# PEDL Research Papers

This research was partly or entirely supported by funding from the research initiative Private Enterprise Development in Low-Income Countries (PEDL), a Foreign, Commonwealth & Development Office (FCDO) funded programme run by the Centre for Economic Policy Research (CEPR).

# Paying for the Truth: The Efficacy of a Peer Prediction Mechanism in the Field

Natalia Rigol (MIT) and Benjamin Roth (MIT)*

April 26, 2016

**Abstract**

We report results from a lab-in-the-field experiment in India in which we test the viability of two kinds of monetary payment rules used to incentivize truth-telling: a novel payment rule that relies on ex-post verification of reports and peer prediction methods (See Prelec, 2004 and Witkowski and Parkes, 2012), which rely only on contemporaneous peer reports. In the experiment, farmers were asked to give reports about their neighbors and were told that these reports would be used to determine cash prizes. We varied whether farmers received incentives for the accuracy of their reports (via the two payment rules) or not. We find that, in the absence of monetary incentives, respondents lie in favor of their family and friends. However, monetary incentives for accuracy improve the quality of reports and both payment rules result in reports of comparable accuracy. This is a reassuring outcome since peer prediction is much easier to implement (though mechanically complex). Importantly, by imposing structure on our data we also find evidence that one peer predictive payment rule, the Robust Bayesian Truth Serum of Witkowski and Parkes (2012), is empirically incentive compatible; respondents maximize their subjective expected utility by reporting truthful answers. Given the broad applicability and the ease of implementation of RBTS, we hope that this experiment will serve as a catalyst to verify its usefulness in other contexts.

*We would like to first thank our team for their tireless work on this project, especially our research manager Sitaram Mukherjee, our project assistant Prasenjit Samanta, and our field manager Sayan Bhattacharjee. We are also very grateful to Rohan Parakh for excellent advice, research assistance, and field management. This research was possible with funding from the Asian Development Bank, PEDL, and Schultz Fund at MIT. We received valuable feedback about this project from Rohini Pande, Benjamin Olken, Abhijit Banerjee, Esther Duflo, Reshmaan Hussam, and Arielle Bernhardt.

# 1 Introduction

It has long been understood that asymmetric information poses a major obstacle to providing services and support to the poor. Principals, such as microfinance institutions (MFIs) or governments, may hope to allocate credit and capital grants to growing or reliable entrepreneurs, to tailor insurance to the risks recipients face, or to target subsidies and welfare programs to marginalized populations. Yet, discerning who does and does not belong in the recipient subgroup is notoriously difficult. Many standard features of microcredit contracts, such as mandatory savings with the MFI, joint liability, and immediate and frequent repayment, are blunt tools to reduce default in the face of an inability to screen new borrowers. But, emerging evidence suggests these tools are not without their costs (see Fischer (2012) and Field, Papp, Pande, and Rigol (2013) for evidence that these contractual features may inhibit efficient investment).

In high income countries, principals may rely on formal information, such as income statements or credit bureau reports, to discern between high and low ability entrepreneurs or between high and low risk borrowers . However, in many low income countries, formal information about the intended targets of services is sparse or non existent. Thus governments, NGOs, and MFIs screening recipients for aid and credit often rely on information embedded in the communities they serve. Relatively little is known, however, about how to reliably and cost-effectively elicit community information. For instance, it is easy to imagine that community members may lie about their peers if they know that their reports will be used to determine the allocation of valuable resources. One obvious method of eliciting truthful reports would be to pay respondents for the accuracy of the information they share. But, such incentive schemes have been slow to catch on in development economics, a trend that Delevande et. al. (2011) attribute to the dearth of payment rules that are both robust to risk preferences and easy to implement. In this paper, we use a field experiment to identify a payment rule that responds to both these needs.

We report experimental evidence from a village in Maharashtra, India that shows that, as a baseline, community members know things about their peers and that they distort their reports when these reports have financial consequences. These two results are necessary to evaluate whether our two payment rules realign incentives for truthfulness: a novel payment rule that is robust to risk aversion and based on ex-post verification, and a class of payment rules known as peer prediction mechanisms that are more complex but highly portable (see e.g. Prelec (2004), and Witkowski and Parkes (2012)). We find that respodnents are responsive to monetary incentives for accuracy. Moreover, comparing our straightforward, easy to understand payment rule based on ex post verification to peer prediction mechanisms we find that the peer prediction mechanisms work surprisingly well in aligning incentives for truthfulness, and maintaining their attractive theoretical properties. Peer prediction mechanisms rely on reports of first order beliefs (beliefs about the likelihood of a given outcome) *and* second order beliefs (beliefs about the distribution of first order beliefs). We generate the first empirical evidence that respondents with low numeracy respond as strongly to peer prediction mechanisms as they do to straightforward payment schemes. Further, we show that one peer prediction mechanism in particular, the Robust Bayesian Truth Serum, is incentive compatible; a non-trivial finding, given the complexity of the payment scheme and its reliance on second order beliefs of community members. Peer prediction mechanisms are extremely easy to implement and appropriate for any setting in which multiple respondents have information about an outcome of interest, so it is our hope that this paper will serve as a catalyst for others to verify and extend their usefulness to other contexts.

The techniques we use to provide monetary incentives for respondents' accuracy are related to the scoring rule and peer prediction literatures. Scoring rules are a class of mechanisms that elicit subjective expectations about future events by paying respondents as a function of their report and the future realization of the event (see e.g. Brier (1950)). We develop a novel scoring rule that is both robust to risk aversion and easy for respondents with low levels of numerical literacy to understand; it requires only that respondents rank the likelihood of events. A number of other authors have devised scoring rules that are robust to risk aversion (see e.g. Karni (2009), Hossain and Okui (2013), and Offerman et. al. (2009)). And Bhattacharya and Roth (2015) propose an alternative rank order scoring rule for respondents with low levels of numerical literacy.

While the scoring rule we develop is easy for respondents to understand, it shares the drawback, common among all scoring rules, that it requires ex-post verification of the outcome of the event in question. In many cases, the outcome is realized after some non-trivial period of time and may be costly to verify. Peer prediction mechanisms bypass this hurdle by paying subjects based only on their own reports and the reports of their peers. However, these mechanisms are extremely complicated– and, therefore, difficult to explain– so their usefulness, while theoretically guaranteed, is an open empirical question. The two peer prediction mechanisms evaluated in this paper are the Bayesian Truth Serum of Prelec (2004), and the Robust Bayesian Truth Serum of Witkowski and Parkes (2012).

The remainder of the paper proceeds as follows. Section 2 describes the setting and our experiment, Section 3 details the payment rules we evaluated, Section 4 outlines our main specifications, Section 5 describes our results, and Section 6 concludes. Tables and a proof of the incentive properties of our novel payment rule are contained in the appendix.

