Can Farmers Create Efficient Networks?

Experimental Evidence from Rural India

A. Stefano Caria* Marcel Fafchamps†

February 28, 2014

*University of Oxford
†Stanford University

Abstract

Peer networks play a crucial rule in the diffusion of innovations such as new agricultural technologies, health schemes or financial products. While theoretical models show that in a variety of cases myopically rational individuals can converge to efficient network architectures after repeated play (Bala and Goyal, 2000), a number of field studies have documented losses in efficiency associated with homophily-the tendency of similar individuals to interact with each other with disproportionate frequency. We run an artefactual field experiment in rural India which tests the extent to which farmers can create efficient network architectures in a repeated link formation game, and whether membership of exogenous-but-salient social groups results in homophily and loss of efficiency. We find that realised efficiency is higher than that achieved by a purely random link formation process, but lower than the level of efficiency which myopically rational play would have achieved. Link formation decisions are consistent with a number of archetypal linking rules derived from payoff maximisation and heterogeneous other-regarding motives. When information about group membership is revealed, social networks become more homophilous, but not significantly less efficient.

1 Introduction

Peer networks play a crucial rule in the diffusion of innovations such as new agricultural technologies, health schemes or financial products. The literature documenting significant interpersonal influence in technology adoption and input decisions is vast. In agriculture, where scarce information is often a powerful barrier to adoption of profitable technologies, many studies have documented peer learning across social networks (Foster and Rosenzweig, 1995; Munshi, 2004; Bandiera and Rasul, 2006; Conley and Udry, 2010; Krishnan and Patnam, 2012). In one study from rural India, for example, more than 50 percent of farmers report discussing with peers about crop prices, weather predictions, crops to plant, cultivation practices and inputs (Fafchamps and Minten, 2012). Interventions that exploit peer networks for the diffusion of information are naturally attracting growing interest among development practitioners and academics (Banerjee et al., 2012; Ben Yishay and Mobarak, 2012).

The structure of peer networks affects the reach and speed of diffusion. The design of peer-to-peer diffusion interventions should thus be informed by an understanding of the social networks they activate. For example, in the case of non-rival goods that can travel the network without decay, efficient network architectures connect all individuals in the population. If efficiency is not achieved in real world networks, interventions should incorporate alternative means of targeting disconnected individuals. Theoretical models actually show that myopically rational individuals forming unilateral, one-way-flow links can converge to efficient architectures after repeated play (Bala and Goyal, 2000). Initial lab evidence on a western student population confirms this prediction (Falk and Kosfeld, 2012). Experimental evidence on the efficiency of the networks formed by subjects in developing countries is however missing. Further, a number of field studies have documented losses in efficiency associated with homophily- the tendency of similar individuals to interact with each other with disproportionate frequency (Berg et al., 2013; Emerick, 2013). The desire to maintain group divisions and identities may indeed cause individuals to sacrifice efficiency.[1] Farmers' networks are often characterized by a high degree of social homophily (Rogers, 2003).

In this paper, we experimentally test the extent to which farmers can form efficient network architectures in a sequential link formation game when players

[1]See also (Golub and Jackson, 2012), who present a theoretical model where homophily slows down the speed of diffusion

identities are private knowledge. We also test whether common knowledge of membership in an exogenous, salient social group increases the number of ingroup links and decreases the efficiency of final architectures.

When group identity is private knowledge, we predict that individuals will play simple, intuitive link formation rules leading to high levels of efficiency. In treatments where individuals are only allowed to form out-links (in-links), this amount to choosing the player with highest out-degree (in-degree). These predictions are borne by standard models of strategic network formation augmented to include social preferences (Bala and Goyal, 2000; Charness and Rabin, 2002).

When group identity is common knowledge, we predict individuals will choose ingroup links more frequently, possibly at the cost of establishing less efficient networks. Previous research has highlighted how social identity generates ingroup favoritism (Tajfel, 1981; Brewer, 1999; Akerlof and Kranton, 2010), affects social, risk and time preferences (Benjamin et al., 2010; Chen and Li, 2009; Kranton et al., 2012), influences behaviour in strategic environments (Yamagishi and Kiyonari, 2000; Charness et al., 2007), and modifies performance (Hoff and Pandey, 2006). Akerlof and Kranton (2000) posit that agents receive utility from following prescriptions associated with their social categories. We document that a prescriptions to restrict links to the ingroup is widely shared among the population in our sample. Hence, when the best partner belongs to the outgroup, individuals face a tradeoff between efficiency and conformity with the social prescription.

In terms of methodology, we take steps against a number of common confounders of experimental inference: low understanding, side payments, wealth effects and experimenter demand effects. We rely on induced, randomised group membership to rule out unobserved covariates that may be correlated with natural groups. As the saliency of induced group membership has been found to influence behaviour in economic experiments (Charness et al., 2007; Eckel and Grossman, 2005), we increase saliency by means of a task that sets the two groups in competition in a different domain.

We run our experiment in the Indian state of Maharashtra. With the many social identities based on caste, religion and class, India offers an appropriate setting to study homophiliy in social networks (Guha, 2008; Dunning and Nilekani, 2013). Recent work on information agents in rural communities indeed suggest that social distance matters and experimental work has shown that priming natu-

ral identities, chiefly caste, affects individual performance and economic outcomes (Hoff and Pandey, 2006; Anderson, 2011; Berg et al., 2013). In India, interest in novel extension approaches that exploit farmers' dense social network activity is also high.

We find that realised efficiency is higher than that achieved by a purely random link formation process, but lower than the level of efficiency which myopically rational play would have achieved. In the treatment where agents form out-links, the simple rule to link with the player with the highest out-degree would deliver 95 percent efficiency. Realised efficiency is about 64 percent, 31 percentage point below what the simple rule could have achieved. Interestingly, when agents are only allowed to form in-links architecture efficiency is not significantly different from efficiency under out-link formation.

When information about group membership is revealed, social networks become more homophilous, but not significantly less efficient. In our experiment, social identity has an impact on network structure, but it is not the cause of observed network inefficiency.

Our contribution is twofold. First, we present findings on network efficiency which are inconsistent with the predictions of existing network formation models. These may be the basis for new modelling of network formation processes, with emphasis on heterogeneity in social preferences and noise. Second, our results document an effect of arbitrary social identity on linking behaviour. This expands our understanding of the settings in which arbitrary social categorization is sufficient to affect behaviour. Our experiment also shows, however, that in our particular context, social identity influences network formation in ways that are not detrimental to efficiency.

The paper is organised as follows. Section 2 presents the design. Section 3 develops predictions and testable hypotheses. Section 4 describes the data. Section 5 reports the results of the analysis. Section 6 concludes.

2 Design

In the experiment subjects unilaterally create one-way-flow links. One player in each session is randomly drawn at the end of the experiment to receive a monetary

prize.[2] Players who in the final network are directly or indirectly connected to the winner of the prize receive a prize of equivalent value.

Play is sequential. The game is divided in two rounds. Each round comprises 6 turns. In every turn, only one player takes a decision. Each player is randomly assigned to one turn per round. Participants are informed of this rule, but do not know the particular order of play which has been drawn for their session. In the first round, players create one link. In the second round, players can rewire their existing link.[3]

Players' decisions are recorded on a network map which is updated after every turn on a white board visible to all players. A number of design features ensure sequential updating takes place without breaking anonymity.[4] Furthermore, the pilot revealed that, when the network map has more than a few links, players find it difficult to calculate quickly the number of connections of each peer. We hence remind the decision maker of the number of connections every peer has in the current network. The counting of connections is done by means of a Java application running on a small laptop operated by the game assistant. After entering a new link, the software produces a table with the number of connections each player has in the current network. This number is written next to the respective player ID on the white board before the next decision maker takes his turn.

The experiment is played by groups of 6 individuals.

The experimental tasks are carried out in the following order. First, players randomly draw a card from an urn which assigns them a letter ID and an experimental group. Second, participants answer three questions on agricultural knowl-

[2]The prize is worth 100 Indian Rupees.

[3]In both rounds players have the options not to form any link. This option is only very rarely used.

[4]Participants record their decisions on a game sheet. Modified cardboard boxes ensure participants cannot see what other players are choosing. However, the boxes do not prevent players to infer from a peer's body movements whether he is updating his game sheet or not. This threatens anonymity as it is possible to determine which participant has the turn by simply checking who is updating his game sheet at a given point in the game. We solve this problem in the following way. At the beginning of each turn, the game assistant publicly calls the ID of the player who has the turn. After allowing some time for reflection, the game assistant then asks all players to make a circle on their game sheet. The player with the turn circles the ID letter of the player to whom he would like to link. The player without the turn makes a circle in an empty box provided on the same page of the game sheet. So players cannot infer the identity of the player with the turn by checking who is updating his game sheet.

edge, which are part of an intergroup contest in agricultural knowledge. At the end of the game, if all players in a group have answered all questions right, the group receives one point and is applauded by everyone. Points are summed across sessions and hence participants are also informed of the overall ranking between the two groups.[5] Third, participants play a simple allocation task, where they have to divide a fixed sum of money between an ingroup and an outgroup recipient randomly drawn from participants in the following session of the experiment. Fourth, participants are given instructions about the link formation game, their answer a number of questions which test their understanding of the game[6], and they play a trial of the link formation game which lasts for 7 rounds. At the end of the trial game, the enumerator randomly draws a participant and shows who would receive the prize if this was the real game. Fifth, the link formation game is played. Finally, participants are asked three questions about their expectations and beliefs and are then administered a short end questionnaire, which collects information on socio-demographic variables and asks participants to explain the motivation behind their decisions in the game.

We rely on a between-subject design. We vary the direction of the flow of benefits associated with a link. In Treatment 1 (henceforth T1), players form out-links. This means that if A chooses B, then A will receive the monetary prize whenever B wins it, but not viceversa. In Treatment 2 (henceforth T2), players form in-links. This means that if A chooses B, then B will receive the monetary prize whenever A wins it, but not viceversa. Consistently with our theoretical predictions in the next section, in T1, as the network is updated, players are reminded of the out-degree of each player. In T2, players are reminded both of the out-degree and in-degree.

We also vary the information about peer group membership available to players during the link formation game. This is cross-cut with T1 and T2. In a first set of treatments, which we call T1no and T2no, individuals have no information about peers' group affiliation. Hence their link formation decisions are by construction unrelated to the groups formed at the beginning of the experiment. In a second

[5]Notice this information is revealed at the end of the game. So, whilst the contest creates the feeling of inter-group competition on a second, unrelated domain, it does not affect the beliefs players have about the levels of knowledge and cognitive ability of players in their group.

[6]The game assistant checks the answers and is instructed to give further explanations of more than one player makes more than one mistake. Hence these can be considered as a lower bound on the level of understanding of players.

Table 1: Summary of treatments

	No identity	Identity
Links take prize	T1no	T1id
Links give prize	T2no	T2id

set of treatments, called T1id and T2id, group identity is common knowledge, as players belonging to different groups are identified with different symbols on the public network map on the whiteboard.[7]

We hence run four treatments, as shown in table 1.

Instructions are framed in terms of a salient example from the local context. The link formation game is presented as a game where one farmer will receive a valuable piece of information about a new agricultural technology, and the network determines who receives help from the farmer with the valuable information. So the choice of a link is presented in terms of choosing who to take help from in case one does get the valuable information or who to help in case one accesses the information. The groups are called the mango and the pineapple group, and are explained with reference to the producer groups which farmers typically form in the areas of our study.

In our design group membership is randomly allocated. The standard protocol in social psychology, on the other hand, relies on groups which are formed on the basis of a trivial preference.[8] While the latter design feature has the potential to increase the saliency of group membership, it also has two disadvantages. First, there is a risk that players' characteristics are associated to what the researchers labels as irrelevant preferences. Second, there is the risk that some players believe that such correlation exists. For example, a player may (erroneously) think that people with a certain preference in art or sport are smarter. In both cases, the effects of common knowledge of social identity would be confounded by those associated with correlated categories and beliefs.

[7]Mango group players are identified with a circle. Pineapple group players are identified with a triangle.

[8]Notice however that ingroup bias in allocation tasks is found also when group membership is determined by the flip of a coin (Tajfel, 1981)

We hence opt for a design which relies on random assignment to social groups and increases the saliency of group identity by means of the contest in agricultural knowledge. This task combines four desirable features:(i) it is linked to the overall framing of the experiment, (ii) it creates a feeling of competition between the two groups on a domain-real agricultural knowledge- which is distinct to the domain of monetary outcomes of the experiment, (iii) the relative position of the two groups in this second domain is only revealed AFTER the link formation game has been played, (iv) every player can have a strong marginal impact on the group's outcome: if a player fails to answer one question correctly, the whole group fails to gain the point for that session. The idea of using contests to increase the salience of group identity has been successfully deployed in the past in experimental studies (Eckel and Grossman, 2005).

To ensure comparability and minimize noise factors during play, we follow a number of established practices in the lab-in-the-field literature. These include extensive piloting, simple standardized instructions that are read out to participants, double translation of all written material, and reliance on physical randomization devices (Barr and Genicot, 2008; Viceisza, 2012).

2.1 Possible Confounders

We take a number of steps against common confounders of experimental analyses.

Low understanding. We test players' understanding before the game starts. Subjects in T1no and T1id are asked 8 understanding questions, while subjects in T2no and T2id are asked 7 understanding questions. The questions test for understanding of the network map and of the incentives that result from the rules of the game. After the questions are asked, enumerators briefly check the answers and give further explanations on the points where players made mistakes. Hence these answers give a lower bound to the level of understanding of players in the game.

Figure 3 in the appendix presents the cumulative distribution of mistakes. In both *self* and *other* treatments, more than 50 percent of players did one mistake or less, and about 80 percent of players did 2 mistakes or less. Given that these mistakes were followed by further explanations, we feel confident that the experiment was well understood.

To further increase understanding, we also run a trial round of the link formation game before the main game is played.[9]

Side-payments. Identities remain anonymous and payments are disbursed privately. This decreases the possibility of side-payments. In particular, it decreases the possibility that network formation decisions will be targeted towards individuals from whom side payments can be more easily extracted.

Wealth effects. Both the allocation and link formation tasks are incentivised with monetary payments. However, there should be no "within-person" wealth effects across the allocation and linking tasks, as the first task involves no costs to the participants. Furthermore, in the allocation tasks individuals allocate money between two peers in a future session of the experiment. Hence there are no"between-person" wealth effects either.[10]

Experimenter demand effects. These arise when subjects, in an attempt to please the experimenter, respond to implicit cues resulting from the experimental design (Zizzo, 2010). For example, the fact that we reveal to players information about their group identities may suggest to them that we expect them to use this information somehow. To minimize such concerns, we rely on a between-subjects design. These designs are thought to be less vulnerable to the EDE critique (Zizzo, 2010). Furthermore, we refrain to give knowledge about players' experimental group identities in the instruction phase and in the trial round, so as to avoid implicitly suggesting homophilous or heterophilous link architectures.

The reminder of the number of connections of each player can be a second source of experimenter demand effects. It could be argued that this feature biases the results in the direction of efficiency, as it increases the saliency of network statistics related to efficiency enhancing strategies. Our aim in including this feature was to exclude the possibility that lack of familiarity with the graphical representation of the network would be driving departures from efficiency. Hence this design features is meant to "give the best shot" to the possibility of efficient networks. In the light of this design feature, our finding that network efficiency is significantly below potential becomes more compelling.

[9]In future work, we plan to incorporate information about the trial round and the understanding questions in the analysis of play.

[10]For example, somebody who has given more to the ingrop in the allocation task has not affected the wealth of ingroup individuals in the current session.

3 Predictions

To generate predictions about the play of the game, we follow much of the existing literature and assume myopic behaviour: the player with the turn chooses his link as if no more turns would follow, ie as if the network which obtains after his link is added will be the final network determining players' payoffs. Recent research has indeed shown that myopic best response is a common strategy in experimental network formation games (Conte et al., 2009).

We define some basic notation following Bala and Goyal (2000). Let N=(1,2,..,n) be the set of players. In T1, each player i choose a (pure) strategy $g_i = (g_{i,1}, g_{i,2}, ..., g_{i,i-1}, g_{i,i+1}, ..., g_{i,n})$, which is a vector of directed links $g_{ij} \in \{0, 1\}$. If $g_{ij} = 1$ i wins the prize whenever j wins it, but not necessarily vice versa. Let Γ_i be the set of possible values of g_i.[11] $\Gamma = \Gamma_1 \times \Gamma_2 \times ... \times \Gamma_n$ is the set of all possible combinations of player strategies. The vector of player strategies $g = (g_1, g_2, ..., g_n)$, drawn from Γ_i, can be represented as a directed network.

In network g, player i can access other players through a direct link, or through a path of links. Let $N_i^d(g) = \{j | g_{i,j} = 1 \ and \ j \in N \setminus i\}$ be the set of players with whom i has a direct link. $\mu_i^d = |N_i^d(g)|$ is thus the number of players with whom i has a direct link. We call this the *out-degree* of player i. $N_{-i}^d(g)$, on the other hand, is the set of players j such that $g_{ji} = 1$. $\mu_{-i}^d = |N_{-i}^d(g)|$ is the *in-degree*: the number of players who have a direct link to i.

A path from player i to player j is a series of links such that $g_{i,y} = g_{y,w} = ... = g_{z,j} = 1$.[12] When there is a path from i to j, i wins the prize whenever j wins it, but not necessarily vice versa. The notation $i \rightarrow^g j$ indicates that in network g there is a path from i to j. If $p_{i,j}(g) = 1$ i wins the prize whenever j wins it. Let $N_i(g) = \{k | i \rightarrow^g k\}$ and $\mu_i(g) = |N_i(g)|$. We call μ_i the *path out-degree*. Let $N_{-i}(g) = \{k | k \rightarrow^g i\}$ and $\mu_{-i}(g) = |N_{-i}(g)|$. We call $\mu_{-i}(g)$ the *path in-degree*.

Network g determines an expected payoff $\pi_i(g)$ for each player. This is simply calculated as the value of the prize, which we normalize to 1, times the probability of winning the prize, which is $\frac{1+\mu_i(g)}{n}$. Let $\pi_{-i}(g) = \{\pi_i, \pi_2, .., \pi_{i-1}, \pi_{i+1}, ..., \pi_n\}$.

[11] As each player is allowed at most one link, there are n possible values of of g_i: n-1 possible links plus the strategy of establishing no links at all.

[12] Notice a direct link is a path of lenght 1.

First assume that in T1 i tries to maximise expected payoff:

$$\max_{j} \ \ \pi_i(g + ij) \tag{1}$$

where g+ij is the network obtained from adding the link $g_{ij} = 1$ to network g. Let $g^*_{i,j}$ be the solution to maximization problem 1. In the first round of the game g_i has all of its entries equal to zero. In the second round, the decision maker has to consider the game as if g_i had only zero entries, as the link specified in the first round is removed once he declares his strategy. So, in both cases, $N_i(g) = \{\varnothing\}$ and $\mu_i(g) = 0$.

When i is evaluating the position in the network of each possible partner j, he is interested in all paths from j that lead to individuals other than i. We define the set of individuals to whom j has a path, excluding the path to i, as $N^{'}_{k,i}(g) = \{z | k \rightarrow^g z \ and \ z \in N \setminus i\}$. And hence $\mu^{'}_{k,i}(g) = |N^{'}_{k,i}(g)|$. Let $N^{maxout}_i(g) = \text{argmax}_k \, \mu^{'}_{k,i}(g)$.

Proposition 1. *Assume i is myopic and chooses her link to maximise equation 1, then* $g^*_{i,j} \in N^{maxout}_i(g)$

Proof. Rewrite equation 1 as: $\frac{1+\mu_i(g+ij)}{n}$. Notice that, as $\mu_i(g) = 0$, $\mu_i(g+ij) = 1 + \mu^{'}_j(g)$. Thus $\pi_i(g + ij) = \frac{2+\mu^{'}_j(g)}{n}$, which is monotonically increasing in $\mu^{'}_j(g)$. □

Propositon 1 says that in T1 a myopically rational agent who want to maximise expected payoff will always create a link towards a player with the maximum path out-degree.

Now let us turn to T2. In this treatment, every player chooses a strategy $\tilde{g}_i = (g_{1,i}, g_{2,i}, ..., g_{i-1,i}, g_{i+1,i}, ..., g_{n,i})$. If $g_{ji} = 1$ j wins the prize whenever i wins it, but not necessarily vice versa. The definition of $\tilde{\Gamma}_i$, $\tilde{\Gamma}$ and $\tilde{g}$ follow naturally the reasoning above.

As $\pi_i(g+ji) = \pi_i(g)$, a purely selfish player would be indifferent between forming and not forming a link, and would equally be indifferent about the consequences of the link he establishes on the other players. On the other hand, previous work on social preferences in experimental economics has shown that individuals care about the payoffs of the other players in heterogeneous, but systematic ways (Charness

and Rabin, 2002; Andreoni and Miller, 2002).[13] Following the literature on social preferences, we assume that players have a utility function that weights concerns for the player's own payoff and the payoff of all other players:

$$u_i(g) = \pi_i(g + ij) + \gamma f(\pi_{-i}(g + ij)) \tag{2}$$

To advance further we have to make some assumptions about the shape of function f. We can explore two archetipal candidates. The first is:

$$u_i(g + ij) = \pi_i(g + ij) + \gamma \sum_{j \in N \setminus i} \pi_j(g + ij) \tag{3}$$

Utility function 3 expresses a concern for aggregate social welfare, which Charness and Rabin (2002) argue to be the model of social preferences with the highest predictive power for dictator game allocations.

In our game, a ji link increases j's probability of winning the prize by letting j access i and the players towards whom i has a path. Notice that some of these players may already be accessed by j. We hence need to define a set of new players that j accesses because of the link to i. Call this set $N''_{i,j} = \{k | i \rightarrow^g k \ and \ j \not\rightarrow^g k\}$. Let $\mu''_{i,j} = |N''_{i,j}|$ be its cardinality.

Furthermore, notice that the in-degree of i in g is zero.[14] Hence $N_{-i}(g + ji) = j \cup N'_{-j,i}(g)$.[15]

We can then express the effect of an ij's link on utility 3 as:

$$u_i(g + ij) - u_i(g) = \gamma \left(\mu''_{i,j} + \sum_{k \in N'_{-j,i}(g)} \mu''_{i,k} \right) \tag{4}$$

[13]Notice that player i's strategy in T1 also has an impact on the payoffs of the individuals who have a path towards i in g. In future work, we will extend this section to include an the analysis of how other-regarding preferences affect behaviour in T1.

[14]The argument is symmetrical to the argument which we have used for T1: in round 1, the in-degree is zero by definition. In round 2, i has to choose as if her in-degree is 0, as the round 1 link will be deleted after her new choice.

[15] $N'_{-j,i}(g) = \{k | k \rightarrow^g j \ and \ k \in N \setminus i\}$, as we need to exclude i from the count of individuals that have a path towards j. In the same way that $N'_{j,i}$, as defined previously, indicates the set of individuals towards whom j has a path, excluding i.

From 4, it is obvious that as the path in-degree of j grows, the number of individuals who benefit from the ij link increases. Furthermore, i would like to choose a partner which minimizes the number of paths that are redundant for individuals in $N_{-i}(g + ji)$. If i has 0 outdegree, or if the number of overlaps is relatively low, this latter concern will be of second order. We hence hypothesize that myopic decision makers with preferences given by 4 will choose links towards the person with the largest path in-degree. In the majority of cases, this heuristics maximises 3.

An alternative paradigm to think about social preference would be that of inequity aversion (Fehr and Schmidt, 1999). Under inequity aversion, a player would feel guilt towards players with a lower out-degree and envy towards players with a higher out-degree. An inequity averse player in the first turn of an T2 session of our experiment would prefer not to form any link, as this would cause him to feel envious of the receiver of the link. This prediction is virtually always falsified in our pilot and main data. We thus do not explore the predictions of the model of inequity averse preferences any further. However, while players may not be have concerns for differences in expected payoff between themselves and their peers, they may still care about the ranking of peer welfare and, in particular, they may attach a special weight to the welfare of the peer who is least well off in the network. The literature in empirical social choice has documented this type of concern (Yaari and Bar-Hillel, 1984), which we can express in a simplified way using the following utility function:

$$u_i = \pi_i(g + ij) + \gamma \min_{j \in N \setminus i} \pi_j(g + ij) \tag{5}$$

Equation 5 is maximised by choosing j with the lowest path out-degree in g. This is the person with the least chance of winning. In case of a tie, we can assume that i will randomise between the partners with the minimum path out-degree.

We simulate link formation processes were all players follow either of the rules above. We have to make an assumption about what happens when more than one option is available that satisfies the rule. In line with Bala and Goyal (2000), we assume in such cases players randomise among their optimal options. We define architecture efficiency as the average path-outdegree in the final architecture over 5.[16] Our simulations show that in T1 the strategy of establishing ij links with

[16]Notice that the cycle architecture, whereby each player has path out-degree of 5 and all players get

the player with the highest path-outdegree delivers average efficiency of about 95 percent.

Further, our simulations show that in T2 the strategy of establishing ji links with the player with the highest path in-degree also delivers average efficiency of about 95 percent. Finally, we simulate a random link formation process and obtain average efficiency of about 48 percent. We hence make the following two hypotheses.[17]

Hypothesis 1. *Architecture efficiency is (i) higher than random and (ii) close to 95 percent.*

Hypothesis 2. *Architecture efficiency in the T2no treatment is not statistically different from architecture efficiency in the T1no treatment.*

On the basis of the discussion above, we also make the following hypothesis regarding individual decisions.

