Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design In The Field*

Reshmaan Hussam (HBS), Natalia Rigol (Microsoft, HBS) and Benjamin Roth (HBS)

January 18, 2018

Please find the most recent version of the paper here.

Abstract

Microentrepreneurs in low-income countries have high marginal returns to capital yet face significant credit constraints. Because returns are highly heterogeneous, the cost of assessing credit worthiness often makes lending to this sector unprofitable. In this paper, we show that (1) community knowledge can help overcome information asymmetries prevalent in poorly developed financial markets and that (2) appropriately designed elicitation mechanisms can extract truthful community reports. We asked entrepreneurs in Maharashtra, India to rank their peers on metrics of business profitability and growth potential. To assess the validity of their reports, we then randomly distributed cash grants of USD 100 to a third of these entrepreneurs. We find that information provided by community members is highly predictive of the marginal return to capital: entrepreneurs ranked in the top tercile earn returns of 23% per month, which is three times the average return within the sample. We horserace community rankings against a machine learning prediction built using entrepreneur characteristics and find that peer reports are predictive over and above observable traits. Yet community information is only useful if it is feasible to collect truthful statements. We experimentally vary the elicitation environment and demonstrate agency problems when community members have incentive to lie: accuracy of community reports decreases by a third when cash grants are at stake. But we also show that tools from mechanism design can be used to address these agency problems. Paying for truthfulness using a peer prediction rule fully corrects for strategic misreporting induced by the high-stakes environment. Public reporting and cross-reporting techniques motivated by implementation theory also significantly improve the accuracy of peer reports.

1 Introduction

Not all entrepreneurs have what it takes to be successful. Both theory and empirical evidence on firm growth emphasize the wide distribution of talent among managers and business owners and, consequently, the significant heterogeneity in expected returns to capital (Lucas, 1978). But, to a lender in a peri-urban area of Maharashtra, India, identifying high-return entrepreneurs — applicants for whom larger or more flexible loans might be profitable for both borrower and creditor — is prohibitively difficult. Consider, for instance, an applicant who runs a convenience store next to her home. She cannot offer collateral and does not have verifiable income to establish a credit history. The shopkeeper is one of a dozen in her neighborhood and all sell a similar mix of grains, pulses, and packaged snacks. Though she might be motivated and highly skilled, the creditor has no means of assessing her potential.

The lender's information asymmetry problem is connected to an important puzzle in development economics. Experimental studies on the returns to capital for microentrepreneurs in low-income countries show that marginal returns to capital are far above standard microcredit interest rates (Fafchamps et al., 2014; McKenzie et al., 2008). Economic theory would suggest that reducing credit constraints should be an important factor in realizing enterprise growth, but the rapid spread of microcredit has had disappointingly low impact: a recent meta-analysis found that, on average, credit has only a modest effect on business profits (Banerjee et al., 2015).

Microcredit's low average impact is less puzzling when the distribution of entrepreneurial ability is taken into account. Numerous studies find that microentrepreneurs' marginal returns are high on average *and* highly heterogeneous. For instance, in a Sri Lanka capital grant experiment, quantile treatment effects imply a marginal return to capital of 0% - 45% per month (de Mel et al., 2008). Similarly, in our experiment, quantile treatment effects from cash grants vary from 0% to 28% per month (Appendix Figure 1). Theoretical and empirical studies of entrepreneurship in developing countries also emphasize that many self-employed individuals are business owners not because of personal ambition but because there are scarce opportunities for wage labor; these individuals are less likely to hire workers or otherwise expand their businesses (Schoar, 2010; De Mel et al., 2010). Yet lenders and policymakers tend to treat microentreprise owners as a relatively homogeneous group: credit and grant programs typically have minimal screening and little to no product differentiation by applicants' capital needs or business capabilities. As the shopkeeper example demonstrates, creditors may be impeded by their inability to differentiate between high and low return applicants. But little is known about how to overcome information asymmetries — or even what information is needed — to identify entrepreneurs with the most capacity to grow.

In the absence of formal financial information, there may be an alternative source of information that banks, governments, and non-profit institutions in developing countries could use to identify entrepreneurs with high potential: entrepreneurs' social network. Consider again the shop-

keeper. Her customer and next-door neighbor might pay attention to how well she markets her business and whether the rice and lentils are clean and good quality. They may take note of her working hours and how fastidiously she keeps her shop floor clean. Several theoretical papers have studied the potential benefits of relying on community members to relax information asymmetries (e.g. Besley and Ghatak, 2005; Varian, 1990). In developing countries, neighbors are also more likely to be engaged in informal risk pooling agreements which require mutual knowledge of one another (Foster and Rosenzweig, 1996; Townsend, 1994).

On the other hand, relying on the community might lead the creditor astray. Among both academics and practitioners, there is a deep and divided literature on the predictors of selection into or success in entrepreneurship.[1] It not clear then that community members would themselves know which parameters to use to assess entrepreneurial ability. And even if community knowledge is accurate, the high stakes involved might introduce an incentive for community members to distort their predictions in favor of their friends and family.

In this paper, we show that community information — the knowledge that neighbors have about one another — is highly predictive of entrepreneurs' marginal returns to capital. Crucially, we also demonstrate that it is possible to collect credible reports even when the provision of important resources is at stake. Using methods from mechanism design theory, we develop a peer elicitation environment which aligns respondents' incentives with truthfulness. By experimentally varying these incentives, we also quantify the magnitude of misinformation when stakes are high and provide evidence on how community members trade off personal gain with benefits to family and peers.

We report on results from a field experiment that we conducted with 1,345 entrepreneurs from Amravati, a city in Maharashtra, India. We assigned respondents and their nearest neighbors to peer groups of 4-6 persons. After collecting detailed baseline data from all respondents, we asked entrepreneurs to rank their peer group members on predicted marginal returns to capital, profits, and other firm, owner, and household characteristics. Once the community reports were complete, we randomly assigned USD 100 grants to one third of entrepreneurs in order to induce growth and assess the accuracy of respondents' predictions. We evaluate the accuracy of community information by comparing how well the rankings predict individuals' true outcomes as reported at baseline or in subsequent follow-up surveys.

Our first main finding is that community members can identify high-return entrepreneurs with stunning accuracy. While the average marginal return to the grant was about 8% per month, our estimates of the marginal returns to capital of entrepreneurs in the top third range from 17% to 27%. Had we distributed our grants using community reports instead of random assignment, we would have more than tripled the total return on our investment.

---

[1] For example, while some studies find a relationship between taste for risk and entrepreneurship, others do not (see Parker (2009) for a review).

To benchmark the value of community information, we compare its predictive accuracy against that of observable entrepreneur characteristics. We build a model to predict entrepreneurs' marginal return to capital using a causal forest (a machine learning technique developed by Wager and Athey (2017) to predict heterogeneous treatment effects). We find that entrepreneurs in the top third of the machine learning model's predicted marginal returns distribution have realized returns of 18% per month. But when we estimate marginal returns based on community information and control for the machine learning prediction, we still find that those in the top tercile of the community prediction distribution earn 17% higher monthly returns than those in the bottom tercile. This finding suggests that community information is valuable above and beyond information that can be captured by observables.

Our second main finding is that strategic misreporting is a first-order concern when eliciting community information. By random assignment, half of respondents were told that their reports would be used only for research purposes (the "*No Stakes*" treatment) and the other half were told that their reports would be used to allocate USD 100 grants to members of their community (the "*High Stakes*" treatment). The correlation between community reports and true outcomes is on average 24% to 35% lower when allocation of resources is at stake, which significantly lowers the value of peer elicitation. We also identify who benefits from misreporting and by how much: we quantify the extent to which participants favor themselves, their family members, and their close friends (as identified by other group members).

Given the importance of strategic misreporting, we explore whether it is feasible to realign incentives to report truthfully. Alongside the "*High Stakes*" treatment, we cross-randomized treatments which varied respondents' immediate benefit (or cost) for truthful responses. Respondents were assigned to report either in private or in a public setting, with their fellow neighbors observing their reports. Participants were also randomly assigned to receive monetary payments based on the truthfulness of their reports. Payments were calculated using the *Robust Bayesian Truth Serum* (RBTS), a peer prediction mechanism which determines participant scores as a function of the contemporaneous reports of other respondents. Importantly, RBTS does not utilize ex-post outcomes (which can be both manipulable and costly to verify) to determine payments.

Our third finding is that methods grounded in mechanism design theory can be used to design a peer-elicitation environment in which truthtelling is incentive compatible. Monetary payments and public reporting do little to improve the accuracy of self-reports. But payments double the predictive power of reports that entrepreneurs make about other group members. We show direct evidence that monetary payments reduce the likelihood that respondents favor their family members or their close friends. Thus monetary payments undo the strategic misreporting induced by shifting from a no-stakes to a high-stakes setting. Using the empirical distributions of reports, we also show that under RBTS truth-telling is empirically incentive-compatible. Finally, we find that public reporting doubles the predictive accuracy of reports about others when there are no stakes, but has no effect in a high-stakes setting. We shed light on this result by exploring two competing forces

4

present in a public elicitation setting: family pressure and community policing.

Our findings contribute to several strands of literature. The idea that social networks — friends, family, colleagues — are a rich source of information has deep roots in development economics. Yet while there is an expansive literature on information diffusion within networks, there have been relatively few empirical studies on information extraction. Though there are many settings in which community knowledge could help private and public-sector actors overcome information asymmetries, the value of this information and the impact of incentives on disclosure are not well-understood. There are a few notable exceptions: first, in the community targeting literature, Alatas et al. (2012) investigate whether villagers in Indonesia can select the village's poorest residents to receive government transfers. They find that community targeting performs worse than a Proxy Means Test for assessing households' level of consumption but better at capturing a household's perception of their own poverty status. Basurto et al. (2017) find that village chiefs in rural Malawi are more likely to target fertilizer subsidies to households that self-report they would benefit from agricultural inputs than the standard PMT method. In the referrals literature, Beaman and Magruder (2012) find that high-quality workers in Kolkata, India can refer other high-quality laborers when incentivized to do so. In contrast, Bryan et al. (2015) find that borrowers in South Africa can do no better than the lending institution in selecting high-quality borrowers among their peers.[2] Lastly, Maitra et al. (2017) show that local traders in India can select microcredit borrowers for whom credit leads to larger increases in production and income than for borrowers selected by standard microcredit, with the caveat that both the selection method (traders' screening versus self-selection into microfinance) and the contract type (individual versus joint liability loans) covary.

Our findings provide new insight into the depth and breadth of social knowledge contained in rural and peri-urban networks. The Alatas et al. (2012) study demonstrates that community members have reliable information regarding observable characteristics (wealth) of persons across their social network. We show that community members can predict marginal returns to capital, a metric that is exceptionally difficult to estimate even using rich observables or expert opinions. This is evidence that community members have accurate knowledge of one another that is much deeper than what has been previously shown. The Beaman and Magruder (2012) and Bryan et al. (2015) studies evaluate the depth of knowledge that individuals have regarding one close peer or family member. We show that community members have more widespread knowledge of their peers: we find that participants can provide accurate reports on their neighbors, not only on persons with whom they have close social ties.

Community knowledge — even if accurate — is only useful for allocative decision-making if those eliciting the information can be confident that they will gather *truthful* reports. And when allocation of resources is at stake, there is reason to be concerned that community members will

---

[2] All referred applicants had to also meet the bank's eligibility criteria and, unlike in our setting, South Africa has a well-functioning credit bureau.

lie. Yet strategic misreporting is not typically addressed in the design of programs which rely on community information to make decisions. For example, community-driven development projects, which leverage community information or community action to make decisions regarding public goods expenditures, are rarely designed to account for strategic behavior (Mansuri and Rao, 2004; King, 2013).

We contribute to a young literature which addresses strategic misreporting in targeting programs. Alatas et al. (2013) examine whether elite capture poses a problem for community reporting, but elites are not the only group with the ability or incentive to lie. Though Alatas et al. (2013) conclude that elite capture is not a significant concern, we find that misreporting is common when community members are told that their reports will influence distribution of grants. Alatas et al. (2012) also elicit community reports in public in order to incentivize truthtelling. However, their experiment is not designed to evaluate the impact of public reporting on the accuracy of reports. Through random variation of the elicitation environment, we show that public reporting is not effective for realigning incentives with truthtelling when allocation of resources is at stake.

Finally, we contribute to an emerging literature which evaluates the implementability of methods developed within the theoretical mechanism design literature. The field of mechanism design offers tools which make truthtelling incentive compatible in theory, but the assumptions underlying these schemes may not hold in practice, and first order first order barriers to implementation are sometimes unmodeled. In this and a companion paper (Rigol and Roth, 2017), we adapt and deploy a peer prediction mechanism to incentivize truthful reporting. To the best of our knowledge, this is the first large-scale setting to use a peer prediction mechanism.

The rest of the paper proceeds as follows. Section 2 introduces the setting and study sample. Section 3 describes our conceptual approach to designing the elicitation environment, Section 4 describes our experimental setting and design, Section 5 describes the data and provides a brief discussion of the randomization, Section 6 discusses how well community members know one another, Section 7 provides our main results, and Section 8 concludes.