# 2 Description of the Experiment

We conducted a census of the village of Telengaon in Amravati District of Maharashtra, India. During the census survey, we asked respondents whether they were engaged in any household enterprise, including farming and livestock activities. After the completion of the census, the 86 households that reported having a household business were assigned to groups of five persons, based on geographical proximity. All five business owners within a group were then invited to a community hall to participate in our study. An the hall, participants each responded in private to a set of questions about themselves (enumerated below) for which they would later be ranking their group members. At the time at which they provided these answers, however, respondents were unaware of the ranking exercise that would follow.

## 2.1 The Ranking Exercise

After the completion of the private data intake, surveyors presented each respondent with five name cards for each of the five members of the group. They then explained the concept of ranking by putting name cards in order of age. Surveyors and respondents first practiced with a few simple exercises such as ranking group members on houserhold size and height. After surveyors had fully explained rankings, the respondents then proceeded to play the following game. They were told that they would provide 10 rankings for the peers in their group. The subjects of ranking were:

- Level of education
- Health expenditures over the previous 6 months
- Number of digits recalled in a digitspan memory test
- Household income in the past 30 days
- Total value of all household assets
- Risk aversion measured by a standard risk aversion game
- Average monthly profits over the past year
- Whether the household had trouble making a loan repayment in the previous 2 months

- Patience measured by a standard patience game
- Marginal return to a Rs.6000 loan[1]

For every question, the person who had the highest average rank in the group would receive a prize of Rs.20. For certain, randomly selected, questions, they also had the possibility of earning an additional personal incentive for the accuracy of their rankings. The incentives were randomly paid according to one of two payment rules: an ex-post accuracy rule we designed specifically for this study and the Bayesian Truth Serum (BTS) of Prelec (2004). Both rules are described fully in the following section. We normalized the payments such that a respondent would earn on average Rs.10, but could earn more than Rs.30 for a fully accurate ranking. For the ex-post accuracy incentive, the surveyors explained with several examples exactly how each respondent's payment would be calculated and showed, via examples with age and height, that more accurate rankings would receive higher incentive payments. For the BTS incentive, respondents were told that their payments would depend on their own rankings and on the rankings given by their peers, but that the rule was too complicated to explain. For both the BTS and the ex-post accuracy payment rules, respondents were assured, however, that the more accurate their rankings were, the higher payment sums they would receive.

The first question posed to respondents was always to rank their peers on their level of education. The question was never incentivized and was only used for the purpose of demonstrating the Rs.20 prize payments. We randomized the order of the questions and block-randomized (groups of 3 questions) the order of the incentives: no incentive, BTS incentives, or ex-post accuracy incentives. Each respondent therefore answered 3 questions with no incentive, 3 questions with the BTS incentive, and 3 questions with the ex-post accuracy incentive.[2] So one group might answer the first 3 questions without incentives, the second 3 questions with the BTS incentives, and the last 3 questions with the ex-post accuracy incentive while another group would answer the first 3 questions with the ex-post accuracy incentive, the next 3 without incentives, and the last 3 with BTS incentives.

In order to avoid coordination, each respondent worked with a surveyor in private to provide his rankings. Once the rankings for a question were completed for all 5 respondents, they were asked to sit next to one another with their ranking sheets. At this point, therefore, all group members would be able

[1] In this question, we elicit the self-reported marginal return to a Rs.6000 loan. In other work conducted by the authors, we utilize data from "cash drop" experiments cited in the introduction which estimate the true average marginal return of microentrepreneurs. The authors of these experiments, before distributing the grants, asked respondents to estimate what they thought their marginal return would be. We find that these self-reported estimates are good predictors of the true marginal return.

[2] Due to weather and logistical constraints, sometimes the games would have to end early. So we don't have the full 9 questions for all respondents.

to observe each others rankings. We designed the game in this manner to intensify the trade-off that respondents faced: in the absence of any personal incentives (BTS and ex-post accuracy), the optimal strategy each respondent should assume was to rank themselves and their close friends and family highly so as to maximize their chance of winning the Rs. 20 prize. This would be even more salient if their peers in the group could observe each person's rankings at the end of each question. On the other hand, if respondents had a personal incentive, then they would maximize their expected private payoff by telling the truth.

Group members' rankings were recorded and incentive and prize payments were disbursed at the end of each question while respondents were still seated as a group. Therefore respondents also received immediate feedback on how the accuracy of their rankings was being rewarded.

### 2.2 Description of the Sample

Our respondents were exclusively male, as the owners of the largest household business in this village happened to all be men. They had an average 8.6 years of formal education. They had earned nearly Rs.8000 in income in the previous month from all household income-generating activities and had average monthly profits of Rs.4300 in the previous year.

## 3 Description of the Payment Rules

### Ex-Post Accuracy Payment Rule

We designed the ex-post accuracy payment rule to be easy to explain to respondents with a low level of numerical literacy and to provide incentives to be truthful regardless of the respondent's degree of risk aversion. Respondents are asked to rank their five peers on the basis of a past or an uncertain future outcome. The ranking is based on whose outcome is most likely to be highest, second highest etc. A bit of notation is required for the formal description of the payment rule.

Suppose there are $n$ respondents in a group, and each is tasked with ranking the entire group along some metric (e.g. entrepreneurial ability). Let $\{X_1, \ldots, X_n\}$ be a set of non negative, independent random variables representing the eventual default status of each member of the group, to be realized at some point in the future (e.g. business profitability)[3]. Let $F_j(\cdot)$ and $f_j(\cdot)$ be the CDF and PDF of $X_j$. Each respondent $i$ submits a permutation $\sigma_i : \{1, \ldots, n\} \to \{1, \ldots, n\}$ that is one to one and onto,

[3]Note, these random variables need not be realized in the future. It would be equivalent that they were realized in the past but that respondents are unaware of their precise realization.

which represents a ranking of his peers. Then, after each random variable is realized, with realizations $\{x_1, \dots x_n\}$ respondents are paid using the rank order payment rule

$$\sum_{j=1}^{n} j \times x_{\sigma_i(j)}$$

That is, respondent $i$ is paid $n$ times the outcome corresponding to the random variable he ranked most highly, $n-1$ times the outcome of the random variable he ranked second most highly and so on. Consider the following assumption.

**MLRP Assumption:**$\{X_1, \cdots, X_n\}$ can be ordered $X_1 < \cdots < X_n$ where $<$ indicates the monotone likelihood ratio property. That is, for $i < j$ and $a < b$ we have $\frac{f_i(a)}{f_i(b)} > \frac{f_j(a)}{f_j(b)}$.

The MLRP assumption is common in contract theoretic applications. If random variables $X_1$ and $X_2$ satisfy the monotone likelihood ratio property, then for any high realization $a$, the ratio of the likelihood $a$ is acheived under $X_1$ relative to $X_2$ is higher than the same ratio for some low realization $b < a$. We have the following theorem.

**Theorem 1:** Under the MLRP assumption, the rank order payment rule is truthful for any expected utility maximizer whose utility is monotone increasing in wealth.

