically relevant for S^j in their mechanism. That is, for any two distinct realizations s^i and $\hat{s}^i$ of S^i, there exists some realization $\hat{s}^j$ of S^j such that $g(s^j|s^i) \neq g(s^j|\hat{s}^i)$.

Now, consider two buyers, i and j. If S^i is stochastically relevant for S^j, then buyer i's signal provides information about the distribution of buyer j's signal. Thus, if it were known that buyer j will truthfully report his signal, then the problem of eliciting buyer i's information is reduced to eliciting his belief about the distribution of j's signal. The elicitation of such belief is exactly the problem considered in statistical decision theory on strictly proper scoring rules.

2.2 Scoring rules

Definition 2.2 *A scoring rule is a function $R(s^j|s^i)$ that, for each possible announcement s^i of S^i, assigns a score to each possible realization s^j of S^j. A scoring rule is strictly proper if the expected score is uniquely maximized at the true value of the parameter S^i.*

In this context, a convenient interpretation is that the scoring rule specifies the payment made by the center to buyer i following each realization S^j. It is strictly proper if buyer i paid according to the scoring rule uniquely maximizes his expected utility by honestly reporting his true signal s^i.

There are a number of strictly proper scoring rules. The three best known are [2]:

- Quadratic scoring rule: $R(s^j_n|s^i_m) = 2g(s^j_n|s^i_m) - \sum_{h=1}^{M} g(s^j_h|s^i_m)^2$,
- Spherical scoring rule: $R(s^j_n|s^i_m) = \frac{g(s^j_n|s^i_m)}{(\sum_{h=1}^{M} g(s^j_h|s^i_m)^2)^{\frac{1}{2}}}$,
- Logarithmic scoring rule: $R(s^j_n|s^i_m) = \ln g(s^j_n|s^i_m)$.

Since the three scoring rules are strictly proper, they can be used to induce truthful revelation by buyer i as long as his private signal S^i is stochastically relevant for some other publicly available signal S^j. However, in the formulation of our problem, each buyer's signal is private information. Therefore, we can only check players' announcements against other players' announcements, not their actual signals.

In this paper, we are only interested in the quadratic and logarithmic scoring rules. The logarithmic and quadratic scoring rules each have advantages. The logarithmic scoring rule is attractive for its simplicity and intuitive appeal, and also for the fact that the payoff assigned to an outcome depends on the probability of that outcome only; the probabilities of outcomes that do not occur do not factor into the payoff. In other words, the payment to buyer i according to the logarithmic scoring rule depends on buyer j's real signal only, not other possible signals for buyer j that do not occur. It has been shown [5] that when there are more than 2 outcomes (i.e. the number of signal types for buyer j is greater than 2), the logarithmic scoring rule (up to a positive affine transformation) is the only strictly proper scoring rule with this property, which has been termed "relevance".

However, the logarithmic score is criticized for how it deals with low-probability events. Small changes in assessments of small probabilities can translate into very large changes in the score (i.e. if $0 < g(s^j_{m_1}), g(s^j_{m_2}) < 1$, then $|\ln g(s^j_{m_1}|s^i_n) - \ln g(s^j_{m_2}|s^i_n)|$ is exponential in $|g(s^j_{m_1}) - g(s^j_{m_2})|$), and in order for the logarithmic score to be strictly proper such changes must be properly assessed by the decision maker. Furthermore, when the distribution $g(s^j_m|s^i_n)$ in question involves low-probability events, the necessary range of payments may become very large (i.e. if there exists a signal s^j_m such that $g(s^j_m|s^i_n) \to 0$, then the corresponding payment $\ln g(s^j_m|s^i_n)$ to buyer i will approach $-\infty$, which is a large penalty). These arguments support quadratic scoring rule as an alternative. The quadratic score has the advantage of not needing to rely on threats of large penalties to induce truthful revelation. However, quadratic scores are more complex, and have the unappealing property that the payoff assigned to an outcome depends on the likelihood of all outcomes (i.e. the payment to buyer i according to the quadratic scoring rule depends on $g(s^j_h|s^i_m)$ for all possible signals s^j_h of buyer j).