Hypothesis 3. *The probability of a link in T1 rises in the partner's path out-degree. The probability of a link in T2 increases in the partner's path in-degree and decreases in the partner's path out-degree.*

In order to make predictions about behaviour when social identity is common knowledge, we have to augment the utility function above. We follow the seminal paper of Akerlof and Kranton (2000) and introduce a positive effect on utility which comes from following a prescription P associated with the group one is member of:

$$u_i(g) = \pi_i(g) + \gamma f(\pi_{-i}(g)) + P_i(g_i) \tag{6}$$

Figures 1 and 2 in the appendix show descriptive evidence from our sample which documents a shared belief that ingroup links are prescribed.[18] Suppose then

the prize, is always achievable. We hence compare reliased efficiency to the highest achievable.

[17]Simulation show that a link formation rule whereby individuals only link to the player with the smallest path-outdegree delivers 75 percent efficiency, which is much higher than what the random link formation rule achieves.

[18]After play, subjects are asked whether they think that in a game like the one that has just been played a player "should" only link to an ingroup peer.

links with ingroup players are prescribed, so that individuals get positive utility whenever they create an ingroup link: $P_i(g_i) = c > 0$ whenever i has linked to an ingroup peer. When group identity becomes common knowledge, we expect two types of effects. First, for any positive value of c, whenever the best response set contains both ingroup and outgroup members, decision makers with preferences given by 6 would link to an ingroup peer. Second, if the benefit c from following the prescription is high enough, we will also observe more ingroup links in cases when the best response set contains only outgroup peers. In this case, simulations suggest that the efficiency of the network will decrease. We hence formulate the following, final hypothesis.

Hypothesis 4. *Common knowledge of group identity generates networks characterised by (i) more ingroup links and (ii) lower efficiency.*

We analyze the data by comparing mean session outcomes in terms of efficiency and ingroup links, and by studying individual decisions. To do the latter we use dyadic regression analysis. In particular, we use models of the following form:

$$l_{ij} = \alpha + \beta X_j + \gamma\, ingroup_{ij} + e_{ij} \tag{7}$$

The unit of observation is all i-j dyads. l_{ij} is a dummy which takes value 1 if player i has chosen to establish a link with player j. The matrix X contains characteristics of j which our theoretical framework predicts will have an influence on i's link formation decisions. These include the path in-degree and path out-degree. The dummy $ingroup_{ij}$ takes a value of 1 when i and j belong to the same group.[19]

We investigate treatments effects by means of simple dummies interacted with the characteristics of interest.

Model (7) will be estimated using OLS, correcting standard errors for arbitrary correlation at the session level. Previous studies have shown that when the number of independent groups of observations is low, which is often defined as less than 42, hypothesis tests based on clustered standard errors over-reject the null. In this

[19]IDs A,C,and E are always assigned to the mango group. IDs B,D,F are always assigned to the pineapple group.

paper we cluster standard errors over 92 independent clusters and we hence do not need to worry about further standard error corrections.

We also present graphs that report the probability that a decision is consistent with four archetypal link formation rules: (i) choose the player with highest out-degree, (ii) choose the player with lowest out-degree, (iii) choose the player with the highest in-degree, (iv) choose the player with the lowest in-degree.[20]

4 Data

We run our field experiment in villages in the rural areas around Pune, in the Indian state of Maharashtra. Villages are situated approximately 1,30 to 3h hours away from Pune. The villages are randomly sampled from a census list of all villages in 4 sub-districts. As the Pune district includes both mountainous and plain areas, we choose to focus on two subdistricts which mostly comprise mountainous areas, and two subdistricts which mostly comprise plain areas.

We rely on a door-to-door random sampling technique. Before reaching the village, our team is shown a Google Earth map of the village. On alternating days, the teams start sampling from the periphery of the village or from the center of the village, which is typically a small square in front of the village temple. We invite to the experiment all male adult farmers who are encountered in the door-to-door visit until we have enough farmers to fill in all planned sessions.

Data collection took place between September and October 2013. In total, we run 81 sessions with 446 subjects. We run 20 sessions of T1no, T1id and T2id, and 21 sessions of T2no. In table 2 in the appendix, we present some regressions that test for covariates balance across treatments. We rely on a small set of covariates which we measure through a short post-play questionnaire. We are unable to find any statistically significant difference in covariates across treatments.

5 Results

We first pool all sessions with no knowledge of group identity together and compare the distribution of average session efficiency to the distribution of average session

[20]From now on, we will drop the qualification "path" to describe path out-degree and path in-degree

efficiency which would obtain if individuals chose their links at random. We obtain the following result, which is represented graphically in figure 4 below:

Result 1. *Architecture efficiency in T1no and T2no is 64 percent, above random play but substantially below potential.*

We can hence find only partial support for our first hypothesis. Efficiency is clearly higher than what random play would have achieved, but 30 percentage points below the mean level for myopic rational play.

< **Figure 4 here** >

Interestingly, final efficiency is very similar in T1no and T2no, in line with what we have hypothesized in our second prediction.

Result 2. *Efficiency in T2no sessions is not significantly different from efficiency in T1no sessions.*

We can explore how this result comes about by means of dyadic regression model 7. We obtain the following result, which confirms hypothesis 3:

Result 3. *In T1no, links are significantly more likely towards peers with a higher out-degree. In T2no, links are significantly more likely towards peers with a higher in-degree, and peers with a lower out-degree.*

As hypothesized, we find that in T1no ties are directed towards peers with a higher out-degree. In T2no, on the other hand, ties are directed towards peers with higher in-degree, or lower out-degree. The effects are of a meaningful magnitude. In T1no, player i is 20 percentage points more likely to choose a peer with path out-degree of 4 than a peer with a path out-degree of 0. In T2no, player i is 13 percentage points less likely to choose the peer with a path out-degree of 4 than a counterpart with path out-degree of 0, and is 16 percentage points more likely to pick a peer with a path in-degree of 4 than a counterpart with a path in-degree of 0.

< **Table 3 here** >

Table 3 reveal a further effect: in T1no player i is more likely to establish a link with a peer with a lower in-degree. This result is difficult to explain within our framework. One possibility is that links carry social value for the receiver. In this case, the utility function has to be augmented with an element reflecting this social value of links. Individuals who choose peers with a low in-degree in T1no could thus be targeting the players who have accumulated the minimum social value in the game so far. We are in position to provide a direct test for this interpretation.[21]

The results of regression analysis are confirmed if we categorise decisions in terms of their consistency with a particular archetypal linking rule. As there are often multiple candidates who satisfy a particular linking rule, rules with a larger number of candidates would appear more frequently even if individuals played completly at random. We calculate confidence intervals around the probability of observing decisions consistent with a given rule in our data, and compare this to the probability of choosing such rule under random play. Results for both the T1no and T2no treatments are reported in table 7 and 8. In T1no, links towards individuals with maximum out-degree and minimum in-degree are observed more often than under random play. In T2no, links towards individuals with maximum in-degree and minimum out-degree are observed more often than under random play. Notice that probabilities do not add up to 1 because a single decision can be consistent with many strategies. For example, when players are arranged on a line, the first player is both the player with highest number of in-degree and the player with least number of in-connections. If we restrict attention to situation where the sets of peers satisfying the two most common linking rules are disjoint, we observe that maximum out-degree is more frequent than minimum in-degree in T1no and maximum in-degree is more frequent than minimum out-degree in T2no. As the number of observations shrink substantially, however, we do not have statistical power to distinguish these probabilities from random play.

Table 9 shows the probability that an individual plays consistently with an archetypal strategies in both rounds. The picture is qualitatively unchanged, with archetypal strategies being played more often than random, as hypothesized.

[21] In the post-play questionnaire farmers are asked the following question: "Do you think that choosing a farmer from your own group is a way of showing respect to him?". 51 percent of farmers answer yes to this question. This evidence is consistent with the view that links carry social value, but represents by no means a credible test.

We next turn to the the effect of social identity. First we show results from the initial allocation task.. Figure 14 shows that the modal allocation was characterised by ingroup bias. Overall 54 percent of individuals showed ingroup bias, while 30 percent were unbiased. This is evidence that the saliency of groups was sufficient to substantially affect behaviour in this different, but related domain.

Second, we ask participants at the end of the game whether they thought that a player in this game "should" only link to ingroup peers. Figures 1 and 2 document that most players answer yes to this question, and expect the majority of peers to also agree with the norm. This is evidence consistent with a social norm of homophily in our field setting.

Our main result on the identity treatments is the following:

Result 4. *In the identity treatments, ingroup links come more frequent but architecture efficiency does not decrease.*

This result confirms the first part of hypothesis 3 and rejects the second part. Figures 10 and 11 show mean levels of homophily and efficiency for different treatments. Mean homophily increases in both T1 and T2. Mean efficiency is essentially unaffected.

< Figures 10 and 11 here >

These effects are confirmed by statistical analysis. The social identity treatment significantly raises the average number of ingroup links, while it does not significantly affect efficiency.

< Table 4 here >

Tables 5 and 6 in the appendix show how ingroup links increase in both T1 and T2 treatments. In T1 this comes about without decreasing the frequency of efficiency enhancing (maximum out-degree) links. However, links towards maximum out-degree ingroup players become much more frequent. Efficiency enhancing (maximum in-degree) links decrease in a much more pronounced fashion in the T2 treatment. Links towards maximum in-degree ingroup players become more frequent.

< **Tables 5 and 6 here** >

We separate decisions where the set of efficiency enhancing best responses and of ingroup peers are disjoint, and decisions where the set intersect. We call these tradeoff and no tradeoff decisions. We observe that in T1 there is an increase in ingroup links both for tradeoff and for no tradeoff decisions. In T2 the increase in homophilous links is concentrated in tradeoff decisions. In both cases, we cannot detect changes significant at the 5 percent level. Figures 12 and 13 in the appendix report these results.

6 Conclusion

Peer networks play a crucial rule in the diffusion of innovations such as new agricultural technologies, health schemes or financial products. While theoretical models show that in a variety of cases myopically rational individuals can converge to efficient network architectures after repeated play (Bala and Goyal, 2000), a number of field studies have documented losses in efficiency associated with homophily- the tendency of similar individuals to interact with each other with disproportionate frequency. Identity economics would explain this tendency by positing that individuals get positive utility from following group-prescribed behaviour and are willing to sacrifice some efficiency to maintain conformity. Experimental evidence on both efficiency and homophily is however scarce. In particular, to our knowledge, no link formation experiment has ever been played with non-student populations.

We devise a sequential, unilateral, one-way-flow link formation game where we can test the predictions of a simple model of strategic link formation and those of an augmented model which includes an identity component. Our design allows us to answer two related questions. First, to what extent can farmers form efficient network architectures? Second, does the introduction of group identity result in architectures characterised by more ingroup links and less efficiency? We run our game as an artefactual field experiment in rural India. The testing domain is closest to the domain of potential application of the findings. Trials of peer-to-peer extension and diffusion campaigns have already been carried out in rural India. Some of these document homophily-related losses in efficiency. A deeper understanding of the efficiency of network formation and the extent to which this is compromised by social prescriptions will be a useful input for the formulation of new intervention designs.

We find that realised efficiency is higher than that achieved by a purely random link formation process, but lower than the level of efficiency which myopically rational play would have achieved. In the treatment where agents form out-links, the simple rule to link with the player with the highest out-degree at that point in the game would deliver 95 percent efficiency. Realised efficiency is about 64 percent, or 31 percentage point below what the simple myopic rule could have achieved. Interestingly, when agents are only allowed to form in-links architecture efficiency is not significantly different from efficiency under out-link formation. When information about group membership is revealed, social networks become more homophilous, but not significantly less efficient. Exogenous social identity, in our experiment, is not the cause of network inefficiency.

References

Akerlof, G. and R. Kranton (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being.* Princeton University Press.

Akerlof, G. A. and R. E. Kranton (2000, August). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Anderson, S. (2011, September). Caste as an impediment to trade. *American Economic Journal: Applied Economics 3*(1), 239–63.

Andreoni, J. and J. Miller (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica 70*(2), pp. 737–753.

Bala, V. and S. Goyal (2000, September). A noncooperative model of network formation. *Econometrica 68*(5), 1181–1230.

Bandiera, O. and I. Rasul (2006, October). Social networks and technology adoption in northern mozambique. *Economic Journal 116*(514), 869–902.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2012, January). The diffusion of microfinance. NBER Working Papers 17743, National Bureau of Economic Research, Inc.

Barr, A. and G. Genicot (2008, December). Risk sharing, commitment, and information: An experimental analysis. *Journal of the European Economic Association 6*(6), 1151–1185.

Ben Yishay, A. and M. Mobarak (2012). Communicating with farmers through social networks. *Working paper*.

Benjamin, D. J., J. J. Choi, and A. J. Strickland (2010, September). Social identity and preferences. *American Economic Review 100*(4), 1913–28.

Berg, E., M. Ghatak, R. Manjula, D. Rajasekhar, and S. Roy (2013, March). Motivating knowledge agents: Can incentive pay overcome social distance? STICERD - Economic Organisation and Public Policy Discussion Papers Series 42, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of social issues 55*(3), 429–444.

Charness, G. and M. Rabin (2002, August). Understanding social preferences with simple tests. *The Quarterly Journal of Economics 117*(3), 817–869.

Charness, G., L. Rigotti, and A. Rustichini (2007, September). Individual behavior and group membership. *American Economic Review 97*(4), 1340–1352.

Chen, Y. and S. X. Li (2009). Group identity and social preferences. *The American Economic Review*, 431–457.

Conley, T. G. and C. R. Udry (2010, March). Learning about a new technology: Pineapple in ghana. *American Economic Review 100*(1), 35–69.

Conte, A., D. D. Cagno, and E. Sciubba (2009, June). Strategies in social network formation. Birkbeck Working Papers in Economics and Finance 0905, Birkbeck, Department of Economics, Mathematics & Statistics.

Dunning, T. and J. Nilekani (2013, 1). Ethnic quotas and political mobilization: Caste, parties, and distribution in indian village councils. *American Political Science Review 107*, 35–56.

Eckel, C. C. and P. J. Grossman (2005, November). Managing diversity by creating team identity. *Journal of Economic Behavior & Organization 58*(3), 371–392.

Emerick, K. (2013). The efficiency of trading in social networks: Experimental measures from india. Technical report.

Fafchamps, M. and B. Minten (2012). Impact of sms-based agricultural information on indian farmers. *World Bank Economic Review 26*(3), 383–414.

Falk and Kosfeld (2012, September). It's all about connections: Evidence on network formation. *Review of Network Economics 11*(3), 1–36.

Fehr, E. and K. M. Schmidt (1999, August). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Foster, A. D. and M. R. Rosenzweig (1995, December). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy 103*(6), 1176–1209.

Golub, B. and M. O. Jackson (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics 127*(3), 1287–1338.

Guha, R. (2008). *India After Gandhi: The History of the World's Largest Democracy.* Ecco book. HarperCollins.

Hoff, K. and P. Pandey (2006, May). Discrimination, social identity, and durable inequalities. *American Economic Review 96*(2), 206–211.

Kranton, R., M. Pease, S. Sanders, and S. Huetter (2012). Identity, group conflict, and social preferences. Working paper.

Krishnan, P. and M. Patnam (2012). Neighbors and extension agents in ethiopia: who matters more for technology dffusion. Working papers, International Growth Center.

Munshi, K. (2004, February). Social learning in a heterogeneous population: technology diffusion in the indian green revolution. *Journal of Development Economics 73*(1), 185–213.

Rogers, E. M. (2003). *Diffusion of innovations (5. ed.)*. Free Press.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. CUP Archive.

Viceisza, A. (2012). Treating the field as a lab: A basic guide to conducting economics experiments for policymaking. Food security in practice technical guide series 7, International Food Policy Research Institute (IFPRI).

Yaari, M. and M. Bar-Hillel (1984). On dividing justly. *Social Choice and Welfare* (1), 1–22.

Yamagishi, T. and T. Kiyonari (2000). The group as the container of generalized reciprocity. *Social Psychology Quarterly 63*(2), pp. 116–132.

Zizzo, D. (2010, March). Experimenter demand effects in economic experiments. *Experimental Economics 13*(1), 75–98.

7 Appendix

Figure 1: "In the link formation game you have just played, do you think a player "should" only link to a peer of his own group?" Percentage of players who answered YES

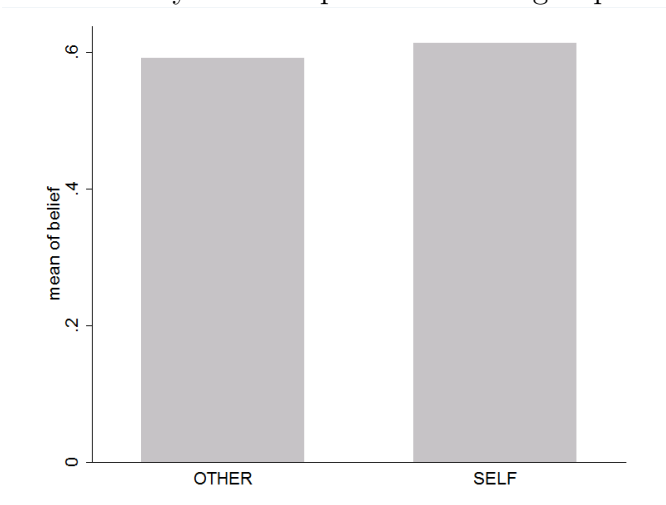


Figure 2: "How many of the other 5 players in the session do you think answered YES to the previous question?" Distribution of expectations

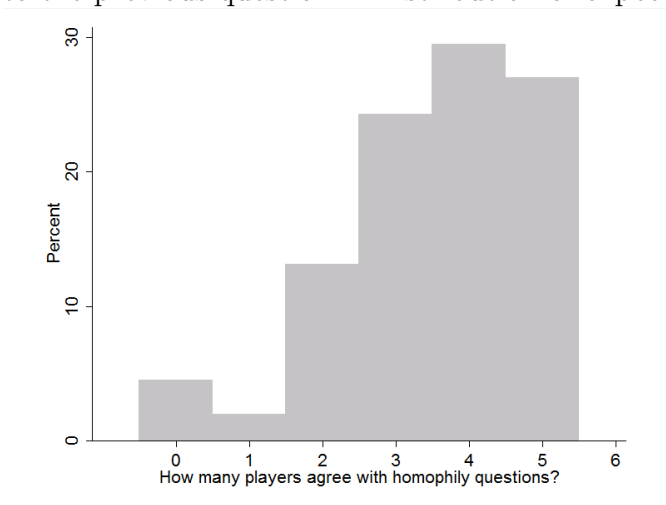


Figure 3: Cumulative distribution of mistakes in understanding questions

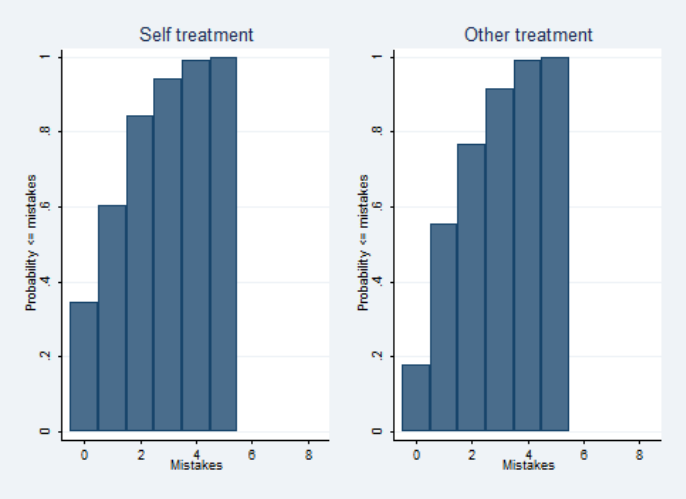


Figure 4: Distribution of architecture efficiency in real and simulated networks

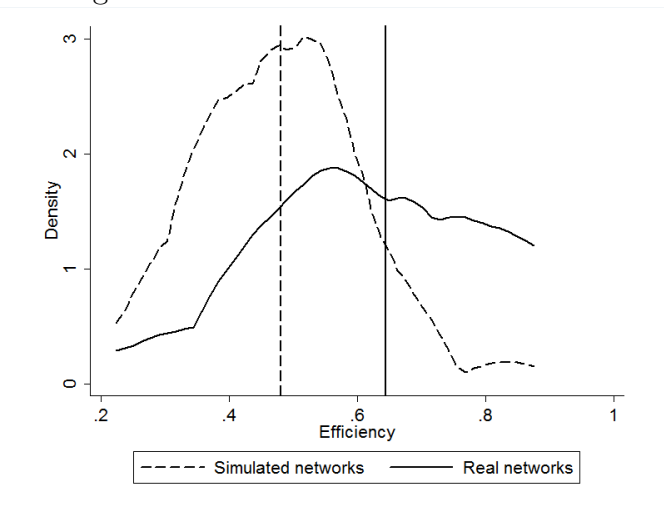


Figure 5: Distribution of efficiency in T1no and T2no

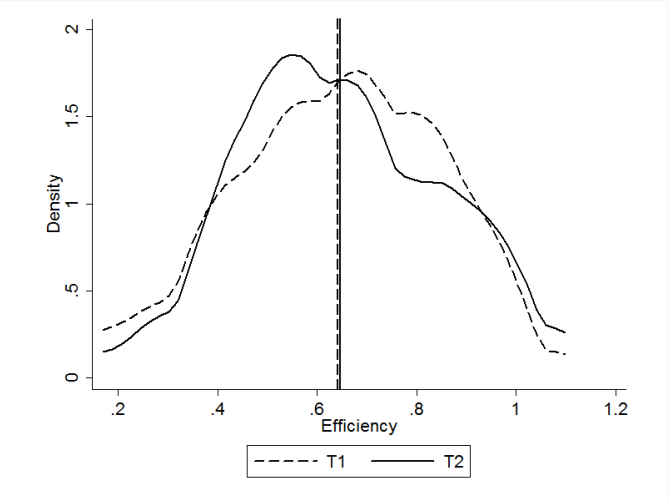
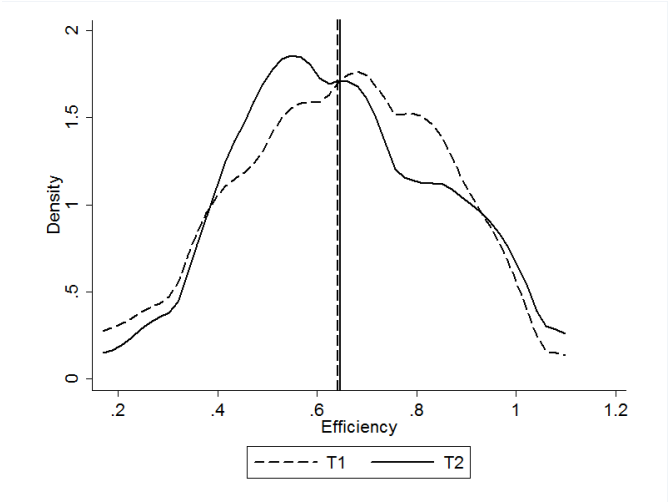


Figure 6: Distribution of efficiency in T1no and T2no



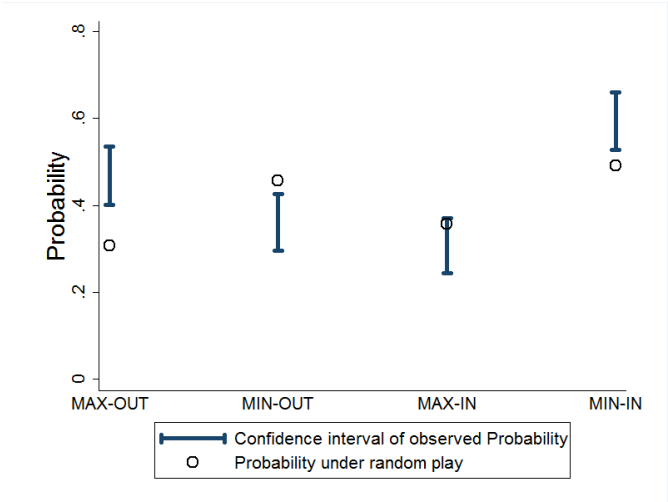


Figure 7: Probability that a decision is consistent with a particular strategy. T1no treatment

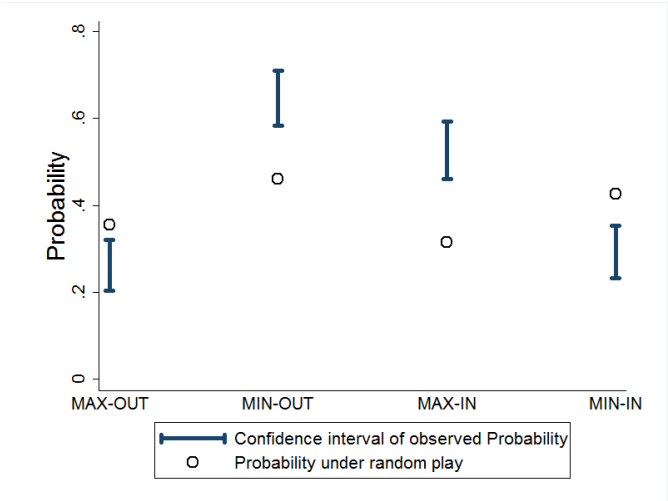


Figure 8: Probability that a decision is consistent with a particular strategy. T2no treatment