We relegate a proof of this theorem to the appendix. To offer some intuition for how the payment rule works, Figure 1 below shows an example we utilized to explain the payment rule to respondents during the experiment. Persons A, B, C, D, and E are all in the same group.[4] We ask respondents to rank all five people from eldest to youngest. In the example below, the group is perfectly ordered according to their age.

| Name | Age | Rank |
|---|---|---|
| A | 39 | 5 |
| B | 29 | 4 |
| C | 28 | 3 |
| D | 26 | 2 |
| E | 24 | 1 |
| Total Points | | 471 |

Figure 1: Ex-Post Accuracy Payment Rule

The payment rule works by giving each respondent a monetary payment in proportion to the total number of points they receive for the exercise. The points are calculated by adding five times the outcome (the age, in this case) of the person they assigned to the highest spot, four times the outcome of the

[4]During the pilot, the names and real ages of the members of each group were utilized.

person assigned to the second highest spot and so on. In Figure 1, the total number of points received for the ranking is 471. Suppose the respondent mis-ranks person A and person B as shown in Figure 2. Then her total number of points would be reduced to 441 and her payment would fall proportionately.

| Name | Age | Rank |
|---|---|---|
| A | 39 | 1 |
| B | 29 | 4 |
| C | 28 | 3 |
| D | 26 | 2 |
| E | 24 | 5 |
| Total Points | | 411 |

Figure 2: Ex-Post Accuracy Payment Rule

Thus it was intuitive to respondents that the truthful rank ordering maximizes their expected payout. However this payment rule relies on the verification of ex-post information, which can be costly to collect. More worryingly, if payments rely on self-reported ex-post outcomes, respondents may collude with other members of their group to misreport ex post outcomes so as to maximize their incentive payments.

## Peer Prediction Payment Rules

Peer prediction methods such as the Bayesian Truth Serum of Prelec (2004) and Robust Bayesian Truth Serum of Witkowski and Parkes (2012) create incentives for respondents to tell the truth without relying on ex-post verification of outcomes. Both payment rules require elicitation of first order beliefs (the ranking that the respondent assigns to each of his peers) and second order beliefs (the probability distribution the respondent assigns to each possible ranking his peers may give one another). A formal description of the payoff rules requires additional notation to describe the environment.

Suppose there is a binary state of the world $t \in (h, l)$ (high, low) representing the entrepreneurial quality of a community member. There are $n$ agents each of whom get a binary signal which is informative of the state of the world. That is each agent receives a signal $s \in \{h, l\}$ which may represent what they observe about their peer (e.g. they appear responsible, smart etc). Suppose further that all agents share a common prior about the state of the world such that they all agree on the prior probability of a high state, and they all agree on the distribution of signals conditional on the state. Let $p_h = P(s_j = h|s_i = h)$ be the probability an agent assigns to one of his peers receiving a high signal conditional on himself receiving a high signal, and analogously let $p_l = P(s_j = h|s_i = l)$. We make the following assumption.

**Assumption (Admissability):** We say the common prior is *admissable* if $p_h > p_l$.

In English this implies that the probability that one's peer receives a high signal is higher if the agent himself receives a high signal. Many natural distributions satisfy this weak requirement.

**The Bayesian Truth Serum**

The BTS is implemented as follows. Every agent states their first order belief (their signal), in a report $x_i \in \{0,1\}$ (imagine $x_i = 1$ corresponding to $s_i = h$). Further they report their second order belief $y_i \in [0,1]$ (this is the fraction of the population they believe will report a high signal, $x_k = 1$ ). Let $\bar{x}$ denote the average first order belief, and $\bar{y}$ denote the average second order belief. The BTS payment is

$$\left[(1-x_i)\,log\left(\frac{1-\bar{x}}{1-\bar{y}}\right) + x_i log\left(\frac{\bar{x}}{\bar{y}}\right)\right] + \left[(1-\bar{x})\,log\left(\frac{1-y_i}{1-\bar{x}}\right) + \bar{x}log\left(\frac{y}{\bar{x}}\right)\right]$$

The main theorem of Prelec (2004) is as follows:

**Theorem (Prelec, 2004):** Under the assumption of admissability, for $n$ sufficiently large, and for risk neutral agents, the Baysian Truth Serum has a Bayes Nash Equilibrium in which all agents report their first and second order beliefs truthfully.

The intuition behind this payment rule is as follows. The second bracketed term compensates the respondent for his stated second order beliefs. A respondent's second order beliefs are evaluated by comparing them to the stated first order beliefs of all respondents. The closer his second order beliefs are to the true distribution of first order beliefs, the more he earns. The first term compensates the agent for his stated first order belief. His first order belief is evaluated by how "surprisingly common" it is relative to stated second order beliefs of his peers. That is, respondents who state first order beliefs whose empirical frequencies are higher than predicted by stated second order beliefs are paid more than respondents who state first order beliefs whose empirical frequencies are lower than predicted.

**The Robust Bayesian Truth Serum**

In order to define the RBTS we must first define the quadratic scoring rule. Let

$$R_q(y,\omega) = \begin{cases} 2y - y^2 & \text{if } \omega = 1 \\ 1 - y^2 & \text{if } \omega = 0 \end{cases}$$

Imagine an agent trying to predict whether some true state $\omega$ is 1 or 0. The quadratic scoring rule has the property that his expected score is maximized by reporting his true belief about the probability the state $\omega$ is 1 (see e.g. Selten, 1998).

The RBTS is implemented as follows. Every agent states their first order belief (their signal), in a report $x_i \in \{0, 1\}$ and their second order belief $y_i \in [0, 1]$ (the fraction of the population they believe will report a high signal, $x_k = 1$ ). For each agent $i$, assign them a peer agent $j$, and a reference agent $k$, and calculate

$$y_i' = \begin{cases} y_j + \delta & \text{if } x_i = 1 \\ y_j - \delta & \text{if } x_i = 0 \end{cases}$$

for arbitrary $\delta$. The RBTS payment for agent $i$ is

$$R_q\left(y_i', x_k\right) + R_q\left(y_i, x_k\right)$$

The main theorem of Witkowski and Parkes is as follows

**Theorem (Witkowski and Parkes, 2012):** Under the assumption of an admissable prior and risk neutral agents, there is a Bayes Nash Equilibrium in which all agents report their first and second order beliefs truthfully.

The intuition behind the payment rule is fairly straightforward. The payment rule has two components. The second component incentivizes the agent to be truthful about his second order beliefs. That is, the agent is paid via the quadratic scoring rule to predict what some reference agent $k$ will anounce as his signal. And by the discussion above, agent $i$ maximizes his expected payment from this component of the scoring rule by truthfully announcing his belief $y_i$ about the likelihood agent $k$ will announce a high signal. In simpler terms, the payment rule rewards agent $i$ for choosing a second order belief as close as posssible to the truth (the realized distribution of first order beliefs).