2.3 Payment rules

We now turn to showing how to construct self-financing (i.e. balanced) transfers that make truthful reporting a Nash equilibrium. Recall that for any two distinct buyers i and j in the mechanism depicted in Section 2.1, S^i is stochastically relevant for S^j. Therefore, paying buyer i based on a strictly proper scoring rule, such as the log-likelihood of the reported realization of buyer j's signal, induces buyer i to truthfully announce his signal, assuming buyer j announces truthfully. To make the transfer balanced, this transfer should be paid by another buyer. However, since the transfer to buyer i depends on buyer j's announcement, buyer j should not pay the transfer to buyer i. Instead, to balance the transfers, we let a third buyer k to pay to buyer i according to the scoring rule. Miller et al. [3] show that these transfers are balanced and under which truthful reporting is a strict Nash equilibrium of the reporting game.

Formally speaking, let R be a strictly proper scoring rule, and a^i be the announced signal of buyer i. For each buyer i, our mechanism chooses another buyer $r(i)$ whose announcement i is asked to predict. If the transfer to i is $R(a^{r(i)}|a^i)$ and is paid by a third player k different from i and $r(i)$ (note that each buyer is paid once and pays once), then these transfers are balanced and induce truthful reporting. Let $b(i)$ be the buyer to whom i pays transfers. Then the net transfer to buyer i is the payment he receives from k minus the transfer he pays to $b(i)$; i.e.,

$$\tau_i(a^i) = R(a^{r(i)}|a^i) - R(a^{r(b(i))}|a^{b(i)}).$$

Although the balanced transfer rule described above induces truthful reporting, it may not necessarily assure ex-post voluntary participation, because the net transfer of an (honest) buyer $\tau_i(s^i)$ could be negative. For example, if the logarithmic scoring rule is used and $0 < g(s^j_m|s^i_n) < 1$, then $R(s^j_m|s^i_n) < 0$. To assure ex-post voluntary participation, we can compensate the net loss of any buyer with money collected in advance. Let $q = \min_{s_m, s_n \in S} R(s_m|s_n)$. Suppose that $q < 0$ (in which case $-q > 0$ is the greatest possible net *loss* of a buyer). If $-q$ is collected in advance from each buyer as a participation fee, and is added to the transfer to each buyer, then each buyer will receive non-negative payments. Therefore, in the worst case, if it happens that a buyer suffers from this greatest possible net loss $-q$, his net payment will be zero after the center adds this amount to the payment to this buyer.

2.4 Sequential interaction

In the reporting game described above, buyers simultaneously report their signals to the seller. The mechanism adapts readily to situations where buyers interact sequentially. In this setting, an infinite sequence of buyers interacts with the seller, and buyer i announces his signal that reflects his prediction on buyer $i+1$'s signal, for whom buyer i's signal is stochastically relevant. Incentives for buyer i can be then provided by a log-likelihood payment (or any other strictly proper scoring rule) made by buyer $i+2$ based on the signal reported by $i+1$.

Suppose the seller interacts with an infinite sequence of buyers, indexed by $i = 1, 2, \ldots$. Initially, the commonly held prior distribution for the seller's type is given by $p(t)$. Let $p_1(t|s^1)$ denote the posterior distribution when buyer 1 receives signal s^1. That is,

$$p_1(t|s^1) = \frac{f(s^1|t)p(t)}{\Pr(s^1)},$$

where $\Pr(s^1) = \sum_{t=1}^{T} f(s^1|t)p(t)$. Buyer 1's belief about the probability that $S^2 = s^2$ is given by

$$g(s^2|s^1) = \sum_{t=1}^{T} \frac{f(s^2|t)f(s^1|t)p_1(t|s^1)}{Z_1}, \tag{1}$$

where the normalizing factor [1] is $Z_1 = \sum_{t=1}^{T} f(s^1|t)p_1(t|s^1)$.