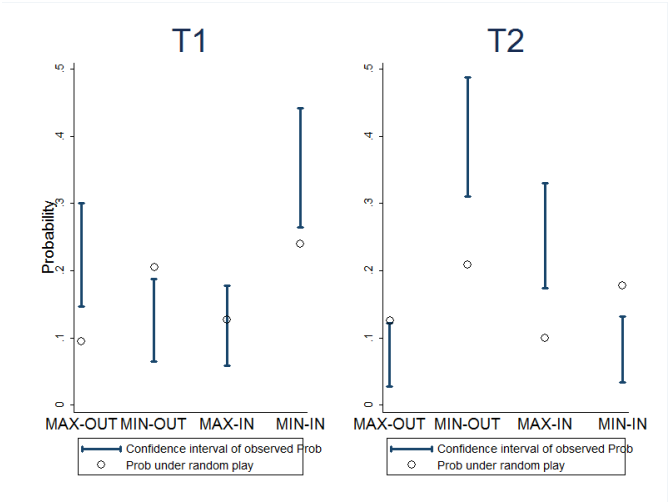


Figure 9: Probability that a individual plays consistently with a particular strategy. T1no and T2no treatments

Figure 10: Mean Homophily

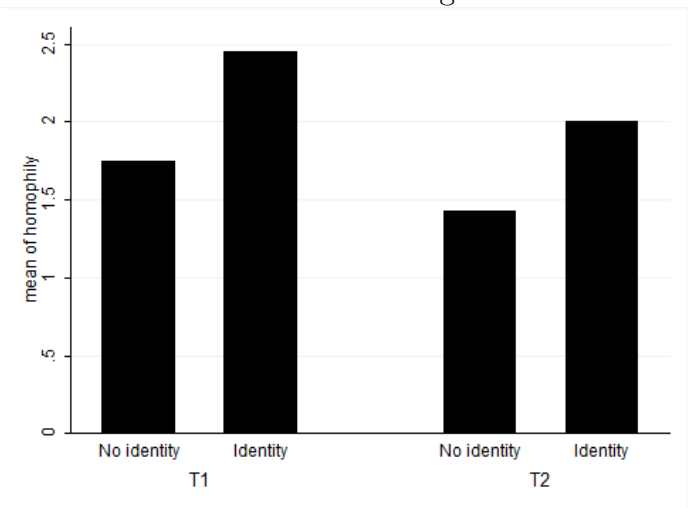
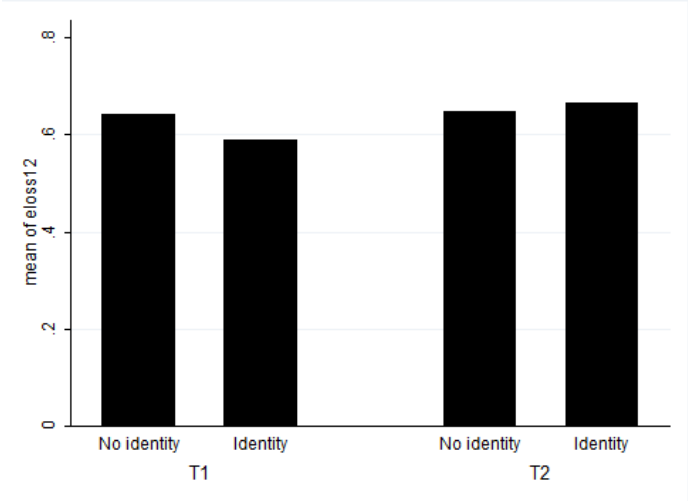


Figure 11: Mean Efficiency



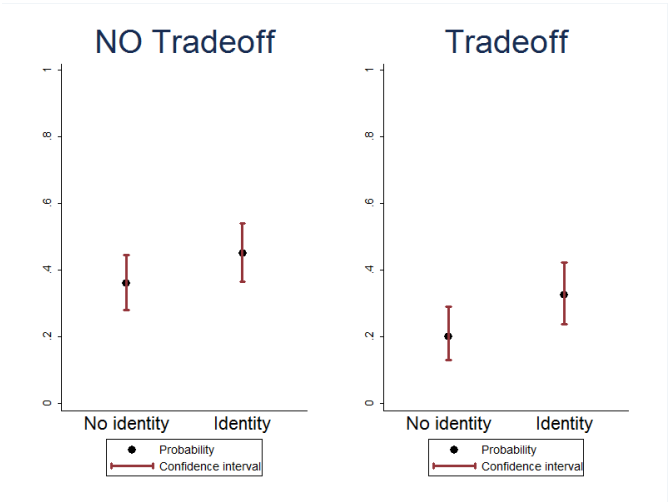


Figure 12: Probability of a ingroup link, for NO Identity and Identity treatments. T1 treatment. Tradeoff = 1 if no ingroup partner is MAX-OUT

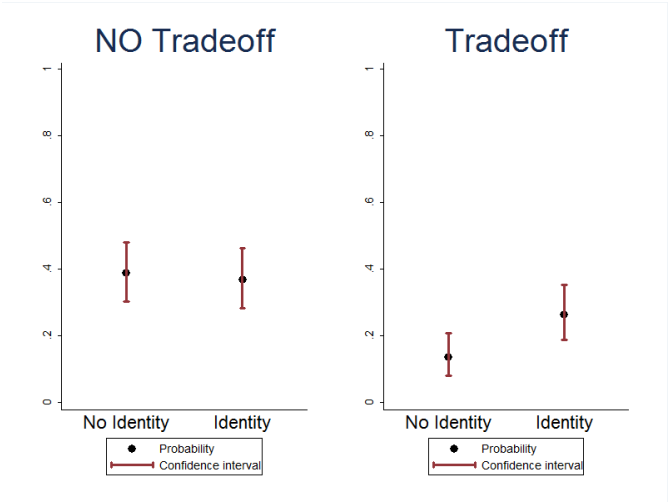


Figure 13: Probability of a ingroup link, for NO Identity and Identity treatments. T2 treatment. Tradeoff = 1 if no ingroup partner is MAX-IN.

Figure 14: Distribution of coin allocations to ingroup partner

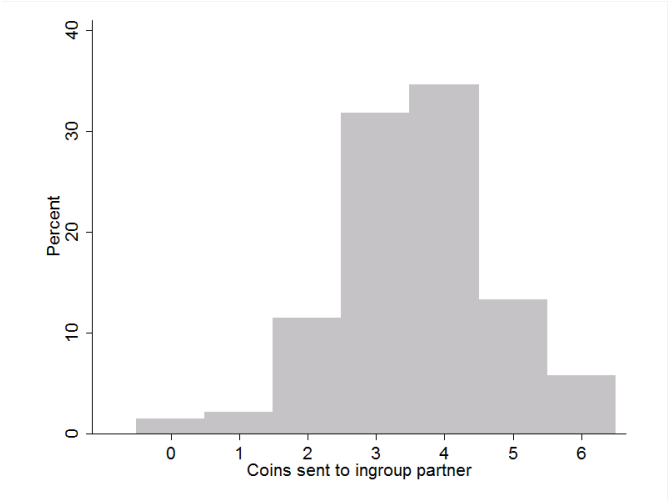


Table 2: OLS regression: Covariate Balance

	T1id	T2no	T2id
	(1)	(2)	(3)
Age	-2.378	-3.664	-1.785
	(2.530)	(2.569)	(2.168)
Education	.100	.043	.025
	(.083)	(.077)	(.072)
UpperCaste	-.067	-.010	-.167
	(.096)	(.096)	(.104)
LandOwned	.077	-.069	-.022
	(.733)	(.608)	(.693)
LandCultivated	.141	-.008	.053
	(.653)	(.546)	(.629)
NetworkSize	.111	2.141	1.099
	(1.095)	(1.765)	(1.381)
SelfRatedUnderstanding	-.008	.166	.008
	(.128)	(.135)	(.134)

OLS regressions. The dependent variable is indicated in the row's name. Upper caste is a variable that takes value of 1 if respondent is not from a schedule caste, a scheduled tribe or an Other Backward Caste. Network size is the self reported number of peers with whom the farmer exchanges advice on agricultural matters. SelfRatedUnderstanding takes a value of 1 if the respondents rates his understanding of the game as "very high". Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors reported in parentheses.

Table 3: Dyadic Linear Probability Model

	T1	T1fe	T2	T2fe
	(1)	(2)	(3)	(4)
j's path out-degree	.051	.052	-.040	-.034
	(.008)***	(.008)***	(.007)***	(.007)***
j's path in-degree	-.030	-.031	.049	.043
	(.008)***	(.008)***	(.008)***	(.006)***
Const.	.175	.240	.187	.249
	(.004)***	(.058)***	(.006)***	(.053)***
Obs.	2400	2400	2460	2460

Dyadic OLS regression. Dependent variable is a dummy which takes a value of one if i chose to establish a link with j. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors corrected for clustering at session level are reported in parentheses. Columns 2 and 4 include dyadic dummies.

Table 4: OLS Regression over sessions

	Homophily	Homophily2	Efficiency	Efficiency2
	(1)	(2)	(3)	(4)
Identity	.640	.700	-.018	-.054
	(.278)**	(.419)*	(.043)	(.061)
T2		-.321		.004
		(.360)		(.062)
T2*Identity		-.129		.073
		(.558)		(.087)
Const.	1.585	1.750	.644	.642
	(.178)***	(.298)***	(.031)***	(.044)***
Obs.	81	81	81	81

OLS regression. The dependent variable in columns 1 and 2 is the number of ingroup links in the final architecture. The dependent variable in columns 3 and 4 is the average out-degree in the final architecture. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors reported in parentheses.

Table 5: Percentage of decisions consistent with a particular strategy. All rounds. T1 treatments

	Mean		
	Ingroup	Max out-degree	Max out-degree and Ingroup
No identity	0.28	**0.47**	0.33
Identity	0.40	**0.47**	0.44

Table 6: Percentage of decisions consistent with a particular strategy. All rounds. T2 treatments

	Mean		
	Ingroup	Max in-degree	Max out-con and Ingroup
No identity	0.26	**0.53**	0.28
Identity	0.31	**0.43**	0.33

A Theory of Good Intentions*

Paul Niehaus
UC San Diego

November 15, 2013

Abstract

Why is other-regarding behavior often misguided? I study a new explanation grounded in the idea that altruists want to *think* they are helping. Frictions arise because perception and reality can diverge ex post when feedback is limited (as for example when donating to international development projects). Among other things the model helps explain why donors have a limited interest in learning about effectiveness; why intermediaries may market based on need, effectiveness, or neither; and why beneficiaries may not be able to do better than accept this situation. For policy-makers, the model implies a generic tradeoff between the quantity and quality of generosity.

*I thank Nageeb Ali, Jim Andreoni, Navin Kartik, Joel Sobel, Adam Szeidl, and seminar participants at Microsoft Research New England, Columbia, UCLA, and NEUDC for helpful comments. Microsoft Research New England provided generous hospitality.

1 Introduction

Other-regarding behavior poses a challenge for social scientists. On the one hand, some people are remarkably generous. Americans give about 2% of GDP to charity each year, for example.[1] This suggests that they care deeply about helping others. Yet in many cases generous people are also quite poorly informed about how to help effectively. For example, only 3% of charitable givers even *claim* to give based on research comparing the effectiveness of alternatives.[2] This pattern is so common that it is embodied in colloquial language, where "well-intentioned" is a euphemism for "poorly informed." Yet if people really are well-intentioned, why don't they *become* well-informed?

Economists have predominantly taken the view that funders want to be effective but find it difficult to learn how. Krasteva and Yildirim (2013) emphasize that the costs of learning may exceed the benefits for small donors. Development economists highlight the role of market failures: information about effectiveness is a public good (Duflo and Kremer, 2003; Levine, 2006; Ravallion, 2009), and communication from practitioners to funders is often distorted by strategic considerations (Pritchett, 2002; Duflo and Kremer, 2003; Levine, 2006). Institutions that produce and disseminate effectiveness research (e.g. CGD, J-PAL, IPA, CEGA) were created in part to address these concerns.

This paper examines an alternative (and complementary) interpretation: funders do not want to be more effective. Instead, they want to *think* that they are effective. Yet perception and reality can diverge. To illustrate the core premise, consider donating to help feed malnourished African children. This induces agreeable thoughts of children eating. Now suppose you learn that the charity in question is ineffective – perhaps an exposé reveals serious fraud. Presumably this reduces your satisfaction. What is more interesting is the counterfactual: if you had not learned of the fraud, you would have continued to experience "warm glow" (Andreoni, 1989) thinking about your impact *even though in reality no such impact existed*. Your altruistic preferences cannot literally be over children's outcomes as these occur on another continent, outside of your experience. Instead, perceptions count. This raises the question: how and how well will learning work in a market where perceptions *are* the product?

I study this question in a model of a single benefactor and beneficiary; the model thus abstracts from public goods issues. The benefactor does not know ex ante how his decisions will affect the beneficiary ex post. The unusual feature of the model is that this uncertainty persists ex post with positive probability. As a result the benefactor may face residual *ambiguity* which he must interpret. For example, a donor may receive no news about whether the charity he gave to was honest and have to decide what this implies. He cannot learn the correct interpretation through repeated experience, precisely because the true state remains unobserved. He therefore adopts the interpretation that maximizes his expected utility. This approach builds on evidence from psychology and economics that people tend to interpret information in a self-serving manner (Mobius et al., 2012).

[1] Author's calculation using data from The Giving Institute (2013) and the Bureau of Economic Analysis (`http://www.bea.gov/national/index.htm#gdp`, accessed 7 August 2013).

[2] See Hope Consulting (2012). The Hope sample over-represents wealthier donors and thus if anything likely overstates the amount of research done by the average donor.

The beliefs this yields have a seemingly innocuous structure: they are (endogenously) Bayesian, and they are consistent with the distribution of all observable data. As a result they are not readily falsifiable. For example, a well-intentioned donor correctly forecasts the probability that he will learn about a scandal involving his chosen charity. On learning of no scandals, however, the same donor assumes that "no news is good news" and views the charity as definitely honest. Because this effect appears only in the presence of ambiguity, the model predicts relatively standard decision-making when outcomes are observable (such as helping a neighbor) but relatively distorted behavior when outcomes are unobserved (such as helping internationally).

Given this interpretation strategy, the benefactor has mixed feelings about learning. On the one hand, he always prefers to avoid ex-post feedback as this constrains his beliefs. A donor who learns that his donation was stolen, for example, is directly worse off as this makes it difficult to believe that it was effective. On the other hand, the benefactor does want to obtain a limited amount of information ex ante, precisely in order to avoid such disappointments. Before donating, for example, a donor would like to know whether an unpleasant scandal will later break. The general result is that the benefactor prefers to do just enough research ex ante to accurately forecast the feedback he will receive ex post, but no more.

These motives in turn shape the marketing strategies that maximize revenue for intermediary organizations such as charities. Critics argue that these organizations provide too little information about effectiveness, with one writing that "useful information about what different charities do and whether it works isn't publicly available anywhere."[3] In the model, however, there is a sense in which this is simply good marketing. Intermediaries see their revenue fall in expectation if they commit to conducting an impact evaluation (formally, generating information about parameters that complement the benefactor's action). The reason is that the benefactor's interests are already aligned with those of the intermediary: he actually *wants* to believe the best about impact, and so further information is more likely to hurt than to help. Conversely, the intermediary benefits from marketing based on need. Formally, revenue increases in expectation with information about parameters that substitute for the benefactor's action. Need is a compelling strategy because of a conflict of interest between the parties: the benefactor wants to believe that things are not that bad, while the intermediary wants him to confront a harsher reality. This may help explain why nonprofit organizations often market using graphic depictions of need (e.g. "poverty pornography" images) and "awareness-raising" campaigns rather than cost-effectiveness claims or research on impact.

The result on effectiveness illustrates a broader theme: a tradeoff between the *quality* and the *quantity* of giving. From the point of view a policy-maker, good intentions are problematic as they may direct resources to relatively ineffective causes. For example, a new approach to poverty reduction with little concrete evidence may capture funders' imagination and attract large sums. The policy-maker could address this by sponsoring rigorous impact evaluation research. If (as expected) the results do not live up to the hype, funders will turn to alternatives. Definitionally, however, funders will be less excited about these alternatives than they originally were about the new approach. As a result, total giving will tend to fall. The

[3] GiveWell, `http://www.givewell.org/about/story`, accessed 10 September 2013.

policy-maker must therefore choose between a larger volume of poorly-informed funding and a smaller flow of better-informed giving. For the same reason even the beneficiary may prefer not to reveal the true state to the benefactor, as it may be better to receive large amounts of help in inefficient ways than to disillusion the giver.

Because it is explicitly built on utility from thoughts and perceptions, the model conveniently organizes a set of facts related to saliance. The link is simply that whatever brings those thoughts to mind tends to raise the return on giving. This helps explain, for example, why donors are more likely to support work on problems that have affected their loved ones (Small and Simonsohn, 2008), and why charities spend money to thank donors repeatedly for past gifts. It may also explain why charities encourage donors to think of their gifts as buying discrete, memorable item (e.g. cows) even when in reality (and in the fine print) they have no influence over fund allocations.

The results above for a "pure" altruist might plausibly set an upper bound on the effectiveness of people with more nuanced motives. Several have been proposed in the literature. Duncan (2004) argues, for example, that some donors care not about beneficiary's welfare per se but about the *impact* of their actions. More recently, Andreoni et al. (2012) present evidence suggesting that *guilt* plays a role. Section 4 applies the good intentions framework to a class of preferences that nest these motives as special cases, depending on the reference point against which the benefactor evaluates outcomes. For the benefactor this leaves matters qualitative unchanged: he continues to do a limited amount of research ex ante and avoid feedback ex post. For other players in the market, however, incentives may reverse. Impact philanthropists are a nonprofit's ideal customers, as they are completely aligned in their desire to believe that donations have a large marginal impact. Guilty givers, on the other hand, pose a challenge; they seek to assuage their guilt by convincing themselves that needs are exaggerated and that nothing they could do would ever really make a difference. They thus provide the sole case in which intermediaries can benefit from marketing based on effectiveness, as well as need.

How broadly applicable are these ideas? The model could be interpreted as describing any sort of other-regarding preference. Empirically the link is tightest to individual charitable giving where, as mentioned above, donors give but do not conduct much research. Donor's qualitative comments further highlight their use of interpretation and assumption. Donors told Hope Consulting (2012), for example, that "with known nonprofits, unless there is a scandal, you assume they are doing well with your money" (p. 38) and that "I don't research, but I am sure that the nonprofits to which I donate are doing a great job." (p. 42), leading the authors to conclude that "this creates a big challenge to getting people to do more research – they see no need to do so." (p. 44) Evidence from laboratory experiments corroborates this. Fong and Oberholzer-Gee (2011) find, for example, that only 1/3 of subjects are willing to pay $1 to learn whether they are playing a $10 dictator game with a disabled person or a drug user. Similarly Dana et al. (2007) find that only 56% of dictators choose to observe *free* information on the relationship between their actions and the recipient's payoffs.[4]

[4]The arguments may also apply to more localized gift-giving. Unwanted Christmas gifts, for example, are so common that there are websites devoted to displaying bad examples: knick-knacks, ugly sweaters, and so on (see for example `www.badgiftemporium.com` or `whydidyoubuymethat.com`). Waldfogel (2009) argues that holiday gift-giving is so wasteful that people should stop it entirely.

For institutional funders data are scarcer, but industry veterans have similar concerns. As recently as 2006 David Levine wrote to argue for "Building Learning into the Global Aid Industry;" by his count, "rigorous evaluations of the impacts of development programs remain rare. In its first 55 years, the World Bank published exactly zero. The U.S. Agency for International Development (USAID) had a better record: that organization funded one randomized study in the 1970s and another one in the 1990s" (Levine, 2006). Pritchett (2002) describes his years in the aid industry as "ignorant armies clashing by night," with "very rarely any firm evidence presented and considered about the likely impact of... proposed actions." Interestingly, Easterly (2006) emphasizes the role of faith and desire: "I feel like kind of a Scrooge... I speak to many audiences of good-hearted believers in the power of Big Western Plans to help the poor, *and I would so much like to believe them myself*" (emphasis added). No doubt this desire is only part of the story, alongside political and organization forces. But it is consistent with the idea that there is something fundamentally different about spending money on others' behalf.[5]

Conceptually the paper draws on and extends two strands of theoretical research. First, it takes quite literally Andreoni's (1989) influential idea that altruists benefit from the "warm glow" that their acts induce. Andreoni has emphasized that "the warm-glow hypothesis simply provides a direction for research rather than an answer to the puzzle of why people give – the concept of warm-glow is a placeholder for more specific models of individual and social motivations" (Andreoni et al., 2012). The present paper offers one such model linking warm glow to perceived outcomes.

Second, it draws inspiration from Brunnermeier and Parker's (2005) theory of optimal expectations. The key technical difference is that, unlike in their model, the decision-maker gets no utility from anticipation or remembrance and faces no tradeoff between anticipatory and flow utility; instead his *sole* objective is to hold pleasant thoughts. As a result he exhibits no cognitive dissonance – that is, no desire to hold beliefs other than those he holds in "equilibrium." More broadly, the paper builds on a tradition that emphasizes the effect of beliefs on well-being (e.g. Akerlof and Dickens (1982)). While this literature has focused on self-regard, its tenets must be at least as important for understanding other-regard.

The rest of the paper is organized as follows. Section 2 presents the framework and characterizes optimal interpretations. Section 3 characterizes learning, beginning with a simple example and concluding with general results. Section 4 extends this analysis to alternative motives for giving, and Section 5 discusses open questions for further research.

2 The Good Intentions Framework

2.1 Timing

There are two players, a benefactor and a beneficiary. Nature initially determines the value of a finite-valued parameter $\theta \in \Theta$ after which the timing of play is as follows:

[5]See also Brigham et al. (2013) who find that micro-finance institutions were unlikely to respond to emails mentioning research that microfinance was ineffective, but significantly more likely to respond to emails that mentioned positive results.

1. A signal $s_1 \in S_1$ is revealed and the benefactor forms subjective ex ante beliefs $\hat{\pi}(\theta, s_2|s_1)$
2. The benefactor chooses a decision $d \in D$
3. A signal $s_2 \in S_2$ is revealed and the benefactor forms subjective ex post beliefs $\hat{\pi}(\theta|d, s_2, s_1)$
4. Payoffs are realized

Let $\pi(\theta, s_2, s_1)$ describe the joint distribution of the observable data (s_1, s_2) and the unobservable parameter θ. No assumption is made that the benefactor knows this distribution, and its relationship to his beliefs is discussed below. The distribution π is fixed for now but will later be endogenized to characterize incentives for learning and communication.

2.2 Payoffs

The beneficiary's payoff depends on the decision d and state θ according to

$$v(d, \theta) \tag{1}$$

In the standard approach to modeling "pure" altruism, the benefactor's payoff would be

$$u(d) + v(d, \theta) \tag{2a}$$

The first term represents the benefactor's private concerns. For example, if $d \in [0, y]$ is a donation to a charitable cause then $u(d) = U(y - d)$ might be the benefactor's consumption utility. The second term represents the utility the benefactor obtains from the beneficiary's outcome. Note that this specification implies that the benefactor is *aware* of the ex-post realization of v. To allow for ex-post ambiguity, the benefactor's payoff must depend on his *perception* of v:

$$u(d) + \mathbb{E}_{\hat{\pi}(\theta|d,s_2,s_1)}[v(d, \theta)] \tag{2b}$$

This perception is captured by $\hat{\pi} \in \Delta(\Theta)$, the benefactor's ex-post subjective belief about the state of the world. The fact that $\hat{\pi}$ may be non-degenerate embodies the idea that uncertainty about θ may not completely resolve by the end of the game.

The altruism described by (2b) is still *pure* in the sense that, conditional on the level of u, the benefactor uses the same function v to assess the beneficiary's well-being as the beneficiary himself. The model thus abstracts from some of the wedges that earlier work has explored. A benefactor might have paternalistic preferences, for example, and care more about keeping the beneficiary from starving than about her other needs (e.g. Garfinkel (1973)). A benefactor might also help in part to signal his type (e.g. Glazer and Konrad (1996), Ali and Benabou (2013)). For simplicity I study pure altruism through Section 3 and then show in Section 4 how the framework can be extended to alternative motives.

2.3 Optimization

Given beliefs, the benefactor's decision-making process is standard: he chooses a decision d to maximize his subjective expected utility. Adopting the shorthand $\hat{\pi}$ for the complete

contingent belief profile $(\hat{\pi}(\theta, s_2|s_1), \hat{\pi}(\theta|d, s_2, s_1))$, we have

$$d^*(\hat{\pi}, s_1) = \arg\max \mathbb{E}_{\hat{\pi}(\theta, s_2|s_1)}[u(d) + v(d, \theta)] \tag{3}$$

The focus of the analysis will be on the evolution of beliefs and their effects on behavior through (3). I restrict the beliefs the benefactor may hold as follows:

Assumption 1 (Admissible beliefs)**.** *Subjective beliefs $\hat{\pi}(\theta, s_2|s_1)$ satisfy*

(a) $\hat{\pi}(\theta, s_2|s_1)$ is a probability measure on $\Theta \times S_2$ for any s_1

(b) $\hat{\pi}(\theta, s_2|s_1) = 0$ if $\pi(\theta, s_2|s_1) = 0$ for any (θ, s_2, s_1)

Subjective beliefs $\hat{\pi}(\theta|d, s_2, s_1)$ satisfy analogous conditions.