2   Study Sample and Context

Our study takes place in Amravati, a city of about 550,000 persons in the state of Maharashtra, India. Households in our sample come from nine neighborhoods along the perimeter of Amravati; we selected these neighborhoods because they have a relatively high proportion of microentrepreneurs.[3] These are densely packed peri-urban slums; in each of these neighborhoods, there are roughly 900 household dwellings in a 500 by 700 square ft. area. In September 2015, we conducted a complete door-to-door census of these neighborhoods, which encompassed 5,573 house-

---

[3]Our selection of neighborhoods was based on advice from local officials in the District Collector's Office. The nine neighborhoods are: Belpura, Vilash Nagar, Mahajan Pura, Akoli, New Saturna, Old Saturna, Wadali, and Pathan Chawk.

holds. Based on households' responses to the census, we determined their eligibility for the study. In line with selection criteria of other recent "cash-drop" experiments (see e.g. de Mel et al. (2008)), all households in our sample have at least one enterprise with (i) USD 1,000 or less in total working and durable capital and (ii) no paid, permanent employees.[4] Almost 30% of households in these neighborhoods owned at least one business and were eligible (1,576 households). Entrepreneurs in 1,345 of these households agreed to participate in our study so our sample population is reasonably representative of the universe of eligible enterprises in Amravati.

Characteristics of Microenterprise Owners. The modal entrepreneur in our sample is 40 years old and has roughly 8 years of formal education. Approximately 60% are male and almost all are married. Most entrepreneurs operate their business close to home, but they operate across a wide range of activities. 30% of sample entrepreneurs work in manufacturing, typically as a tailor or stitcher. Another 30% work in services, mainly in food preparation and hair salons. Within the retail sector (30% of the sample), the most common business type is a grocery shop. Outside of these three sectors, entrepreneurs are spread evenly across construction and livestock rearing. On average, sample entrepreneurs earn profits of Rs. 4500 per month (USD 2.5 per day), which accounts for roughly half of their household income. Entrepreneurs also face a significant amount of risk: between the baseline and one year follow-up survey, about 10% of businesses in control group households were shut down. In over a third of these cases, the reason given for enterprise closure was illness of the business owner. Correspondingly, medical expenses make up a large fraction of household spending: on average, respondents report spending nearly 30% of their monthly earnings on health-related expenditures. Perhaps as a means of insuring against risk, households diversify across types of income-generating activities: in half of sample households, there is at least one fixed salary or daily wage worker and one fifth of households own more than one business.

Characteristics of Microentrepreneurs' Peer Networks. In order to elicit entrepreneurs' knowledge of one another, we assigned study participants to peer groups of roughly five persons based on geographical proximity. Peer groups are the unit of information collection: entrepreneurs are asked to report on only themselves and their other group members, not on the entire community. Importantly, we find that peers know their group members well. On average, peers reported that they visited another group member on 22 occasions in the previous 30 days. Respondents were not able to identify another group member in less than 1% of cases. Two-thirds of respondents identify at least one other group member as a family member or close friend. In 70% of groups, at least two people operate a business in the same (broad) industry category. Entrepreneurs also actively maintain strong social ties within their group: over 50% of respondents reported that they regularly discuss private family and business matters with at least one other group member. And, entrepreneurs have at least some knowledge of *every* group member: 87% of respondents correctly identified for all other group members whether that person owned a motorcycle (half of

---

[4]Following de Mel et al. (2008)'s selection criteria, we excluded farmers and self-employed service persons, such as domestic helpers and teachers. If there were multiple business owners in the household, we required that the household have at most USD 2000 in combined business capital.

respondents are motorcycle owners) and 80% correctly identified who among their peers had young children living in their home. In Section 6, we evaluate how well community members can predict household and enterprise characteristics.

3   Mechanisms to Incentivize Truthful Revelation

Agents' knowledge is only valuable for decision-making if it is incentive compatible for agents to report truthfully. When the allocation of resources is at stake, strategic misreporting may be a first-order concern. Mechanism design offers an array of tools which make truthtelling incentive compatible in theory, and one of our goals is to understand which of these tools work to realign incentives in practice. In this section, we describe our conceptual approach for designing and evaluating the peer ranking elicitation environment.

Public Reporting. Fear of public reprisal is a powerful deterrent to socially undesirable behavior. This insight has been applied to incentivize costly actions across a number of settings (notable examples include using public notification of individuals' voting record (Gerber et al., 2008) or electricity usage (Allcott and Rogers, 2014) to encourage behavioral change). Intuitively, conducting peer elicitation in public may reduce strategic misreporting because participants care about their reputation for honesty. At the same time, publicity may exacerbate pressure to rank one's family, friends, and influential members of the community more highly. Because manipulating the observability of reports is cheap and straightforward to implement, resolving this ambiguity in practice may yield substantial benefit in disciplining community reports. To assess the relative strength of these competing effects, we randomly vary whether the peer elicitation exercise takes place in a private or public setting.

Paying for Truthfulness. Explicit monetary incentives for accuracy offer a promising deterrent to misreporting. One straightforward way to implement monetary incentives would be to pay respondents based on the closeness of their reports with an ex-post measure of accuracy. But often ex-post measures of accuracy are unavailable, or prohibitively costly to collect (such as in the case of estimating marginal returns to capital, which can never be confirmed for an individual entrepreneur). Even when signals of ex-post accuracy exist, using them necessitates a time-lag between the moment of elicitation and subsequent payment for reports. In settings with weak institutions, where trust in outsiders is minimal, respondents may demand to be paid contemporaneously with their reports. To circumvent these concerns we evaluate monetary incentives delivered via a peer prediction scheme, which rewards respondents based exclusively on their own reports and the contemporaneous reports of their peers. The particular payment rule we use is the *Robust Bayesian Truth Serum,* described in detail in the next section.

Zero-sum Elicitation.   During our peer elicitation exercise, entrepreneurs rank one another on metrics of business growth and profitability. Respondents are assigned to groups based on geographical proximity and each person ranks herself and the other members of her peer group

8

(see Section 4.1 for details on our elicitation intervention). Within each $4 - 6$ person group of entrepreneurs, we evaluate two forms of community rankings: rankings relative to the particular members of the group, and reports placing each entrepreneur in quintiles relative to the community at large. The former has a zero-sum nature, in which promoting someone's position necessitates diminishing another's, and may therefore be more effective at inducing truthful reports (a respondent cannot merely place everyone in the highest position). However, if group members have correlated attributes, then these rankings may be less informative than rankings that assess each entrepreneur relative to the broader community. By examining both mechanisms we investigate which of these concerns dominates in practice.

Cross-Reporting. In the spirit of cross-reporting techniques which play a prominent role in mechanism design and implementation theory (see Maskin (1999)), we ask respondents to identify each group member's closest peer in the group, with the intention of exploring whether group members identified as close peers distort their reports to favor one another. We also ask respondents to identify who in their peer group has the most accurate information regarding each ranking metric.

3.1 The Robust Bayesian Truth Serum

Peer prediction mechanisms, including Witkowski and Parkes' (2012) *Robust Bayesian Truth Serum* (RBTS), incentivize truthful reporting of beliefs without reference to ex-post measures of accuracy.[5] Instead, these mechanisms determine payments as a function of the contemporaneous reports of several respondents.

We implemented a variant of RBTS, which requires elicitation of agents' first order beliefs (the ranking that an agent assigns to each of his peers) and second order beliefs (the probability distribution the agent assigns to each possible ranking his peers may give one another). RBTS rewards an agent's second order beliefs based on their proximity to the empirical distribution of stated first order beliefs. First order beliefs are evaluated based on how "surprisingly common" they are relative to other agents' stated second order beliefs. That is, agents are compensated for first order beliefs that have empirical frequencies higher than predicted by other agents' stated second order beliefs. Witkowski and Parkes (2012) show that under the assumption of a common and admissible prior, truthful reporting is a Bayesian Nash Equilibrium. See Appendix for details on RBTS as well as an intuition for its incentive compatibility.

Implementation of the Robust Bayesian Truth Serum. Peer prediction methods are attractive because they make truthtelling incentive compatible and circumvent the need for ex-post verification of outcomes. The principal challenge to implementation of RBTS is its complexity. It is infeasible to describe RBTS (and its incentive comparability) to respondents in our setting who are largely innumerate. This is a challenge shared by many mechanisms implemented in practice

---

[5]See Prelec (2004) for a seminal contribution to this literature.

(most notably, two-sided matching algorithms, versions of which are commonly used in education and entry-level labor markets). A common tactic, which we take in this study, is simply to assert to respondents that they can do no better than to tell the truth.[6]

In Rigol and Roth (2017) we provide evidence that this is a reasonable tactic. We report on an experiment among a sample drawn from a very similar population to that of our current study, in which compare the accuracy of peer reports when paying agents for truthfulness using a straightforward payment rule based on ex-post accuracy and when paying agents using peer prediction mechanisms. Surveyors carefully and completely explained the ex-post payment rule to respondents. For the peer prediction method, surveyors simply asserted to respondents that they would maximize their incentive payments by telling the truth. We elicit information regarding borrower reliability and entrepreneurial ability and we find that the additional accuracy induced by the simple ex-post incentive is statistically and economically indistinguishable from that induced by the peer prediction method. Both payment methods led to significantly more accurate reports than elicitation without monetary payments.

That respondents believe our assertion that they should tell the truth is reassuring, but it may nevertheless be desirable to verify that RBTS's theoretical properties hold in practice. While RBTS is incentive compatible in theory, it may be that given the empirical distribution of beliefs, respondents can indeed increase their payoff with deceptive reports. In Rigol and Roth (2017), we verify that the payment method is incentive compatible in practice. To do so, we estimate the higher order beliefs of respondents in the sample and used these beliefs to determine respondents' subjective expected payments from RBTS. Details of this exercise are replicated in the Appendix of this paper.

That RBTS is incentive compatible in practice is encouraging 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, that assertion will only be reinforced with repeated use — because RBTS is incentive compatible, agents will receive experiential feedback over time that truth-telling is the highest paying strategy.

---

[6]The National Resident Matching Program, which matches new physicians to residency spots in the United States, has a video explanation of the steps involved in the mechanism and advises physicians that "To make the matching algorithm work best for you, create your rank order list in order of your true preferences, not how you think you will match." The video explanation and accompanying instructions do not attempt to explain why truthtelling is a dominant strategy. The website is: staging-nrmp.kinsta.com/matching-algorithm. For the Boston Public Schools matching system, parents are told "List a number of choices (BPS recommends at least five) and order them in the true order of preference to increase the chances of getting the school that you want."

4 Experimental Design

4.1 Design of the Peer Elicitation Exercise

Recruitment. In October 2015, we visited the 1,576 eligible households and invited them to participate in our study. At the time of recruitment, households were told that a research team was conducting a project to study entrepreneurship and business growth.[7] In December 2015 - April 2016, we conducted baseline surveys of the 1,345 sample households. Separately, we also assigned respondents to groups of five based on geographic proximity, for a total of 274 groups across all neighborhoods.[8] Once all baseline surveys in a given neighborhood were complete, surveyors returned to sample households to invite respondents to a meeting at the local town hall. Respondents were not given any information regarding the content of the meeting, or that they would be placed into groups with their peers. They were told, though, that to thank them for their participation in the study the research team would conduct a public lottery where some participants would be awarded a USD 100 grant.

Explanation of the Exercise to Respondents. Upon arrival at the town hall, respondents were each given 20 lottery tickets. They were told that, at the end of the activity, all people present would put their lottery tickets into an urn and grant winners would be selected by drawing lottery tickets. Participants were then separated and individually paired with a surveyor. Surveyors explained to participants that they would be asked to provide information about themselves and their neighbors. In order to ensure that participants were introduced to the elicitation exercise in a clear and consistent way, we created animated videos to introduce respondents to the concepts covered in the rankings questions and to guide them through the activity. When explaining the concept of marginal returns, we used examples to emphasize to respondents that an entrepreneur's projected marginal returns corresponds to their expected *change* in profits in response to the grant, and not their *level* of profits. After watching the videos, participants completed a series of quizzes to test their understanding of the activity and concepts. The introduction and subsequent ranking activity took place behind a privacy screen. The screen was there to ensure that coordination of responses would not be possible (as explained below, respondents in the public reporting treatment were later randomly assigned to complete a subset of their rankings among their peers). Surveyors also told participants which of their neighbors they would be ranking and gave them four to six placards, each with the name of a group member.

Questions Asked in the Ranking Exercise. First, we asked participants to rank themselves and their peers on predicted marginal returns to a USD 100 grant. We then asked respondents to rank themselves and their peers across several additional entrepreneur characteristics: educational

---

[7]No information regarding the community information nature of the project was disclosed to respondents at this time.