Given the distribution specified in (1), buyer 1 can be induced to truthfully reveal s^1 using a strictly proper scoring rule. Buyer 1 is asked to predict the distribution of buyer 2's announcements. Payments to buyer 1 can then be made by buyer 3 after buyer 2 announces.

By recursively updating beliefs about the seller's type using Bayes' rule, each buyer's announcement feeds into the information used by subsequent buyers. We have

$$p_i(t|s^1, \ldots, s^i) = \frac{f(s^i|t)p_{i-1}(t|s^1, \ldots, s^{i-1})}{\Pr(s^i)}, \tag{2}$$

where $\Pr(s^i) = \sum_{t=1}^{T} f(s^i|t)p_{i-1}(t|s^1, \ldots, s^{i-1})$. Hence

$$g(s^{i+1}|s^i) = \sum_{t=1}^{T} \frac{f(s^{i+1}|t)f(s^i|t)p_i(t|s^1, \ldots, s^i)}{Z_i}, \tag{3}$$

where the normalizing factor [2] is $Z_i = \sum_{t=1}^{T} f(s^i|t)p_i(t|s^1, \ldots, s^i)$.

The sequential interaction reporting game proceeds as follows. Beginning with buyer 1, each buyer i observes s^i and is asked to predict the distribution of the signal of buyer $i+1$. Transfers constructed according to a strictly proper scoring rule, such as the logarithmic rule $R(s_n^j|s_m^i) = \ln g(s_n^j|s_m^i)$, using the conditional distribution g specified in (3) elicit truthful announcement from buyer i. This announcement then becomes common knowledge and is used to update beliefs about the seller according to (2). Incentives to buyer $i+1$ are then constructed using a strictly proper scoring rule that incorporates these updated beliefs. Transfers to buyer i are paid by buyer $i+2$ after buyer $i+1$ announces.

3 Experiments

In this experimental study, we simulate auctions with sequential interaction involving 500 buyers, 5 discrete signal types, and 5 discrete seller types. Payment distributions are calculated across 10,000 auctions, for the quadratic and logarithmic scoring rules. Buyers are assumed to have one of the following dispositions: truthful, noisy with respect to additive Gaussian noise, or dishonest by a multiplicative factor of their true signal. By holding all but one of the buyers to be truthful, we can comparatively evaluate a noisy buyer's payment distribution with respect to his noise variance and a dishonest buyer's payment distribution as a function of his lying factor. We do not implement reporting rings, as our only concern is to analyze the payment distributions for particular behaviors. This allows us, for empirical purposes, to fix individual roles and ignore the practical truth-incentives provided by rings. In addition, we can make the impact of ignoring the last buyer statistically insignificant, by simulating with a large number of buyers, N. As this number tends toward infinity, the effect of buyer N's absence on the payment distribution becomes vanishingly small.

[1] There is a typo in the formula for $g(s^2|s^1)$ on page 14 of [3], namely, that the normalizing factor (i.e. the denominator in the summation) in the formula should be Z_1 stated above, not $\Pr(s^2|s^1)$ as stated in [3]. For a correct derivation of the formula for $g(s^2|s^1)$, see the proof of Lemma 1 on page 9 of [3]. The normalizing factor Z_1 plays no active role in our experimental study, since it is simply a constant and so has no impact on the payment distributions, in which this paper is interested. It does affect the actual payments, but in real-life applications of scoring rules, the output value of the scoring rule is scaled (by an affine transformation) to obtain the actual payment [6].

[2] There is a typo in the formula for $g(s^{i+1}|s^i)$ on page 14 of [3], namely, that the normalizing factor should be Z_i stated above, not $\Pr(s^{i+1}|s^i)$ as stated in [3]. The normalizing factor Z_i plays no active role in our experimental study. See footnote 1.