Part (a) of this assumption simply says that beliefs are well-defined. Part (b) is substantive and imposes a degree of logical consistency: the benefactor understands that some compound events are impossible and does not hold beliefs that are clearly incompatible with the facts. Beyond this, however, the relationship between probabilistic events may be ambiguous. For example, if the set $\{\theta : \pi(s_2, s_1|\theta) > 0\}$ has more than one element for some given (s_2, s_1) then it is unclear how the benefactor should weight their relative likelihood. Moreover, this problem does not go away with repetition of the game: because the benefactor does not observe θ ex post, he cannot learn about $\pi(\theta|s_2, s_1)$ regardless of how many i.i.d. draws of (s_2, s_1) he observes. I resolve this indeterminacy by studying beliefs that are optimal in the sense that they maximize expected utility.

$$\max_{\hat{\pi}} \mathbb{E}_{\pi}\Big[u(d^*(\hat{\pi}, s_1)) + \mathbb{E}_{\hat{\pi}(\theta|d^*, s_2, s_1)}[v(d^*(\hat{\pi}, s_1), \theta)]\Big] \text{ such that } \hat{\pi} \text{ is admissible} \tag{4}$$

Note the distinct roles played here by ex ante and ex post beliefs: while the former determine the mapping from signals s_1 into actions, the latter determine how the benefactor interprets the consequences of those actions.

2.4 Interpretation & Discussion

The "good intentions" framework departs from standard modeling techniques in two ways. First, the benefactor holds preferences over beliefs as well as over outcomes. This idea builds on a literature dating at least as far back as Akerlof and Dickens (1982), who model a employee who prefers to believe that his risk of workplace injury is low. More recently Caplin and Leahy (2001) study the effects on decision-making of anxiety about future payoffs, while Brunnermeier and Parker (2005) study the general problem of optimal beliefs when expectations about the future affect current happiness. As these examples illustrate the literature has focused on self-regarding beliefs; the argument here is that thoughts or beliefs are at least as important for understanding other-regard. When giving to Africa, for example, it is hard to see how anything *other* than beliefs could matter.

Second, the model explicitly endogenizes beliefs through optimization, in the spirit of Akerlof and Dickens (1982) and Brunnermeier and Parker (2005). A natural question is whether this leads to beliefs that are coherent either internally or with what the benefactor observes.

To examine this, note first that the benefactor's ex post belief $\hat{\pi}(\theta|d, s_2, s_1)$ affects his payoffs only through $\mathbb{E}_{\hat{\pi}(\theta|d,s_2,s_1)}[v(d,\theta)]$. He will therefore choose to be as optimistic as possible ex post about the beneficiary's situation. Formally, optimal beliefs put full weight on the state

$$\overline{\theta}(d, s_2, s_1) = \arg \max_{\theta \in \Theta : \pi(\theta|s_2,s_1)>0} [v(d,\theta)] \tag{5}$$

which is the best state of the world consistent with the information history. Given this, the benefactor's ex ante problem reduces to

$$\max_{\hat{\pi}} \mathbb{E}_{\pi} \Big[u(d^*) + v(d^*, \overline{\theta})] \Big] \tag{6}$$

where I have suppressed arguments for brevity. This says that the benefactor holds ex ante beliefs that induce optimal behavior, given that he will ultimately take the optimistic interpretation $\overline{\theta}$. One can then show that optimal beliefs are, without loss of generality, Bayesian.

Lemma 1 (Baysian Updating)**.** *There exist optimal subjective beliefs satisfying Bayes' rule, i.e.*

$$\hat{\pi}(\theta, s_2|s_1)\hat{\pi}(s_1) = \hat{\pi}(\theta, s_2, s_1)$$
$$\hat{\pi}(\theta|d, s_2, s_1)\hat{\pi}(s_2, s_1) = \hat{\pi}(\theta, s_2, s_1)$$

for all (θ, s_2, s_1)*.*

The proof (see Appendix A) is constructive and shows that beliefs derived as conditional probabilities from the prior

$$\hat{\pi}(\theta, s_2, s_1) = 1(\theta = \overline{\theta}(d^*(s_1), s_2, s_1))\pi(s_2, s_1) \tag{7}$$

are optimal. The interpretation of this specification is that the benefactor holds an unbiased view $\pi(s_2, s_1)$ of the *likelihood* of the various kinds of feedback he might receive, but chooses to *interpret* this feedback as proving that an appealing state of the world $\overline{\theta}$ has been realized. This has four noteworthy implications.

First, optimal beliefs have the usual mathematical properties of beliefs: for example, they behave as martingales. This implies that an empirical researcher cannot identify beliefs as "well intentioned" without ancillary data such as the empirical distribution π.

Second, optimal beliefs are consistent with observable data. Formally, the marginal distribution over (s_2, s_1) implied by (7) is the empirical distribution $\pi(s_2, s_1)$. This implies that the beliefs of a benefactor with unbounded time to learn about the model environment through repeated experience could converge to optimal beliefs. It is a corollary that optimal beliefs differ from the objective distribution only in describing data that are unobservable, i.e. the conditional distribution of θ given (s_2, s_1). Optimization is in this sense a mild assumption here relative to the literature, which has argued that people maintain optimistic interpretations even when these directly conflict with observable data. Brunnermeier and Parker (2005) argue, for example, that "psychological theories provide many channels through which the human mind is able to hold beliefs inconsistent with the rational processing of objective data"

(p. 1093). Mobius et al. (2012) show empirically that subjects interpret data about their ability with self-serving biases even when the data generating process is specified unambiguously and beliefs are elicited incentive-compatibly. In contrast, our focus here is on ambiguous questions such as the likelihood that a nonprofit executive is corrupt conditional on the absence of scandal, which provide even greater scope for the imagination.

Third, optimal beliefs are self-consistent: a benefactor holding them would not wish to alter them. To see this note that if the agent believes the true distribution is some $\hat{\pi}$ satisfying (7), and then uses (7) to re-calculate optimal beliefs, he arrives again at $\hat{\pi}$. (Note also that this need not hold for the empirical distribution π.) This property is one point of distinction between the model and others such as Brunnermeier and Parker (2005) in which agents hold self-inconsistent beliefs, reflecting the tension between between utility from actions and utility from beliefs. Here there is no such tension.

Fourth, the model nests the benchmark case of preferences over outcomes. To see this, consider evidence (s_2, s_1) that is consistent with only a single state $\theta : \pi(\theta|s_2, s_1) > 0$. For such evidence the only admissable interpretation is $\hat{\pi}(\theta|s_2, s_1) = \pi(\theta|s_2, s_1)$. Next, call feedback *fully revealing* if it always uniquely identifies the state, i.e. $\{\theta \in \Theta : \pi(\theta|s_2, s_1) > 0\}$ is single-valued for any (s_2, s_1) such that $\pi(s_2, s_1) > 0$. Then the following holds:

Lemma 2 (Role of Feedback)**.** *Beliefs derived via Bayesian updating from the prior $\pi(\theta, s_2, s_1)$ are optimal if feedback is fully revealing.*

In other words, the good intentions framework and the standard one coincide precisely when the benefactor expects no ex-post ambiguity about θ.[6] Intuitively, such cases are like decisions the benefactor makes which affect only himself. In these cases he directly experiences the consequences, which we can think of as a way in which he "learns" the realization of θ.

3 Effective Giving

How much will the benefactor learn in equilibrium when choosing how to help? I first illustrate the main ideas in an example and then provide generalizations in Section 3.5. For concreteness the narrative describes charitable giving.

3.1 An Example

Don, a marketing executive in Manhattan, considers giving to an NGO working to help Ben, a farmer in Africa. Don can donate any amount d up to total income y. Ben's welfare depends both on this donation and on other exogenous factors such as the level of rainfall or the effectiveness of the NGO. For simplicity, the situation is either Good ($\theta = \theta^g$) or Bad ($\theta = \theta^b$), where Ben's utility v satisfies $v(\theta^g, d) > v(\theta^b, d)$ for all d. Don's prior is that $\pi(\theta = \theta^g) \equiv \gamma \in (0, 1)$. Don genuinely wants to see Ben better-off, but since Ben is thousands of miles away this desire is reflected in preferences over *thoughts* about what is happening in

[6] The antecedent can be made both necessary and sufficient by adding appropriate sensitivity conditions.

Africa. Formally, Don maximizes

$$y - d + \hat{\gamma}_2 v(\theta^g, d) + (1 - \hat{\gamma}_2) v(\theta^b, d) \tag{8}$$

where $\hat{\gamma}_2$ is his subjective ex-post assessment of the likelihood that the situation is good. In each period he either observes θ or learns nothing. For example, interpreting θ as a measure of NGO effectiveness, he might or might not learn about an impact evaluation of its work. Interpreting θ as growing conditions, he might or might not read news about the state of African agriculture. Let p be the probability that he learns the truth before donating, and q the conditional probability that he learns it after donating if he had not learned it before.

If Don learns θ before donating then this pins down beliefs and he chooses

$$d^*(\theta) \equiv \arg\max_d y - d + v(\theta, d) \tag{9}$$

In the more interesting case where he does not learn before donating, he anticipates the views he will hold in the future. With probability q he will learn the true state, while with probability $1 - q$ he will obtain ambiguous information which he will interpret as meaning that all is well ($\theta = \theta^g$). His future perception is thus $\hat{\gamma}_2 = 0$ with probability $q(1 - \gamma)$ and $\hat{\gamma}_2 = 1$ with probability $1 - q(1 - \gamma)$. Given this, he optimally interprets the absence of news at time $t = 1$ to mean that matters in Africa are good with probability $\hat{\gamma}_1 = 1 - q(1 - \gamma)$[7] and gives

$$d^*(\emptyset) \equiv \arg\max_d y - d + \hat{\gamma}_1 v(\theta^g, d) + (1 - \hat{\gamma}_1) v(\theta^b, d) \tag{11}$$

3.2 Learning to Help

Don's tendency to take an optimistic view of things shapes his motives for learning. Consider first the effect on his payoffs of learning the truth ex post. If he already knew it then of course it has no effect. If it is news to him, however, then it cannot be welcome news. The reason is that when uninformed Don optimally reasons that "no news is good news" and believes all is well ($\theta = \theta^g$), while becoming informed may force him to confront the reality that things are in fact not well ($\theta = \theta^b$).

Observation 1. *Don's expected payoff strictly decreases in the probability that he becomes informed after donating.*

This observation highlights the idea that information is a *constraint*, ruling out hypotheses that formerly were plausible. Yet somewhat paradoxically, the constraining nature of ex post information can also make ex ante information endogenously valuable. To see this, suppose that Don knew he would definitely learn the truth ex post, and consider his demand for

[7] To see this note that this belief uniquely ensures

$$\arg\max_d y - d + \mathbb{E}_{\hat{\gamma}_1}[v(\theta, d)] = \arg\max_d y - d + \mathbb{E}_{(1-q(1-\gamma))}[v(\theta, d)] \tag{10}$$

Note that $\hat{\gamma}_1 = \mathbb{E}_\pi[\hat{\gamma}_2]$ so that the evolution of Don's beliefs satisfies the law of iterated expectations and with it Bayes' rule.

information ex ante. In this case his expected payoff is

$$(\gamma)\left[\max_d y - d + v(\theta^g, d)\right] + (1-\gamma)\left[\max_d y - d + v(\theta^b, d)\right] \tag{12}$$

when informed and

$$\max_d y - d + (\gamma)v(\theta^g, d) + (1-\gamma)v(\theta^b, d) \tag{13}$$

when uninformed. It follows directly from optimization and continuity that the former is strictly greater than the latter, so that Don values information. But now suppose that Don expects not to learn the truth ex post. In this case his payoff when informed ex ante is again given by (12), but his payoff when uninformed ex ante is

$$\max_d y - d + v(\theta^g, d) \tag{14}$$

He thus obtains a benefit from being *uninformed* proportional to

$$\max_d \left[y - d + v(\theta^g, d)\right] - \max_d \left[y - d + v(\theta^b, d)\right] \geq \max_d \left(v(\theta^g, d) - v(\theta^b, d)\right) > 0 \tag{15}$$

The intuition here, just as for ex post learning, is that information constrains the imagination. Absent any threat of real consequences, Don prefers maximum scope to "think positive."

Observation 2. *Don's payoff increases (decreases) in the probability he learns the truth before donating when he will (will not) learn the truth after donating.*

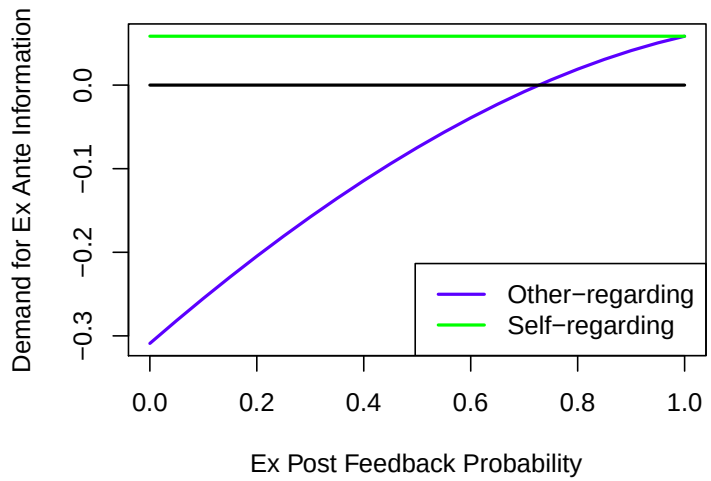
This observation summarizes a novel way of thinking about learning. The primal role of information is as a constraint on the imagination: it limits what thoughts one can reasonably entertain about the world. This makes it undesirable. On the other hand, given that such constraints are to be encountered, there is some value in knowing now what tomorrow's thoughts may be and acting so as to avoid disappointment. This generates positive demand.

Figure 1 illustrates the tension between the costs and benefits of learning with a parameterized example. When the probability that Don will learn the truth ex post is low he is strongly averse to learning the truth ex ante, as in all likelihood this will simply constraint his beliefs. As the probability of ex post learning rises his demand for ex ante research rises correspondingly until, past some threshold, it becomes positive. At all interior points his demand is strictly lower, however, than would be the case if he were making the decision for himself rather than for Ben.

Note that this result implies ex post feedback can stimulate demand for ex ante research. This is consistent with economicsts' arguments that *measuring* outcomes is necessary in order to force those spending money to pay attention to them. For example, Muralidharan (2012) writes of education policy in India that

> The Indian state has done a commendable job in improving the education indicators that were measured (including school access, infrastructure, enrollment, and inclusiveness in enrollment) but has fallen considerably short on the outcome indicators that have not been measured (such as learning outcomes). While independently measuring and administratively focusing on learning outcomes will not by itself

Figure 1: Demand for Ex Ante Information on Effectiveness



Notes: plots Don's willingness to pay for information for the case where $v(d, \theta) = \theta \log(d)$, $\theta^g = 2$, $\theta^g = 1$, and $\gamma = 0.2$, as a function of the probability he will learn the truth ex post.

> lead to improvement, it will serve to focus the energies of the education system on the outcome that actually matters..."

3.3 Intermediaries

Don's ambivalent attitude towards learning in turn shapes the incentives of other players in the market. In this section I focus on a revenue-maximizing intermediary seeking to obtain donations from Don – for example, a charity. What marketing strategies maximize these donations? I focus here on the expected returns to generating various kinds of information which will then be disclosed to the public. This might correspond, for example, to commissioning an academic study by J-PAL.

Consider first the impact of better ex-post outcome measurement:

Observation 3. *Ex post feedback increases (decreases) expected generosity if v is submodular (supermodular).*

The probability of ex post feedback affects Don's decision only in the case where he is uninformed ex ante, so that his donation is given by (11). The comparative static is

$$\frac{\partial d}{\partial q} = \frac{(1-\gamma)[v_d(\theta^g, d) - v_d(\theta^b, d)]}{(1 - q(1-\gamma))v_{dd}(\theta^g, d) + q(1-\gamma)v_{dd}(\theta^b, d)} \tag{16}$$

which shares the sign of $v_d(\theta^b, d) - v_d(\theta^g, d)$. Next consider ex-ante research:

Observation 4. *Suppose that ex ante information does not affect expected generosity when ex post feedback is perfect. Then ex ante information strictly increases (decreases) expected generosity if v is submodular (supermodular) and feedback is limited.*

To see this, first consider the case of perfect feedback. Define $d^*(\gamma)$ as

$$d^*(\gamma) \equiv \arg\max_d y - d + \gamma v(\theta^g, d) + (1-\gamma)v(\theta^b, d) \tag{17}$$

If feedback is perfect ($q = 1$) then Don gives $d^*(\gamma)$ when uninformed, $d^*(1)$ if he obtains good news ex ante, and $d^*(0)$ if he learns bad news ex ante. Ex ante information thus has no average effect if $d^*(\gamma) = \gamma d^*(1) + (1-\gamma)d^*(0)$. Suppose this holds. Now consider the case with imperfect ex post feedback. If informed ex ante Don's expected donation is again $\gamma d^*(1) + (1-\gamma)d^*(0)$. If uninformed his donation solves (11). The solution to this equation is decreasing (increasing) in q if v is supermodular (submodular), and hence Don gives less (more) than $d^*(\gamma)$ when uninformed.

The mechanism underlying both these results is that Don prefers to believe that things are going well, so that information generally forces him to revise his beliefs negatively. How this affects his donation d then depends on whether giving is more or less impactful when the situation θ is bad. If θ complements donations – for example, if it measures effectiveness – then forcing Don to confront reality will lower his perception of marginal returns and depress giving. The intermediary has no incentive to do this. If, on the other hand, θ substitutes for donations – for example, if it measures Ben's baseline income – then forcing Don to confront reality will raise his perception of marginal returns and increase giving. Put another way, Don wishes to believe Ben is doing well, but the charity needs him to realize that Ben is desperately needy.

These results may help explain nonprofit marketing practice. Critics often lament how little rigorous information nonprofits provide about what they do and how impactful it is. Yet the model predicts that ambiguity on these dimensions is actually helpful, since it leaves space for donors to imagine the best. On the other hand, nonprofits often present information about need or use "awareness-raising" campaigns; these will be especially effective when altruists have a generic bias towards believing that others are doing better than they really are.

More broadly, the negative result for effectiveness research highlights a generic tradeoff in the model between the *quantity* and *quality* of altruistic activity. This is easiest to see from the perspective of a social planner seeking to maximize beneficiary well-being and choosing whether or not to sponsor research on effectiveness. While the research has the potential to increase the effectiveness of a *given* dollar of funding, it will also tend (according the result above) to reduce the total number of dollars given. It is thus unclear whether the beneficiary benefits. This has obvious implications for policy-makers allocating funds to development research. It also explains why the beneficiary may choose not to disillusion a well-intentioned donor even when given the chance (see Appendix B for a formal result).

3.4 Salience and Charitable Giving

By shifting emphasis from outcomes to thoughts, the good intentions model also provides a helpful framework for organizing some features of charitable marketing and giving related to salience that are hard to accomodate in standard models. To illustrate this, consider extending the model trivially by introducing a parameter $\rho \in (0, 1)$ which measures the probability that Don thinks about Ben ex post. Then his expected payoff is

$$y - d + \rho \left[\hat{\gamma}_2 v(\theta^g, d) + (1 - \hat{\gamma}_2) v(\theta^b, d)\right] \tag{18}$$

This has several direct implications.

1. Donors give more to causes that are more memorable for them (higher ρ). This may help explain why people are more likely to give to issues that have affected friends and loved ones (Small and Simonsohn, 2008). For example, a donor who has lost a loved one to cancer is more likely to remember a gift supporting anti-cancer research through the associate property of memory (e.g. Tulving and Schacter (1990)).
2. As a corollary, charities can increase donations by making them more memorable. The most direct such strategy is of course to frequently remind the donor of his gift, and indeed "thank-you" notes are generally considered a good marketing practice.[8] Less obviously, charities can enhance recall of a gift by associating it with something specific and memorable. Linking a donation to an "identifiable victim" is one such strategy and has been show to increase giving (Jenni and Loewenstein, 1997). The use of "gift catalogues" may play a similar role; these allow donors to visualize their donation as leading to the provision of some specific, tangible thing (e.g. a goat) which they themselves "chose."[9]

3.5 General Functional Forms

This section generalizes the observations made using specific functional forms above. Doing so requires language to compare the information content of signals: a sense in which two signals are the same, and the standard Blackwell sense in which one is more informative than the other.

Definition 1 (Information equivalence)**.** *Random variables X and Y are informationally equivalent if there exists a bijection f such that $Y = f(X)$.*

Definition 2 (Blackwell garbling)**.** *Let $h(x, y, z)$ give the joint distribution of the random variables (X, Y, Z). X is a Blackwell garbling of Y with respect to Z if $h(x|y, z)$ is independent of z.*

[8] See for example `https://www.blackbaud.com/files/resources/downloads/WhitePaper_RecurringGiving.pdf`. Note that in the model Don's taste for reminders is ambiguous because v has no absolute unit: intuitively, thinking about Ben may make Don either happy or sad. Modifying Don's preferences along the lines suggested by Duncan (2004), so that Don cares about the *difference* his contribution made, resolves this ambiguity in favor of reminders.

[9] Gift catalogues are harder to rationalize as mechanisms for control, for two reasons. First, altruistic donors should not want control as they are unlikely to have good information about which interventions are most needed. Second and more importantly, donors' "choices" are typically not legally binding, as the accompanying fine print makes clear that the nonprofit will do whatever it wants with the donation. See for example `http://philanthropy.com/article/Holiday-Gift-Catalogs-Are/64374/`.

The shorthand $X \succsim Y$ indicates that the benefactor's expected payoff is weakly greater when he observers the random variable X than when he observers Y. We can now generalize Observation 2 and show that the benefactor prefers as little ex post feedback as possible.

Proposition 1. *Let random variable S_2' be a garbling of S_2 with respect to (S_1, θ). Then $S_2' \succsim S_2$.*

As above, the intuition is that feedback constrains the benefactor without helping him make decisions.

Proposition 2.

- *Let S_1 be informationally equivalent to S_2. Then $S_1 \succsim S_1'$ for any S_1'.*
- *Let S_1 be a garbling of S_2 with respect to θ and let S_1' be a garbling of S_1 with respect to S_2. Then $S_1 \succsim S_1'$.*

This generalizes Observation 2. The first part states that the benefactor's weakly prefers to observe ex ante what he will eventually observe ex post. In particular, he has no demand for information prior to making his decision that he will not subsequently learn after that decision. The second part states that, among signals that are strictly less informative than what he will observe ex post, the benefactor weakly prefers more informative ones. It is a corollary that he places a (weakly) positive value on such signals, since a white-noise signal is trivially a member of this set.

Generalizing Observation 4 requires a bit more work, as we need a generalization of the idea that ex ante information does not affect expected generosity under standard preferences (or equivalently, when ex post feedback is perfect).

Definition 3. *Suppose d is real-valued. The benefactor's preferences respect expectation if*

$$\arg\max_d \mathbb{E}_\mu[u(d) + v(d, \theta)] = \mathbb{E}_\mu[\arg\max_d u(d) + v(d, \theta)] \tag{19}$$

holds for any $\mu \in \Delta(\theta)$.

This condition says that, while particular realizations of θ may influence generosity one way or another, disclosure of θ neither increases nor increases generosity *in expectation*. We can now state and prove a general result on complementary and substitutability:

Proposition 3. *Suppose that Θ is ordered, D is real-valued, and $v(\theta, d)$ is monotone increasing in both arguments.*

- *Let S_2' be a garbling of S_2 with respect to θ. Then $\mathbb{E}_\pi[d]$ is higher (lower) under S_2' than under S_2 if v is supermodular (submodular).*
- *Let S_1' be a garbling of S_1 with respect to (S_2, θ) and suppose that the benefactors preferences respect expectation. Then $\mathbb{E}_\pi[d]$ is higher (lower) under S_1' than under S_1 if v is supermodular (submodular).*

Like Observation 4, this result implies that generosity tends to increase when information about needs is disclosed, but tends to decrease when information about effectiveness is disclosed.

4 Alternative Motives for Giving

The results above describe an altruist who is "pure" in the sense that she cares about (her perceptions of) the beneficiary's welfare. One might reasonably guess, then, that they are an upper bound on the effectiveness of an altruist with less aligned motivations. This section examines that guess. I consider a family of *reference-dependent* preferences of the form

$$u(d) + \mathbb{E}_{\hat{\pi}(\theta|d,s_2,s_1)}[v(d,\theta) - v(\overline{d},\theta)] \tag{2b'}$$

This specification extends that in (2b) by allowing the benefactor to care about her perceptions, not of the beneficiary's payoff $v(d,\theta)$ per se, but of the difference between that payoff and some reference payoff $v(\overline{d},\theta)$. The reference payoff is itself determined by a reference decision $\overline{d} \in D$ the benefactor could have made. In other words, the beneficiary thinks about the difference between what she *did* and something that she *could have done*.

This family nests at least two cases of interest from the literature. The first is the model of "impact philanthropy" proposed by Duncan (2004). In Duncan's model, a charitable giver cares about the difference between the outcome obtained when he gives and the counterfactual outcome that would have obtained had he given nothing. These preferences can be represented using as a reference point the decision $\overline{d} = \arg\max_d u(d) \in D$, or "doing nothing." In this case the benefactor effectively asks himself, how much better off is the beneficiary than if I had simply pursued my own private interests? His payoff is proportional to the (perceived) size of this gap.

The second case is "guilty giving." Andreoni et al. (2012), among others, have argued that other-regarding behavior is often motivated by a desire to close the gap between what one is doing and what one feels one could or should do. One simple way of capturing this idea is to let $\overline{d}$ measure what could or should be done. Loosely speaking, the benefactor then experiences pride when she does "more" than $\overline{d}$ but guilt when she does "less." We can isolate the latter motive by letting $\overline{d}$ represent a maximally generous action. If D is real-valued, for example, let $\overline{d} = \max D$. Note that this case describes an opposite extreme to the impact philanthropy model; together the two cases thus bookend the set of possible reference points.

Despite their variety, it turns out that the models described by (2b') share most properties of the base "pure altruism" case.