The first component of the payment rule incentivizes the agent to be truthful about his first order beliefs. The term $y_i'$ takes an arbitrary person $j$'s second order belief $y_j$ and either raises or lowers it depending on $i$'s report $x_i$. RBTS pays agent $i$ $R_q\left(y_i', x_k\right)$, and so $i$ wants $y_i'$ to be as near as possible to the true distribution of responses in the population. The admissability assumption guarantees that if person $j$ were to know that person $i$'s signal were high, then person $j$ would increase his assessment as to the number of people in the group who received high signals. Likewise, if $j$ were to learn that $i$'s signal were low, $j$ would lower his assessment about the number of people in the group who received high signals. In effect the mechanism raises or lowers $j$'s assessment based on $i$'s report, and then pays $i$ based on the closeness of this modified report to the truth. Thus $i$ can do no better than to tell the truth.

**Implementing Peer Prediction**

We used the BTS in the field to incentivize rank order responses about members of each group, and evaluated both rules using simulations and the data we collected. The model and payment rule, however, were designed for binary responses. Thus while responses contain a rank ordering of 5 people, we treat each ranking as a composite response to 25 yes/no questions of the form "Is person $i$ the highest ranking individual in the group?", "Is he the second highest?" and so on. We elicited second order beliefs of the form "How many people will say person $i$ is the highest ranking individual in the group?" "How many will say he is the second highest?" and so on. From there we directly applied the payment rule, callibrated so that the expected difference between payments arising from truthful and deceptive answers was large. Note that the accuracy of responses across various questions in a single ranking were correlated, but under the assumption of risk neutrality (which is maintained throughout the peer prediction literature and may be empirically reasonable with respect to moderate sums of money), these correlations are irrelevant.

The peer prediction methods are cheap to implement as they circumvent the need for ex post verification, but are extremely complicated and thus impossible to describe to the respondents. Therefore rather than explain the payment rule, we told participants that their payments would be based on their own rankings and the rankings their group members, but that the exact payment scheme was too difficult to explain. As with the ex-post verification rule, however, we reassured them that they would maximize their personal payments by telling us the truth about what they knew.

We address the tradeoff between simplicity of explanation and the lack of necessity for ex-post verification in the analysis that follows and show that little is lost by using peer prediction rather than the ex-post accuracy rule. Moreover, we show that the RBTS may actually be incentive compatible according to the empirically generated data and argue that this is a desirable property for mechanisms used in the long term even if their precise mechanics may not be understood by participants.

# 4 Specification

Before we can evaluate whether and how incentives improve the accuracy of responses, we need to understand whether the rankings given by group members have predictive power. Our main econometric specification asks how well a respondent's rankings of his peer matches with his peer's rank ordering based on true outcome realizations. To ground the concept of the true outcome rank ordering, we provide a simple example in Figure 3. Persons' A-E self-reported incomes are presented. The true outcome rank is the ordering of the income values from highest to lowest.

| Name | Income-True Outcome Value | True Outcome Rank |
|---|---|---|
| A | 5000 | 5 |
| B | 4000 | 4 |
| C | 3000 | 3 |
| D | 2000 | 2 |
| E | 1000 | 1 |

Figure 3: True Outcome Values vs. True Outcome Ranks

This specification is of interest as it allow for comparability across questions:

$$Outcome_{ikj} = \alpha_0 + \alpha_1 Rank_{ikj} + \alpha_2 Question_{ikj} + X_{ikj} + \gamma_j + \epsilon_{ikj} \quad (1)$$

In the regression above, $i$ indexes the person being ranked, $j$ indexes his group, and $k$ indexes the person doing the ranking. Depending on the table, $Outcome_{ikj}$ will represent either a measure of either person $i$'s outcome value (column 2 of Figure 3) or person $i$'s rank position in group $j$ (column 3 of Figure 3). $Rank_{ikj}$ is the rank that person $k$ gives person $i$ in group $j$. The highest rank for a particular question is 5 and the lowest is 1. If $Outcome_{ikj}$ represents person $i$'s rank position in group $j$, then $\alpha_1 = 1$ if respondents are able to perfectly predict the true outcome value of their peers. $Question_{ikj}$ is a dummy variable for the question. $X_{ijk}$ is a control for person $i$'s level of education and $\gamma_j$ is a group fixed effect.[5] Standard errors are clustered at the group level.

The primary interest of this study is to evaluate what types of incentives can disuade respondents from behaving strategically when reporting their rankings when they know that allocative decisions will be made based on their responses. To estimate whether offering incentives improves the accuracy of responses, this specification stacks the within-group rank order of outcome values as reported by each client $i$ in group $j$ across all questions and regresses these true rank-order outcomes on the rank given to person $i$ by person $k$.

[5]We include a group fixed effect in this specification because we had 3 groups with only 4 members.

$$Outcome_{ikj} = \alpha_0 + \alpha_1 Rank_{ikj} * Incetive_{ikj} + \alpha_2 Rank_{ikj} + \alpha_3 Incentive_{ikj} + \alpha_4 Question_{ikj} + X_{ikj} + \gamma_j + \epsilon_{ikj} \quad (2)$$

The rank variable, $Rank_{ikj}$ is interacted with whether respondent $k$ received an incentive ($Incentive_{ikj}$)- pooled across BTS and ex-post accuracy- for the responses given for the question. $\alpha_1$ is our main coefficient of interest. If incentives produce more accurate answers, then $\alpha_1 > 0$. The remainder of the variables are as above and we cluster our standard errors at the group level.

If incentives improve accuracy, we test whether the rankings reported when respondents are given incentives via BTS or ex-post accuracy differed significantly in their accuracy. To address this we estimate the following model

$$Outcome_{ikj} = \beta_0 + \beta_1 Rank_{ikj} * BTSIncentive_{ikj} + \beta_2 Rank_{ikj} * EPIncentive_{ikj} + \beta_3 Rank_{ikj} + \beta_4 BTSIncentive_{ikj} + \beta_5 EPInc \quad (3)$$

where $BTSIncentive_{ijk}$ is a dummy for whether the question was incentivized with BTS and $EPIncentive_{ijk}$ is a dummy for whether it was incentivized with the ex-post payment rule. The coefficients $\beta_1$and $\beta_2$ reveal how each incentive type influences rankings versus no incentives. An F-test of $\beta_1$-$\beta_2$ will indicate whether the accuracy of the ranks is significantly different between these two payment rules.

# 5 Results

## 5.1 Do Incentives Improve Accuracy?

In Table 1, we present the results of regression model 1. The outcome variable is the rank order of true values as reported by the members in the group: for 5-member groups, this is a number between 1 and 5 with 5 being the highest value in the group and 1 being the lowest. In column 1, we present the results controlling only for the question for which the rank was given while in column 2 we add the education control. If *Rank* predicted outcomes perfectly, the coefficient $\alpha_1$ would be 1. We find that, in the no controls specification, being ranked 1 higher by a group member $k$ increases the true outcome rank of group member $i$ by 0.1. Adding the education control improves a precision slightly.