[8]We organized respondents into groups that would minimize the geographic distance between study households. The total number of respondents per neighborhood was not always a multiple of 5, so some groups had 4 or 6 clients.

attainment; average number of hours spent at work per week; performance in a digit span memory test; and, projected monthly profits 6 months post-grant disbursal, if the business owner were to receive a USD 100 grant. We also asked about a number of household-level characteristics: average monthly income over the past year, total value of assets; total medical expenses in the past 6 months; and, loan repayment trouble over the previous year. Finally, we asked respondents to report on whom they thought was most deserving of the grant. We deliberately did not provide any additional description or criteria to this question and instead emphasized that respondents should select based on criteria that they thought were important for this metric. Note that to minimize respondent fatigue peer groups completed the ranking exercise only for a randomly assigned subset of these metrics (but all respondents completed the marginal returns ranking). For details on the assignment of ranking questions by treatment group, see the Appendix. And, participants completed both relative and quintile rankings for questions on marginal returns, business profits, and household income and assets, but only relative rankings for the remaining questions (this was also done to reduce fatigue). Finally, respondents were asked to cross-report on their peers: they identified one another's closest peer in the group and, for each ranking question, respondents identified the group member they believed would have the information required to answer the question most accurately.

4.2 Description of Treatments

Respondents were cross-randomized (at the group level) to give their ranking reports under the following three treatment conditions, for a total of eight treatment cells: *No Stakes* vs. *High Stakes* (S_0 vs. S_1), *Private* vs. *Public* (P_0 vs. P_1), and *No Payments* vs. *Payments* (T_0 vs. T_1). We also randomly selected one-third of our sample to receive USD 100 grants. Grant randomization occurred at the individual level and was stratified by group. See Figure 1 for the randomization design.

High Stakes Environment (S_0 vs. S_1). For this treatment, participants were told that their responses in the ranking exercise would help determine the winner of the lottery that would occur at the completion of the activity. All participants across treatment groups were given twenty lottery tickets upon arrival at the town hall. Respondents in the high stakes treatment were told that, for each question, the peer ranked highest (on average) by group members would receive extra lottery tickets, and so would have a better chance of winning.[9] In order to ensure that we would have sufficient power to test predictions from the marginal returns rankings, all participants completed this ranking in a no-stakes setting (the marginal return ranking occurred prior to other rankings and to any mention of the high stakes treatment).[10]

---

[9]We did not tell participants how many extra lottery tickets would be awarded to the person ranked highest; in order to keep the randomization as close to uniform as possible, we awarded only one extra lottery ticket per ranking. Respondents were in a high stakes setting for four ranking questions, and so a person in this treatment group could win at most four extra lottery tickets. Participants completed all rounds of ranking questions prior to the disbursal of the extra lottery tickets.

[10]Measures of profits among microentrepreneurs in settings like this one are notoriously noisy (see, for instance,

Public Reporting (P_0 vs. P_1). Participants in both the Public and Private Reporting groups responded to each ranking question behind a privacy screen, in the presence of only their surveyor. But in the Public treatment, after completing each ranking question, peers came to the center of the room and sat in a circle with their response clipboard in front of them. Participants were told that they were doing this so that the survey coordinator could record their responses, but the true purpose was to give them the opportunity to observe one another's rankings.[11] Crucially, participants understood ahead of doing the ranking exercise that their peers would see their responses. This was described to them in their introductory animation video and, to ensure that participants understood the set-up, groups performed several practice rounds. In the privacy treatment, respondents completed all ranking questions before interacting with peers and, even after the activity was completed, group members did not see each other's individual responses.

Payments for Truthfulness (T_0 vs. T_1). The introductory video for participants in the monetary incentives group explained that respondents would be paid per ranking question, based on the truthfulness of their responses. As explained in Section 3, we did not explain the details of the RBTS scoring rule to participants. Instead, participants were told that people who reported what they truly believed would receive an extra Rs. 100 on average (which is equivalent to 2/3 of the average daily wage). Payments were calibrated using the empirical distribution of beliefs from Rigol and Roth (2017) to maximize strength of the incentive to tell the truth while adhering to a project budget constraint. Since RBTS incentive payments require respondents to report second-order beliefs, the introductory video also explained this concept to participants. Respondents expressed their beliefs about the distribution of first-order beliefs in the community by allocating coins to quintile bins. There were 20 coins to distribute, representing 20 hypothetical community members. For example: if $respondent_i$ believed that 8 out of 20 community members would rank $peer_j$ in the top quintile for household assets, she would place 8 coins into the top quintile bin, and so on. Groups that were not in the monetary payments treatment were not asked to report second order beliefs and were not paid for each ranking report; instead, they were given a lump sum payment to compensate them for their time.

Enterprise Grant. Upon completion of the peer elicitation exercise, group members came to the center of the room and placed their lottery tickets into an urn. One respondent was blindfolded and then drew tickets to award USD 100 grants to one or two group members (the number of winners per peer group was determined by random assignment). When surveyors earlier visited respondents to tell them the date of their town-hall meeting, they also gave respondents a business plan worksheet. Surveyors reminded respondents that grants were meant to be used for business purposes (respondents had been told about the grant lottery at the time of recruitment) and instructed them to describe in the worksheet what they would do with the grant money, if they

---

De Mel et al. (2009)). Due to budget constraints, our experiment is just powered to detect how well marginal returns rankings predict realized marginal returns when accuracy of reports is not confounded by the incentive to lie present in a high-stakes setting.

[11]Surveyors report that respondents did in fact almost always look at their peers' rankings.

won. Participants then brought their completed worksheets to the town-hall meeting and winners were again reminded of the grant purpose (but we did not enforce that winners put their winnings towards their business). Grant money was distributed to winners via bank transfer.

4.3 Identification Strategy

Random assignment allows us to use the difference between post-period profits of grant winners and post-period profits of grant losers as an estimate of the true average marginal return to the grant. We therefore identify the informational value of community members' reports by testing the predictive power of respondents' marginal return rankings against true marginal returns.

Next, we assess whether community information extraction is susceptible to strategic misreporting when allocation of resources is on the line. We measure accuracy by comparing peer reports to self-reported values that participants provided at the time of the baseline survey.[12] By comparing accuracy of peer reports for participants in the *No Stakes* and *High Stakes* groups (S_0 vs. S_1), we identify the effect on strategic misreporting of shifting the elicitation environment to one in which reports can have consequences for allocation of grants.

Finally, we measure the efficacy of mechanisms to realign incentives for truthful reporting: a comparison of the accuracy of peer reports in the *Private* versus *Public* treatments (P_0 vs. P_1), or in the *No Payments* versus *Payments for Truthfulness* treatments (T_0 vs. T_1), identifies the effect each of these mechanisms has on respondents' truthfulness. Because we cross-randomize treatments, we can separately identify the strength of these mechanisms in the benchmark, *No Stakes* setting, and in the *High Stakes* setting, where respondents have a counteracting incentive to lie.

5 Data and Randomization Checks

Description of the Data. Our main analysis uses data from respondents' peer rankings during the elicitation exercise and from respondent surveys. Baseline surveys were conducted between December 2015 and April 2016, and three follow-up surveys were conducted between May 2016 and March 2017. For all survey rounds, each business owner in the household completed a detailed business module about her own enterprise and answered questions about her well-being. The business module included questions on enterprise costs; revenues; profits; seasonality; inventories; labor inputs; assets; and business history. At baseline, entrepreneurs also completed a digit span test and a set of psychometric questions.[13] In each survey round, the study respondent also provided

---

[12] In order to ensure that we would have sufficient power to test predictions from the marginal returns rankings, all participants completed this ranking in a no-stakes setting.

[13] Respondents answered each psychometric question in the module by providing their agreement with the given statement, where agreement was rated on a scale of one to five, with five indicating strong agreement and one indicating strong disagreement. A detailed description of the psychometric assessment module is in the Appendix. The psychometric module questions are organized according to categories developed by industrial psychologists: polychronicity measures the willingness to juggle multiple tasks at the same time (Bluedorn et al. 1999); impulsiveness

information regarding her household's finances. The household-level module included questions on income, health expenditures, credit history and loan repayment issues, and assets. For the asset section, the respondent indicated whether the household owned a particular type of asset and its current resale value. Surveyors were trained to visually verify that the household owned each of the assets about which they reported. At baseline, the respondent also completed a full household roster with education and labor history for each household member. For a complete timeline of the project and data, please see Figure 2.

Randomization Checks. In Appendix Table 1, we present the randomization check of baseline characteristics by treatment. To check for balance we estimate the model,

$$Characteristic_{ij} = \tau_0 + \tau_1 Treatment_j + \epsilon_{ij}$$

where i indexes the individual and j indexes the group. $Treatment_j$ is a dummy for whether the group was assigned to the $No\,Stakes$" vs. $High\,Stakes$ treatment (columns 1 and 2), the $No\,Payments$ vs. $Payments$ treatment (columns 3 and 4), and the $Private$ vs. $Public$ treatment (columns 5 and 6).

The odd columns 1-7 show the average of each characteristic for the control group in each block. So column 1 shows the means of characteristics for groups that were assigned to $No\,Stakes$. The even columns show τ_1 for each treatment (the difference between treatment and control characteristics) . The characteristics in *Panel A* are about the entrepreneur who was ranked during the ranking exercise and in *Panel B* are about her primary business. In *Panel C*, we show household level baseline measures. The variables "Value of Business Assets" and "Avg. Monthly Profits" are shown as aggregates over all household businesses. So if the the ranked entrepreneur is the only business owner in the household, these reflect the values of only her businesses.

The majority of entrepreneur and household characteristics are balanced across treatment groups. Entrepreneurs assigned to *Payments* report lower household monthly income and entrepreneurs assigned to *Public* report lower value of household assets. At the bottom of the table, we presents the p-value from an F-test of whether the treatment group coefficients are jointly equal to zero. All of the joint tests of equality are rejected, suggesting that the randomization was effectively implemented.

We do not see significant differences between lottery winners and losers on income, profits, and assets, which is important since for the $High\,Stakes$ group the grant was partly allocated using rankings for those three variables. Nonetheless, as we discuss below, to account for this feature of the design, we weigh all regressions that exploit variation between lottery winners and losers

is a measure of the speed at which a person makes decisions and savings attitudes (Barratt Impulsiveness Scale); tenacity measures a person's ability to overcome difficult circumstances (Baum and Locke 2004); achievement is a measure of satisfaction in accomplishing a task well (McClelland 1985); and locus of control measures a person's willingness to put themselves in situations outside of their control (Rotter 1996).

by the inverse number of lottery tickets that each person received. We also present the results of a joint test of statistical significance and cannot reject that all groups are drawn from the same population.

6   Background Results - Entrepreneurs' Community Knowledge

We begin our empirical analysis by investigating the depth of community members' knowledge of one another. As discussed in Section 2, entrepreneurs have close social ties with peers in their neighborhood. Over half of respondents regularly discuss private family and business matters with at least one other group member and on average group members visit each other 22 times per month. In this section, we show that community members' frequent interactions and social ties lead them to have accurate knowledge about one another's concurrent household finances and enterprise characteristics. In our main empirical analysis (Section 7), we will argue that community members also have the depth of knowledge required to make predictions about entrepreneurs' marginal returns.

During the ranking exercise, community members reported on their peers' average monthly household income; predicted monthly profits if they were to receive a USD 100 grant; total value of household assets; household medical expenses over the previous six months; average weekly work hours; and, predicted performance on a working memory test.[14] At baseline, we asked each entrepreneur to self-report answers to these same questions (at the time of the baseline survey, respondents had no knowledge of the purpose of the study or of the peer ranking activity). To evaluate the accuracy of community reports, we estimate the distance between entrepreneurs' self reports and community members' reports for that person. We use the following regression model:

$$Y_{ijq} = \beta_0 + \beta_1 \overline{Rank}_{ijq} + \gamma_n + \theta_m + \tau_s + \epsilon_{ijq}, \tag{1}$$

where $Rank_{ijkq}$ is the rank that person k in group j assigns to person i (also in group j) on question q. $\overline{Rank}_{ijq} = \sum_{k=1}^{n} \frac{1}{n} * Rank_{ikjq}$, where n is the total number of group members in group j. So $\overline{Rank}_{ijq}$ is the average rank assigned to person i by the members of group j on question q. Y_{ijq} is the corresponding outcome (baseline survey self report) for question q of person i. For example, if the outcome is household income ($Y_{ijincome}$), the corresponding average rank is the average household income ranking given to person i by her peers ($\overline{Rank}_{ijincome}$). To improve precision, we add neighborhood (γ_n), survey month (θ_m), and surveyor (τ_s) fixed effects. Standard errors are clustered at the group level.

In Table 1, we present the estimates of Specification 1.[15] To allow for comparability of esti-

---

[14]We use a digit span test, which is a commonly used test for working memory. Respondents are shown flashcards with an increasing number of digits and asked to recall the numbers from memory. The surveyor records the total number of digits that the respondent correctly repeated back.

[15]In Table 1, we pool across all treatment groups: *No Stakes* vs. *High Stakes* treatment, the *No Payments* vs.

mates across questions, in *Panel A* we convert each outcome and the corresponding average rank for each question into percentiles. So, a 1 percentile increase in $\overline{Rank}_{ijq}$ is associated with a β_1 percentile increase in the outcome variable Y_{ijq}. In *Panel* B, we present results in levels of the outcome and the average rank, so that a 1 unit increase in $\overline{Rank}_{ijq}$ is associated with a β_1 increase in the value of the outcome variable Y_{ijq}.