Proposition 4. *Lemmas 1 and 2 and Propositions 1 and 2 continue to hold replacing (2b) with (2b').*

The proof is by simple redefinition: let

$$\tilde{v}(d,\theta) \equiv v(d,\theta) - v(\overline{d},\theta) \tag{20}$$

and the proofs go through as before replacing v with $\tilde{v}$, since nothing in them relies on anything special about the structure of v. Conceptually the point here is that the disinterest in learning captured by these propositions, which may help explain ineffective altruism, is a product of perceptions management per se. It does not matter (qualitatively) whether the perceptions at stake are perceptions of outcomes, of impact, or of something else; the beneficiary will seek to

learn things she must eventually learn anyway, but otherwise will prefer to avoid information.

What about the market? Here it turns out – as professional fundraisers will attest – that knowing your donor's motives is essential for effective persuasion. Information has very different average effects on impact philanthropists and guilty givers, both of whom are different in turn from the "pure" altruist described above. The most ironic result is that the optimal strategy for a fundraising intermediary marketing itself to an impact philanthropist is to provide *no information at all.* The intuition is quite simple. An impact philanthropist wants to believe that the marginal impact of his dollar is as high as possible. He therefore interprets any ambiguous information as implying that the need is great and the available means of helping effective. Any information the intermediary provides will tend to lower his expectation of marginal impact, which in turn leads him to lower his donation.

For guilty givers, on the other hand, the reverse is true. To minimize guilt, these donors want to believe ex-post that there is nothing they could possibly have done that would have made any difference. Such a donor might convince himself, for example, that the need is not very great, or that all foreign aid is corrupt so that none of his donation would actually reach people in need. Such beliefs would enable him to give very little without experiencing guilt over missed opportunities. A fundraiser pitching such a donor could thus benefit by providing incontrovertible evidence of both need and efficacy. The donor, of course, would do his best to avoid this pitch.

The following Proposition formalizes these points. Note that this result is consistent with Proposition 3; the latter cannot be applied directly here since $v(d,\theta) - v(\overline{d},\theta)$ need not be increasing in θ even if v itself is.

Proposition 5. *Suppose that the benefactor's preferences are as in (2b'), Θ is ordered, D is real-valued, $v(d,\theta)$ is increasing in both arguments, and $v_d(d,\theta)$ is monotonic in θ.*

- *Let S_2' be a garbling of S_2 with respect to θ. Then $\mathbb{E}_\pi[d]$ is higher (lower) under S_2' than under S_2 if $\overline{d} = \min D$ ($\overline{d} = \max D$).*
- *Let S_1' be a garbling of S_1 with respect to (S_2,θ) and suppose that the benefactors preferences respect expectation. Then $\mathbb{E}_\pi[d]$ is higher (lower) under S_1' than under S_1 if $\overline{d} = \min D$ ($\overline{d} = \max D$).*

The impact philanthropy model case may also help explain the success of "matching grant" vehicles in fundraising. In a typical matching setup, an organization obtains a promise from a large funder to match subsequent smaller donations. The puzzle for economists is why such arrangements are credible: if the small donations do not materialize, will the large funder – who was clearly excited about funding the organization – really refrain from giving? This is exactly the sort of question an economist would ask – but exactly the sort of question a well-intentioned donor would *not* ask. An impact donor wants very much to believe that the large funder's commitment is credible, since this increases his marginal impact. He can do so, moreover, as long as there is ambiguity about counterfactual states. After donating himself, the donor simply needs to believe that the large funder would not have contributed if he had not. Fortunately for him, there is unlikely to be unambiguous evidence to the contrary.

5 Conclusion

Standard models of other-regarding behavior model benefactors with preferences over a beneficiary's outcomes. This approach is unrealistic as it posits that the decision-maker has preferences over events he never experiences. I study an alternative framework in which the benefactor has preferences over his beliefs about the beneficiary's outcomes. This framework nests the standard model in the special case where the benefactor obtains complete ex post information about the beneficiary's outcomes; absent perfect feedback the models' predictions diverge. Consistent with the motivation for the framework, the benefactor in the model endogenously prefers to avoid ex post feedback and also avoids ex ante information about the beneficiary except to avoid subsequent disappointment. The results may help explain a range of puzzles about effective giving ranging from poorly chosen holiday gifts to misspent charitable donations and foreign aid.

While static, the framework developed here is dynamically consistent in the sense that the benefactor holds beliefs that match the true distribution of observable variables. Formally modelling a dynamic extension could potentially shed further light on the evolution of altruism. Two specific conjectures seem worth examining. First, benefactor behavior will be self-perpetuating. A benefactor who takes an arbitrary action at time t will be motivated to believe this action was effective at time $t + 1$, which will in turn motivate him to repeat the action. This may explain why nonprofits place such priority on the initial acquisition of donors. Second, benefactors may tend to become "jaded" over time as the accumulation of evidence increasingly constrains the extent to which they can "think positive."

References

Akerlof, George A and William T Dickens, "The Economic Consequences of Cognitive Dissonance," *American Economic Review*, June 1982, *72* (3), 307–19.

Ali, Nageeb and Roland Benabou, "Image versus Information," Technical Report, UC San Diego 2013.

Andreoni, James, "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, December 1989, *97* (6), 1447–58.

_ , Justin Rao, and Hannah Trachtman, "Avoiding The Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving," Technical Report, UC San Diego June 2012.

Brigham, Matthew, Michael Findley, William Matthias, Chase Petrey, and Daniel Nelson, "Aversion to Learning in Development? A Global Field Experiment on Microfinance Institutions," Technical Report, Brigham Young University March 2013.

Brunnermeier, Markus K. and Jonathan A. Parker, "Optimal Expectations," *American Economic Review*, September 2005, *95* (4), 1092–1118.

Caplin, Andrew and John Leahy, "Psychological Expected Utility Theory And Anticipatory Feelings," *The Quarterly Journal of Economics*, February 2001, *116* (1), 55–79.

Che, Yeon-Koo, Wouter Dessein, and Navin Kartik, "Pandering to Persuade," *American Economic Review*, February 2013, *103* (1), 47–79.

Dana, Jason, Roberto Weber, and Jason Kuang, "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness," *Economic Theory*, October 2007, *33* (1), 67–80.

Duflo, Esther and Michael Kremer, "Use of randomization in the evaluation of development effectiveness," Technical Report, World Bank 2003.

Duncan, Brian, "A theory of impact philanthropy," *Journal of Public Economics*, August 2004, *88* (9-10), 2159–2180.

Easterly, Bill, *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*, Oxford University Press, 2006.

Fong, Christina and Felix Oberholzer-Gee, "Truth in giving: Experimental evidence on the welfare effects of informed giving to the poor," *Journal of Public Economics*, 2011, *95* (5), 436–444.

Garfinkel, Irwin, "Is In-Kind Redistribution Efficient?," *The Quarterly Journal of Economics*, May 1973, *87* (2), 320–30.

Glazer, Amihai and Kai A Konrad, "A Signaling Explanation for Charity," *American Economic Review*, September 1996, *86* (4), 1019–28.

Hope Consulting, "Money for Good: The US Market for Impact Investments and Charitable Gifts from Individual Donors and Investors," Technical Report, Hope Consulting May 2012.

Jenni, Karen and George Loewenstein, "Explaining the Identifiable Victim Effect," *Journal of Risk and Uncertainty*, 1997, *14* (3), 235–257.

Kalai, Ehud and Ehud Lehrer, "Rational Learning Leads to Nash Equilibrium," *Econometrica*, September 1993, *61* (5), 1019–45.

Krasteva, Silvana and Huseyin Yildirim, "(Un)Informed Charitable Giving," *Journal of Public Economics*, 2013, *106*, 14–26.

Levine, David, "Learning What Works – and What Doesn't: Building Learning into the Global Aid Industry," Technical Report, UC Berkeley 2006.

Milgrom, Paul and Chris Shannon, "Monotone Comparative Statics," *Econometrica*, January 1994, *62* (1), 157–80.

Mobius, Markus, Muriel Niederle, Paul Niehaus, and Tanya Rosenblat, "Managing Self-Confidence: Theory and Experimental Evidence," Technical Report, UC San Diego August 2012.

Muralidharan, Karthik, "Using Evidence for Better Policy The Case of Primary Education in India," Technical Report, UC San Diego 2012.

Pritchett, Lant, "It pays to be ignorant: A simple political economy of rigorous program evaluation," *Journal of Policy Reform*, 2002, *5* (4), 251–269.

Ravallion, Martin, "Evaluation in the Practice of Development," *World Bank Research Observer*, March 2009, *24* (1), 29–53.

Small, Deborah A. and Uri Simonsohn, "Friends of Victims: Personal Experience and Prosocial Behavior," *Journal of Consumer Research*, October 2008, *35* (3), 532–542.

The Giving Institute, *Giving USA 2013*, Giving USA Foundation, 2013.

Tulving, E. and D. L. Schacter, "Priming and human memory systems," *Science*, January 1990, *247* (4940), 301–306.

Waldfogel, Joel, *Scroogenomics: Why You Shouldn't Buy Presents for the Holidays*, Princeton University Press, 2009.

A Proofs

Proof of Lemma 1

Consider the following family of history-contingent subjective beliefs:

$$\hat{\pi}(\theta, s_2, s_1) = 1(\theta = \overline{\theta}(d^*(s_1), s_2, s_1))\pi(s_2, s_1) \tag{21}$$

$$\hat{\pi}(\theta, s_2|s_1) = 1(\theta = \overline{\theta}(d^*(s_1), s_2, s_1))\pi(s_2|s_1) \tag{22}$$

$$\hat{\pi}(\theta|d, s_2, s_1) = 1(\theta = \overline{\theta}(d, s_2, s_1)) \tag{23}$$

where

$$d^*(s_1) = \arg\max_d \mathbb{E}_{\pi(s_2|s_1)}[u(d) + \mathbb{E}_{\hat{\pi}(\theta|d,s_2,s_1)}[v(d,\theta)]] \tag{24}$$

is the action the benefactor takes given these beliefs. It is straightforward to verify that the beliefs thus defined satisfy Bayes rule following any signal realizations. Intuitively, the benefactor retains objective beliefs about the distribution of signals (s_2, s_1) but distorts their *interpretation*, i.e. what these signals reveal about θ. To show that these beliefs also maximize the benefactor's payoff we need to show that they satisfy two conditions. First, if $\Theta(s_2, s_1)$ denotes the set of admissible beliefs upon observation of (s_2, s_1) then $\hat{\pi}(\theta|d, s_2, s_1)$ must solve

$$\max_{\tilde{\pi} \in \Theta(s_2,s_1)} \mathbb{E}_{\tilde{\pi}}[v(d,\theta)] \tag{25}$$

which it evidently does by definition. Second, $\hat{\pi}(\theta, s_2|s_1)$ is optimal if (though not necessarily only if) it induces the action that is optimal, i.e.

$$\arg\max_d \left[u(d) + \mathbb{E}_{\hat{\pi}(s_2|s_1)}[v(d,\theta)]\right] = \arg\max_d \left[u(d) + \mathbb{E}_{\pi(\theta,s_2|s_1)}\mathbb{E}_{\hat{\pi}(\theta|d,s_2,s_1)}[v(d,\theta)]\right] \tag{26}$$

This condition holds if

$$\hat{\pi}(\theta|s_1) = \mathbb{E}_{\pi(s_2|s_1)}[\hat{\pi}(\theta|d, s_2, s_1)] \tag{27}$$

$$= \mathbb{E}_{\pi(s_2|s_1)}[1(\theta = \overline{\theta}(d, s_2, s_1))] \tag{28}$$

$$= \sum_{s_2} 1(\theta = \overline{\theta}(d, s_2, s_1))\pi(s_2|s_1) \tag{29}$$

which follows from the definition of $\hat{\pi}(\theta, s_2|s_1)$ above.

Proof of Lemma 2

Proof. Suppose (s_2, s_1) is fully revealing; then we can write $\theta = f(s_2, s_1)$ for some function f. This implies that $\overline{\theta}(d, s_2, s_1) = f(s_2, s_1)$ and also that $\pi(\theta, s_2, s_1) = 1(\theta = f(s_2, s_1))\pi(s_2, s_1)$. We can now apply the construction used to prove Lemma 1 to show that beliefs derived via Bayesian updating from $\hat{\pi}(\theta, s_2, s_1) = 1(\theta = f(s_2, s_1))\pi(s_2, s_1) = \pi(\theta, s_2, s_1)$ must be optimal. □

Proof of Proposition 1

Fix a realization s_1. The benefactor's expected payoff if he observes S_2 is

$$u(d^*) + \sum_{s_2} \left[\max_{\theta \in \Theta(s_2, s_1)} \{v(d^*, \theta)\} \right] \pi(s_2|s_1) \tag{30}$$

where d^* is a decision that maximizes this expression. Now suppose instead he observes the realization of S_2'. Since d^* remains a feasible decision his payoff cannot be less than

$$u(d^*) + \sum_{s_2} \sum_{s_2'} \left[\max_{\theta \in \Theta(s_2', s_1)} v(d^*, \theta) \right] \pi(s_2'|s_2, s_1)\pi(s_2|s_1) \tag{31}$$

Now consider some realization (s_2', s_2, s_1, θ) observed with positive probability such that $\pi(s_2, s_1, \theta) > 0$ so that $\theta \in \Theta(s_2, s_1)$. We can write

$$\begin{aligned} \pi(s_2', s_2, s_1, \theta) &= \pi(s_2'|s_2, s_1, \theta)\pi(s_2, s_1, \theta) \\ &= \pi(s_2'|s_2)\pi(s_2, s_1, \theta) \\ &> 0 \end{aligned}$$

where the second step follows from the fact that S_2' garbles S_2 with respect to (S_1, θ) and the third from the fact that s_2' is observed. Thus for any realization we have $\Theta(s_2, s_1) \subseteq \Theta(s_2', s_1)$. This implies that the maximum in (31) is at least as great as that in (30) for any particular (s_2', s_2) and hence (31) is also greater in expectation. Since (31) is a lower bound on the benefactor's payoff when observing S_2, his actual payoff must also be weakly greater.

Proof of Proposition 2

Proof. **Part 1**. Fix the distribution of S_2. First note that because the benefactor chooses d after observing s_1 but then chooses $\overline{\theta}$ after observing both s_2 and s_1, his payoff is bounded above by

$$U(s_2, s_1) \equiv \max_{d, \theta \in \Theta(s_2, s_1)} u(d) + v(d, \theta) \tag{32}$$

which is the payoff he would obtain if he could choose d after observing both signals. Next, observe that when S_1 is equivalent to S_2 then the benefactor achieves this upper bound. Finally, note that when S_1 is not equivalent to S_2 then

$$\begin{aligned} \Theta(s_2, s_1) &= \{\theta \in \Theta : \pi(\theta|s_2, s_1) > 0\} && (33) \\ &\subseteq \{\theta \in \Theta : \pi(\theta|s_2) > 0\} && (34) \\ &= \Theta(s_2) && (35) \end{aligned}$$

and hence the constraint in (32) is weakly tighter than when S_1 is equivalent to S_2, so that $U(s_2, s_1)$ is weakly lower. Since this is an upper bound on the benefactor's payoff it implies that his realized payoff must also be weakly lower than when S_1 is equivalent to S_2.

Part 2. The proof follows the standard argument showing that information weakly im-

proves decision-making, with the caveat that we must also establish that observing a garbling of S_2 does not impose any additional constraints on beliefs.

Fix a realization s_1 of S_1. The benefactor's payoff when he observes this is

$$u(d^*) + \sum_{s_2} v(d^*, \overline{\theta}(d^*, s_2, s_1)\pi(s_2|s_1) \tag{36}$$

where d^* is the decision that maximizes this expression. If instead the benefactor were to observe s_1' then his payoff, again conditional on the (unobserved) value of s_1, is

$$u(d(s_1')) + \sum_{s_2} v(d(s_1'), \overline{\theta}(d(s_1'), s_2, s_1')\pi(s_2|s_1', s_1) \tag{37}$$

where $d(s_1')$ is the optimal decision given s_1'. To simplify this expression note that

$$\begin{aligned}
\pi(s_2|s_1', s_1) &= \frac{\pi(s_1'|s_2, s_1)\pi(s_2|s_1)\pi(s_1)}{\pi(s_1', s_1)} \\
&= \frac{\pi(s_1'|s_1)\pi(s_2|s_1)\pi(s_1)}{\pi(s_1', s_1)} \\
&= \pi(s_2|s_1)
\end{aligned}$$

where the key second step follows since s_1' is a garbling of s_1 with respect to s_2. Note also that

$$\begin{aligned}
\Theta(s_2, s_1) &= \{\theta : \pi(\theta, s_2, s_1) > 0\} \\
&= \{\theta : \pi(s_1|s_2, \theta)\pi(s_2, \theta) > 0\} \\
&= \{\theta : \pi(s_1|s_2)\pi(s_2, \theta) > 0\} \\
&= \{\theta : \pi(s_2, \theta) > 0\}
\end{aligned}$$

where the third step follows since s_1 is a garbling of s_2 with respect to θ and the last since $\pi(s_1|s_2) > 0$ for any observed realization. This implies that $\overline{\theta}(d, s_2, s_1)$ does not depend on s_1. An analogous argument shows that $\overline{\theta}(d, s_2, s_1')$ does not depend on s_1'. Exploiting these two facts we can rewrite (37) as

$$u(d(s_1')) + \sum_{s_2} v(d(s_1'), \overline{\theta}(d(s_1'), s_2, s_1)\pi(s_2|s_1) \tag{38}$$

which must by definition be weakly less than (36) since d^* is defined as the decision that maximizes that expression. □

Proof of Proposition 3

Proof. **Part 1**. Conditional on s_1, we can write the benefactors objective function as

$$f(d, \{x(s_2', s_2, s_1)\}) \equiv u(d) + \sum_{s_2} \sum_{s_2'} v(d, x(s_2', s_2, s_1))\pi(s_2'|s_2)\pi(s_2|s_1) \tag{39}$$

where

$$x(s_2', s_2, s_1) = \max\{\theta : \pi(\theta, s_2, s_1) > 0\} \tag{40}$$

in the case where he observes S_2 and

$$x(s_2', s_2, s_1) = \max\{\theta : \pi(\theta, s_2', s_1) > 0\} \tag{41}$$

in the case where he observes S_2'. (Note that we can write the distribution of S_2' in this separable form because it garbles S_2 and that x does not depend on d since v is monotone in θ.) Examining f, its latter argument is an element of a lattice with dimension support$(S_2) \times$ support(S_2'); moreover since S_2' garbles S_2 we have $\max\{\theta : \pi(\theta, s_2', s_1) > 0\} \geq \max\{\theta : \pi(\theta, s_2, s_1) > 0\}$ for any realization (s_2', s_2), so that S_2' induces a weakly larger element of this lattice than S_2. It then follows from the monotone comparative statics theorem (Milgrom and Shannon, 1994) that the solution is weakly greater (smaller) under S_2' if v is supermodular (submodular).

Part 2. Conditioning on any realization s_1' of S_1', the expected effect of observing S_1 instead can be written as

$$\sum_{s_1} \left[\arg\max_d u(d) + \sum_{s_2} v(d, \overline{\theta}(s_2, s_1))\pi(s_2|s_1) \right] \pi(s_1|s_1') \\ - \arg\max_d u(d) + \sum_{s_2} v(d, \overline{\theta}(s_2, s_1'))\pi(s_2|s_1') \tag{42}$$

Note that this statement exploits the fact that S_1 is finer than S_1' to write $\pi(s_2|s_1, s_1') = \pi(s_2|s_1)$ and $\overline{\theta}(s_2, s_1, s_1') = \overline{\theta}(s_2, s_1)$. By adding and substracting we can decompose this difference further as follows:

$$\sum_{s_1} \left[\arg\max_d u(d) + \sum_{s_2} v(d, \overline{\theta}(s_2, s_1))\pi(s_2|s_1) \right] \pi(s_1|s_1') - \sum_{s_1} \left[\arg\max_d u(d) + \sum_{s_2} v(d, \overline{\theta}(s_2, s_1'))\pi(s_2|s_1) \right] \pi(s_1|s_1') \\ + \sum_{s_1} \left[\arg\max_d u(d) + \sum_{s_2} v(d, \overline{\theta}(s_2, s_1'))\pi(s_2|s_1) \right] \pi(s_1|s_1') - \arg\max_d u(d) + \sum_{s_2} v(d, \overline{\theta}(s_2, s_1'))\pi(s_2|s_1') \tag{43}$$

This decomposition highlights two distinct effects of information. The first is the constraint effect: observing S_1 rather than S_1' places additional restrictions on what the benefactor can reasonably believe ex post. The second is a prediction effect: observing S_1 gives the benefactor a more precise prediction of S_2. The proof proceeds by showing that (a) the constraint effect has the sign predicted by the theorem, and (b) the prediction effect is zero when the benefactor's preferences respect expectation.

(a) It is enough to show the result for any particular realization (s_1, s_1'). Consider therefore

$$\arg\max_d u(d) + \sum_{s_2} v(d, \overline{\theta}(s_2, s_1))\pi(s_2|s_1) - \arg\max_d u(d) + \sum_{s_2} v(d, \overline{\theta}(s_2, s_1'))\pi(s_2|s_1) \tag{44}$$

By the same argument used above to prove part 1 of the proposition this difference is negative (positive) if v is supermodular (submodular). Intuitively, information tends to force the donor to hold a less optimistic view of θ, which increases generosity if and only if d and θ are substitutes.

(b) The prediction effect can be written as

$$\mathbb{E}\left[\arg\max_d u(d) + \mathbb{E}[v(d,\overline{\theta})|S_1]\right] - \arg\max_d u(d) + \mathbb{E}\left[v(d,\overline{\theta})\right] \tag{45}$$

for appropriate priors (which I suppress for brevity). Since preferences respect expectation we know that

$$\mathbb{E}\left[\arg\max_d u(d) + v(d,\overline{\theta})\right] = \arg\max_d u(d) + \mathbb{E}\left[v(d,\overline{\theta})\right] \tag{46}$$

Moreover since this property holds for any prior we can apply it a second time after conditioning on a realization s_1 to show that

$$\mathbb{E}\left[\arg\max_d u(d) + v(d,\overline{\theta})|s_1\right] = \arg\max_d u(d) + \mathbb{E}[v(d,\overline{\theta})|s_1] \tag{47}$$

Taking expectations of both sides over S_1 yields

$$\mathbb{E}\left[\arg\max_d u(d) + v(d,\overline{\theta})\right] = \mathbb{E}\left[\arg\max_d u(d) + \mathbb{E}[v(d,\overline{\theta})|S_1]\right] \tag{48}$$

which together with (46) implies that (45) is zero.

□

Proof of Proposition 5

Proof. **Part 1**. Given d and the realization (s_2, s_1) the benefactor's ex-post problem is

$$\max_{\theta\in\Theta(s_2,s_1)} v(d,\theta) - v(\overline{d},\theta) \tag{49}$$

Since $v_d(d,\theta)$ is monotone in θ, the solution to this problem must also solve $\max_{\theta\in\Theta(s_2,s_1)} v_d(d,\theta)$ for *any* d if $d \geq \overline{d} = \min D$, and $\min_{\theta\in\Theta(s_2,s_1)} v_d(d,\theta)$ for *any* d if $d \leq \overline{d} = \max D$. It follows that further constraining the benefactor's ex-post beliefs by revealing additional information will decrease (increase) the expected value of $v_d(d,\theta)$ for any d, and thus weakly decrease (increase) his expected donation, when $\overline{d} = \min D$ ($\overline{d} = \max D$).

Part 2. The argument proceeds exactly as in the proof of Part 2 of Proposition 3. The effect of coarser information has two effects, a constraint effect and a prediction effect; the prediction effect is zero when preferences respect expectation, while the sign of the constraint effect depends on $\overline{d}$ as in Part 1 above.

□

B Communication

At the heart of the preceding analysis is the idea that other-regarding behavior is qualitatively different from self-regarding behavior because of the lack of directly experienced consequences. Benefactors do not experience the effects they produce for beneficiaries but instead learn about them indirectly. One channel for this indirect learning is of course communication between benefactor and beneficiary. For example, givers and receivers of holiday gifts may talk beforehand about the kinds of things the receiver likes, and often talk afterwards about the suitability or desirability of the gift chosen – the giver hoping to hear the receiver say that it was "just what I wanted."

To better understand good intentions in settings where such direct communication is possible it is necessary to model strategic communication between benefactors and beneficiaries. This section does so in an extended and adapted version of the parable of Don and Ben. Specifically, I enrich Don's choice set so that he decides between alternative methods of helping, and also allow Ben to communicate ex ante with Don.

B.1 An Example, Continued

Don, the Manhattan marketing executive, is again contemplating a donation to help Ben, the African farmer. Don has become aware of two different NGOs both of which work in Ben's village but which provide different services, and must decide how much to donate to each. Let $d = (d^a, d^b)$ represent his giving, where $d^a, d^b \geq 0$ and Don's budget constraint is $d^a + d^b \leq y$. Ben's preferences are represented by

$$v(\theta, d) = \theta^a d^a + \theta^b d^b \tag{50}$$

The interpretation is that θ^i measures the marginal impact of intervention i on Ben's welfare. Don is uncertain about these impacts, knowing only that they are drawn from distribution π with support on $[\underline{\theta}^a, \overline{\theta}^a] \times [\underline{\theta}^b, \overline{\theta}^b]$ where $\underline{\theta}^a > 0$, $\underline{\theta}^b > 0$. Don does want to help in the way he perceives to be most effective; he seeks to maximize

$$u(y - d^a - d^b) + \mathbb{E}_{\hat{\pi}}[\theta^a d^a + \theta^b d^b] \tag{51}$$

Don does not anticipate any feedback on the impact his donations have. Before he gives, however, Ben has an opportunity to send him a costless message m from some arbitrary set M.