Respondents are able to predict some of the variation in the true outcome order, but the accuracy of that prediction when pooled across questions and incentives is modest. We therefore break down our results further to better understand what information respondents have about different domains of their peers' economic lives. We present the regressions by each question in Table 2.[6] Respondents are best able to predict income, assets, and marginal returns rank orders. Being ranked 1 higher by a group member, on average, implies an increase in the true income rank order by 0.3. Although noisy, they also predict some of the variation in digitspan, profits and health expenditures. When we pool across and incentives and no incentives, it appears that respondents are not able to predict risk and time preferences, as well as whether their peers have trouble making loan repayments.

We have shown that respondents have valuable information about their neighbors across important domains. But can rankings be improved if respondents are provided incentives? In Table 3, we separate the effects by whether respondents received incentives (pooling across BTS and ex post accuracy payment rules) as specified in regression model 2. Without incentives, when we pool across all questions, the rank has no predictive power. The coefficient on the interaction, however, is large and significant in both columns 1 and 2, implying that monetary incentives have a powerful effect on collecting accurate reports. As before, we disaggregate the regressions by individual questions in Table 4. First, notice that respondents are completely unable to predict time (Column 8) and risk preferences (Column 9), with or without incentives. It is possible that respondents do not have information about how risk averse or patient their peers are. Alternatively, it is possible that the games we used to measure risk and time preferences are ineffective at predicting their respective outcomes. Although we do not have a diffinitive resolution to these questions, we do find (in regressions not shown) that our patience and risk outcomes are not predictive of other outcome variables with which they should in theory be correlated such as income, profits, or health risk (Levitt and List 2006).

For self-reported marginal returns, health, loan repayment ability, and digitspan rankings without incentives are also completely uninformative. But giving respondents an incentive improves accuracy dramatically. In the case of loan repayment, incentives increase predictiveness by as much as 0.44 over groups that did not receive incentives for this question. In the case of income, assets, and profits, respondents are able to predict well even without incentives. One possible explanation for this is that respondents knew that these were relatively more observable measures for both researchers and others in

[6]The number of observations vary because not all groups answered all questions due to logistical constraints in the field.

the group. They may have inferred that purposely misreporting too severely on these questions would have been too transparent.

While we know that respondents have information about one another along most dimensions tested and that incentives matter in obtaining accurate rankings, a natural next step is to quantify the information that respondents have in magnitudes of the outcome variable. This will provide an economic measure of how much variation between group members ranks can predict. Rather than the true outcome rank being the left hand side variable, we present regressions where the left hand side variable is in outcome values in Table 5. With incentives, as can be seen by adding the coefficients on $IncomeRank$ and $IncomeRank * ReceivedIncentive$ in column 3, the predicted increase in income from being ranked 1 higher by the group is Rs. 1400. When we, on the other hand, regress the true income values on the rank order if group members had ranked their peers perfectly, as presented in Appendix Table A.2,[7] the coefficient on rank is approximately 3700. This tells us that the true average increase in income from moving up one rank is Rs.3700. Thus community reports capture a substantial fraction of the total variation in our sample.

### 5.2 Do Payments Calculated via Peer Prediction Mechanisms Incentivize Truth-Telling?

Having documented that respondents have information about one another and that incentives can help improve the accuracy of the information, we can ask the principal answer of this study: does the payment rule that we use to provide incentives affect to the quality of responses obtain? Respondents incentive payments were calculated using either an ex-post accuracy rule, which paid by comparing how a respondent ranked a peer with the information that the peer self-reported prior to the experiment, or via BTS. The most salient distinction between the two payment rules during the game was in the detail with which each payment scheme was described. At the point in the game when respondents were to be paid for questions via the ex-post accuracy rule, surveyors spent a considerable amount of time explaining exactly how incentives would be calculated. When it was time to give rankings that would be rewarded via BTS, surveyors simply told respondents that their payments would be based on their own ranks and the rankings of their peers, but that it was too complicated to explain the payment rule exactly. With both ex-post accuracy and with BTS, surveyors emphasized to respondents that the way to maximize their payments was to be as truthful about what they knew as possible.

[7] We regress the true outcome value on the true outcome rank.

Did the level of explanation affect the accuracy of the responses? Rather than pooling across BTS and ex-post accuracy, we examine these two payment rules separately and present the results in Table 6. First, notice that both BTS and ex-post incentives substantially improve the accuracy of rankings. At the bottom of the table, we provide an F-test of whether BTS and ex-post incentivized significantly different responses. We find that we cannot reject that the quality of responses induced by each payment rule is the same, although the point estimate on the difference is large. If anything, BTS appears to lead to more accurate rankings than ex-post accuracy. Although we have no way of testing this hypothesis formally, one possible reason BTS may have performed better is that our respondents could have felt overwhelmed with the mathematical explanations of the ex-post payment rule.[8]

How does the provision of incentives affect who is highly ranked? In Table 7, we interact whether the person being ranked is a family member with whether a respondent receives an incentive. In the absence of incentives, respondents rank family members on average 0.85 ranks higher. The interaction term, however, suggests that when a respondent receives incentives, the skewing towards family members is almost completely negated. This is the case even after controlling for how well a respondent knows a family and a non-family peer in his group, implying that this is unlikely to be the result of group members knowing their family members better. This result is consistent with the interpretation that in the absence of personal incentives, respondents simply want to allocate the Rs. 20 prize to persons they are closest to in the group. When a personal incentive is given, however, they adjust their rankings to maximize their own payoff.

### 5.2.1 Verifying the Theoretical Properties of BTS in the Data

Given that BTS is just as effective in eliciting truthful responses as the straightforward payment rule, it is worthwhile to verify that the payment method works as theoretically predicted. That the peer prediction mechanism should work as predicted is desirable for several reasons. First, we do not want to deceive respondents when we tell them they can do no better than to tell the truth. Second, even if respondents do not know the precise mechanics of the payment rule, if it does not provide strong incentives for truthfulness respondents may learn over time that deceptive answers result in similar payment distributions.

The first step to verifying that the BTS works as predicted is to verify that, on average, people have informative second order beliefs. One should not expect second order beliefs to be perfectly informative about the distribution of first order beliefs, as some respondents are misinformed about the true state

[8] As described in Section 3, surveyors used ages and heights of group members to calculated hypothetical incentive payments and to illustrate how the rule paid.

of the world (for example, how much income group members really earn). But reasonable predictive power is reassuring that respondents understand the task and have informed second order beliefs. To ask whether reported second order beliefs have any predictive power over reported first order beliefs we estimate the following model

$$FOB_{nij} = \alpha_1 \overline{SOB}_{nijk} + \alpha_2 Question_{ikj} + X_{ikj} + \gamma_j + \epsilon_{ikj}$$

In the regression above, $FOB_{nij}$ is variable that captures the number of group members in group $j$ who said that person $i$ is in rank position $n$. So this outcome captures the distribution of first order beliefs in the group. And $\overline{SOB}_{nijk}$ measures the frequency with which respondent $k$ in group $j$ thought person $i$ would be ranked in spot $n$. A positive coefficient on $\alpha_1$ indicates that average second order beliefs are meaningful predictors of first order beliefs within a group.