Entrepreneurs report on their peers' household and enterprise characteristics with remarkable precision. For example, in Column 3 of *Panel A*, a 1 percentile increase in the assets rank is associated with a 0.23 percentile increase in the distribution of actual household assets. They can also accurately assess even difficult to observe characteristics: for instance, a one unit increase in the average rank level is associated with a 0.63 extra digits recalled in the Digit Span Memory Test (Column 5 of *Panel B*): moving from the 5th percentile to the 95th percentile in average digitspan rank is associated with a doubling of the total number of digits an entrepreneur recalls. To contextualize the size of these estimates, we regress the business profits percentile on the percentile of the education of the entrepreneur: a 1 percentile increase in the education distribution is associated with a 0.12 percentile increase in the distribution of business profits. In other words, the community information is nearly twice as useful in predicting business profits as as the entrepreneur's own level of education.

7 Main Results

7.1 Entrepreneurs' Average Marginal Returns to Capital

In the next section, we will investigate whether community members can accurately predict one another's returns to the grant. First, we assess the average impact of the intervention on entrepreneurs' profits. Following de Mel et al. (2008), we estimate average marginal returns to the grant with the primary specification,

$$Y_{ijt} = \alpha_0 + \alpha_1 Winner_{it} + \gamma_i + \sum_{t=1}^{3} \delta_t + \theta_m + \tau_s + \epsilon_{ijt}. \tag{2}$$

Y_{ijt} measures either total household business profits or household income of person i in survey round t.[16] We measure business profits by asking entrepreneurs the following question: "Now that you have thought through your sales and your expenses from the past 30 days, I would like you to think about the profits of your business. By business profits, I mean taking the total income received from sales and subtracting all the cost of producing the items (raw material, wages to employees, fixed

---

Payments treatment, and the *Private* vs. *Public*. In Sections 7.5 and 7.6, we break these estimates up by treatment.

[16]Bernhardt et al. (2017) reanalyze data from several cash-drop experiments with microentrepreneurs and find that measures of returns to capital differ substantially when analyzed at the household versus enterprise level. We therefore aggregate profits of all household businesses, for all specifications.

costs, etc). Can you tell me your business profits in the past 30 days?"[17] Household income is also measured using a single question: "What is your total household income over the past 30 days from all income generating activities?" Like de Mel et al. (2008), we remove the outliers of the household income and total profits distributions (levels) by trimming the top 0.5% of both the absolute and percentage changes in profits measured from one period to the next. We also estimate regression Specification 2 for $log(Y_{ijt} + 1)$ of income and profits, using the untrimmed distributions.[18] In the main specification, we utilize three rounds of follow-up surveys, so t ranges from 0 (baseline) to 3. [19]

$Winner_{it}$ is an indicator for whether person i won a grant at or before survey round t. Note that $Winner_{it}$ is 0 at period $t = 0$ for all persons i. We also include the following fixed effects: person (γ_i), survey round (δ_t), survey month (θ_m), and surveyor (τ_s). Standard errors are clustered at the group level. The coefficient of interest in regression Specification 2 is α_1, which measures average marginal return to the grant in the sample.

In the *No Stakes* treatment group, assignment of grant winners was uniformly random: all participants received twenty lottery tickets and each group member was equally likely to have their tickets drawn from the urn. But, as described in Section 4.2, respondents in the *High Stakes* group were eligible to receive up to four extra lottery tickets, based on whether their peers ranked them highest for the treatment questions.[20] To account for this, we weigh all regressions by the inverse propensity score.(Rosenbaum, 1987). In Appendix Figure 2, we plot the distribution of lottery tickets in the sample.

Table 2 presents results from estimating Specification 2. We find that the grant had a large positive effect on household income and total household profits. On average, households that win grants report an extra Rs. 422.3 in household income and an extra Rs. 507.7 in total household profits over households that were not awarded grants. These gains in household income and profits represent very high marginal returns to the grant: on average, households earn returns of $7.6\% - 8.6\%$ per month.[21] These estimates are in line with average returns estimated from cash grants in other settings: de Mel et al. (2008) find that profits increase by 7.6% per month in response to a USD 100 grant and Fafchamps et al. (2014) show that profits increase by 9.7% per month in

---

[17](De Mel et al., 2009) find that asking one aggregate summary measure (rather than for the components) reduces noise in the estimation of profits.

[18]The results remain nearly identical whether we log-transform the trimmed or untrimmed income and profits distributions.

[19]The month before we began our fourth (last) round of data collection, the Indian government removed from circulation two currency notes - the Rs. 1000 and Rs. 500 bills - overnight. The result was a tremendous shock to the formal and informal economy. As Banerjee and Kala (2017) report, traders experienced a 20% drop in sales due to demonetization. In fact, in the last round of surveying, over 50% of our sample reported being adversely affected by demonetization. For this reason, we exclude the post-demonetization wave of data from the analysis presented in the main tables. We replicate all the main tables with all five data rounds in Appendix Tables 16-19. The results are qualitatively identical but marginally noisier in a few specifications.

[20]For a more detailed description of the *High Stakes* treatment, please refer to Section 4.2.

[21]We arrive at this number by dividing the marginal increase in monthly income and profits by the size of the grant (Rs.6000).

response to a USD 120 grant.

The average yearly return among entrepreneurs in our sample – between 91.2 and 103.2% – greatly exceeds the market interest rate for microfinance loans in Maharashtra. In 2016, the average yearly APR for microcredit in India was 24%.[22] And, we find high demand for credit: 92% of entrepreneurs in our sample report having a desire to borrow. There is also deep microfinance penetration in Maharashtra: in 2015, there were 27 MFIs operating there, the highest level of any North Indian state.[23] Still, fewer than 20% of our sample report having borrowed from a microfinance institution.

The persistent frictions in Maharashtra's microfinance market might be linked to the large variation in entrepreneurs' returns. Though the average marginal return in our sample is very high, there is also evidence to suggest substantial heterogeneity. We plot quantile treatment effects from the grant and find that returns from 0% to 28% per month (Appendix Figure 1).[24] We replicate this exercise using data from de Mel et al. (2008) and Fafchamps et al. (2014) and find similar results. Among the de Mel et al. (2008) sample, monthly returns range from 0% to 45% and in the Fafchamps et al. (2014) sample they vary between 0% and 30% per month. But MFIs do not currently have a cost-effective method of screening applicants and, instead, they typically offer all borrowers a standard contract which involves a small loan and frequent repayments (Banerjee, 2013).

Lending institutions' asymmetric information problem motivates our main empirical analysis. In the next section, we ask: can the variation in entrepreneurs' returns be predicted ex-ante? Do community members have the depth of knowledge required to identify entrepreneurs with the most potential to grow? Next, we investigate whether community reports can accurately predict the marginal returns of entrepreneurs.

7.2 Can Communities Predict Entrepreneurs' Marginal Returns To Capital?

We have shown that community members have detailed and accurate information about one another's household finances and enterprise attributes. Now, we assess whether entrepreneur peers can accurately identify whom among them has the most potential to grow.

Our measure of community knowledge is entrepreneurs' average marginal returns rank.[25] In

---

[22]This estimate comes from the Bharat Microfinance Report (2016).

[23]Religare Capital Markets, India Microfinance Sector Market Report 2015.

[24]Though quantile regressions provide suggestive evidence that there may be heterogeneity in returns in the sample, they cannot be interpreted causally without imposing strong assumptions. See Abadie et al. (2002) for a detailed discussion on the interpretation of quantile regressions.

[25]We collect both zero-sum and quintile community ranks. Results are qualitatively similar with both ranking methods but, because there is heterogeneity in peer groups' average marginal return to capital, we find that quintile ranks are a more accurate assessment of an entrepreneur's returns relative to the community. All analysis in Section 7.2 uses the quintile community rankings. Results using the zero-sum rankings are in Appendix Tables 20-22. Section 7.5 contains a more detailed discussion of the two ranking methods.

the peer ranking activity, entrepreneurs were placed into groups with four to six of their nearest neighbors. Respondents were asked, "Could you please rank your group members in order of who you think had the highest marginal returns to the Rs. 6,000 grant? In other words, who would gain the most in monthly profits, or who would grow their business the most, from receiving a Rs. 6,000 grant?" An entrepreneur's average marginal returns rank is the mean of every rank assigned to her by her group members. We plot the distribution of average rank, which takes on values between one and five, in Appendix Figure 3. Since group members are in full agreement about an entrepreneur's rank in fewer than 15% of cases, the distribution of average marginal return rank values is relatively smooth.

Figure 1: Marginal Returns to the Grant by Percentile of the Community Ranks Distribution
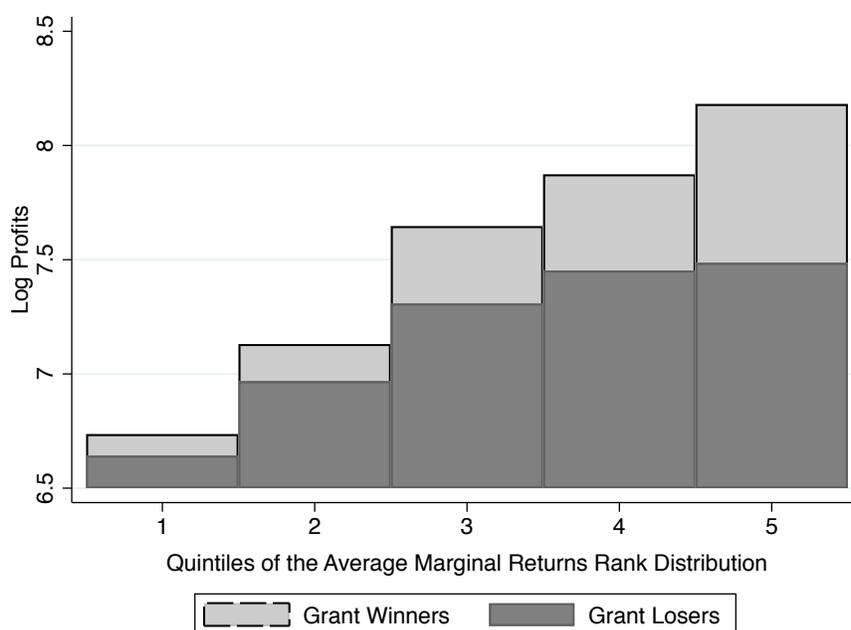
Notes: This figure plots the log of average post-grant profits (y-axis) by quintiles of the average marginal returns rank distribution (x-axis). The dark gray bars correspond to log profits of entrepreneurs who did not win grants and the light gray bars correspond to log profits of entrepreneurs who did win grants. Marginal returns rank is the rank assigned by a peer when asked "Could you please rank your group members in order of who you think had the highest marginal returns to the Rs. 6,000 grant? In other words, who would gain the most in monthly profits, or who would grow their business the most, from receiving a Rs. 6,000 grant?" Average marginal returns rank is the mean of the marginal returns ranks assigned to an entrepreneur by her peers and by herself.

In Figure 1, we plot the log of post-grant profits by grant treatment assignment and by quintile of average marginal returns rank. Each bar corresponds to average post-grant profits for entrepreneurs in a given quintile of the marginal returns rank distribution. Dark gray bars are profits of grant losers and light gray bars are profits of grant winners. We find that the gap in post-period profits between grant winners and grant losers – in other words, entrepreneurs' marginal

20

return to the grant – is increasing in the community's rank report.

Figure 1 suggests both that there is significant heterogeneity in returns to the grant and that community members are accurately able to identify the ordering of their peers' heterogeneous returns ex-ante. We further illustrate the striking accuracy of community members' predictions in Figure 3. In that figure, we plot kernel-weighted local polynomial regressions (degree 1) of log profits at follow-up for grant winners and for grant losers on average marginal return rank percentile. We find that an entrepreneur's marginal returns rank is strongly correlated with her realized profits in response to the grant: below the 35th percentile of the ranks distribution, post-grant profits for winners and losers are statistically indistinguishable. But above the 35th percentile, the distance between treatment and control profits increases with marginal returns rank – this increasing distance is a measure of respondents' prediction accuracy.

Our main specification is a difference-in-differences estimation of the relationship between community ranks and marginal returns to the grant. We extend the model from Specification 2 to incorporate peer ranks:

$$Y_{ijt} = \alpha_0 + \alpha_1 Winner_{it} + \alpha_2 Winner_{it} \times \overline{Rank}_{ij} + \gamma_i + \sum_{t=1}^{3} \delta_t + \theta_m + \tau_s + \epsilon_{ijt}. \tag{3}$$

$Rank_{ikj}$ is the rank that person k in group j assigns to person i (also in group j). $\overline{Rank}_{ij} = \sum_{k=1}^{n} \frac{1}{n} * Rank_{ikj}$, where n is the total number of group members in group j. So $\overline{Rank}_{ij}$ is the average marginal returns rank assigned to person i by the members of group j. The coefficient α_2 identifies the average additional marginal return to capital associated with a one unit increase in marginal return rank. The difference-in-differences specification estimates α_2 for a model in which marginal return increases linearly in the value of average rank. We also estimate a non-linear model in which the ranks distribution is divided into terciles and rank tercile is interacted (as above) with $Winner_{it}$. In Appendix Table 2, we show that the sample is balanced across rank terciles and grant treatment groups at baseline. In Appendix Figure 4, we replicate Figure 3 with baseline profits and show that differences in marginal returns to the grant are not driven by baseline differences in profits.