Because he does not anticipate any feedback, Don finds it optimal to hold the same beliefs about the effectiveness of each intervention both before and after donating. In particular if he chooses to fund intervention i then he will optimally interpret Ben's message m to mean that

$$\hat{\pi}(\theta^i = x|m) = 1(x = \max\{\theta^i : \mathbb{P}(m|\theta^i) > 0\}) \tag{52}$$

In other words, Don holds the most optimistic view of the intervention he is funding that is

also consistent with Ben's message. Denoting by

$$\overline{\theta}^i(m) = \max\{\theta^i : \mathbb{P}(m|\theta^i) > 0\} \tag{53}$$

the most optimistic view of intervention i given message m, Don thus donates to intervention

$$i^*(m) = \arg\max_{i \in \{a,b\}} \{\overline{\theta}^i(m)\} \tag{54}$$

and gives a total donation $d^*(m)$ characterized by

$$u'(y - d^*(m)) = \overline{\theta}^{i^*(m)}(m) \tag{55}$$

Given this, Ben's problem is to choose a message m solving

$$\max_{m \in M} d^*(m)\theta^{i^*(m)} \tag{56}$$

This expression highlights the fact that Ben's communication decisions must trade off two goals: he wants to steer Don towards the more effective intervention, but also wants to encourage Don to give generously to whichever intervention he chooses.[10] His credibility on these topics, however, is very different. Don knows that Ben has no direct incentive to lie about *which* kind of help he prefers. He does have a direct incentive to mislead Don about the effectiveness of this intervention, since he would always prefer that Don give more, while Don trades off this help against his private benefits of consumption.

Formally, it follows immediately from inspection of (56) that any equilibrium must be action-equivalent to an equilibrium in which Ben chooses at most one message that induces Don to donate to each intervention. The reason is simply that if two messages m, m' both induced intervention a (say) and $d^*(m) < d^*(m')$ then Ben would always prefer to send message m'. Hence we can without loss of generality restrict attention to equilibria in which Ben sends at most two messages with positive probability, m^a inducing a or m^b inducing b. This in turn lets us characterize a unique recipient-optimal equilibrium. To do so define $\overline{\theta}^i = \max\{\theta^i\}$ as the most optimistic view about intervention i given priors π. Then we have

Observation 5. *There exists a unique equilibrium in which Don gives $d^*(\overline{\theta}^a)$ to a if $\theta^a d^*(\overline{\theta}^a) \geq \theta^b d^*(\overline{\theta}^b)$ and gives $d^*(\overline{\theta}^b)$ to b otherwise.*

Proof. By the argument above, in any equilibrium strategy Don either gives $d^*(m^a)$ to a or $d^*(m^b)$ to b. Ben's problem thus amounts to choosing between the payoffs $\theta^a d^*(m^a)$ and $\theta^b d^*(m^b)$. It follows that in any equilibrium Ben sends message m^a if and only if

$$\frac{\theta^a}{\theta^b} \geq \frac{d^*(m^b)}{d^*(m^a)} \tag{57}$$

[10] Provided $\theta^i \geq 0$. Consider this case for now.

Given this, Don's optimal donation level d^a on observing m^a must satisfy

$$u'(y - d^*(m^a)) = \max\left\{\theta^a : \exists\theta^b \text{ such that } \pi(\theta^a, \theta^b) > 0 \text{ and } \frac{\theta^a}{\theta^b} \geq \frac{d^*(m^b)}{d^*(m^a)}\right\} \tag{58}$$

$$= \overline{\theta}^a \tag{59}$$

where the second step follows from the assumption that π has full support on an interval in $\mathbb{R}^2$. Similarly, Don's donation on observing m^b is given by $u'(y - d^*(m^b)) = \overline{\theta}^b$. This uniquely determines $\frac{d^*(m^b)}{d^*(m^a)}$. If this quantity lies within $\left[\frac{\underline{\theta}^a}{\overline{\theta}^b}, \frac{\overline{\theta}^a}{\underline{\theta}^b}\right]$ then it defines a unique interior equilibrium; in this case there is some communication in equilibrium. If on the other hand it is greater than $\frac{\overline{\theta}^a}{\underline{\theta}^b}$ then Ben only sends m^b, while if it is less than $\frac{\underline{\theta}^a}{\overline{\theta}^b}$ then Ben only sends m^a; in these cases nothing is communicated in equilibrium. □

This equilibrium generically features a distortion away from the most effective intervention. To see this, consider the most interesting case in which there is non-trivial communication in equilibrium. In order to maximize effectiveness Ben would like to recommend intervention a if and only if $\theta^a \geq \theta^b$. In equilibrium, however, he gets intervention a when $\theta^a d(\overline{\theta}^a) > \theta^b d(\overline{\theta}^b)$. These conditions coincide only if $\theta^a = \theta^b$; otherwise they diverge, and Ben is either too likely to get one or the other intervention.

The basic issue here is intuitive. For any given amount Don spends, he and Ben would both prefer that he spend it on the most effective intervention. This motivates Ben to inform Don if the intervention he is considering is not in fact the best. Ben also realizes, however, that if Don is excited about the potential of one intervention then disillusioning him may not only affect *how* he helps but also *how much*. He may therefore optimally allow Don to retain a mistakenly optimistic view of some "pet" intervention, preferring a lot of somewhat useful help to a smaller amount of more impactful giving.[11]

The result indicates that the size of this distortion depends on the relative magnitude of $\overline{\theta}^a$ and $\overline{\theta}^b$. If the two interventions allow similar scope for optimism or have similar "upside potential" then distortions will be minimized. For example, there should be little bias in conversations about the best way to achieve some fixed goal. If not then there will be a bias towards the intervention with more upside potential at the expense of the one with the higher expected return; in extreme cases where $\underline{\theta}^a d(\overline{\theta}^a) > \overline{\theta}^b d(\overline{\theta}^b)$ communication breaks down entirely. Note that because bias is driven by upside this implies that donors will tend to be biased towards relatively new, untested interventions whose potential upside is still very high at the expense of older, more tested interventions whose effects are well-known – a bias which gives rise in a natural way to "fads."

[11] While the details differ, the basic tension here parallels that in Che et al. (2013). They study a model in which an agent advises a decision-maker on which of several discrete projects to implement. Given perfect information the decision-maker and agent have identical preferences over these projects, but the decision-maker also places positive value on an "outside option" which is worthless to the agent. This tension introduces distortions in communication, with the better-informed agent sometimes recommending inferior projects in order to prevent the decision-maker from exercising his outside option.

Economic Circumstances and Decision-Making

Leandro S. Carvalho (University of Southern California)
Stephan Meier (Colombia University)
Stephanie W. Wang (University of Pittsburgh)

December 2013

Abstract

This paper contributes to the debate on the effects of poverty by exploring, in an U.S. field setting, the causal effect of financial strain on cognitive and economic decision-making. Specifically, we exploit the sharp discontinuity in financial resources at payday for a low socioeconomic-status sample and randomize whether the participants received an online survey shortly before or shortly after payday. The survey collected measures of cognitive function and administered incentivized risky choice and (monetary and non-monetary) intertemporal choice tasks. The study design was effective in generating variation in economic circumstances: participants assigned to the before-payday group had on average 52% less cash and 48% lower expenditures than the after-payday group. The results show that before-payday participants behaved as if they were more present biased when making choices over monetary rewards. However, this difference was not observed for intertemporal choices in a *non*-monetary real effort task. These outcomes taken together suggest that the difference in the monetary intertemporal choice is most likely due to liquidity constraints rather than poverty reducing self-control. Our results support the "rational adaptation" view that the poor do the best they can given their economic circumstances. Additionally, we do not find differences in key aspects of cognitive functions such as cognitive control and working memory (measured by the Flanker, Simon, and Cognitive Reflection Test), the willingness to take risks, the likelihood of making decision-making errors (measured as violations of GARP), or the proneness to heuristic judgments (gambler's fallacy and framing) across the treatments.

1. Introduction

The poor often behave differently from the non-poor. They are for example more likely to make use of expensive payday loan (Bertrand and Morse 2011, Dobbie and Skiba 2013) and check-cashing services (Rhine et al. 2006), to play lotteries (Haisley et al. 2008), and to repeatedly borrow at high interest rates (Ananth et al. 2007). A debate about the reasons underlying such differences has a long and contentious history in the social sciences with two opposing views that either the poor are rational adapters who are making optimal decisions for their economic environment (e.g., Schultz 1964) or that a "culture of poverty" makes them more prone to mistakes (e.g., Lewis 1965).[1] Among economists, a lingering question has been whether the poor are more impatient (e.g., Lawrance 1991; Carvalho 2013; Haushofer et al. 2013), more risk averse (e.g., Tanaka et al. 2010; Gloede et al. 2012), and have lower self-control (e.g., Banerjee and Mullainathan 2010; Spears 2011), which could trap them into a cycle of poverty (e.g., Bernheim et al. 2013). A third view has emerged from the work of Mullainathan, Shafir, and co-authors (e.g., Shah et al. 2012; Mani et al. 2013), who argue that scarcity, defined as "having less than you feel you need" (Mullainathan and Shafir 2013 pp. 4), impedes cognitive functioning, which in turn would lead to decision-making errors and myopic behavior—a hypothesis that is consistent with a number of studies documenting an association between cognitive ability and economic choices (e.g., Burks et al. 2009, Dohmen et al. 2010, Benjamin et al. forthcoming).

There are, however, major challenges in isolating the causal effects of economic circumstances on decision-making. There may not be only a reverse causality bias – the economic decisions one makes determine one's economic circumstances – but also unobserved individual characteristics, such as innate cognitive ability, could confound the relationship

[1] See Bertrand, Mullainathan and Shafir (2004) and Duflo (2006).

between economic circumstances and decision-making. The identification of the effects of poverty on time preferences is further complicated by the possibility that poverty may affect credit constraints and arbitrage opportunities, which in turn could influence intertemporal choices elicited to measure time preferences (Pender 1996; Coller and Williams 1999; Frederick et al. 2002; Stahl 2010).[2]

This paper uses changes in economic circumstances at payday to investigate whether there is a causal effect of resource scarcity on economic decision-making. Previous work has documented that the expenditures and the caloric intake of social security and food stamp recipients increase sharply at payday (Stephens 2003; Huffman and Barenstein 2004; Shapiro 2005; Hastings and Washington 2010). This pattern is observed only for households with little or no savings (Mastrobuoni and Weinberg 2009), which suggests that some households struggle to make ends meet as the end of the pay cycle approaches.

To exploit the sharp change in economic circumstances at payday, we designed and ran online surveys in which 1,098 participants with annual family income below $40,000 were randomly assigned to a group that was surveyed shortly *before* payday (hereafter, the before-payday group) or to a group that was surveyed shortly *after* payday (hereafter, the after-payday group). We collected measures of cognitive function and administered incentivized risky choice and (monetary and non-monetary) intertemporal choice tasks to investigate whether the before-payday group, who at the time had supposedly worse economic circumstances, would behave differently from the after-payday group.

We show the study design was effective in generating variation in economic circumstances: participants assigned to the before-payday group were more short on money than participants

[2] E.g., the poor may less willing to delay gratification if they have outstanding debt with an interest rate that is higher than the interest rate offered in the experimental task. In particular, they would choose an immediate reward over a larger, more delayed reward and use it to pay part of their outstanding debt.

assigned to the after-payday group. In particular, the first group had on average 52% less cash and expenditures 48% lower than the latter group. Those patterns are, as expected, more pronounced for participants whose economic circumstances one would expect to change more sharply at payday, namely those who self-reported to live from paycheck to paycheck; with an annual family income below $20,000; who received only one payment per month (which presumably made smoothing more difficult); and who had experienced substantial financial hardships.

Our results, which we discuss next in the context of three strands of literature, do not support the hypothesis that resource scarcity *per se* impedes cognitive function and worsens the quality of decision-making (Spears 2011; Mullainathan and Shafir 2013). Even though the study design produced economically meaningful variation in economic circumstances, we find that participants surveyed before and after payday had similar performances on a slew of cognitive function tasks that involve cognitive control and tax the cognitive load, such as the Flanker task, a working memory task and the cognitive reflection test (CRT, Frederick 2005). Furthermore, we find no difference in the likelihood of heuristic judgment in tasks involving framing or the gambler's fallacy for example. As important, there were no significant differences across the two groups in the quality of their intertemporal choices, measured in terms of the consistency of such choices with the General Axiom of Revealed Preference (Choi et al. 2007a, 2007b). These results also relate to an emerging literature on heterogeneity in decision-making (e.g., Korniotis and Kumar 2011; Besedeš et al. 2012; Agarwal and Mazumder 2013), contrasting with previous findings that the better-off are better decision-makers (Choi et al. 2013).

The findings also contribute to the literature on the effects of poverty on time and risk preferences, suggesting there is no such relationship. We find the two groups made similar

intertemporal choices over costly tasks (namely choosing between taking a shorter survey sooner or a longer survey in the future). And even though the before-payday group behaved as if they were more present-biased when making intertemporal choices over monetary rewards (Andreoni and Sprenger's Convex Time Budget (CTB) method, 2012), this is most likely explained by differences in liquidity constraints given that the two groups made similar intertemporal choices over *non*-monetary rewards. They also made similar choices in a risk choice task (Eckel and Grossman, 2005), suggesting that economic circumstances do not affect risk aversion. More generally, these findings speak to burgeoning literatures on whether preferences are stable over time or across contexts (Meier and Sprenger 2010; Straznicka 2012; Chuang and Schechter 2013) and on the determinants of time and risk preferences (e.g., Nagel and Malmendier 2011; Nguyen 2011; Callen et al. 2013; Cameron and Shah 2013; Carvalho et al. 2013). Finally, our results are consistent with an interpretation that the poor are rational and make their best given their circumstances (e.g., liquidity constraints).

The paper is structured as follows. In section 2 we discuss the study design. Section 3 presents the results followed by a concluding discussion.

2. Study Design

We collected data using the RAND-USC American Life Panel (ALP), an ongoing Internet panel with respondents aged 18 and above living in the U.S. About twice a month, respondents receive an email with a request to visit the ALP site and complete questionnaires. Respondents with no Internet access at the time of recruitment are provided laptops and an Internet access subscription, which partly mitigates concerns about selection due to Internet access. The panel has been recently expanded to include 2,000+ members from vulnerable populations drawn from zipcodes with a large fraction of low-income, low-education, or minority populations. We

restricted the study sample to panel members with an annual household income of $40,000 or less.

The study consisted of one baseline and one follow-up survey. The baseline survey collected information used to determine participants' paydays. Depending on whether a given participant was randomized into the before-payday or the after-payday group, the follow-up survey was timed to open for this participant few days before or few days after her payday. The follow-up survey measured various aspects of the decision-making of the two randomly assigned groups.

2.1. The Baseline Survey and Study Sample

The baseline survey collected data on the arrival dates of all payments the participant (and his/her spouse) expected to receive during January 2013.[3] Respondents were asked to provide information about the date and the amount of each payment. They also reported the method of payment (e.g., direct deposit, cash, etc.). See Appendix XXX for screenshots of the baseline survey. The study focuses on subjects who provided information about the number of expected payments and the dates of these payments, and on those who anticipated receiving less than five payments from a single income source in January 2013.[4]

Of the study participants, about 44% expected to receive wages and salaries, 41% anticipated social security or disability insurance payments, 6.6% welfare payments, and 2.4% unemployment compensation. Thirty-eight percent anticipated receiving one payment only in January 2013 while 41.7% expected receiving two payments. The median payment (among all sources of income) was $1,379. For wages and salaries, the median payment was $1,500 while the median payment for social security, disability insurance, or retirement income was

[3] To test the survey design, we conducted a pilot in May of 2010 with about 200 respondents, in which it was randomly assigned whether a participant was surveyed before or after payday.
[4] The rationale for dropping respondents who anticipated 5 payments or more from a single income source is that their income should be sufficiently spread out over time, making it easier for them to smooth consumption.

approximately $1,000. Direct deposit and receiving a check in person were the most common methods of payment for wages and salaries. Social security, disability insurance, welfare payments, unemployment compensation, and retirement income were typically paid either through direct deposit or through a check in the mail. See Appendix XXX for more details about the payments.

By design, we targeted low-income households (annual household income of $40,000 or less). Results from the baseline survey confirm that the study sample was indeed mostly low-income and had suffered hardships in the past because of shortage of money. Eighteen percent of our sample reported being unemployed, which is more than two times the 7.8% U.S. unemployment rate in December of 2012. One-fifth reported being disabled and only 37% reported that they were working at the time. A third of our sample had an annual family income of $15,000 or less. More than half reported that because of a shortage of money at least one of the following had happened to them in the 12 months before the survey: could not pay electricity, gas or phone bills, or for car registration or insurance; pawned or sold something; went without meals; were unable to heat home; sought assistance from welfare or community organizations, friends, or family; took a payday loan.

2.2.Randomization and Treatment Compliance

The opening dates of the follow-up survey, which were specific to each study participant, depended on the participant's payday and her random assignment. In particular, the follow-up survey opened 7 days before payday for participants in the before-payday group and 1 day after

payday for participants in the after-payday group.[5] Participants were sent emails informing them when the survey was available.[6]

The payday of each participant was identified using the data collected in the baseline survey. If the largest expected payment was preceded by an interval of two weeks or more without payments, the payday was set as the date of this largest payment. Otherwise, the payday was set as the date of the payment that followed an interval of 14 days or more without payments. 208 participants whose payments were all less than 2 weeks apart were dropped from the study sample. See Appendix XXX for the flow of participants through the study, which gives more details about sample restrictions and survey nonresponse.

The remaining 1,191 study participants were then randomly assigned to the before-payday or after-payday groups using a stratified sampling and a re-randomization procedure (see Appendix XXX for more details). We stratified on whether participants strongly agreed at baseline with the statement "I live from paycheck to paycheck" and on whether they anticipated receiving one payment only in January 2013, as we planned to check whether the effects would be any different for those participants whose economic circumstances one would expect to change more sharply at payday. The randomization was indeed successful in making assignment to the before- and after-payday groups orthogonal to observable baseline characteristics (see Appendix XXX).

More importantly, the study design was effective in generating variation on the time at which participants started and finished the survey. The median respondent assigned to the before-payday group started the survey 2.4 days before payday and completed it 1.5 days before

[5] Take for example a participant whose payday fell on January 15th. If s/he were assigned to the before-payday group, the survey would be available starting from January 8th at 12:00:01am. If s/he were assigned to the after-payday group, the survey would be available starting from January 16th at 12:00:01am.

[6] Participants whose survey opened on a given day were sent an email at 12:00:01am of that given day to inform them that there was a new survey ready for them (a remainder email was sent 5 days later).

payday.[7] The median respondent assigned to the after-payday group started the survey 4.4 days after payday and completed it 5 days after payday. The differences across the two groups are all statistically significant at 1%. See Appendix XXX for more detailed results.

Notice that the study design allowed us to manipulate when the survey was made available to a participant, but we did not have control over when the participant started (or finished) the survey. Consequently, there was imperfect compliance in the sense that some of the participants assigned to the before-payday group effectively started (or finished) the follow-up survey after payday. About 70% of participants assigned to the before- payday group started the survey before payday while 63% completed the survey before payday.[8] By construction all participants assigned to the after-payday group started the survey at least one day after payday.

In our analysis we estimate intention-to-treat effects, exploiting the random assignment to the before-payday group as a source of exogenous variation in starting the survey before payday.

2.3.The Follow-up Survey

The follow-up survey included experimental tasks, tests and questions on (1) economic decision-making, (2) cognitive function, and (3) economic circumstances, which we discuss here briefly. For more details and screenshots of the follow-up survey, see Appendix XXX.

(1) Economic Decision-Making. The follow-up survey included three experimental tasks to measure economic decision-making: two intertemporal choice tasks – one with monetary rewards and one with non-monetary rewards – and a lottery choice task. In the monetary intertemporal task, which was a variant of Andreoni and Sprenger's Convex Time Budget (CTB)

[7] Because participants were not required to complete the survey in just one sitting, the time interval between the time at which a respondent started the survey and the time at which she completed the survey may have been much greater than the time it would effectively have taken her to complete it had she done it uninterruptedly.

[8] There is no statistically significant difference across the before-payday and after-payday groups in the likelihood of starting the survey within 7 days of its opening date. Our results are very similar if we restrict the sample to "compliers" who started the survey within seven days of its opening. See Appendix XXX.

(2012), participants had to allocate an experimental budget of $500 between two payments with pre-specified dates. The amount allocated to the second payment was paid with interest. Participants were asked to make twelve of these choices in which we varied the experimental interest rate (0%, 0.5%, 1%, or 3%), the date of the first payment (today or in 4 weeks) and the time delay between the two payments (4 weeks or 8 weeks). Approximately one percent of participants were selected at random to be paid one of their choices (the choice for which the participant was paid was randomly selected among the participant's twelve choices).

We also administered a task in which participants made intertemporal choices over real effort (similar to Augenblick, Niederle, and Sprenger 2013) to address concerns about the use of monetary rewards to measure time discounting (e.g., Frederick, Loewenstein, and O'Donoghue 2002). More specifically, participants had to choose between completing a shorter survey within 5 days and a longer 30-minutes survey within 35 days. There were asked to make five of these choices, in which the length of the sooner survey was gradually increased (10, 15, 20, 25, and 29 minutes). Five similar choices followed in which the deadlines were shifted to 90 days (shorter) and 120 days (longer). Approximately one percent of participants were selected at random to have their choices implemented (one among the participant's ten choices was randomly selected for implementation). "Implementation surveys" were sent to those selected participants.[9]

To analyze the willingness to take risks, we used a lottery-choice task by Eckel and Grossman (2002, 2007), in which participants were asked to choose one among six lotteries. Each lottery had a 50-50 chance, based on a coin flip, of paying either a lower or higher reward. The five (higher; lower) pairings were ($28,$28), ($36,$24), ($44,$20), ($52,$16), ($60,$12),

[9] If they completed the survey before the deadline, they received a $50 Amazon gift card and $20 was added to the quarterly check they regularly receive for answering surveys. The dates of these payments were fixed and thus did not depend on when respondents finished the implementation surveys (as long as if completed before the deadline).

and ($72,$0). Approximately one percent of participants were selected at random to be paid according to their choices.

Two additional tasks measured loss aversion as in Fehr and Goette (2007) and simplicity seeking using the task in Iyengar and Kamenica (2010).[10] While the latter task was incentivized, the former was not.

(2) Cognitive Function. We measured cognitive function using the Flanker task, a working memory task, and the Cognitive Reflection Test (CRT). In the flanker task, a well-established cognitive control task that is part of the NIH toolbox (Zelazo et al. 2013), subjects were supposed to focus on a central stimulus while inhibiting attention to distracting stimuli (Ericksen and Ericksen 1974).[11] In the working memory task participants were asked to recall a sequence of colors that gradually increased in length if the participant could successfully repeat the sequence.[12] The CRT measures one's ability to suppress an intuitive and spontaneous incorrect answer and give the deliberative and reflective correct answer (Frederick 2005). In addition to the measures of cognitive function we administered, we also have several other measures of study participants' cognitive abilities, including both fluid and crystallized intelligence, which were collected in previous ALP surveys. Appendix XXX shows that our measures of cognitive function are strongly correlated with other measures of cognitive ability.

---

[10] In the simplicity seeking task participants had to choose among a number of lotteries that paid different amounts depending on the roll of a dice. One of the options was a simpler dominated lottery that paid $0 or $10 with a 50-50 chance. Half of participants were randomly assigned to choose one option among 11 lotteries. The other half had to choose one among 3 lotteries. As Iyengar and Kamenica (2010), we find that those who had to choose among 11 options were more likely to choose the simpler dominated lottery than those who had 3 options.

[11] In each one of the twenty trials of the flanker task, participants were shown five arrows in a row that pointed either to the left or the right. Participants were supposed to determine – as quick as possible without making mistakes – the direction of the center arrow while ignoring the directions of the four arrows that surrounded the center arrow. Performance was measured in terms of response time and number of correct trials.

[12] We adapted the memory game *Simon* made by Milton Bradley for our web-based survey. The participants saw a circle with four quadrants of four different colors. There were four buttons of these same colors below the circle. The quadrants lighted up in a random sequential order and the participants had to reproduce the order by pressing the buttons. If the participant could successfully repeat the sequence, an additional color was added to the sequence. The measure of working memory was the length of the longest sequence the participant was able to reproduce.

We also used the "disease problem" (Tversky and Kahneman 1981) to measure sensitivity to framing and a task proposed by Topak et al. (2011) to measure belief in the "Law of Small Numbers" (Rabin 2000), that is, an error in which individuals exaggerate how likely it is that a small sample resembles the overall population from which it is drawn.

(3) Financial circumstances. Finally, the follow-up survey included questions on cash holdings, checking and savings accounts balances, and expenditures that permit examining if the study design generated variation in economic circumstances.

3. Empirical Results

The presentation of the empirical results is structured as follows. First, we show that the study design generated substantial differences in the economic circumstances of the before-payday and after-payday groups. Second, we examine if these differences in economic circumstances were accompanied by differences in the intertemporal choices and the risk choices of the two groups. We then proceed to investigate if there were any differences in the cognitive function of such groups. Lastly, we look into whether subgroups who are the most strained before payday show different behavioral patterns.