All code was written in Matlab. The seller's type is fixed, as assumed by Miller et al. We set the actual type to be $(T+1)/2$, where T is the number of seller types and T is odd; in our simulations this value is 3. The prior probability of a seller type being t, $p(t)$, is governed by a Gaussian with mean equal to the seller's actual type and variance of 1. The buyers' signals are also drawn from a normal distribution with mean equal to the seller's actual type and a variance of 1. We should note that the buyer signals do not have to be drawn from the same distribution as the seller's prior; this occurs because we happen to set the signal variance to also be 1. The probability of a buyer signal s_m given seller type t, $f(s_m|t) = \Pr(S^i = s_m|t)$, is described by a Gaussian with mean equal to seller type t and variance of 1. In effect, the peak of the normal distribution is shifted according to the value of t. Continuous normal distributions are discretized by first drawing a uniformly random continuous event, and then mapping it to the discrete normally distributed event corresponding to the Gaussian bin in which it falls. Values that exceed the smallest or largest non-continuous marker are mapped back toward the edge; discrete perimeter events accumulate the probability mass of extreme continuous events.

Now that we have defined the seller's prior $p(t)$ and the signal probability $f(s_m|t)$ conditioned on the seller type, we can describe the experimental simulations. An auction simulation proceeds as follows. First, 500 signals are drawn from the aforementioned Gaussian and stored in an array, one signal for each buyer/round. Each signal corresponds to a particular buyer's observation of the seller's type. For each round, a signal is "observed" by accessing the corresponding array element. Upon revealing a new signal, the previous buyer's feedback payment is computed. Initially, it is straightforward to calculate $p_1(t|s^1)$ and $g(s^2|s^1)$ using the formulas supplied in Section 2.4. The subsequent values $p_i(t|s^1,\ldots,s^i)$ and $g(s^{i+1}|s^i)$ are slightly more complicated, as they are recursively calculated using previously computed results. In such a way, information from previous buyers' announcements is incorporated into later buyers' beliefs about the seller. Using the $g(s^{i+1}|s^i)$ values, it is simple to determine the quadratic and logarithmic payments. Our experiments use a modified $g'(s^{i+1}|s^i)$ function with the denominator $\Pr(s^i)$ omitted. Since $\Pr(s^i)$ is a normalizing constant, this modification does not affect the resulting payment distributions. After calculating the feedback payments, two transformations are performed. First, the net payments are determined by subtracting the amount a buyer must pay another player from the payment he receives. We assume that only buyers 1 through $N-1$ receive feedback payments, and that buyer $i+2$ pays i, except for buyers $N-2$ and $N-1$ who are paid by buyers 1 and 2 respectively. Secondly, to assure voluntary participation, negative net payments are transformed into positive net payments. In our implementation, the absolute value of the most negative actual payment is added to all payments. Section 2.3 specifies that the added magnitude should be of the theoretically smallest payment, since these refer to bonds that are collected in advance. Since we are only concerned with payment distributions, this slight modification is innocuous. Finally, the payment distribution for each buyer is averaged across 10,000 auctions.

Up until now we have assumed all buyers to be truthful, i.e. they report exactly the signals they observe. However, we are interested in comparing the feedback payment distributions for honest reporting with two other buyer types: poor evaluators and liars. The former are less skillful at determining the seller's type. To simulate this, we add random noise to the drawn buyer signal. This noise is normally distributed around mean zero with variance v. Dishonest buyers, on the other hand, have the ability to evaluate but choose to incorrectly report their observations. Their lying strategy is to scale the observed signal by a factor k. We should point out that additive noise and dishonest scaling are incorporated before the continuous-to-discrete-signal mapping.