In Table 8 we find respodents' beliefs about what others believe predict the empirical distribution of reported ranks. Column 1 presents the results without controls, and Column 2 presents the results with controls. We see that respondents do have accurate second order beliefs and that the accuracy of these beliefs improves with controls. Could this be mechanically true if respondents simply put the most weight in the second order beliefs on their own first order beliefs? In Appendix Table A.3, we show that this is not the case. We generate a second order beliefs distribution that puts 100% weight on a respondents own first order belief and 0 weight on any other second order belief. When we regress the first order beliefs distribution ($FOB$) on this new SOB, we see that the second order belief is less predictive than when the true second order beliefs are analyzed.

We next investigate whether people with accurate second order beliefs have accurate first order beliefs. Prelec and Seung (2006) note the theoretical connection between these two objects. Those with accurate second order beliefs have accurate knowledge of the state and thus also have accurate first order beliefs. The motivation behind this question is that if the answer is in the affirmative in practice, then this provides a cost-effective method to determine which respondents have accurate information. To answer it, we first calculated the correlation, at the ranker by rankee by question level, of the second order beliefs and the true distribution of first order beliefs. A higher correlation indicates that second order beliefs are more predictive of the observed distribution of rankings. We then regressed the true outcome rank order on the predicted rank order (the first order belief), the accuracy of second order beliefs (the correlation between first and second order beliefs), and the interaction of these two variables. The coefficient on the interaction between the the accuracy of second order beliefs and the rankings given by each respondent tells us whether respondents who have more accurate second order beliefs also have more accurate first order beliefs. In Table 9, we present the results of this regression with and without controls. We find that respondents with more accurate second order beliefs also have more accurate first order beliefs. Increasing the correlation between second order beliefs and true rankings by one standard deviation (0.53), improves the accuracy of rankings by 100%.

### 5.2.2 Incentive Compatibility

In this section we evaluate the incentive compatibility of the peer prediction payment schemes. While we calculated payments using BTS in our experiment, we have found that the payments it induces are extremely noisy. This is because in practice, it was very common for respondents to announce zeroes for their second order beliefs. That is, respondends would frequently guess that no one would rank people in certain positions. Because BTS evaluatse the ratio of one's own answer and community expectations, responses near zero cause the mechanism to output extremely large positive and negative values that don't correspond closely with accuracy. So in what follows, we evaluate the incentive compatability of the Robust Bayesian Truth Serum, which does not suffer from sensitivity to extreme responses in the same manner. We are able to do so because the inputs required for the two payment rules are identical.

Throughout the experiment, we told respondents that they would maximize their personal payoffs if they reported the ranks truthfully. While RBTS is truthful under certain reasonable assumptions about how beliefs are formed, its incentive compatability under the empirical distribution of beliefs in practice remains an open question. We therefore test two related, but distinct, questions: first, are larger RBTS payments associated with more accurate reports? In other words, do respondents get higher payouts following correct answers than they do following incorrect answers? Note that this is not vacuously satisfied as respondents are paid only based on the reported first and second order beliefs of their peers, and not on the ex post accuracy of their responses. Second, are respondents maximizing their subjective expected utility by telling the truth?[9]

To answer the first question, we estimate the following stacked regression using the questions for which group $j$ received the RBTS incentive.

$$Correct_{nikj} = \alpha_1 RBTSpayment_{nikj} + Question_{ikj} + X_{ikj} + \gamma_j + \epsilon_{ikj}$$

In the regression above, $Correct_{nikj}$ is a dummy for whether person $k$ correctly identified that person

[9]This is the standard theoretical notion of incentive compatability.

$i$ either is or is not in position $n$ in the true rank ordering. Thus if person $k$ said that person $i$ was (not) in spot $n$ and person $i$ was indeed (not) in spot $n$ then $Correct_{nikj}$ is 1. Else it is 0. $RBTSpayment_{nikj}$ measures the payment that person $k$ received for his stated first order belief about whether person $i$ was in spot $n$. A positive coefficient $\alpha_1$ indicates that respondents can expect to earn more money from correct answers than incorrect answers. We present these results in Appendix Table A.1 and find that indeed a higher RBTS payment is associated with correct answers: a person that receives an extra RBTS point is 7 percentage points more likely to have given a correct ranking.

More truthful answers, however, are not necessarily more accurate. Our second goal is to therefore test whether more truthful answers are associated with higher payments. Due to the coarsness of our elicited measures of belief, we cannot verify directly whether or not their priors are admissable. However we can determine the distribution of payoffs respondents expect to receive under alternative responses to see whether they succeeded in maximizing their subjective expected utility. Respondents' payments depend on the distribution of first order beliefs (i.e. the empirical distribution of signals) and on the distribution of second order beliefs. Therefore to determine whether it is incentive compatible to tell the truth, we must understand what the respondent believes are the distributions of first and second order beliefs in the population. We obtain the former for free: Respondents' beliefs about the distribution of first order beliefs are their second order beliefs, and we elicited these in our survey. We did not, however, elicit their beliefs about the distribution of second order beliefs: their third order beliefs. We must therefore construct those. The intuition is presented in the following three steps:

1. We assume that respondents hold a common prior. If so, we can back out their third order beliefs from (a) the distribution of second order beliefs conditional on each first order belief and (b) their belief about how probable each first order belief is. The latter corresponds to her second order beliefs. [10]

2. We approximate the distribution of second order beliefs in the population conditional on any given first order belief with the (sparse) empirical distributions we observe.

3. Given second and third order beliefs, we can calculate a respondent's subjective expected utility from reporting the truth (her stated first order belief) and from any other report.[11] Specifically, we assume that the report the respondent has given is her true belief and calculate her payment.

[10] If agents have common priors then conditional on the signal they receive, they would update to have the same posterior belief. We stress here that we elicited ranks and not signals. Therefore two agents who report the same rank do not necessarily have the same posterior as the rank is a coarse measure of a signal.

[11] Notice that we only utilize incentivized data since it is the only time we collected second order beliefs.

Holding constant her own second order belief and the first and second order beliefs of her peers, we then calculate her payments for every other possible report she could have given.

The results from this exercise are presented in Figure 4 below. Column 1 of the figure depicts the percentage of the time that telling the truth gives the largest payment, column 2 depicts the percentage of the time that telling the truth results in the second largest payment, etc. Taking the graph at face value, telling the truth maximizes the respondent's subjective expected utility about 35% of the time and it minimizes her subjective expected utility only about 10% of the time. An ideal graph would place all of its weight in the first column.