Table 3 shows results of the difference-in-differences estimation of respondents' ability to predict true marginal returns to capital. Outcome variables are household income and total household profits, in both levels and logs. For the linear-in-rank version of the estimation (*Panel A*), the coefficient α_2 is large and positive for all four outcome variables. Coefficients for income and log income are both significant at the 5% level; for profits, levels are noisy and not significant but the coefficient for log profits is significant at the 1% level. An extra unit of average rank is associated with increases in profits and income of between Rs. 283.2 and Rs. 848.1 per month. These amounts translate to increases in monthly returns to the grant of between 4.7% and 14%. Average marginal return to capital in the sample is about 7.1% per month and an entrepreneur ranked one standard deviation above the mean has monthly marginal return to capital of 16.4% (the mean and standard

deviation of the marginal return rank are 3.46 and 0.66, respectively). For an entrepreneur ranked two standard deviations above the mean, monthly returns to capital are 25.7%.

Panel B in Table 3 shows results from the non-linear, tercile rank version of the difference-in-differences estimation. Consistent with results from the local polynomial regressions in Figure 3, we cannot reject that the entrepreneurs in the bottom tercile of the marginal returns rank distribution have zero returns to the grant. For three of the four outcome variables (all but level of household profits), the coefficient on $Winner_{it}$ actually implies a negative return to the grant. Also consistent with Figure 3, the coefficients on log income and log profits for the middle tercile are positive, but not significant, and the level effects are almost precisely zero.[26] The strongest treatment effects of the grant are concentrated among entrepreneurs in the top tercile of the average rank distribution: depending on whether we use household income or profits, the coefficients on $Winner_{it} \times Top\overline{Tercile}_{ij}$ imply that monthly returns to the grant for the top tercile range from 16.6% to 26.7%. We can statistically reject that the grant has the same effect for entrepreneurs in the middle and top tercile.

Regression Specification 3 identifies the treatment effect of the grant off of the within-person differences in profits and income in the pre- and post- grant disbursal periods for grant winners and losers. As a robustness check, we also present results using an alternative specification in which the treatment effects are identified by comparing the cross-sectional differences between treatment and control groups in the post-grant disbursal periods, controlling for the baseline value of the outcome characteristic. Our specification is:

$$Y_{ijt} = \beta_0 + \beta_1 Winner_{ijc} + \beta_2 Winner_{ijc} \times \overline{Rank}_{ijc} + \beta_3 \bar{Y}_{ijPRE} + \sigma_c + \theta_m + \tau_s + \epsilon_{ijt}, \quad (4)$$

where Y_{ijt} are post-treatment outcomes (so t ranges from 1 to 3 rather than 0 to 3 as in Specification 3) and $\bar{Y}_{ijPRE}$ is the pre-treatment (time 0) value of the outcomes. σ_c is a neighborhood cluster fixed effect. Standard errors are clustered at the group level. We present the analogue of Table 2 using Specification 4 in Appendix Table 4 and the analogue of Table 3 in Appendix Table 5. Results in the robustness specification are qualitatively very similar in terms of the size and significance of coefficients.

Throughout our main analysis, our calculation of respondents' average rank includes their self-rank. The impact of including respondents' self-rank on community rank accuracy is ex-ante ambiguous. We might expect entrepreneurs to have better knowledge about themselves than they do about others. But it is also reasonable to assume that respondents will be more likely to strategically misreport in favor of themselves than when reporting about others. In Figure 4, we investigate the impact of self-rank on the community's accuracy. We replicate the local polynomial regression of log profits at follow-up on marginal return rank percentile (as in Figure 3) with three

---

[26]Mechanically, since the middle tercile is fixed, the difference between the level and log results occurs because there are some extreme right-tail observations in the distribution of income and profits for the middle tercile ranks. The weight of these outliers in the regression is diminished when the distributions are log-transformed.

specifications of the rank variable: (i) average rank including self-rank (*Panel 1*), (ii) average rank excluding self-rank (*Panel 2*), and (iii) only self-rank (*Panel 3*).[27] Results shown in *Panel 1* and *Panel 2* are very similar, which indicates that entrepreneurs have strong knowledge of their peers and that community rank accuracy is not driven by the information contained in self-rank. We also see in *Panel 3* that entrepreneurs are able to predict their own marginal returns to the grant, though fewer entrepreneurs give themselves low rank values and so the correlation between the self-rank marginal returns prediction and actual marginal returns (the vertical distance between the profits of grant winners and losers) is weaker than it is with the average rank prediction. Finally, in Appendix Table 3, we replicate the results of Table 3 but exclude self-rank from the calculation of average rank. We find that results are nearly identical to those presented in Table 3, which again indicates that peers do indeed have important and valuable information about one another. In Sections 7.5 - 8, we further discuss the knowledge that entrepreneurs have about themselves and manipulation in self-reports.

Across specifications, we find that communities have deep knowledge of entrepreneurs' growth potential. Importantly, community members' predictions map to economically significant differences in returns. Lending institutions would have good reason to target top-ranked entrepreneurs for credit: in 2016, the average yearly APR for microcredit in India was 24%. Entrepreneurs in the top tercile of community ranks earn *monthly* returns of 16.6 to 26.7%.[28] But the point estimates in Table 3 also imply that entrepreneurs in the bottom tercile (and perhaps also middle tercile) may not have been able to cover the cost of a USD 100 loan without reducing their total household consumption (since these entrepreneurs do not increase their profits in response to the capital intervention). Thus for a lending institution, distributing credit without screening while maintaining a constant interest rate could lead to a substantial increase in risk.

In the next section, we explore whether differences in entrepreneurs' investment decisions can help explain the large gaps in returns that we observe. For ease of exposition, the remaining main tables show only the rank tercile specification. All tables with the linear-in-rank value specification can be found in the Appendix.

7.3 Who are the Top-Ranked Entrepreneurs?

In the previous section, we showed that the large variation in marginal returns to capital observed in our sample and in other similar cash-drop studies can be predicted ex-ante. Why, then, are some entrepreneurs likely to earn such high returns while others are not? In this section, we

---

[27]Unlike average rank, which is the mean of 4-6 reports, the self-rank value is the result of a single report. As such, the self-rank variable only takes on integer values. For consistency across regressions in the three panels, we use rank value (rather than rank percentile as we did in Figure 3). As can be seen in Appendix Figure 3, there are few observations with a rank value below two. We therefore bottom code all three measures of rank.

[28]There are many possible reasons why a loan might have induced different selection and investment patterns, but it is useful to benchmark entrepreneurs' returns against market rates. See (Fiala, 2013) for an experiment which randomly allocates loans or grants to entrepreneurs.

study how top-ranked entrepreneurs' investment behavior and demographic characteristics differ from those of low-ranked entrepreneurs.

Entrepreneurs' Investment Decisions in Response to the Grant. In follow-up rounds of data collection, we asked grant winners to report on whether and how they had invested the grant money. Expenditures of the grant money were divided into business expenses (inventory, durable assets, labor, and other) and non-business expenses (loan repayment, giving out loans, household repairs, and other household expenses). We also asked respondents if they had supplemented the grant money with their own funds to make a business purchase. In Appendix Table 6 we examine the relationship between self-reported investment decisions and marginal returns rank. To do so we regress grant expenditures in each category (the sum of which is Rs. 6,000) on whether an entrepreneur is in the top or middle tercile of the marginal returns ranks distribution. The coefficients on the top and middle terciles indicate the difference in grant expenditures between entrepreneurs in those groups and entrepreneurs in the bottom tercile (the omitted group). Business owners in the top tercile invest an extra Rs. 903.1, or 25%, more of their grants in their enterprise than those in the bottom tercile. (We are under powered to detect statistically significant differences between the top and middle tercile.) Most of the gap between top and bottom tercile investments is due to expenditures in inventories. Both the top and middle terciles spend significantly less money on "Other Household Expenses" – medical expenses, education, food consumption, etc. – and are less likely to have saved their grant money for a future use.

Self-reports of grant expenditures suggest differences in investment behavior, but, since money is fungible, the observed effects might simply be due to mental accounting. To investigate whether grant investments translate to real increases in business inputs, we use regression specification 3 to compare inventories, business assets, and labor outcomes of grant winners and losers. Results are shown in Table 4. Consistent with the pattern of investment observed in the previous table, we find that the grant induces top and middle ranked entrepreneurs to accumulate higher capital stocks: top tercile grant winners report an extra Rs. 1181.5 worth of inventory and an extra Rs. 9041.5 of durable assets. The treatment increases the capital stock (inventory plus durable assets) by approximately 150% of the grant amount. This treatment effect is within the confidence bound of increases in capital stock found in McKenzie et al. (2008).

The grant also induces increases in inputs complementary to capital: own, household, and non-household labor. In Columns (1) and (2), we show that grant winners in the top tercile spend an extra 5 hours per week and an extra 1.5 days per month working when compared to their untreated counterparts. The treatment also has an impact on the amount of household and non-household labor. At baseline, 21% of enterprises in our sample employ household labor. Household workers in these enterprises contribute an average of 30 hours per week and almost none of them are officially paid a wage. Nine percent of households report using non-household labor in at least one of their businesses at baseline.[29] Among these businesses, the average weekly wage bill at baseline is Rs.

---

[29]For single-enterprise households, our eligibility criteria specified that businesses could not employ non-household

3221. The grant induces top-ranked entrepreneurs to be 6.6% more likely to have a household laborer and 6.0% more likely to have a non-household laborer at follow-up when compared to their untreated counterparts.

We find that top ranked entrepreneurs' investment behavior is markedly different from that of bottom ranked entrepreneurs: they invest a higher proportion of their grant into their business, turn those investments into higher business stock, and devote more time to working in their business. Do high-return entrepreneurs differ in other observable ways? In the next section, we investigate whether top-ranked entrepreneurs' demographic characteristics are significantly different from those of entrepreneurs in the bottom tercile.

Demographic Characteristics of Top-Ranked Entrepreneurs. In Table 5, we compare baseline characteristics of households and entrepreneurs in all three terciles of the marginal returns ranks distribution. In Column 1, we present the mean of each characteristic for the bottom tercile group. We then estimate the following model:

$$Y_{ijc} = \beta_0 + \beta_1 (Middle\,Tercile)_{ijc} + \beta_2 (Top\,Tercile)_{ijc} + \sigma_c + \theta_m + \tau_s + \epsilon_{ij} \tag{5}$$

In Columns 2 and 3, we present the coefficients from regressions of each baseline characteristic on whether the respondent is ranked in the middle (β_1) or top (β_2) terciles, respectively. Coefficients can be interpreted as the difference in each characteristic associated with being in one of the upper terciles relative to being in the bottom tercile.

Top-ranked entrepreneurs are 8 percentage points more likely to be male; about 2 years younger; and, are less likely to be married than entrepreneurs in the bottom tercile. Entrepreneurs in the bottom and top terciles have roughly the same number of years of education, yet those who are top ranked remember an average of 0.57 digits more in the digit span memory test. Top-ranked business owners work an extra 6.8 hours per week and 1.8 days per month. We asked business owners how much a salaried job would have to pay per month in order for them to exit self-employment. Top ranked entrepreneurs report that they would require 22% higher monthly wages to leave their businesses. Top-ranked entrepreneurs are slightly more likely to be engaged in a food preparation business and less likely to engage in livestock than bottom-ranked entrepreneurs, but otherwise the industry distribution is similar across terciles.

Households with a top-ranked entrepreneur have the same total number of businesses as households in the lower terciles. But these households have enterprises that are 52% larger in terms of assets and earn 40% higher profits per month. They also earn 13.3% higher monthly income. Household labor composition is very similar across all three groups, but top and middle ranked households are slightly less likely to employ a household daily wage worker.

labor at baseline. But households with multiple enterprises were eligible as long as there was at least one enterprise that met our eligibility criteria. Almost all households that report using non-household labor fall into this latter category. See Section 2 for a detailed explanation of eligibility criteria for households with multiple enterprises.

For the most part, entrepreneurs in the middle tercile have baseline characteristic means that lie between the means of the bottom and top ranked entrepreneurs. Two notable exceptions are that they have higher levels of education and business assets.

7.4 Benchmarking the Value of Community Information Against Observables

We showed in the previous section that top-ranked entrepreneurs differ from low-ranked entrepreneurs across several observable demographic characteristics. These findings raise the question: are community members simply using observable information to rank one another? If all valuable information about an entrepreneur's potential was contained in characteristics that could be verified by a lending institution, then peer elicitation might not be a good screening tool. Instead, lending institutions might be better off collecting observable data and developing an algorithm (credit scorecard) to optimally combine these observable measures. In this section, we benchmark the value of community information against the value of observables. First, we investigate whether community information remains valuable for predicting high-return entrepreneurs even after controlling for baseline characteristics. Next, we compare the predictive power of each source of information. These questions are related but distinct: community members may use information that is orthogonal to information captured by observables, but the accuracy of community reports may still be lower than the accuracy of a selection mechanism based on observable characteristics.