4 Results

In our experiments we define 499 honest buyers and fix a single poor evaluator or liar. We then compare the payment distribution of the latter two buyer types with the average payment distribution for the 499 honest reporters. All statistics are further averaged across 10,000 auctions. Ten poor evaluator simulations are run using the following noise variance values: 0.1, 0.3, 0.5, 0.7, 0.9, 1, 2, 3, 4, 5. Similarly, ten liar simulations are conducted with the lying factors 0.1, 0.3, 0.5, 0.7, 0.9, 1, 1.5, 2, 2.5, 3 (a lying factor of 1 means the liar does not lie). Tables 1 and 2 show statistics for the noisy and honest buyers' feedback payments, and the ratio of the noisy and honest payment averages. Tables 3 and 4 display similar information for the lying and honest buyers. Figures 1 through 8 plot the data from Tables 1 through 4 in a more readable fashion.

QUADRATIC PAYMENTS					
noise var	noisy avg	noisy var	honest avg	honest var	noisy/honest ratio
0.1	0.2772	0.0138	0.2798	0.0136	0.990708
0.3	0.2692	0.0136	0.2799	0.0136	0.961772
0.5	0.2659	0.0139	0.2799	0.0136	0.949982
0.7	0.2604	0.0139	0.2799	0.0136	0.930332
0.9	0.2575	0.0138	0.2799	0.0136	0.919971
1	0.2553	0.0138	0.2799	0.0136	0.912111
2	0.2441	0.0139	0.2799	0.0136	0.872097
3	0.2353	0.0131	0.2799	0.0136	0.840657
4	0.2293	0.0132	0.28	0.0136	0.818929
5	0.2254	0.013	0.2799	0.0136	0.805288

Table 1: Statistics for noisy versus honest buyers' payments with quadratic scoring

LOGARITHMIC PAYMENTS					
noise var	noisy avg	noisy var	honest avg	honest var	noisy/honest ratio
0.1	3.7665	1.5197	3.7955	1.4635	0.992359
0.3	3.6536	1.5792	3.7911	1.4677	0.963731
0.5	3.597	1.6719	3.7909	1.4627	0.948851
0.7	3.524	1.7025	3.7899	1.4633	0.92984
0.9	3.4883	1.7212	3.7947	1.4629	0.919256
1	3.4733	1.7712	3.798	1.465	0.914508
2	3.3099	1.8493	3.7985	1.4633	0.87137
3	3.2124	1.8514	3.7935	1.4652	0.846817
4	3.0916	1.8702	3.7967	1.4614	0.814286
5	3.0346	1.8759	3.7973	1.463	0.799147

Table 2: Statistics for noisy versus honest buyers' payments with logarithmic scoring

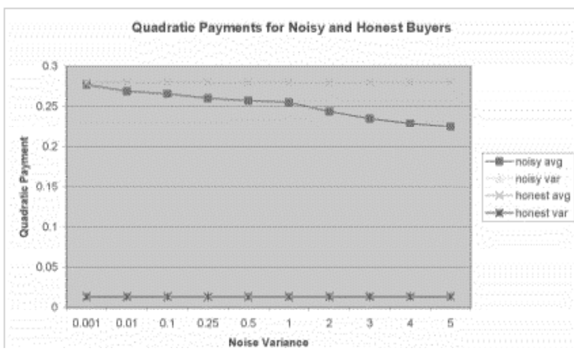

Figure 1: Graph of noisy/honest buyers' payments with different noise variances for quadratic scoring

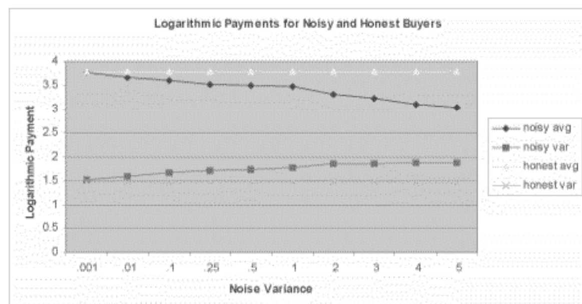

Figure 3: Graph of noisy/honest buyers' payments with different noise variances for logarithmic scoring