Figure 4: Incentive Compatibility of RBTS Using Collected Data

We compare this graph to one produced using the same data, but paying respondents via the Bayesian Truth Serum. In Figure 5, we observe that unlike the graph produced with RBTS incentive payments, the BTS graph shows that the BTS payment rules gives the largest payment to any answer with nearly equal probability. If anything, it rewards the answer furthest from the truth the most. We therefore rely on RBTS as it is more incentive compatible in our data than BTS.

The observed departure from this ideal, even in RBTS, may be due to the failure of our assumptions required by RBTS holding in practice, or by our noisy approximation of third order beliefs. To evaluate this, we perform a simulation in which we generate data that perfectly abides by all of the assumptions required for the incentive compatability of RBTS. We describe the specific steps in the appendix, but we provide a brief intuition here. We generate groups of agents and to each agent $i$ we assign a true skill

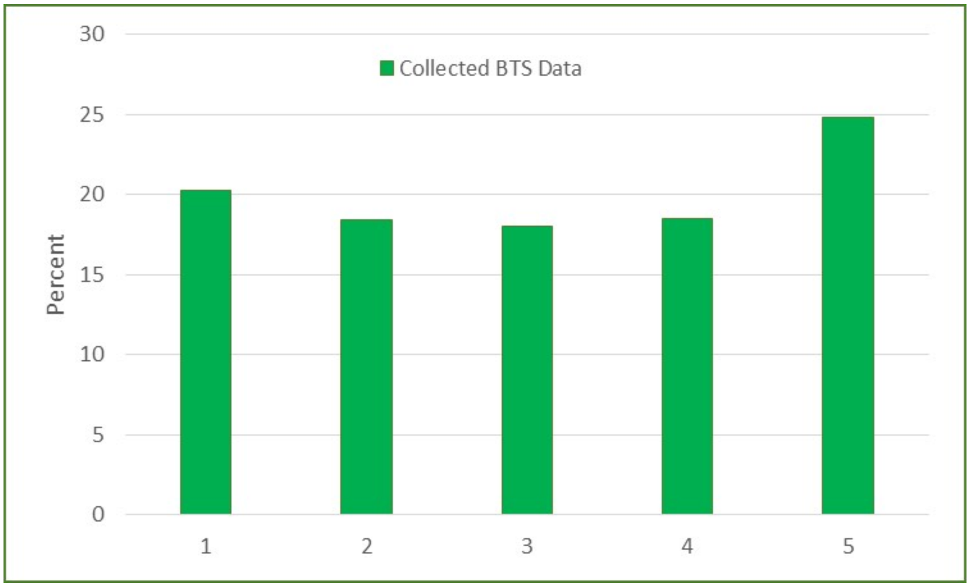


Figure 5: Incentive Compatibility of BTS Using Collected Data

level. Each agent also receives informative signals about the skill levels of each of his group members. Agents update their priors based on these signals and these form the basis of their second order beliefs. We can thus compute each agent's complete second and third order beliefs.

Because the data is generated to be perfectly consistent with the assumptions of RBTS, the agent is always best off telling the truth. Next we compress our simulation data to correspond exactly to the data we collected from our respondents: just first and second order beliefs about group rank. In the process we, of course, lose a tremendous amount of richness, but this allows us to have two data sets (collected and simulated) that contain identical level of detail. We then generate the same graph as we did for our collected data and present it in Figure 6.


Figure 6: Incentive Compatibility of RBTS Using Collected and Simulated Data

The graph produced with the collected and with the simulated data are strikingly similar. We therefore argue that our test yields the most favorable result for incentive compatability and that telling respondents that they will maximize their expected payments by reporting truthfully is good advice.

# 6 Conclusion

In this paper we present results from an experiment measuring what people know about one another in a village community in Maharashtra, and we quantify the extent to which their deceptive reports can be corrected via monetary incentives. We find that respondents have substantial information about one another, and that the quality of this information is improved when provided with monetary incentives for truthfulness. Moreover we show that complicated but very easy to implement payment rules known as peer prediction mechanisms perform just as well as simpler mechanisms whose implementation is more burdensome and that one of these mechanisms, RBTS, is truthful. Given the applicability of peer prediction to a wide variety of settings we see potential for it to be incorporated in many settings. In ongoing work, we are exploring the usefulness of several other insights from mechanism design theory including ensuring the privacy of respondents when they report, asking respondents to cross-report on who is likely to have valuable information and who their peers are likely to lie about, and eliciting information in a zero sum (where someone's gain is necessarily someone else's loss) versus a non-zero sum manner. Our ongoing work also encompasses real-world stakes (whether a respondent receives a large grant for business purposes). It is our hope that the insights from this research agenda can inform the methods used by practitioners to leverage community information.

# References


- Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., & Tobias, J. (2012). Targeting the Poor: Evidence from a Field Experiment in Indonesia. The American Economic Review, 102(4), 1206-1240.

- Alatas, V., Banerjee, A., Chandrasekhar, A. G., Hanna, R., & Olken, B. A. (2012). Network structure and the aggregation of information: Theory and evidence from Indonesia (No. w18351). National Bureau of Economic Research.

- Attanasio, O., Augsburg, B., De Haas, R., Fitzsimons, E., & Harmgart, H. (2011). Group lending or individual lending? Evidence from a randomised field experiment in Mongolia.

- Beaman, Lori, and Jeremy Magruder. "Who gets the job referral? Evidence from a social networks experiment." The American Economic Review (2012): 3574-3593.

- Bhattacharya, V. and Roth, B. (2015). Rank-Based Scoring Rules for Binary Events. Working Paper

- Brier, Glenn W. "Verification of forecasts expressed in terms of probability."Monthly weather review 78.1 (1950): 1-3.

- Bryan, G. T., Karlan, D., & Zinman, J. (2012). Referrals: Peer screening and enforcement in a consumer credit field experiment (No. w17883). National Bureau of Economic Research.

- Carpena, F., Cole, S., Shapiro, J., & Zia, B. (2013). Liability structure in small-scale finance: evidence from a natural experiment. The World Bank Economic Review, 27(3), 437-469.

- Delavande, Adeline, Xavier Giné, and David McKenzie. "Measuring subjective expectations in Developing Countries: A Critical Review and new evidence." Journal of Development Economics 94.2 (2011): 151-163. APA

- Ghatak, Maitreesh. "Group lending, local information and peer selection."Journal of development Economics 60.1 (1999): 27-50.