Combining Community Information with Observable Characteristics. Is the value of peer reports diminished when we hold constant entrepreneurs' baseline characteristics? We consider two sets of entrepreneur and household characteristics: first, we control for characteristics that could feasibly be observed and verified by a loan officer. In a setting like ours, where trustworthy third-party sources of information such as credit bureaus, tax returns, or business audits are not available, the scope of verifiable information is limited. In the set of loan officer controls we include the entrepreneur's gender, marital status, age, education, digit span memory test, number of salaried and wage workers, and business type. Next, we add harder to observe and verify characteristics, including household income, the value of household assets, hours worked, the value of business assets, average yearly profits at baseline[30], and the other measures presented in the randomization balance check (Appendix Table 1).

In Table 6, we present results of the main marginal returns specification (Specification 3) with the addition of the interaction of the baseline controls with $Winner_{it}$. We find that community information is almost orthogonal to information captured by the loan officer controls (odd numbered columns); the estimates in Table 6 are strikingly similar to those obtained without controls (Table 3). Moreover, controlling for the full set of baseline characteristics (even numbered columns) only increases the size of the coefficients on community reports. This is because marginal returns rank

---

[30]We asked respondents at baseline to tell us what had been their monthly profits for each month for the previous 12 months. This section of the survey (seasonality) is distinct from the section of the survey in which we ask respondents to report about business activities (including profits) in the previous 30 days. So the average yearly profits value is distinct from the profits in the previous 30 days, which is the outcome variable in Table 3.

is positively correlated with baseline profits and baseline profits are negatively correlated with marginal return to capital (implying that there are diminishing returns to capital).[31] In Appendix Table 9, we include the same set of controls and estimate the value of community information using the robustness specification (Specification 4); results are qualitatively similar. Finally, since psychological characteristics such as tenacity, polychronicity, and optimism, have also been shown to be predictive of credit worthiness and entrepreneurial aptitude (see Klinger, Khwaja, and Carpio 2013), we assess the value of entrepreneurs' responses to a psychometric test. We find that the value of community information remains almost identical to the original results presented in Table 3 (results from this estimation are presented in Appendix Table 10).[32]

We have shown that community members rely on knowledge that is residual to a wide range of observable entrepreneur and household characteristics. Respondents' opinions of their peers are informed by behaviors, personality traits, or other attributes that are not captured by extensive financial and psychometric profiles. This result sheds light on the depth of social ties in our study setting. The findings discussed in this section also motivate a further question: which source of information – community reports or observable characteristics – is *more valuable* for identifying high-return entrepreneurs? We have learned that community reports and observable characteristics convey distinct information about entrepreneurs' quality, but the analysis thus far does not assess their relative accuracy. We investigate this question in the next section.

Predicting High-Return Entrepreneurs. We aim to compare the relative value of community reports versus observable characteristics for identifying entrepreneurs with high growth potential. From our analysis in Section 7.2, we have an estimate of the value of community information. Next, we estimate the value of observable characteristics. This is a *prediction* problem, and not a parameter estimation problem: our goal in this section is not to understand the relationship between individual covariates (baseline characteristics) and entrepreneurs' returns. Instead, we seek to combine the information contained in all covariates to produce a prediction of these returns. We will estimate $\hat{y}$ rather than $\hat{\beta}$. Machine learning techniques, which are built to flexibly combine covariates without overfitting, are well suited for this task. We adapt a prediction technique developed by Athey and Imbens (2015) and Wager and Athey (2017) to form a marginal returns ranking of entrepreneurs based on their baseline characteristics. We then compare the predictive power of this ranking to that of the community reports ranking. In the sub-sections below, we describe our method, data, implementation, and then the results of this exercise.

---

[31]To see this, we regress income and profits on the interaction of baseline profits with winner and winner with the fixed effects included in Specification 3. We find that the coefficient on the interaction between winner and baseline profits is negative and significant at the 1%. Regression results available upon request from the authors.

[32]Regressors are labeled according to the psychological trait for which they are meant to proxy (the specific wording of the statement is found in the Appendix). There are two traits that are strongly predictive of marginal returns: optimism and achievement. We find that optimism negatively predicts marginal returns: business owners who are more likely to agree with the statements "In times of uncertainty I expect the best" and "I'm always optimistic about the future" and those who are more likely to disagree with "If something can go wrong with me, it will" have lower self-reported marginal returns. People who agree with the statement "Part of my enjoyment in doing things is improving my past performance" tend to have higher marginal returns.

I. Machine Learning Method.

Researchers who seek to use machine learning methods to identify treatment effect heterogeneity in randomized controlled trials face a unique problem: predictive models are developed through training on a dataset which contains both covariates, x, and outcomes, y, for each observation in the sample. But the outcome we aim to predict – each entrepreneur's marginal return to capital – is an *individual* treatment effect. As is the case with all randomized controlled trials, our data only allows us observe the *average*, or conditional average, treatment effect.[33] We therefore adopt a set of tools developed by Athey and Imbens (2015) and Wager and Athey (2017), which modify standard regression tree and forest techniques to allow for identification of heterogeneity in treatment effects.

The standard regression tree algorithm works by recursively partitioning the data into "leaves" of observations with the same set of characteristics. Data is partitioned so as to minimize mean squared error across splits. A prediction $\hat{y}$ is then formulated by taking the average value of y within each leaf. A regression forest extends the regression tree algorithm by growing and averaging across many trees. Athey and Imbens (2015) and Wager and Athey (2017)'s causal forest algorithm instead computes the average treatment effect, $\hat{\tau}$, within each leaf by estimating $\bar{y}_{treat} - \bar{y}_{control}$ for that leaf. Data is then partitioned so as to maximize treatment heterogeneity, i.e. maximize the difference in treatment effects across splits. In the Appendix, we provide a more detailed description of this prediction technique.

II. Data.

A second challenge in applying machine learning techniques to randomized control trials is the issue of sample size. Due to budget constraints, most RCTs – including our own – are just sufficiently powered to detect treatment effects. Machine learning, however, requires splitting the data into a training sample (to generate the model) and a test sample (to evaluate the out-of-sample goodness of fit). To extend our sample size, we use McKenzie et al. (2008)'s data. Our study uses the same sample selection criteria as McKenzie et al. (2008) and their study takes place in Sri Lanka, which is physically and economically proximate to India. Our study and the Sri Lanka experiment data share many of the same covariates and, in Appendix Table 11, we show that the two samples are similar across household and enterprise baseline characteristics. Note, though, that there are important differences between the two studies: McKenzie et al. (2008) gave out both in-kind and cash grants, while we only give cash grants. Additionally, they randomize \$100 and \$200 grants, while we only give \$100 grants. But McKenzie et al. (2008) analyze the differences in treatment effects from in-kind versus cash grants and \$100 and \$200 grants and find that they cannot reject equivalency of treatment effects. In our exercise, we consider an entrepreneur as treated if she

---

[33]Using notation from the Rubin Causal Model, let $W_i \in \{0, 1\}$ be an indicator of treatment, where $W_i = 0$ indicates that person i did not receive the treatment (say, a grant) and $W_i = 0$ indicates that person i did receive the treatment. For each individual i, we would like to observe an individual treatment effect $\tau_i = Y_i(1) - Y_i(0)$- this would allow us to train the model on τ_i as the ground truth. But each person only has one realized outcome ($Y_i(1)$ or $Y_i(0)$). So we can only observe the average treatment effect (ATE) for the population $E[Y(1) - Y(0)]$ or the conditional average treatment effect (CATE) $E[Y(1) - Y(0)|X = x]$.

received any of the four treatments.

In Sri Lanka, the authors did not collect rankings information. Since the goal of this exercise is to horserace our community information model with a machine learning prediction, we train the model on the Sri Lanka data and test it on our India sample data. We discuss robustness checks after the main analysis.

III. Implementation.

In the Appendix, we give a step-by-step instruction of our implementation of the causal forest algorithm. Here we highlight a few important features of the implementation. Machine learning techniques typically include penalization of model complexity so as to avoid overfitting. Following Katz and Roth (2017), we build a cross-validation method to determine the minimum node size at which the tree should no longer split. The minimum node size is a proxy for model complexity since smaller terminal node sizes imply a larger number of splits. We implement a k-fold cross validation method and the results of this exercise are shown in Appendix Table 12. We fit the full model on all of the Sri Lanka data with these parameters and predict the marginal returns of entrepreneurs in the India data.

As an additional robustness exercise, we also do an in-sample fit using the India data. We follow the same technique as with Sri Lanka: we conduct a k-fold cross-validation exercise and present the results to choose the optimal minimum node size in Appendix Table 13. Using the minimum node size from the cross-validation exercise, we generate a prediction using the India trained model on the India data. The goal of this exercise is not to use the India model to generate an out-of-sample prediction. Rather, this exercise is a conservative estimate of the "best fit" of our covariates. Because an in-sample test will very likely have very strong predictive power (due to overfitting), comparing community information to this in-sample estimate is a high bar to meet.

IV. Results.

In Table 7, we present the results of the machine learning exercise. In Columns 1 and 2, we first replicate the results shown in Column 6 of Table 3 and Column 6 of Table 6, respectively.[34] The machine learning exercises (from the model trained in Sri Lanka and the model trained in India) produce a numerical prediction of the marginal returns of each individual in the sample based on their baseline characteristics. For comparison with our main specification for the community rankings estimates, we divide the predictions into terciles. In Columns 3-6 we test how well the machine learning estimation predicts true marginal returns in our sample. In Appendix Table 14, we show the linear specification.

In Columns 3 and 4, we present the results of the machine learning prediction trained with the Sri Lanka. As shown in Column 3, the top tercile of entrepreneurs as identified by the Sri

---

[34]The estimates (and number of observations) differ slightly to ensure a comparable sample with the machine learning exercise. So in the replication of Table 6 Column 6, we only control for the variables that we use in the machine learning exercise (a subset of the variables used in Table 6).

Lanka prediction earn an extra Rs. 879.4 in marginal returns to the grant over the bottom tercile of entrepreneurs. In Column 4, we add the community information prediction. First note that the coefficient on $Winner * TopTercile$ is large and significant at the 5% level. Furthermore, the machine learning prediction estimate for the top tercile gets noisier and a bit smaller, but remains a good prediction of the best entrepreneurs. The correlation between community rank and the machine learning prediction is 0.1. Taken together, these results indicate that community members are using additional information to rank beyond (detailed) covariates that are observable to the researcher and that peer information is very valuable in identifying high-ability entrepreneurs. Columns 5 and 6 demonstrate the same point: despite the fact that the model is clearly overfit in-sample, community ranks continue to be predictive above and beyond the machine learning prediction.

What would have been our monthly return on investment had we used either community information or the machine learning prediction to allocate capital? In Figure 4, we separately depict the return on investment for each allocation mechanism: (i) random allocation (the equivalent of the population average return), (ii) allocation using the community information, and (iii) allocation using the Sri Lanka machine learning prediction.[35] The top tercile of entrepreneurs using the community rankings method earn monthly returns of 22.5%, while those in the top tercile of the Sri Lanka machine learning prediction distribution earn monthly returns of 18.0%.

7.5   Do Peers Distort Their Responses When There Are Real Stakes?

The analysis in Tables 2-7 has shown that communities are well informed about members' marginal returns. Why don't lending institutions leverage this information? One reason may be that acquiring *truthful* community reports is challenging. When a highly desirable resource (such as a grant) is at stake, community members may alter their reports to benefit or hurt particular individuals in their social network. In this section, we quantify whether and by how much peers misreport in high stakes settings.

We showed in Section 6 that community members are very well-informed about one another's household finances and enterprise characteristics. In that section, we evaluated the accuracy of community reports by estimating the distance between entrepreneurs' self reports at baseline and community members' reports for that person (at baseline, respondents were not told anything about the purpose of the study or about the ranking exercises). Because we cannot observe a marginal return for each individual in the sample, in Sections 7.2 to 7.4 we evaluated the "accuracy" of community information by estimating whether the true marginal returns of more highly ranked people were higher than the true marginal returns of less highly ranked entrepreneurs.[36] For the remainder of the paper, we return to the method of evaluating the accuracy of reports presented

---

[35]Note that this regression tells us the returns for each tercile separately, while the regressions in Table 2, for example, tell us the difference in marginal returns between the top/middle terciles and the bottom tercile.

[36]See a detailed discussion of this problem in Section 4.3.

in Section 6 but break up the estimates by treatment group.[37]

In order to assess whether and how peers lie when there is incentive to strategically misreport, after the completion of the marginal returns ranking exercise, half of our sample was informed that their rankings would affect the probability that their peers (or themselves) would win the USD 100 grant (this is the *High Stakes* group). Respondents in the *No Stakes* group continued to believe that their ranking responses would only be used for research purposes. We assess strategic misreporting in Table 8 by amending Specification 1 to compare accuracy in the *High Stakes* and *No Stakes* groups:

$$Y_{ijq} = \alpha_0 + \alpha_1 \overline{Rank}_{ijq} + \alpha_2 Stakes_j + \alpha_3 Stakes_j \times \overline{Rank}_{ijq} + \gamma_n + \theta_m + \tau_s + \delta_q + \epsilon_{ijmq} \qquad (6)$$

The model includes the following fixed effects: neighborhood (γ_n), survey month (θ_m), and surveyor (τ_s). Standard errors are clustered at the group level. α_1 captures the accuracy of the report in the control group (*No Stakes*). α_3 tells us whether the rankings are differentially informative when respondents are told their reports will be used to help determine grant allocation.[38] [39] To increase power, we stack the percentilized outcomes and ranks across all 6 columns presented in *Panel A* of Table 1 and add a question fixed effect (δ_q) to the regression model.