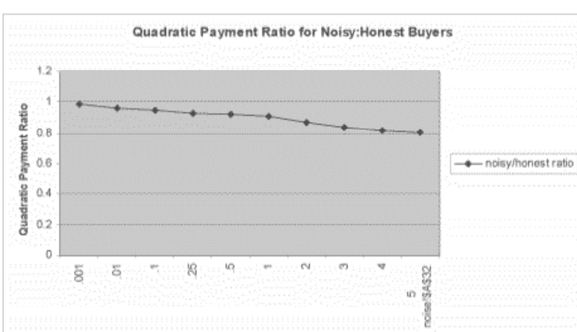

Figure 2: Noise-to-honest payment ratios with different noise variances for quadratic scoring

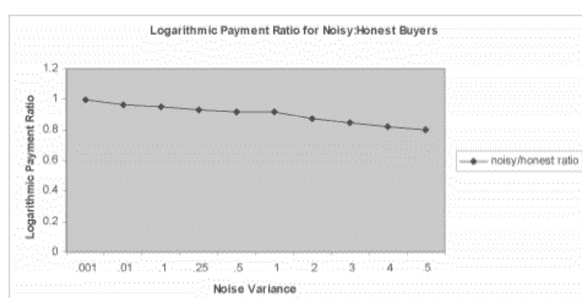

Figure 4: Noise-to-honest payment ratios with different noise variances for logarithmic scoring

QUADRATIC PAYMENTS					
lying factor	lying avg	lying var	honest avg	honest var	lying/honest ratio
0.1	0.1727	0.007	0.28	0.0136	0.616786
0.3	0.1783	0.0075	0.28	0.0136	0.636786
0.5	0.243	0.0111	0.2799	0.0136	0.868167
0.7	0.2535	0.0133	0.2799	0.0136	0.905681
0.9	0.2789	0.0138	0.2798	0.0136	0.996783
1	0.2816	0.0139	0.2798	0.0136	1.006433
1.5	0.2147	0.0134	0.28	0.0136	0.766786
2	0.2007	0.0094	0.28	0.0136	0.716786
2.5	0.1815	0.0087	0.28	0.0136	0.648214
3	0.1813	0.0087	0.2801	0.0136	0.647269

Table 3: Statistics for lying versus honest buyers' payments with quadratic scoring

LOGARITHMIC PAYMENTS					
lying factor	lying avg	lying var	honest avg	honest var	lying/honest ratio
0.1	2.2656	1.1125	3.8053	1.4578	0.59538
0.3	2.3497	1.217	3.8005	1.4619	0.618261
0.5	3.3466	1.6319	3.7919	1.4618	0.882565
0.7	3.4409	1.7337	3.791	1.463	0.90765
0.9	3.7862	1.4956	3.7922	1.4618	0.998418
1	3.8034	1.4874	3.7959	1.4642	1.001976
1.5	2.8218	1.8735	3.802	1.4624	0.742188
2	2.7225	1.5859	3.7947	1.4599	0.717448
2.5	2.382	1.2815	3.8014	1.4569	0.626611
3	2.3779	1.3215	3.8047	1.4592	0.62499

Table 4: Statistics for lying versus honest buyers' payments with logarithmic scoring

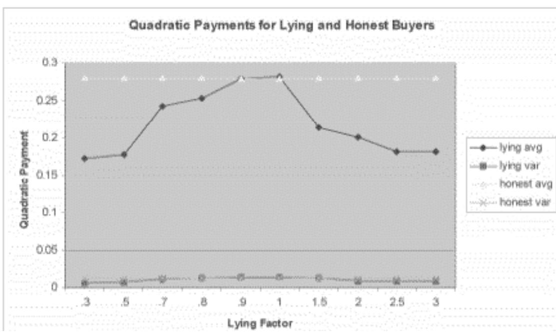

Figure 5: Graph of lying/honest buyers' payments with different lying factors for quadratic scoring