- Ghatak, Maitreesh. "Screening by the company you keep: Joint liability lending and the peer selection effect." The Economic Journal 110.465 (2000): 601-631.
- Giné, X., & Karlan, D. S. (2014). Group versus individual liability: Short and long term evidence from Philippine microcredit lending groups. Journal of Development Economics, 107, 65-83.
- Hossain, T., and R. Okui (2013): The Binarized Scoring Rule," Review of Economic Studies, 80(3), 984-1001
- Karni, E. (2009): A Mechanism for Eliciting Probabilities," Econometrica, 77(2), 603-606.
- Levitt, Steven D., and John A. List. "What do laboratory experiments measuring social preferences reveal about the real world?." The journal of economic perspectives (2007): 153-174.
- Offerman, T., J. Sonnemans, G. Van De Kuilen, and P. P. Wakker (2009): A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes," Review of Economic Studies, 76(4), 1461-1489.
- Prelec, Dražen. "A Bayesian truth serum for subjective data." science 306.5695 (2004): 462-466.
- Witkowski, Jens, and David C. Parkes. "A Robust Bayesian Truth Serum for Small Populations." AAAI. 2012.


# Appendix I

In this section we provide a proof of Theorem 1 that the straight forward payment rule is incentive compatible for any expected utility maximizer.

**Proof of Theorem 1:** We want to show that ranking truthfully and being paid according to the proper scoring rule described above is a dominant strategy for an expected utility maximizer with a utility function that is monotone increasing in wealth. This is equivalent to showing that truthful ordering of the events results in a payment distribution that first order stochastically dominantes (FOSD) the payment distribution resulting from any other ordering. Consider an arbitrary, potentially untruthful ordering $\sigma$ where $\sigma_i$ is the item that is ranked $n-(i-1)$ highest. (That is, the item a respondent ranks in spot $n$ is the highest item, the item a respondent ranks in spot 1 is the lowest item)

Let $A_j = \{x = (x_{\sigma_1}, \cdots, x_{\sigma_n}) : \sum_{i=1}^{i=n} i \times x_{\sigma_i} < j\}$. That is, $A_j$ is the set of sets of realizations such that if each realization in some set $x \in A_j$ occurs, then the agent is paid some amount less than j.

For $x \in A_j$ let $P_\sigma(x) = \prod_i f_{\sigma_i}(x_{\sigma_i})$ is the probability that precisely the realizations in $x$ occur.

Let $X$ be the random variable with CDF $F_\sigma$ that corresponds to the total payout resulting from the ordering $\sigma$.

Then $F_\sigma(j) = P_\sigma(X < j) = \sum_{x \in A_j} P_\sigma(x)$ is the probability that $X < j$.

Consider finding $k, l$ s.t. $\sigma_k < \sigma_l$ but $X_{\sigma_k} > X_{\sigma_l}$ That is $\sigma_k$ and $\sigma_l$ is a pair that is misordered in $\sigma$.

Clearly, if no such pair exists then $\sigma$ is truthful. We will show that if such a misordered pair exists, then flipping that pair such as to create a new ordering $\sigma'$ such that $\sigma'_l = \sigma_k$, $\sigma'_k = \sigma_l$ and $\sigma'_i = \sigma_i$ for $i \neq k, l$ results in a payment distribution for $\sigma'$ that FOSD the payment distribution for $\sigma$. After showing this, the result follows, because starting at any aribtrary untruthful ordering, the truthful ordering will be arrived at after a finite number of such switches, at which point no such switches remain.

To show FOSD we will show that $F_\sigma(j) > F_{\sigma'}(j)\ \forall j$. Consider an arbitrary element $x \in A_j$, There are three cases.

- Case 1: $x_{\sigma_k} < x_{\sigma_l}$

In this case there is another $x'$ in $A_j$ s.t.$x'_{\sigma_l} = x_{\sigma_k}$ and$x'_{\sigma_k} = x_{\sigma_l}$ and $x'_{\sigma_i} = x_{\sigma_i} \forall i \neq l, k$. So $P_\sigma(x) + P_\sigma(x') = P_{\sigma'}(x) + P_{\sigma'}(x')$ so the probability that one of these two events occurs is unchanged.

- Case 2: $x_{\sigma_k} = x_{\sigma_l}$

Then$P_\sigma(x) = P_{\sigma'}(x)$

- Case 3: $x_{\sigma_k} > x_{\sigma_l}$

Then $\frac{P_\sigma(x)}{P_{\sigma'}(x)} = \frac{f_{\sigma_k}(x_{\sigma_k}) f_{\sigma_l}(x_{\sigma_l})}{f_{\sigma_k}(x_{\sigma_l}) f_{\sigma_l}(x_{\sigma_k})} > 1$ because $\frac{f_{\sigma_k}(x_{\sigma_k})}{f_{\sigma_k}(x_{\sigma_l})} > \frac{f_{\sigma_l}(x_{\sigma_k})}{f_{\sigma_l}(x_{\sigma_l})}$ by $X_{\sigma_k} > X_{\sigma_l}$. $\square$

# Appendix II

This model simulates the truthfulness exercise of RBTS. Imagine there are 5 people to be ranked based on their "skill."

1. Skill is $\theta_i \in \Theta_i = \{1, \cdots, 3\}$. There is a common uniform prior.

2. Given a skill $\theta_i$ each ranker receives an independent signal $s_j \in S_j = \{1, \cdots, 3\}$. The distribution of signals given a state $\theta_i$ is $Pr(s_j|\theta_i) = \begin{cases} .6 & \text{if } \theta_i = s_j \\ .2 & \text{else} \end{cases}$

3. From this we can derive the state posterior $Pr(\theta_i|s_j) = \frac{Pr(\theta_i \cap s_j)}{Pr(s_j)} = \begin{cases} .6 & \text{if } \theta_i = s_j \\ \\ .2 & \text{else} \end{cases}$

4. We can also derive the signal posterior $Pr(s_{j'}|s_j) = \frac{Pr(s_{j'} \cap s_j)}{Pr(s_j)} = \frac{\int_{\theta_i} Pr(s_{j'} \cap s_j|\theta_i) Pr(\theta_i)}{\frac{1}{3}} = \begin{cases} .44 & \text{if } s_{j'} = s_j \\ \\ .28 & \text{else} \end{cases}$

5. **FIRST ORDER BELIEFS:**

   (a) Each ranker $j$ gets a signal $s_{j,i}$ for each person to be ranked $i \in \{1, \cdots, 5\}$. He can use these to form a ranking of the 5 people to be ranked. We break ties randomly.

6. **SECOND ORDER BELIEFS**

   (a) We compute ranker $j$s second order belief, which will be a distribution over the 120 possible rankings. For a random person there are $3^5 = 243$ possible signal combinations he received. For each potential signal combination a player receives we calculate the probability a random other ranker $j'$ received any other signal combination.

   (b) Given that each signal realization corresponds to a rank ordering we can use the probabilities calculated above to determine the probability the agent assigns to another arbitrary agent announcing each possible ranking.

7. We divide agents into groups of 5 at random, and repeat the above steps for each group so that each group has first and second order beliefs that mimic our experimental data.

8. We run perform the same IC exercise as with the experimental data.