3.1.Cash and Expenditures

Table 1 below indicates that overall the before-payday group was more short on money than the after-payday group. On average the before-payday group had roughly 20% less cash than the after-payday group (Panel A). The before-payday group also reported having spent less money in the previous seven days (Panel B). The median expenditure for the before-payday group was

22% lower than the median expenditure of the after-payday group.[13] A Wilcoxon rank-sum test that the two samples are drawn from the same distribution can be rejected at 5% for both cash and total expenditures (not shown in the table).

Cash Holdings and Total Expenditures

		Mean		*Median*	
	N	After Payday	Before-after % diff.	After Payday	Before-after % diff.
Group		**Panel A: Cash**			
All	1,044	$103	-20% **	$45	-22% ***
Live from Paycheck to Paycheck	554	$78	-22% *	$30	-17%
Annual Income below $20,000	465	$84	-18% ***	$40	-33% ***
One Payment	415	$97	-30%	$45	-33% **
Financial Hardship	543	$86	-30% ***	$30	-20%
Group		**Panel B: Total Expenditures**			
All	1,056	$1,156	-48% ***	$200	-20% ***
Live from Paycheck to Paycheck	557	$1,619	-64% ***	$250	-27% ***
Annual Income below $20,000	470	$1,735	-70% ***	$200	-25% **
One Payment	423	$926	-41% **	$200	-40% ***
Financial Hardship	547	$1,621	-65% ***	$230	-33% ***

Note: The table shows the mean and the median of cash (Panel A) and total expenditures (Panel B) for the after-payday group. It also reports the before-after difference in percentage terms calculated using the delta method. For the calculation of the means the top 1% of the overall distribution of cash and total expenditures were trimmed. Results are estimated for the following sub-groups: i) participants who strongly agreed at baseline with the statement "I live from paycheck to paycheck"; ii) participants with an annual family income of $20,000 or less; iii) participants who expected to receive one payment only throughout January 2013; and iv) participants who because of a shortage of money had experienced in the 12 months before the baseline survey at least one of the following: could not pay electricity, gas or phone bills; could not pay for car registration or insurance; had pawned or sold something; had gone without meals; had been unable to heat home; had sought assistance from welfare or community organizations; had sought assistance from friends or family; or had taken a loan from a payday lender.

[13] The before-after difference in median expenditures of 20% is comparable to what previous studies have found. Mastrobuoni and Weinberg (2009) find that Social Security recipients with less than $5,000 in savings consumed 25% calories in the week after payday relative to the week before payday while Shapiro (2005) finds that the caloric intake of food stamp recipients reduces by 15% over the food stamp month.

These patterns are even starker for participants whose economic circumstances were expected to change more sharply at payday, namely those who at baseline reported that they strongly agreed with the statement "I live from paycheck to paycheck", those with an annual family income below $20,000, those who received only one payment per month (which presumably made consumption smoothing more difficult), and those who had gone through financial hardships in the previous 12 months. Among the latter group, for example, participants assigned to the before-payday group spent on average 65% less than the after-payday group.

In sum, the economic circumstances of the before-payday and after-payday groups at the time of the follow-up survey were substantially different, and the financial strain the before-payday group was experiencing may have affected their decision-making and behavior at the time. In what follows, we investigate if the before- and after-payday groups made different intertemporal and risk choices and/or had different performances in cognitive function tasks.

3.2.Intertemporal Choices

At least since Fisher (1930), economists have debated whether the poor have higher discount rates (e.g., Lawrance 1991, Carvalho 2013; Haushofer et al. 2013). The fact that the poor are more likely to be liquidity constrained makes it particularly challenging to test such hypothesis. All else equal credit constrained individuals are less willing to delay gratification (Pender 1996). Because a credit constrained individual may not be able to borrow against future income, her marginal utility of $1 today may be higher than her marginal utility of $1 in the future, which could be confounded with a high discount rate (Frederick, Loewenstein, and O'Donoghue 2002).

In our results we find that the before-payday group, who had at the time worse economic circumstances than the after-payday group, behaved as if they were more present-biased when making intertemporal choices over monetary rewards. As shown in the first column of Table 2,

Intertemporal Choices over Monetary Rewards and over Costly Tasks

	Dependent Variable	
	Monetary	**Non-Monetary**
Independent Variables	Sooner Reward / 500	{Longer-later Survey}
{Before Payday} * {Today}	0.02	-0.02
	[0.008]***	[0.021]
{Before Payday} * Interest rate	0.54	0.02
	[0.647]	[0.013]*
{Before Payday} * Delay time	0.00	-
	[0.002]	
{Before Payday}	-0.01	-0.01
	[0.020]	[0.026]
{Today}	-0.01	0.08
	[0.006]*	[0.014]***
Interest rate	-9.46	-0.11
	[0.467]***	[0.009]***
Delay time	0.00	-
	[0.001]	
Constant	0.61	0.39
	[0.014]***	[0.018]***
Joint Test (before-after differences = 0)		
{Before Payday} and interaction terms	0.02	0.30
excluding {Before Payday} * {Today}	0.21	0.19
N	12,720	10,250

Note: In the intertemporal choice task with monetary rewards subjects had to allocate $500 between pre-specified sooner and later dates. The amount allocated to the later date was paid with interest. They made 12 of these choices in which we varied the sooner date, the delay time between the sooner and the later dates, and the experimental interest rate. In the intertemporal choice task with non-monetary rewards subjects had to choose between answering a shorter-sooner survey and a longer-later survey of 30 minutes. They made 10 choices in which we varied the duration of the shorter-sooner survey (i.e., the interest rate for procrastinating survey completion) and whether the shorter-sooner survey had to be completed in the next 5 days or in the next 90 days. The deadline for the longer-later survey was always 30 days after the deadline of the shorter-sooner survey. The dependent variable in the first column is the sooner reward divided by 500. The dependent variable in the last column is an indicator variable for whether the respondent chose to answer the later-longer survey. The standard errors are clustered at the individual level.

the before-payday group increased the amount allocated to the sooner check by 2 percentage points in response to a change in the sooner date from four weeks to today. This difference is statistically significant at 1%. There are no other statistically significant differences in the choices across the two groups. In particular, there are no differences in how the two groups responded to a change in delay time, which suggests the before-payday group did not have a lower discount rate than the after-payday group.

While the result that the before-payday group behaved as if they were more present-biased is consistent with an interpretation that scarcity or financial hardship reduces self-control, it is also possible that before-after differences in liquidity constraints could account for such behavior. To test this hypothesis, we analyze intertemporal choices over *non-monetary* rewards for which liquidity constraints should not matter.

We find the two groups behaved similarly in the task in which they made intertemporal choices over a costly real-effort task, namely choosing between answering a shorter survey sooner and a longer survey later. Both groups displayed behavior *consistent* with present bias: Participants were 8 percentage points more likely to choose the longer-later survey when the shorter-sooner survey had to be completed within 5 days (as opposed to having 90 days).[14] However, there were no differences across the two groups; they responded similarly to the change in the deadline of the shorter-sooner survey.[15] There is no evidence therefore of differential present bias in the task with non-monetary rewards, suggesting that liquidity constraints explain why the before-payday group behaved as if they were present-biased when make intertemporal choices over monetary rewards. These results also provide further evidence

[14] There is no statistically significant difference on how much time the time the before- and after-payday groups took on average to complete the follow-up survey, which does not support the hypothesis that the marginal utility of time was different across the two groups.

[15] The before-payday group is marginally (10% level) less responsive to the "interest rate" charged on postponing the real-effort task (i.e., the length of the shorter-sooner survey).

that the before-payday group had worse economic circumstances at the time than the after-payday group.

3.3. **Risk Choices**

According to standard utility theory, the more financially constrained before-payday group is in a steeper part of their utility function and consequently should be less willing to take risks. However, it is well documented that the poor are more likely to purchase lottery tickets (Haisley et al. 2008; Beckert and Lutter 2012), and there are studies that investigate whether the poor have in fact different risk preferences (e.g., Tanaka et al. 2010; Gloede et al. 2012).

As Table 3 shows, in our study the before-payday and after-payday groups made similar choices in the lottery-choice task. The before-payday group was somewhat more willing to take risks, but the differences are small and are not statistically significant.

Cumulative Distribution of Choices in Lottery-Choice Task

	Dependent Variable: Fraction Chose Lotteries at least as Risky as (Heads,Tails)					
	(70,2)	(60,12)	(52,16)	(44,20)	(36,24)	(28,28)
{Before-Payday}	0.00	0.02	0.04	0.00	0.02	-
	[0.024]	[0.028]	[0.030]	[0.030]	[0.027]	
Constant	0.18	0.29	0.39	0.59	0.72	1
	[0.017]***	[0.020]***	[0.021]***	[0.021]***	[0.019]***	
P-value Wilcoxon			0.42			

Note: The table reports the distribution of choices in a lottery-choice task in which subjects chose one among six lotteries that paid different amounts depending on a coin toss. The dependent variable is the fraction of participants who chose lotteries as risky or riskier than the lottery listed in the column. Robust standard errors in brackets. N = 1,064.

The before-payday and after-payday groups also made similar choices in two additional lottery-choice tasks: there were no differences across the two groups in both the loss aversion and simplicity seeking experimental tasks.

3.4. Violations of the General Axiom of Revealed Preference

To investigate whether scarcity affects the quality of decision-making, as suggested by Mullainathan and Shafir (2013), we follow two approaches: first, we look at violations of the General Axiom of Revealed Preference (GARP) as a measure of quality of decision-making—an approach that has been used in several recent studies (e.g., Choi et al. 2007a, 2007b, 2013).[16] Second, we look at performance in cognitive function tasks, because changes in cognitive functioning could possibly interfere with decision-making (e.g., Benjamin et al. forthcoming).

We focus on violations of GARP in the intertemporal choice tasks. In the task with monetary rewards, a violation corresponded to a decrease in the later reward in response to an increase in the experimental interest rate. In the task with non-monetary rewards a violation corresponded to a switch from a choice of the longer-later survey to the shorter-sooner survey in response to an increase in the duration of the latter (the later survey was always kept constant at 30 minutes). Our outcome of interest is the number of GARP violations in each task.

[16] See Kariv and Silverman (2012) for an argument for using consistency with rationality as a measure of quality of decision-making.

Number of Violations of General Axiom of Revealed Preference (GARP) in Intertemporal Choice Tasks

	Monetary	Non-monetary
{Before Payday}	0.17	0.03
	[0.115]	[0.047]
Constant	1.44	0.28
	[0.078]***	[0.032]***
N	1,060	1,025

Note: In the intertemporal choice task with monetary rewards subjects violated GARP when they decreased the later reward in response to an increase in the experimental interest rate. In the intertemporal choice task with non-monetary rewards subjects violated GARP when they switched from the longer-later survey to the shorter-sooner survey in response to an increase in the duration of the shorter-sooner survey. Robust standard errors shown.

Table 4 shows that the after-payday group had 1.44 violations. The before-payday group had on average 0.17 more violations, but this difference is not statistically significant. There were also no statistically significant differences in the non-monetary intertemporal task. The p-value of Wilcoxon rank-sum test was respectively 0.24 and 0.70.[17]

In terms of heuristics, there were also no differences in how sensitive the two groups were to framing (Tversky and Kahneman 1981) and how likely they were to succumb to the gambler's fallacy—i.e., the incorrect expectation that after one particular realization of a random variable the next realization of this same random variable will be different. These results are shown in Appendix XXX. We look next at cognitive functioning.

[17] We also do not find a difference in the number of violations of GARP in the loss aversion task. The mean for the after-payday group is 0.222 and the before-after difference is 0.016 with a p-value of 0.62. We chose to not include in the table because the loss aversion task was not incentivized.

3.5. Cognitive Function

As shown in Table 5, the before-payday and after-payday groups had similar performances in the tasks and tests used to measure cognitive function. On the Flanker task, a well-established cognitive control task that is part of the NIH Toolbox, participants assigned to the before-payday group were on average two percent slower in their response time than other participants, but they were also one percentage point more likely to respond correctly. None of these differences are statistically significant at 10%. The before-payday group had slightly better performances in the working memory task and in the Cognitive Reflection Test (Frederick 2005), but again, these differences are not statistically significant.

Cognitive Function

	Flanker		*Simon*	*Cognitive Reflection Test*		
	Ln(Response Time)	Fraction Correct	# Rounds	Bat and ball	Widgets	Lily pads
{Before Payday}	0.02	0.01	0.02	0.00	0.02	0.01
	[0.028]	[0.010]	[0.239]	[0.013]	[0.023]	[0.021]
Constant	7.16	0.87	4.69	0.04	0.16	0.13
	[0.029]***	[0.011]***	[0.164]***	[0.009]***	[0.016]***	[0.015]***
N	20,557	20,565	1,038	1,047	1,055	1,053

Note: In each one of the 20 trials of the flanker task, participants were shown five arrows in a row that pointed either to the left or the right. Participants were supposed to determine – as quick as possible without making mistakes – the direction of the center arrow while ignoring the directions of the four arrows that surrounded the center arrow. Response time was measured in miliseconds. The flanker regressions include trial-specific dummies. The working memory task was based on the memory game Simon made by Milton Bradley. Participants saw a circle with four quadrants of four different colors. There were four buttons of these same colors below the circle. The quadrants lighted up in a random sequential order and the participants had to reproduce the order by pressing the buttons. If the participant could successfully repeat the sequence, an additional color was added to the sequence. The measure of working memory is the length of the longest sequence the participant was able to reproduce. The Cognitive Reflection Test, which is formed by three questions, measures one's ability to suppress an intuitive and spontaneous incorrect answer and give the deliberative and reflective correct answer (Frederick 2005). Robust standard errors shown.

These results contrast with the findings of Mani, Mullainathan, Shafir and Zhao (2013), who document that sugar cane farmers in India have worse performance in related cognitive function tasks during the pre-harvest season, when they supposedly have fewer resources, than during the post-harvest season.

One possibility for why we do not find an effect on cognitive function is that short-term variations in economic circumstances may not translate into short-term variations in the *feeling* of scarcity. The participants in our study may have been experiencing a constant feeling of scarcity that did not subside with the payment arrival. To the extent that it is possible to measure this subjective perception of scarcity, we show *suggestive* evidence in the appendix that is consistent with such interpretation: The before-payday and after-payday groups had equal subjective assessments about how hard it would be for them to cover the expenses they expected to have in the 5 days following the follow-up survey and they also reported similar levels of stress with their personal finances and satisfaction with their financial situation at the time (see Appendix XXX).

While our design is similar to the design used in Mani, Mullainathan, Shafir and Zhao (2013), that analyze cognitive function before and after harvest of sugar cane farmers in India, the longer interval between before and after harvest – 4 months – may have enabled them to capture variations in the feeling of scarcity that we may have not been able to. By contrast, the shorter interval used in our study – fewer than 7 days on average – seems to lend more credibility to the assumption that "after" provides a good counterfactual to "before."

3.6.Subgroups Analysis

One could worry that individuals participating in our study may not be poor enough to trigger a psychology of poverty, which could explain why our findings do not align with the results of

Mani et al. (2013). To look into this hypothesis, we estimated separate results for the subgroups whose economic circumstances changed more sharply at payday as shown in Table 1.

We find that the results for these subgroups mirror the results for the overall sample (detailed results shown in Appendix XXX). In particular, among participants of a given subgroup (e.g., those making less than $20,000 per year), those assigned to the before-payday group behaved as if they were more present-biased than those assigned to the after-payday group when making *monetary* intertemporal choices, but the two groups made similar intertemporal choices over *non-monetary* rewards. There were also no before-after differences in risk choices, violations of GARP, or cognitive functioning.

In sum, even for the very constrained among our already relatively poor sample, no differences in decision-making are observed.

4. Conclusions

The debate about the causes and effects of poverty has a long and contentious history in the social sciences. This discussion has gained new life with recent studies arguing that poverty impedes cognitive function, worsens the quality of decision-making, and leads to myopic behavior among other effects, which could explain why the poor often seem to make decisions that reinforce their conditions of poverty.

Using the sharp discontinuity in available resources at payday for a low-SES population, we look at the causal effects of scarce resources on decision-making. This makes our study the first that we are aware of to provide experimental evidence on whether financial strain affects the economic decision-making of poor U.S. families.

We show the study design was effective in generating exogenous variation in economic circumstances. We find, as previous studies have documented, that individuals run out of money as payday approaches and that economic circumstances improve precipitously at payday. The phenomenon is even more pronounced for subgroups whose economic circumstances one would expect to change more sharply at payday, such as those who receive one payment only per month and those in the bottom half of the income distribution of our low-SES study sample.

Our results indicate that scarce resources can indeed affect the willingness to delay gratification. We find that before payday participants behaved as if they were more present biased when making choices over monetary rewards. However, present-biased behavior was the same before and after payday when participants made choices over *non*-monetary rewards. These outcomes taken together suggest that the difference in the monetary intertemporal choice is most likely due to liquidity constraints rather than poverty reducing self-control. Our results support the "rational adaptation" view that the poor do the best they can given their circumstances. In particular, credit market imperfections that prevent the poor from borrowing (Banerjee 2001) can explain why they are less willing to delay gratification when they are short on cash. In other words, our findings are not consistent with the hypothesis that poverty alters time preferences.

Our results also do not support the hypothesis that financial strain by itself worsens the quality of decision-making. Even though there are substantial differences in economic circumstances before and after payday, we find no evidence that the willingness to take risks, the likelihood of making decision-making errors (measured as violations of GARP), or the proneness to heuristic judgments (gambler's fallacy and framing) differs across the before-payday and after-payday groups. Furthermore, we do not find differences in key aspects of cognitive functions such as cognitive control and working memory (measured by the Flanker, Simon, and

Cognitive Reflection Test) across the treatments. These cognitive and behavioral differences are not present in various subsamples where one would expect the financial constraints to be the starkest, such for example those in our study with an annual family income of $20,000.

Our findings in conjunction with the previous literature suggest that much more needs to be done to understand the effects of the interplay between long-term socioeconomic status and short-term economic circumstances on cognitive functions, subjective well-being, and economic decision-making. We showed that financial constraints do not deterministically lead to cognitive deficits and decision-making mistakes in contrast with what previous studies suggest. Our study focused on one of the most prevalent and systematic variations in economic circumstances, namely the pay cycle. Perhaps the participants had partially adapted to this variation and are inured to its effects psychologically. Alternatively, the short-term increase in economic scarcity may have been dwarfed by the long-term persistent deprivation such that they experience a constant feeling of scarcity that does not subside with the payment arrival. As Mullainathan and Shafir (2013) suggest, "[t]he feeling of scarcity is distinct from its physical reality. Physical limits, of course, play a role—the money in our savings account, the debts we owe…But so does our subjective perception…." (pp. 11)

Our study found that the primary factor leading to differences in economic decision-making was the difference in economic circumstances, specifically the presence of liquidity constraints. These findings suggest caution when drawing the direct link from resource constraints to cognitive constraints to economic mistakes. They point to future work that can refine the framework for thinking about the effect of poverty and scarcity on economic decision-making: when do economic factors dominate, when do psychological factors dominate, and how do they interact with each other. Gathering more evidence will bring us closer to understanding which

cognitive functions can be degraded or even enhanced and which are unaffected by economic resource constraints. Further exploration of the underlying mechanisms that connect such cognitive functions and economic decision-making is also warranted.

References

Agarwal and Muzumder, "Cognitive Abilities and Household Financial Decision Making," *American Economic Journal: Applied Economics,* V (2013), 193-207.

Ananth, Bindu, Dean Karlan, and Sendhil Mullainathan, "Microentrepreneurs and Their Money: Three Anomalies," mimeo, 2007.

Andreoni, James and Charles Sprenger, "Estimating Time Preferences from Convex Budgets," *American Economic Review*, CII (2012), 3333-56.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger, "Working over Time: Dynamic Inconsistency in Real Effort Tasks," mimeo, 2013.

Banerjee, Abhijit and Sendhil Mullainathan, "The Shape of Temptation: Implications for the Economic Lives of the Poor," mimeo, 2010.

Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro, "Who is 'Behavioral'? Cognitive Ability and Anomalous Preferences," *Journal of the European Economic Association,* XI (2013), 1231-1255.

Bernheim, B. Douglas, Debraj Ray, and Sevin Yeltekin, "Poverty and Self-Control," mimeo, 2013.

Bertrand, Marianne, and Adair Morse, "Information Disclosure, Cognitive Biases, and Payday Borrowing," *Journal of Finance,* LXVI (2011), 1865-1893.

Bertrand, Marianne, Sendhil Mullainathan, and Eldar Shafir, "A Behavioral-Economics View of Poverty," *American Economic Review Papers and Proceedings,* XCIV (2004), 419-423.

Besedeš, Tibor, Cary Deck, Sudipta Sarangi, and Mikhael Shor, "Age Effects and Heuristics in Decision Making," *Review of Economics and Statistics,* ICIV (2012), 580-595.

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini, "Cognitive Skills Affect Economic Preferences, Strategic Behavior, and Job Attachment," *Proceedings of the National Academy of Sciences of the United States of America,* CVI (2009), 7745-7750.

Callen, Michael, Mohammad Isaqzadeh, James D. Long, and Charles Sprenger, "Violent Trauma and Risk Preference: Artefactual and Experimental Evidence from Afghanistan," *American Economic Review,* forthcoming.

Cameron, Lisa and Manisha Shah, "Risk-Taking Behavior in the Wake of Natural Disasters," mimeo, 2013.

Carvalho, Leandro S., “Poverty and Time Preference,” mimeo, 2013.

Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv, “Revealing Preferences Graphically: An Old Method Gets a New Tool Kit,” *American Economic Review Papers and Proceedings,* IXVII (2007), 153-158.

Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv, “Consistency and Heterogeneity of Individual Behavior under Uncertainty,” *American Economic Review*, IXVII (2007), 1921-1938.

Choi, Syngjoo, Shachar Kariv, Wieland Mueller, and Dan Silverman, “Who is (More) Rational?” *American Economic Review,* forthcoming.

Chuang, Yating and Laura Schechter, “Stability of Experimental and Survey Measures of Social, Risk, and Time Preferences Over Multiple Years,” mimeo, 2013.

Coller, Maribeth and Melonie B. Williams, “Eliciting Individual Discount Rates,” *Experimental Economics,* II (1999), 107-127.

Dobbie, Will, and Paige Marta Skiba, "Information Asymmetries in Consumer Credit Markets: Evidence from Payday Lending," *American Economic Journal: Applied Economics*, V (2013), 256-282.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde, “Are Risk Aversion and Impatience Related to Cognitive Ability?” *American Economic Review,* C (2010), 1238-1260.

Duflo, Esther, “Poor but Rational?” *Understanding Poverty*, 2006.

Eckel, Catherine and Philip Grossman, “Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk,” *Evolution and Human Behavior*, XXIII (2002), 281-295.

Eriksen, Barbara A. and Charles W. Eriksen, “Effects of Noise Letters upon the Identification of a Target Letter in a Nonsearch Task,” *Perception and Psychophysics*, XVI (1974), 143-149.

Fehr, Ernst and Lorenz Goette, “Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment,” *American Economic Review,* XCVII (2007), 298-317.

Frederick, Shane, “Cognitive Reflection and Decision Making,” *Journal of Economic Perspectives,* XIX (2005), 25-42.

Frederick, Shane, George Loewenstein, and Ted O’Donoghue, “Time Discounting and Time Preference: A Critical Review,” *Journal of Economic Literature,* XL (2002), 351-401.

Gloede, Oliver, Lukas Menkhoff, and Hermann Waibel, “Shocks, Individual Risk Attitude, and Vulnerability to Poverty among Rural Households in Thailand and Vietnam,” mimeo, 2012.

Haisley, Emily, Romel Mostafa, and George Loewenstein, "Subjective relative income and lottery ticket purchases," Journal of Behavioral Decision Making, XXI (2008), 283-295.

Hastings, Justine and Ebonya Washington, "The First of the Month Effect: Consumer Behavior and Store Reponses," American Economic Journal: Economic Policy, II (2010), 142-162.

Haushofer, Johannes, Daniel Schunk, and Ernst Fehr, "Negative Income Shocks Increase Discount Rates," mimeo, 2013.

Huffman, David and Matias Barenstein, "A Monthly Struggle for Self-Control? Hyperbolic Discounting, Mental Accounting, and the Fall in Consumption Between Paydays," mimeo, 2005.

Iyengar, Sheena and Emir Kamenica, "Choice Proliferation, Simplicity Seeking, and Asset Allocation," *Journal of Public Economics,* XCIV (2010), 530-539.

Kahneman, Daniel and Amos Tversky, "The Framing of Decisions and the Psychology of Choice," *Science,* CCXI (1981), 453-458.

Kariv, Shachar and Dan Silverman, "An Old Measure of Decision-Making Quality Sheds New Light on Paternalism," *Journal of Institutional and Theoretical Economics,* CLXIX (2013), 29-44.

Korniotis, George M. and Alok Kumar, "Do Older Investors Make Better Investment Decisions?" *Review of Economics and Statistics*, XCIII (2011), 244 – 265

Lawrance, Emily C., "Poverty and the Rate of Time Preference: Evidence from Panel Data," *Journal of Political Economy*, IC (1), 54-77.

Lewis, Oscar, *LaVida: A Puerto Rican Family in The Culture of Poverty,* New York: Random House, 1965.

Malmandier, Ulrike and Stefan Nagel, "Depression Babies: Do Macroeconomic Experiences Affect Risk-Taking?" *Quarterly Journal of Economics,* CXXVI (2009), 373-416.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao, "Poverty Impedes Cognitive Function," *Science,* CCCXLI (2013), 976-980.

Mastrobuoni, Giovanni and Matthew Weinberg, "Heterogeneity in Intra-monthly Consumption Patterns, Self-Control, and Savings at Retirement," *American Economic Journal: Economic Policy*, I (2009), 163-89.