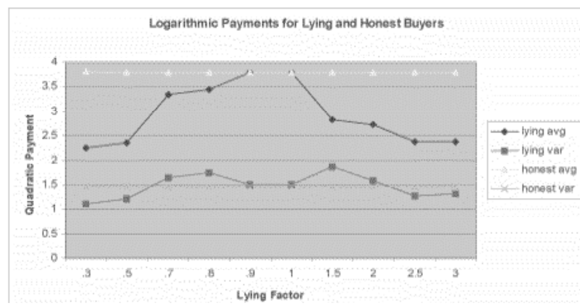

Figure 7: Graph of lying/honest buyers' payments with different lying factors for logarithmic scoring

Figure 6: Lying-to-honest payment ratios with different lying factors for quadratic scoring

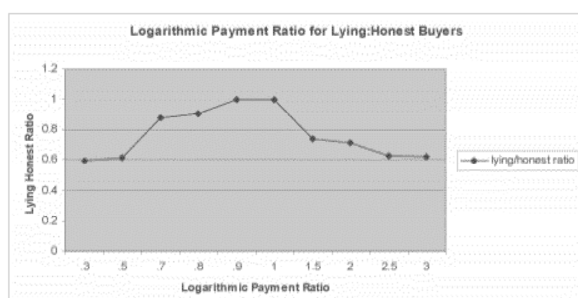

Figure 8: Lying-to-honest payment ratios with different lying factors for logarithmic scoring

5 Discussion

First, we analyze the payment distributions for the noisy and honest buyers. Ideally, for the poor evaluators, we would like to see diminishing feedback payments as the noise variance increases. This would indicate that buyers with low evaluation ability receive less money. Another way of interpreting this is that buyers have an incentive to spend sufficient effort when rating the seller. We would also expect that the noisy buyer payments be less than those for honest reporting. Looking at Table 1 we see that the quadratic payments follow the aforementioned desired behavior. The noisy quadratic payments are inversely proportional to the noise variance, and the noisy to honest payment ratios are less than one. Table 2 indicates the same relationship between noise variance and noisy logarithmic payments. In fact, the noisy to honest payment ratios for logarithmic scoring are almost identical to those for quadratic scoring. This suggests that, holding all other factors constant, both scoring rules are equally effective at inducing evaluation effort. Both payment ratio plots fall along roughly linear curves.

For the liar payment distributions we would hope to see diminishing feedback payments as the lying factor moves away from unity. This would suggest that the more dishonest a buyer is, the lower his payment will be. Therefore, it is in a buyer's interest to abide by a policy of honest reporting. Both quadratic and logarithmic payments follow this desired trend nicely. For each plot in Figures 6 and 8, we see a peak in the middle, with decreasing payments as the lying factor moves away from 1. The rate at which the payment ratios fall is not as uniform as that of the noisy case. This implies that the degree of payment decrease is not merely a function of the amount of increased dishonesty, but is also affected by the previous level of honesty. For both scoring rules, the ratio plots are almost identical, the only difference being slightly smaller logarithmic ratios for the four most extreme lying factors. This makes sense, because in the presence of low-probability events, the range of logarithmic payments can become large. Therfore, in the case of extreme lying strategies, logarithmic scoring is slightly better at inducing truthful reporting.