Respondents may have idiosyncratic preferences for misreporting about certain peers in their group and may otherwise make idiosyncratic errors. One way to reduce noise is to average across all reports given about a particular group member.[40] So in Columns 1-3 of Table 8, we show the regressions at the ranker-rankee level of observation ($Rank_{ijmq}$) and Column 4-6 are the regressions with the average rank ($\overline{Rank}_{ijq}$). We observe that the average predictiveness of ranks in the (*No Stakes*) group increases significantly when reports are averaged: in Column 1, a 1 percentile increase in the rank distribution is associated with a 0.162 shift in the outcome distribution in the individual regressions and a 0.252 shift in the average regression (Column 4). Averaging reports is therefore a costless way to nearly double the predictiveness of community reports.

Do respondents misreport in high stakes settings? We find that the coefficient on *Rank* × *High Stakes* is large, negative, and significant. We should note that this was not ex-ante obvious: the *High Stakes* treatment may have had a positive effect since introducing stakes may have caused respondents to focus or take the exercise more seriously. The regression implies that responses are significantly less accurate when respondents have an incentive to behave strategically: in the pooled individual regression in Column 1, the responses become 34.6% less accurate in the *High Stakes*

---

[37] As explained in Section 4.2, we did not randomize the the *High Stakes* and *No Stakes* treatments until after the marginal returns ranking was completed due to power considerations. As a result, we cannot evaluate misreporting by observing variation in ability to predict marginal returns.

[38] To reduce clutter in the regression tables, we have omitted the *High Stakes* coefficient from the regression report as it does not contain information relevant for the interpretation of results, but rather simply adjusts the constant,

[39] In this section, we pool across the *Public* and *Payments* treatments.

[40] In Table 1, all reports are averaged.

group.

Lastly, we asked respondents to rank their peers relative to others in the group (zero-sum ranking) and also relative to the community by reporting the quintile of the neighborhood distribution that they believe the peer to be in (quintile ranking). We hypothesized that quintile ranks could contain more valuable information about rankings because entrepreneurs are compared to the community more broadly than only the group. But they could also be more susceptible to misreporting: unlike with zero-sum ranks, respondents could, for example, place all of their peers in the top quintile of the distribution and claim that everyone is equally excellent.

To compare these two elicitation methods, in Columns 2-3 and 5-6, we show the results by separately stacking zero-sum and quintile rankings. In all four columns, the outcome variable is the same (percentile of Y_{ijq}). What changes is the method of reporting. In Columns 2 and 5, the regressor is the percentile in the (individual or average) quintile rank distribution. In Columns 3 and 6, the regressor is the percentile in the (individual or average) zero-sum rank distribution.[41]

The coefficients on *Rank* in the individual (Columns 2 and 3) and the average regressions (Columns 5 and 6) are very similar, implying that in the absence of high-stakes, the value of information from relative and quintile ranks is very similar. While the coefficient on *Rank* × *High Stakes* in the quintile regressions is larger in magnitude both in the individual and average models, we cannot reject that respondents misreport by the same amount in either type reporting method. This implies that, for this set of rankings, the relative and quintile rankings are equally informative.

Whom do respondents lie in favor of? At the start of the ranking exercise, we asked respondents to report their relationship with each peer in the group. We also asked each respondent to report who is person *i*'s closest peer in the group. The *cross-reported peer* is the peer that is most frequently reported as person *i*'s closest friend in the group. To analyze who respondents lie in favor of, in Table 9 we re-estimate regression Specification 7 but limit the sample to a respondent's reports about herself, her family members, and her cross-reported peers in the group. First, notice that the rankings of all three categories of persons, particularly family, are quite accurate in the *No Stakes* group. When stakes are introduced, however, the rankings become between 31% and 58% less accurate. There does not appear to be any differential pattern in misreporting by quintile or relative question. But the fact that accuracy dramatically decreases in the quintile ranks implies that *High Stakes* does not simply increase the general error rate due to re-rankings. As further corroboration, we analyze how the rankings themselves (not just accuracy) are affected by proximity between peers in Appendix Table 15. We see that respondents up-rank themselves, family members, and cross-reported peers relative to other peers in the group in the *No Stakes* treatment but that

---

[41]In Table 1, we stacked the zero-sum and quintile ranks by question. So in Column 1 of Table 1, the outcome variable is the household income and the regressors are the income quintile and zero-sum ranks, with a fixed effect for ranking type. Notice that the outcome variable is the same (household income) whether the regressor is a quintile or zero-sum ranking.

family members are ranked even more highly when *High Stakes* are introduced.

Overall, we find that misreporting is a first order problem if a principal wants to use community information to make valuable allocations.

7.6   Can Mechanism Design Tools Correct Incentives to Misreport?

Monetary Incentives and Public Reporting. In the previous section, we provided evidence that respondents distort their reports when they have a strategic incentive to do so. These distortions have a substantial impact on the accuracy of reports. Can we use tools from mechanism design to generate incentives for truthful reporting? And, are these tools effective even in high-stakes settings? We test the efficacy of two tools: payments for the accuracy of reports and reporting in public versus private.

In Table 10, we provide evidence of the *Public* and *Payments* treatments on the accuracy of reports. Again, following Specification 1 we estimate,

$$
\begin{aligned}
Y_{ijq} = {} & \eta_0 + \eta_1 \overline{Rank}_{ijq} + \eta_2 Public_j \times \overline{Rank}_{ijq} + \eta_3 Payments_j \times \overline{Rank}_{ijq} \\
& + \eta_4 Public_j \times Payments_j \times \overline{Rank}_{ijq} + \eta_5 Public_j + \eta_6 Payments_j \\
& + \eta_7 Public_j \times Payments_j + \gamma_n + \theta_m + \tau_s + \delta_q + \epsilon_{ijmq}.
\end{aligned}
\tag{7}
$$

The coefficient η_1 identifies the accuracy of reports in groups in which respondents do not receive incentive payments and report in private. The coefficients on the first three interaction terms tell us the additional accuracy due to reporting (i) in public without monetary payments (η_2), (ii) in private with monetary payments (η_3), and (iii) in public with monetary payments (η_4).[42]

To determine how these tools perform in a high stakes setting, we split results by *No Stakes* (odd columns) and *High Stakes* (even columns). Given that respondents lie substantially in favor of themselves (Table 9), we also split the results by whether a respondent is reporting about herself (Columns 1 and 2) or about her peers (Columns 3 and 4).

We find that community members are both more accurate and less responsive to incentives for truthfulness when reporting about themselves. Putting respondents in a high-stakes setting decreases the accuracy of self-reports by 26%. Moreover, neither payments for truthfulness nor public reporting have any impact on the accuracy of self-reports. Note, though, that the accuracy of their self-reports (0.158 in Column 2) in the high-stakes setting is approximately the same as the accuracy of reports about others in the group in the private and no payments treatment (0.141 in Column 4).

When reporting about others, incentives for truthfulness can have a large impact on respon-

---

[42]To reduce clutter in the regression tables, we have omitted the coefficients $Public_j \times Payments_j \times \overline{Rank}_{ijq}$, $Public_j$, $Payments_j$, $Public_j \times Payments_j$ from Table 10 as they do not contain information relevant for the interpretation of results, but rather simply adjust the intercept.

dents' accuracy. First, in the *No Stakes* setting, the *Payments* and *Public* treatments both double the accuracy of reports (they each lead to increase in accuracy between 0.14 - 0.17. The coefficient on the treatment in which respondents receive monetary incentives and report in public is large and negative (*Average Rank × Payments × Public*). But, we can reject at the 10% level that the accuracy of information in this group is the same as in the private reporting and no monetary incentives group. We therefore interpret the negative coefficient as an indication that monetary payments and public reporting are substitutes. We find that incentives improve accuracy even in the *No Stakes* treatment. We interpret this result as evidence that strategic misreporting may be an important concern in programs which involve elicitation of community information or opinions, even when respondents are not directly affected by the decision at hand. We discuss this finding further in the conclusion (Section 8).

Remarkably, the monetary payments treatment is just as effective when allocation of resources is at stake: the *Payments* treatment still improves accuracy by 0.14, which is an increase in accuracy of over 100%. So, providing monetary payments corrects nearly all of the strategic misreporting that is induced by asking respondents to report in a high stakes setting.

In the *High Stakes* setting, we find that the *Public* treatment no longer has a significant impact on accuracy. As discussed in Section 3, the impact of public reporting on accuracy is ambiguous ex-ante. There may be pressure for respondents to up-rank their family members, but there may also be pressure from non-family members and other peers to be truthful. When we introduce stakes, both of these pressures are intensified: family members and close friends want the respondent to sway the grant allocation in their favor, but it may also be especially important to the community that members be truthful when there are high stakes.In Table 11, we shed light on how respondents trade off these allegiances, although we caution that this is an exploratory exercise that was not specified in our Pre-Analysis Plan. The outcome variable in this table is the rank that respondent i assigned to peer j in the group. *Family* indicates whether respondent i is related to peer j (we exclude self-ranks). Compared to non-family peers (the omitted group), family members receive an extra 0.46 and 0.25 ranks on average when reports are given in private in the no-stakes (Column 1) and high stakes settings (Column 2), respectively. So family members are favored over non-family members in either setting.

The coefficient on *Family × Public* corresponds to the disciplining effect of being in public when ranking a family member. When there are no stakes, the *Public* treatment fully corrects the incentive to up-rank family members (Column 1). This is consistent with the result in Column 3 of Table 10, which shows that public reporting in a no-stakes setting improves the quality of reports. But when respondents are asked to rank with stakes, we cannot reject that the *Public* treatment has no effect: respondents continue ranking their family members more highly than non-family members (Column 2). This is consistent with the result in Column 4 of Table 10 which shows that public reporting in a stakes setting has no differential impact over private reporting in a stakes setting.

Cross-Reporting. As shown in Table 9, community members are capable of successfully identifying persons for whom a peer is likely to lie in favor. We also asked respondents to name the person who would be best able to predict who would provide the most accurate reports on average. In Table 12, we interact rank with whether a respondent has been selected by her group as the one who would provide the most accurate answers. In column 1, pooling across all questions, we see persons who are selected provide information that is 50% more accurate than information provided by the standard respondent in the group.

8   Conclusion

We find that community members have residual information about their peers that is valuable for targeting. Not only can community members identify characteristics of their peers' enterprises, they can also predict which of their peers have high returns to capital. But community information is also susceptible to strategic misreporting. In particular, we identify a tendency for respondents to favor their friends and family members in their reports. Moreover misreporting is exacerbated when respondents are told that their reports will influence the distribution of grants. If we assume that stakes would have reduced the accuracy of the marginal returns ranking by a third (the estimated average reduction in accuracy across the metrics evaluated in the high stakes treatment), then the marginal returns prediction for the top third of entrepreneurs would drop from 22.5% to 14.8% per month.[43]

However we also find that a variety of techniques motivated by mechanism design theory are effective in realigning incentives for truthfulness. Relatively small monetary payments for accuracy, eliciting reports in public rather than in private, and cross reporting techniques all substantially improve the accuracy of reports.

Is it worth it for principals to invest in collecting community information and providing incentives to respondents? We calibrated the payment rule to pay, on average, Rs. 25 per question per respondent. In total, we paid Rs. 17000 in incentives for the marginal returns question. If a lender were distributing 450 loans (as we did with grants), this would increase the cost on each loan by approximately Rs. 40 per month. In Section 7.2, we estimated that the cost of interest that an MFI would charge per grant is Rs. 570 per month. Adding the incentives costs (transferring it to the borrower) implies that the cost of the loan to each respondent per month would be Rs. 610. Using the returns estimate from our preferred specification (Table 2 *Panel B*, Column 3), borrowers would still earn a net return of Rs. 388 per month if the full cost of the monetary incentives were passed on to the borrowers.

Our aim is that the peer elicitation method identified in this paper can be useful for targeting

---

[43]In Appendix Table 22, we re-estimate the same model as in Table 7, but analyze the results question by question rather than pooling across questions. We see that for 7 out of the 9 characteristics presented in Table 7, the reduction in accuracy due to *High Stakes* is between one and two thirds vis-a-vis the *No Stakes* group.

in poorly developed financial markets in low-income countries, where information asymmetries are prevalent.

References

Abadie, A., J. Angrist, and G. Imbens (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica 70*(1), 91–117.

Alatas, V., A. Banerjee, R. Hanna, B. A. Olken, R. Purnamasari, and M. Wai-Poi (2013). Does elite capture matter? local elites and targeted welfare programs in indonesia. Technical report, National Bureau of Economic Research.

Alatas, V., A. Banerjee, R. Hanna, B. A. Olken, and J. Tobias (2012). Targeting the poor: evidence from a field experiment in indonesia. *The American Economic Review 102*(4), 1206–1240.