Looking more closely at Tables 1 through 4, we observe some interesting variance attributes. The first is that the variances of quadratic payments are more than a factor smaller than those for logarithmic payments. This can be explained by examining the scoring rule equations. Logarithmic payments are calculated using the probability of a single outcome, and hence exhibit greater variability. Quadratic payments are more stable because they depend on the sum of all possible outcome likelihoods. This aggregate term does not fluctuate as much as the single outcome likelihood. Reading Table 2, we see that the variance of noisy logarithmic payments increases with the noise variance. This is not surprising, since increasing the added noise variance results in a greater variety of outcomes, which translates into a greater variety of payments. Quadratic payment variances, however, do not exhibit this behavior. For this case, the sum term is relatively stable and dominates any variance introduced by the actual outcome. On the other hand, Table 3 indicates that quadratic payment variances decrease as the lying factor moves away from 1. Observe that, as the multiplicative factor increases or decreases, more signals are mapped to the extreme events. The end result is a decreased variety of outcomes and, hence, payments. The quadratic sum term is affected here, since multiplying the observed seller type more dramatically decreases the payment range, by shifting and truncating the set of possible outcomes. Table 4 shows roughly the same relationship for the logarithmic payment variances and the lying factors. The problematic data point pairs $(0.9, 1)$ and $(2.5, 3)$ can be understood through an analysis of the multiplicative mapping. First, our experiments draw buyer signals from the set $\{1, 2, 3, 4, 5\}$. As we apply the lying factor to the original set, they are transformed into the lying signal sets $\{1, 2, 3, 4, 5\}$ and $\{3, 5\}$, for $(0.9, 1)$ and $(2.5, 3)$, respectively. Both problematic pairs map to the same set of lying signals. Therefore, it is conceivable that these adjacent points would have similar values and by chance deviate from the expected variance relationship. Finally, we observe that the quadratic payments are approximately one factor smaller than the logarithmic payments. As discussed above, the logarithmic scoring rule depends on a single likelihood and as a result has a larger variance. The range of logarithmic payments is also larger, thereby increasing the potential payment sizes. We should also note that the simulations that use logarithmic scoring run much faster than those that implement quadratic scoring. This is because the logarithmic rule only needs a single outcome likelihood, while the quadratic rule requires summing over all possible outcome likelihoods.

6 Conclusion

In [3], Miller et al. propose a mechanism for eliciting truthful evaluations by providing monetary incentives. They discuss some of the theoretical benefits of the various scoring rules - the logarithmic scoring rule is attractive because an outcome's payoff depends on the probability of that outcome only, quadratic scoring is more appropriate for continuous signals with small density, etc. In this experimental study we ran simulations involving poor evaluators and liars for both the quadratic and logarithmic scoring rules. Our data has verified that both quadratic and logarithmic payments can induce evaluation effort and motivate honest reporting. We found that both rules are equally effective for the former problem, while logarithmic payments are slightly better at promoting truth-telling in the case of extreme lying strategies. The quadratic rule has the advantage of smaller payment variances for noisy, lying, and honest buyers. This means that honest buyers are more consistently rewarded for their virtue and aptitude, while poor evaluators and prevaricators are more reliably punished for their incompetence and ignobility. Although quadratic payments are roughly a factor smaller than logarithmic payments, this is not significant. Their ratios are most relevant, as a scaling constant can always be applied to the actual payments. Finally, logarithmic scoring is inherently more intuitive and less computationally demanding, resulting in faster simulation runs. Our empirical results support the claims put forth by Miller et al. They also provide some basis for choosing between the quadratic and logarithmic scoring rules.

References

[1] R. Axelrod. *The Evolution of Cooperation*. Basic Books, New York, 1984.

[2] R. Cooke. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press, New York, 1991.

[3] N. Miller, P. Resnick, and R. Zeckhauser. Eliciting Honest Feedback in Electronic Markets. *Working paper for the SITE02 workshop*, 2002. Available at `http://www.si.umich.edu/~presnick/papers/elicit/`.

[4] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation Systems: Facilitating Trust in Internet Interactions. *Communications of the ACM 43(12): 45-48*, 2000.

[5] E. Shuford, A. Albert, and H. Massengil. Admissible Probability Measurement Procedures. *Psychometrika, 31: 125-145*, 1966.

[6] J. Sonnemans and T. Offerman. Is the Quadratic Scoring Rule Behaviorally Incentive Compatible? *Working paper in CREED, University of Amsterdam*, 2001. Available at `http://www1.fee.uva.nl/creed/pdffiles/qspaper5.pdf`.