Allcott, H. and T. Rogers (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *The American Economic Review 104*(10), 3003–3037.

Athey, S. and G. W. Imbens (2015). Machine learning methods for estimating heterogeneous causal effects. *stat 1050*(5).

Banerjee, A., D. Karlan, and J. Zinman (2015). Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics 7*(1), 1–21.

Banerjee, A. V. (2013). Microcredit under the microscope: what have we learned in the past two decades, and what do we need to know? *Annu. Rev. Econ. 5*(1), 487–519.

Basurto, P. M., P. Dupas, and J. Robinson (2017). Decentralization and efficiency of subsidy targeting: Evidence from chiefs in rural malawi. Technical report, National Bureau of Economic Research.

Beaman, L. and J. Magruder (2012). Who gets the job referral? evidence from a social networks experiment. *The American Economic Review 102*(7), 3574–3593.

Bernhardt, A., E. Field, R. Pande, and N. Rigol (2017). Household matters: Revisiting the returns to capital among female microentrepreneurs. Technical report, National Bureau of Economic Research.

Besley, T. and M. Ghatak (2005). Competition and incentives with motivated agents. *American Economic Review 95*(3), 616–636.

Bryan, G., D. Karlan, and J. Zinman (2015). Referrals: Peer screening and enforcement in a consumer credit field experiment. *American Economic Journal: Microeconomics 7*(3), 174–204.

de Mel, S., D. McKenzie, and C. Woodruff (2008). Returns to capital in microenterprises: Evidence from a field experiment. *The Quarterly Journal of Economics 123*(4), 1329–1372.

De Mel, S., D. J. McKenzie, and C. Woodruff (2009). Measuring microenterprise profits: Must we ask how the sausage is made? *Journal of development Economics 88*(1), 19–31.

De Mel, S., D. J. McKenzie, and C. M. Woodruff (2010). *International Differences in Entrepreneurship*, Chapter Who are the Microenterprise Owners?: Evidence from Sri Lanka on Tokman v. de Soto, pp. 63–87.

Fafchamps, M., D. McKenzie, S. Quinn, and C. Woodruff (2014). Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in ghana. *Journal of development Economics 106*, 211–226.

Fiala, N. (2013). Stimulating microenterprise growth: Results from a loans, grants and training experiment in uganda.

Foster, A. D. and M. R. Rosenzweig (1996). Technical change and human-capital returns and investments: evidence from the green revolution. *The American economic review*, 931–953.

Gerber, A. S., D. P. Green, and C. W. Larimer (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review 102*(1), 33–48.

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, 508–523.

Maitra, P., S. Mitra, D. Mookherjee, A. Motta, and S. Visaria (2017). Financing smallholder agriculture: An experiment with agent-intermediated microloans in india. *Journal of Development Economics 127*, 306–337.

Maskin, E. (1999). Nash equilibrium and welfare optimality. *The Review of Economic Studies 66*(1), 23–38.

McKenzie, D., S. de Mel, and C. Woodruff (2008). Returns to capital: Results from a randomized experiment. *Quarterly Journal of Economics 123*(4), 1329–72.

Parker, S. C. (2009). *The economics of entrepreneurship*. Cambridge University Press.

Prelec, D. (2004). A bayesian truth serum for subjective data. *Science 306*(5695), 462–466.

Rigol, N. and B. Roth (2017). Paying for the truth: The efficacy of peer prediction in the field. *Working Paper*.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association 82*(398), 387–394.

Schoar, A. (2010). The divide between subsistence and transformational entrepreneurship. *Innovation policy and the economy 10*(1), 57–81.

Townsend, R. M. (1994). Risk and insurance in village india. *Econometrica: Journal of the Econometric Society*, 539–591.

Varian, H. R. (1990). Monitoring agents with other agents. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift Für Die Gesamte Staatswissenschaft*, 153–174.

Wager, S. and S. Athey (2017). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* (just-accepted).

APPENDIX

Details for Robust Bayesian Truth Serum

This discussion is based on Rigol and Roth (2017).

Theory and Intuition

In this appendix section we discuss the details of the Robust Bayesian Truth Serum, an intuition for the underlying incentive properties, and our implementation of the payment rule in the field. The following discussion of the model is based on Witkowski and Parkes (2012).

Suppose there is a binary state of the world $t \in (h, l)$ (high, low) representing the entrepreneurial quality of a community member. Agents 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 say the common prior is *admissible* if $p_h > p_l$, which in English 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.

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 ω 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 ω 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 $x_i \in \{0, 1\}$ (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, $x_k = 1$). 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 δ. The RBTS payment for agent i is

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

The main theorem of Witkowski and Parkes is that under the assumption of an admissible 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 announce 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 possible 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 admissibility 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.

Practical Implementation

We used this payment rule in the field to incentivize rank order responses about members of each group. 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, calibrated 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.

Incentive Compatibility Exercise for RBTS

Throughout the experiment we told respondents that they would maximize their personal payoffs if they reported truthfully. While RBTS is truthful under certain reasonable assumptions about how beliefs are formed, its incentive compatibility under the empirical distribution of beliefs in practice remains an open question. We therefore evaluate whether respondents are maximizing their subjective expected utility by telling the truth.

Due to the coarseness of our elicited measures of belief, we cannot verify directly whether or not the respondents' priors are admissible However we can determine the distribution of payoffs respondents can 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 responses about the question of interest) and on the distribution of second order beliefs. Therefore, to determine whether truth telling is incentive compatible, 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 behind the construction 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. [44]

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.[45] Specifically, we assume that the report the respondent has given is her true belief and calculate her payment. 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 2 below. Column 1 of the figure depicts the

---

[44]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.

[45]Notice that we only utilize incentivized data since it is only for these data that we collected second order beliefs.

percentage of instances in which telling the truth gives the largest payment, column 2 depicts the percentage of instances 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 in about 35% of instances and it minimizes her subjective expected utility in only about 10% of instances. An ideal graph would place all of its weight in the first column.

Figure 2: Incentive Compatibility of RBTS Using Collected Data

The observed departure from this ideal 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 compatibility of RBTS. We generate groups of artificial agents, each of whom holds a common prior and receives a signal about the skill level of their peers. Agents update their priors based on these signals and these form the basis of their second and third order beliefs, each of which we can compute.

Because the data is generated to be perfectly consistent with the assumptions of RBTS, the agent always maximizes his expected utility by 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. 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 3.
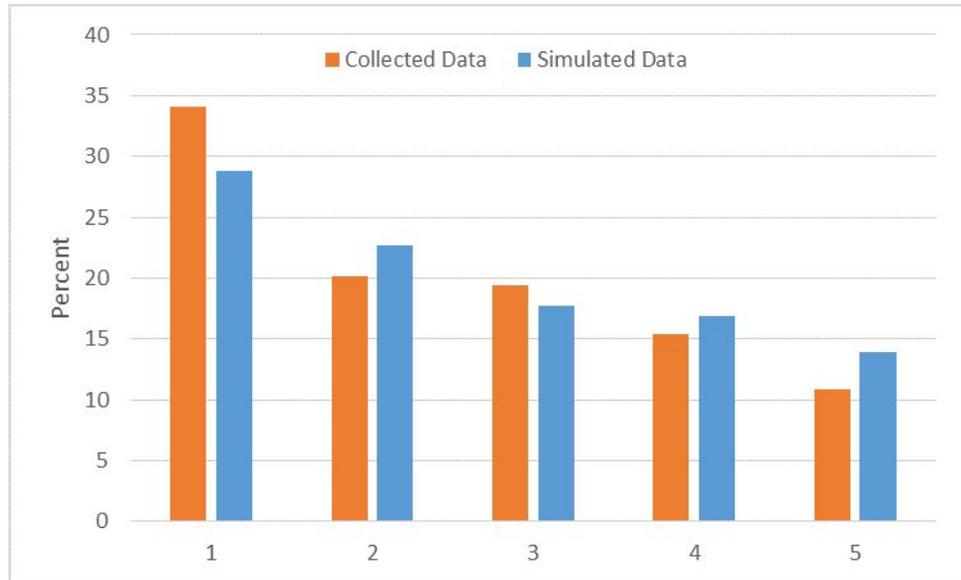
Figure 3: Incentive Compatibility of RBTS Using Collected and Simulated Data

The graph produced with the collected and with the simulated data are strikingly similar. We therefore conclude that our noisy approximation of third order beliefs could be to blame for the observed weights in columns 2 through 5, and argue that our test yields the strongest evidence in favor of the incentive compatibility that could be achieved via this method. Therefore, telling respondents that they will maximize their expected payments by reporting truthfully may indeed be good advice.

Entrepreneurial Psychology

Impulsiveness:

- I plan tasks carefully.

- I make up my mind quickly

- I save regularly.

Optimism:

- In uncertain times I usually expect the best.

- If something can go wrong for me, it will.

- I'm always optimistic about my future.

- Generally speaking, most people in this community are honest and can be trusted

Locus of Control

- A person can get rich by taking risks.

- I only try things that I am sure of.

Tenacity

- I can think of many times when I persisted with work when others quit

- I continue to work on hard projects even when others oppose me.

Polychronicity:

- I like to juggle several activities at the same time

- I would rather complete an entire project every day than complete parts of several projects.

- I believe it is best to complete one task before beginning another.

Achievement

- Part of my enjoyment in doing things is improving my past performance

- If given the chance, I would make a good leader of people.

Organized person:

- My family and friends would say I am a very organized person

Intuition for Machine Learning Prediction

Regression Tree/Forest   The regression tree algorithm has two major tasks: (1) decide how to split the data at each step and (2) decide when to stop splitting. Broadly, the algorithm will consider every value of every covariate as a possible split point. To decide on the first split, the algorithm systematically partitions the data at each value of each of the covariates and computes the goodness of fit criterion for either side of the partition. It picks the split where the goodness of fit criterion is optimized. As a specific example, suppose there are j covariates and the goodness-of-fit criterion is the mean square error $\frac{1}{N} \sum_{i=1}^{n} (\bar{y} - y_i)^2$ where $\bar{y}$ is the average value of y within a partition.[46] The algorithm will choose covariate j and $X_{ij} = s$ such that the following function is minimized

$$\frac{1}{N} \sum_{i:X_{ij}<s} (\bar{y} - y_i)^2 + \frac{1}{N} \sum_{i:X_{ij} \geq s} (\bar{y} - y_i)^2$$

For the second split, the algorithm will perform the same search but separately for $X_{ij} < s$ and $X_{ij} \geq s$. In each subsequent node, the data in the leaf will be split in the same manner to minimize mean squared error until a limit is reached. One important thing to note is that the algorithm is "greedy:" it does not search for partitions to globally minimize the mean squared error of all X_{ij}.

---

[46]$\bar{y} = argmin_{\hat{y}} \left(\frac{1}{N} \sum_{i=1}^{n} (\hat{y} - y_i)^2 \right)$

To form a prediction, the algorithm computes $\bar{y}$ at each terminal leaf. [47] To deal with the problem of overfitting , the researcher can use cross-validation techniques to determine the optimal place to stop splitting.

A regression forest extends the regression tree algorithm using a bootstrapping technique. The algorithm selects (with replacement) a subset of all covariates and a subset of the data to grow a tree. This process is repeated for a user-selected number of trees. If there are B trees grown and observation i appears in $\{b : b \leq B\}$ trees, then to make a prediction for observation i the algorithm computes $\frac{1}{b} \sum_{n=1}^{b} \bar{y}_{in}$ where $\bar{y}_{in}$ is the average value of y in the leaf of tree n where observation i ends up.

Cross Validation Exercise

We partition the training (Sri Lanka or India) data into 5 non-overlapping parts. A very important part of this partitioning is that we have a panel data set and so all observations for one person have to fall in the same partition. We build a tree using the first 4 "training" folds and estimate model fit on the 5th "test" fold for a range of minimum node sizes. To estimate model performance, we estimate the following linear regression

$$Y_{ijt} = \beta_0 + \beta_1(\hat{MR} * Winner)_{ij} + \beta_2(\hat{MR})_{ij} + \beta_3(Winner)_{ij} + \epsilon_{ij} \tag{8}$$

where Y_{ijt} are the observed post-treatment profits for the "test" fold and $\hat{MR}$ is the predicted marginal return for each individual in the "test" fold (using the model estimated in the "training" folds). For each node size, we record β_1. We begin this process leaving out 1 fold and training on the other 4 until the model has been tested on all 5 folds (so we have 5 total iterations of this process with a 5-fold partition). For each node size, we compute $\bar{\beta}_{1n} = \frac{1}{5} \sum_{f=1}^{5} \beta_{1nf}$ where β_{1nf} is the value of β_1 for minimum node size n in test fold f. We select the minimum node size in which the $\bar{\beta}_{1n}$ is highest. The intuition is that we would like to pick the model that allows us to estimate the biggest treatment effect in a test dataset. The largest $\bar{\beta}_{1n}$ exists for minimum node size 15.

---

[47]For example, the minimum number of observations within a leaf. The fewer the minimum allowable number of observations within a leaf, the greater the number of splits the algorithm can perform.