Meier, Stephan and Charles Sprenger, "Stability of Time Preferences," mimeo, 2010.

Mullainathan, Sendhil and Eldar Shafir, *Scarcity: Why Having Too Little Means So much,* New York: Times Books, 2013.

Nguyen, Quang, “Does Nurture Matter: Theory and Experimental Investigation on the Effect of Working Environment on Risk and Time Preferences,” *Journal of Risk and Uncertainty*, XLIII (2011), 245–270.

Pender, John L., “Discount Rates and Credit Markets: Theory and Evidence from Rural India,” *Journal of Development Economics,* L (1996), 257-296.

Rhine, Sherrie L. W., William H. Greene, and Maude Toussaint-Comeau, “The Importance of Check-Cashing Businesses to the Unbanked: Racial/Ethnic Differences,” *Review of Economics and Statistics*, LXXXVIII (2006), 146-157.

Schultz, Theodore W., *Transforming Traditional Agriculture,* New Haven: Yale University Press, 1964.

Shah, Anuj K., Sendhil Mullainathan, and Eldar Shafir, “Some Consequences of Having Too Little,” *Science,* CCCXXXVIII (2012), 682-685.

Shapiro, Jesse, “Is There a Daily Discount Rate? Evidence from the Food Stamp Nutrition Cycle,” *Journal of Public Economics*, LXXXIX (2005), 303-325.

Spears, Dean, “Economic Decision-Making in Poverty Depletes Behavioral Control,” *B.E. Journal of Economic Analysis & Policy*, XI (2011), 1-44.

Stahl, Dale O., "Intertemporal Choice and Liquidity Constraints: Theory and Evidence, *Economic Letters,* CXVIII (2013), 101-103.

Stephens, Mel, “3rd of tha month: Do Social Security Recipients Smooth Consumption Between Checks?” *American Economic Review*, XCIII (2003), 406-422.

Straznicka, Katerina, “Temporal Stability of Risk Preference Measures,” mimeo, 2012.

Tanaka, Tomomi, Colin F. Camerer, and Quang Nguyen, “Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam,” *American Economic Review,* C (2010), 557-571.

Toplak, Maggie E., Richard F. West, and Keith E. Stanovich, The Cognitive Reflection Test As A Predictor of Performance on Heuristics-and-Biases Tasks, *Memory and Cognition*, XXXIX (2011), 1275–1289.

Tax Me if You Can

Evidence on Firm Misreporting Behavior and Evasion Substitution*

PRELIMINARY, COMMENTS WELCOME

Paul Carrillo† Dina Pomeranz‡ Monica Singhal§

October 2013

Abstract

This paper examines the multitasking problem in the context of tax enforcement. A recent literature has argued that the ability of the tax authority to verify taxpayer reports is crucial to reducing evasion and building fiscal capacity. However, the effectiveness of such policies may be limited if taxpayers can evade along multiple margins with differing degrees of verifiability. We examine these issues in the context of firm tax compliance in Ecuador. We present a simple conceptual framework to examine evasion substitution. We then test the predictions of the framework by exploiting a natural experiment in which the tax authority notified selected firms of discrepancies between their declared revenues on previously filed corporate income tax returns and revenues estimated from third-party information. We find that firms revised their returns to increase declared revenues to match third-party reports but also increased costs by almost the same amount, resulting in only minor increases in total tax collection.

JEL codes: H25, H26, O23

*We thank Raj Chetty, John Friedman, Edward Glaeser, James Hines, Lawrence Katz, Wojciech Kopczuk, Michael Kremer, Asim Kwaja, Ben Olken, Andrei Shleifer, Joel Slemrod, and participants at the conferences of the American Economics Association, the National Tax Association, and at Harvard research seminars for helpful comments and discussions. We are exceedingly grateful to the Ecuadorian Tax Authority for outstanding collaboration and to the Harvard Weatherhead Center for research support.

†George Washington University, pcarrill@gwu.edu.

‡Harvard University and NBER, dpomeranz@hbs.edu.

§Harvard University and NBER, monica_singhal@harvard.edu

1 Introduction

A challenge faced by principals in monitoring agents is that increasing incentives or enforcement on one margin of behavior may cause agents to substitute adverse behavior to other margins. This classic multitasking problem (Holmstrom and Milgrom, 1991) is relevant not only in the context of effort but also when thinking about corruption and crime. For example, increased monitoring of one type of illegal activity may lead individuals to substitute their criminal efforts elsewhere.

We consider the multitasking problem in the context of tax enforcement. The canonical model of tax compliance, the Allingham-Sandmo model (Allingham and Sandmo, 1972), sets up a Beckerian framework in which taxpayers trade off the benefits of tax evasion against the probability of being audited and facing a penalty. In practice, however, audits are often costly and may be subject to corruption. In addition, it is difficult to reconcile low observed audit rates and penalties with the generally low levels of tax evasion observed in developed countries.

A growing literature therefore argues that understanding information flows is central to effective taxation. When governments imperfectly observe transactions, important differences emerge between forms of taxation that are equivalent in standard models of taxation but that differ in the information they generate for the government (Slemrod, 2008). In particular, several recent studies have focused on the importance of third-party information: the ability of the government to verify the reports of the taxpayer against other sources, such as an employer report of salary or the reports of a firms transaction partners (e.g., Kleven et al., 2010; Pomeranz, 2011; Kumler et al., 2013).

If this type of information verification is a primary form of tax enforcement, the multitasking problem is immediately relevant. Since third-party information is almost always partial in nature, third-party cross-checks may not be effective if taxpayers can evade along multiple margins with differing degrees of verifiability. In this case, taxpayers may simply substitute their evasion behavior away from the monitored margins towards less verifiable line items or transactions. Understanding this type of "evasion substitution" has important implications for the design of optimal audit and enforcement strategies, particularly for developing countries attempting to improve state capacity and reduce distortions from tax evasion.

In this paper, we set up a conceptual framework to examine the effects of third-party

reporting on firm tax evasion and exploit a natural policy experiment in Ecuador to test the predictions of the model. In our framework, which builds on Kleven et al. (2009), firms can affect reported profits (and therefore tax liability) by misreporting revenues and by misreporting costs. We show that if the audit probability is a decreasing function of the reported profit rate, firms will have an incentive to "look small," under-reporting revenues and potentially under-reporting costs. We show that the introduction of third-party reporting on revenues creates a floor on reported revenues and will cause firms to partially offset by increasing reported costs.

To test these theoretical predictions, we use a natural policy experiment in which the Ecuadorian Tax Authority (Servicio de Rentas Internas, SRI) notified a sample of almost 8,000 firms about discrepancies detected on their reported revenues for previously filed corporate income tax returns from the years 2008-2010. Three rounds of notifications were sent. In the second and third rounds, information about the exact revenue discrepancy was included. Firms were encouraged to revise and re-file these returns. Since the intervention involves previously filed returns, the observed responses must be reporting responses rather than "real" behavioral responses. Detailed line item and transaction level data allow us to examine both the direct effects of the notifications as well as possible substitution of evasion to other margins.

We demonstrate three main sets of empirical findings. First, evidence on revenue and cost discrepancies from the full universe of economically active firms provides support for the assumptions of the model. We observe bunching of reported revenues and costs at the third-party reported amounts. In addition, a significant share of firms (about 20%) report costs that are *lower* than third-party reported costs. This type of cost under-reporting would not be predicted under many alternative models of firm evasion and has implications beyond the scope of this paper. For example, if firms have incentives to under-report costs, self-enforcement of the VAT may be undermined.

Second, when firms are provided with a tax authority estimate of revenues, revised revenues bunch sharply at this amount, with over 35% of amending firms matching the SRI estimate exactly. In the first round, when firms are told that their reported revenues are lower than the SRI estimate but are not given a specific revenue estimate, we observe substantially less bunching. This provides strong evidence that the observed responses are misreporting rather than firms simply correcting errors and converging to a "true" estimate of revenue.

Finally, we find that firms substantially increase their reported costs to offset the increase in reported revenues. A large share of firms match their cost adjustment very closely to their revenue adjustment, even when revenue adjustments are quite large. As a result, increases in tax liability are a small fraction of what they would have been had firms only adjusted reported revenues.

Our paper contributes to the existing literature along several dimensions. Kleven, Kreiner and Saez (2009) present a model of third-party reporting in the context of taxes on wages and show that third-party reporting will be ineffective if workers can offset required increases in reported salary income by adjusting self-reported income. They then argue that third-party reporting can increase revenue if the tax authority targets individuals based on self-reported income or losses, common practices in developed countries. Our framework is complementary in that we show that under certain conditions, common in many developing countries, substitution effects may be very large. In addition, we extend their framework to consider tax evasion by firms, generating additional testable predictions on cost under-reporting and limits to enforcement.

We see two main contributions of our empirical results. First, to the best of our knowledge, this is the first study to examine responses to cross-checks of tax information directly, except for Alm et al. (2009), who study this issue in a lab setting and find that cheating increases as individuals earn larger shares of income that are not perfectly detectable. Several studies (e.g. Slemrod and Yitzhaki, 2002, Engstrom and Hesselius, 2007, Kleven et al., 2010; Pomeranz, 2011; Fellner et al., forthcoming) have considered responses to general deterrence messages or audit announcements. Kleven et al. (2010) and Pomeranz (2011) find indirect evidence for the relevance of cross-checks, in large-scale randomized studies in Denmark and Chile respectively, but do not examine behavioral responses to cross checks directly.

Second, our paper contributes to the very limited existing literature on evasion substitution. Klepper and Daniel (1989) use observational data from the Taxpayer Compliance and Monitoring Program (TCMP) and find evidence consistent with substitution. For example, firms that have large amounts of income that are not subject to information reporting have higher compliance on "inferior" evasion margins. However, these data are observational; there is no exogenous variation in the existence or degree of actual or perceived information reporting. Kopczuk (2012) finds that the introduction of a flat tax in Poland leads to increases not only of declared business revenues, but also business costs. And Yang (2008) shows empirical evidence of displacement effects in a different context:

when a customs reform in the Philippines increased enforcement on a particular method of duty avoidance, there was substantial displacement to alternative avoidance methods. Our rich data allow precise measures of substitution responses at the firm-category level.

The remainder of the paper proceeds as follows. Section 2 provides a brief conceptual background for thinking about evasion substitution. Section 3 describes the Ecuadorian tax system and empirical predictions. Section 4 discusses the data and methods. Section 5 presents the results and Section 6 concludes.

2 Conceptual Framework

In the following section, we develop a conceptual framework to examine the effects of third-party reporting on tax evasion. We begin with a brief review of Kleven, Kreiner and Saez (2009) – henceforth, KKS 2009 – who embed third-party reporting in the standard Allingham-Sandmo framework. We then apply this intuition to the case of tax evasion by firms.

2.1 Third-Party Reporting in the Allingham-Sandmo Framework

In Allingham and Sandmo (1972), the taxpayer has true income W and chooses her level of reported income, $\widehat{W}$. She pays taxes on her declared income at rate τ. If she is caught evading, which happens with a probability of detection p, she must pay the owed tax as well as a penalty. We follow Allingham and Sandmo (1972) and Yitzhaki (1974) in assuming that the penalty takes the form of an additional incremental penalty rate θ levied on the evaded tax. The taxpayer then maximizes her expected utility, with the first order condition of the following maximization problem implicitly defining $\widehat{W}^*$.

$$EU = (1-p)U(W - \tau\widehat{W}) + pU(W - \tau W - \theta\tau(W - \widehat{W}))$$

In KKS (2009), W is comprised of two components: W_T (income that is third-party reported) and W_S (income that is self-reported). The tax payer chooses the reported values of these income components: $\widehat{W_T}$ and $\widehat{W_S}$. The detection probability on third-party reported income is 1, so the taxpayer will set $\widehat{W_T} = W_T$.

However, if there is no constraint on $\widehat{W}_S$, and in particular if $\widehat{W}_S$ can be negative, third-party reporting will be irrelevant to tax collection. Since the taxpayer is optimizing over total $\widehat{W}$, $\widehat{W}^*$ remains unchanged. The taxpayer will fully substitute misreporting to $\widehat{W}_S$, or, in the terminology of KKS (2009), there will be full crowd out. They go on to show that this irrelevance result will fail to hold if self-reported losses are disallowed or if the audit rate depends on the level of self-reported income.

2.2 The Case of Firms

We now build on this intuition to examine the effect of third-party reporting on tax evasion in the firm context. We model firms as having revenues and costs and paying a flat tax on profits. Both revenues and costs are comprised of components that are third-party reported and components that are self-reported. We therefore have:

$$\begin{aligned}
R &= R_T + R_S = \text{revenues}, \ \widehat{R} = \text{reported revenues} \\
C &= C_T + C_S = \text{costs}, \ \widehat{C} = \text{reported costs} \\
\pi &= R - C = \text{profits}, \ \widehat{\pi} = \widehat{R} - \ \widehat{C} = \text{reported profits} \\
e &= (\pi - \widehat{\pi}) = \text{evaded profits} \\
Tax &= \tau\widehat{\pi}
\end{aligned}$$

This framework corresponds well to our empirical setting: the corporate tax in Ecuador is essentially a flat profit tax and the Ecuadorian tax authority has partial but not complete third-party information on both revenues and costs. The revenue discrepancy notifications can be thought of as a positive shock to R_T relative to firm priors. We return to a more detailed discussion of the predictions of the model in Section 3.3.

2.2.1 Asymmetry of Third-Party Reporting on Revenues and Costs

We first note that there is an inherent asymmetry in the effects of third party reporting on revenues and costs. Third-party reporting creates a "floor" on both of these variables: the tax authority knows that the firm has at least R_T in revenues and at least C_T in costs. For revenues, this means that $\widehat{R} \geq R_T$. But this is not the case for costs: if the firm reports $\widehat{C} \leq C_T$, the government knows that these costs are legitimate. However,

if it reports $\widehat{C} > C_T$, these costs may or may not be legitimate. Therefore, third-party reporting creates the possibility of detecting under-reporting of revenue but does not allow detection of over-reporting of costs unless 100% of costs are third-party reported. As we discuss below, certain auditing rules create incentives for firms to *under*-report costs, in which case third-party reporting of costs does become relevant.

2.2.2 Evasion Substitution

We now examine the taxpayer decision under alternative assumptions about the information available to the tax authority. We model firms as risk averse, which allows interior solutions for $\widehat{\pi}$. This is a realistic assumption for Ecuador, in which a large share of firms are sole proprietorships or owned by a single family and corresponds more generally to a context in which firms dislike volatility on profits. We also assume throughout that $\pi \geq 0$ and $\widehat{\pi} \geq 0$. In Ecuador, as in many countries, declaring negative profits may trigger additional scrutiny. We can therefore think of firms as being able to report negative profits only if they report truthfully.

Case 1: The tax authority has *no* information about the firm or about the true distributions of any of the underlying variables.

In this case, the only limiting factor on evasion is the possibility of being randomly audited and facing a penalty. Audits must be random (p is a scalar) because the tax authority has no knowledge on which to base any targeted audits.

The firm optimization problem is then as follows:

$$EU = (1-p)U(\pi - \tau\widehat{\pi}) + pU(\pi - \tau\pi - \theta\tau(\pi - \widehat{\pi}))$$

The firm chooses $\widehat{\pi} = \widehat{R} - \widehat{C}$, and $\widehat{\pi}^*$ is the optimal choice of reported profits. Suppose that we introduce third-party reporting on revenues, so that firms must now set $\widehat{R} \geq R_T$. It is immediately clear that third-party reporting will not result in a decrease in evasion or an increase in tax revenue. Firms optimize over $\widehat{\pi}$, which is only a function of the difference between reported revenues and reported costs. Therefore, even if 100% of revenues are third-party reported, firms can fully offset by adjusting costs in order to return to their initial optimal $\widehat{\pi}^*$.

In KKS (2009), the "irrelevance" of third-party reporting can be overcome if self-

reported losses are disallowed. This is a reasonable policy if true self-reported losses are rare. However, this type of policy is not easily applicable to the case of firms, since all firms will have true costs. Irrelevance can also be overcome if the audit rate is a function of self-reported income. We consider the effect of targeted audit rules below.

Note that in Case 1, the levels of reported revenues and reported costs are indeterminate: even given the constraint that $\widehat{R} \geq R_T$, there are an infinite number of $(\widehat{R}, \widehat{C})$ pairs that will produce the same $\widehat{\pi}$. This is consistent with the motivating assumption that the tax authority can only do random audits.[1] Kleven, Kreiner and Saez (2009) formalize an alternative, in which the probability of detection is increasing in the level of evasion. Specifically, the audit function is specified as $p(e)$ where $p' > 0$. This is obviously a reduced form approximation, since the detection probability cannot be based directly on the (unobserved) level of evasion. Underlying the $p(e)$ function, then, is the idea that the tax authority has some direct information about the true levels of the relevant firm variables or signals of this information. In the following two cases, we consider specific types of information that may be available to the tax authority and effects on the firm optimization decision.

Case 2: The tax authority has no "external" information about the firm. There is no third-party reporting and it cannot observe any other characteristics of the firm (e.g., how big the factory is). However, the tax authority can observe whether the firm self-reports seem "internally consistent" with each other. The detection probability can therefore be a function of firm self-reported variables.

Specifically, we now assume that the tax authority has some information about the true distribution of profit rates. For example, if a firm reports \$100 in profits on \$1000 in revenue, that may be plausible, but if the firm reports \$100 in profits on \$1000000 in revenue, that is more suspicious. We assume that the audit probability is $p(\frac{\widehat{\pi}+\varepsilon}{\widehat{R}})$ where $p' < 0$ and ε is a small number greater than zero. The addition of ε is simply to differentiate among cases where $\widehat{\pi} = 0$, so that declaring zero profits on a large amount of revenue is more likely to trigger an audit than declaring zero profits on a small amount of revenue. This specification of the audit rule is plausible for our empirical context. In a number of field interviews, tax authority staff indicated that the reported profit rate

[1] Note that the indeterminacy could be broken if penalties are a function of misreporting on individual variables rather than on evaded tax. This would require the tax authority to care about honest reporting above and beyond the effect on overall tax under-reporting. This might occur if, for example, a firm's own reports generate information about other firms. We do not model inter-firm spillover effects of enforcement here.

is one of the characteristics they consider when determining whether to audit, and firms and accountants reported that they believe that the tax authority pays close attention to this variable.

The firm optimization problem then becomes:

$$EU = (1 - p(\frac{\widehat{\pi} + \varepsilon}{\widehat{R}}))U(\pi - \tau\widehat{\pi}) + p(\frac{\widehat{\pi} + \varepsilon}{\widehat{R}})U(\pi - \tau\pi - \theta\tau(\pi - \widehat{\pi}))$$

Proposition 1. *Since $p' < 0$, firms will choose the lowest level of declared revenue consistent with their declared profits. Specifically, as $\varepsilon \to 0$, $\widehat{R}^* = \widehat{\pi}^*$. This then implies that $\widehat{C}^* = 0$.*

Proof. At any given level of $\widehat{\pi}$, the firm always prefers to minimize p. Since $\widehat{R} \geq \widehat{\pi}$ and $p' < 0$, the firm choice that minimizes p is $\widehat{R} = \widehat{\pi}$, which implies that $p = p(1)$. Define $\widehat{\pi}^*(p(1))$ as the optimal choice of reported profits at this audit probability, and call this $\widehat{\pi}^{**}$. Now note that the firm can change $\widehat{R}$ while keeping $\widehat{\pi}$ fixed (by adjusting $\widehat{C}$), and this does not affect expected utility at a given p. This defines $\widehat{R}^{**} = \widehat{\pi}^{**}$. Intuitively, the firm will maximize its reported profit rate to 100% in order to minimize its audit probability and chooses the optimal level of reported profits given this audit rate. The optimal choice of $\widehat{C}$ is then zero. □

This result arises from two assumptions. First, we have assumed that the tax authority has no direct information about true revenues or costs, an assumption which we relax in Case 3 below. Second, we have assumed that the p function is monotonic, so that the tax authority is always less likely to audit as the reported profit rate increases.[2] Note that in Case 2, firms will do *all* evasion on the revenue margin and will not even report their legitimate costs. This is because they can get to their desired level of profits by adjusting either reported revenues or reported costs, but under-reporting revenues gives the added benefit of reducing the audit rate.

Case 3: The tax authority follows the audit rule above but also has third-party information on revenues. We model this as in KKS (2009) by assuming that the detection

[2]This assumption could be modified so that a reported profit rate that is "too high" also appears suspicious. Since our main focus is on the effects of third-party reporting, which creates a *lower* bound on reported revenues, we retain the assumption of a monotonic p function for simplicity.

probability on third-party reported revenue (R_T) is 1.[3]

Proposition 2. *Define $\widehat{R}^*$ as the optimal firm choice in the absence of third party reporting. If $R_T \leq \widehat{R}^*$, third-party reporting will have no effect. If $R_T > \widehat{R}^*$, then $\widehat{R}^{*\prime} = R_T$.*

Proof. Since the detection probability on third-party reported revenue is 1, the firm must report $\widehat{R}' \geq R_T$. Suppose that the firm sets some $\widehat{R}' > R_T$ and then maximizes to get an optimal level of reported profits, $\widehat{\pi}^{*\prime}$, which implies an optimal level of reported costs, $\widehat{C}^{*\prime}$. Now suppose δ is a small constant greater than zero, and the firm instead chooses $\widehat{R}'' = \widehat{R}' - \delta$ and $\widehat{C}'' = \widehat{C}^{*\prime} + \delta$. The new level of reported profits is $\widehat{R}'' - \widehat{C}''$, which is equal to $\widehat{\pi}^{*\prime}$. However, the audit probability of the firm decreases and expected utility therefore increases. Thus, $\widehat{R}^{*\prime} = R_T$ □

This implies that $\widehat{C}^{*\prime} = R_T - \widehat{\pi}^{*\prime}$, but note that $\widehat{C}^{*\prime} \lesseqgtr C$. Firms will adjust their reported costs, but the new level of reported costs may be larger or smaller than their true costs. Intuitively, the audit rule creates incentives for firms to "look smaller" than they are, and the restriction that $\widehat{\pi}$ must be positive means that firms may under-report costs. Our field interviews suggest that firms do in some cases leave real costs "off the table." We return to direct empirical tests for cost under-reporting in Section 5.

We can now examine the response of $\widehat{\pi}$ to R_T. Define $Y_N \equiv \pi - \tau\widehat{\pi}$ (after tax profits in the non-audited state) and $Y_A \equiv \pi - \tau\pi - \theta\tau(\pi - \widehat{\pi})$ (after tax profits in the audited state).

Taking the first order condition with respect to $\widehat{\pi}$ we have:

$$p(\frac{\widehat{\pi} + \varepsilon}{R_T})U'(Y_A)\theta\tau - (1 - p(\frac{\widehat{\pi} + \varepsilon}{R_T}))U'(Y_N)\tau - \frac{1}{R_T}p'(\frac{\widehat{\pi} + \varepsilon}{R_T})(U(Y_N) - U(Y_A))$$

The first and second terms capture the "standard" Allingham-Sandmo tradeoff: higher evasion results in higher utility in the non-audited state but lower utility in the

[3]In practice, the tax authority may also observe noisy signals about true revenues. For example, it might observe the size of the firm factory. In this case, we could think of the probability of detection increasing in the gap between estimated and reported revenues. Third-party reporting as described above would correspond to a specific case in which the signal of the minimum level of true revenue is precise. For simplicity and to best match our empirical context, this is the case we consider here.

audited state. The third term captures the fact that higher evasion increases the probability of being audited. We would expect close to full substitution ($\frac{\partial \widehat{\pi}}{\partial R_T} \approx 0$) as $p' \to 0$. Intuitively, this corresponds to the audit probability function being fairly flat with respect to the reported profit rate. This might, for example, be reasonable if firms could legitimately have a wide range of profit rates, so that the variance of the $\frac{\pi}{R}$ distribution is high.

2.2.3 Third-Party Reporting of Costs

We have so far considered third-party reporting of revenues. As demonstrated, requiring firms to report revenues at or above the third-party threshold will result in increased tax collection, even if the magnitude of this increase is small due to offsetting cost adjustments. What may be less intuitive is the effect of third-party reporting for costs. We have seen that in the absence of third-party revenue information, firms will report zero costs. With third party revenue information, firms may still choose to report costs that are lower than their true costs. Suppose that the firm reports $\widehat{C} < C_T$, so that the tax authority knows that the firm's costs are under-reported. It is possible to show that in some circumstances, it may actually be desirable for the tax authority to require the firm to report $\widehat{C} \geq C_T$, even though this means that the firm will increase its reported costs. This is because forcing firms to increase their reported costs causes both their evasion level and audit probability to increase. Firms will therefore have an incentive to increase reported revenues.

2.2.4 Adding Limits to Enforcement

So far, we have focused on how information affects the probability that the tax authority can detect misreporting. However, the taxpayer should also care about the enforcement probability: the chance that the tax authority takes action to collect the owed tax. In practice, the tax authority may face constraints on its ability to enforce tax collection, *even conditional on having full information* (Aparicio et al., 2011; Gadenne and Singhal, 2013). We model the possibility of limits to enforcement by introducing q, the probability that the tax authority takes action to collect the tax owed. The relevant risk for the taxpayer from evading is then captured by pq. We have so far assumed $q = 1$; we now consider the case where $q < 1$. Note that when $q < 1$, we cannot simply rewrite a detection probability of one on third-party reported revenues as a constraint: the taxpayer could

choose $\widehat{R} < R_T$ if she believes that the probability of enforcement is low. However, if the taxpayer chooses to do this, the tax authority knows with certainty that she is evading (if the taxpayer is not underreporting profits, there is no incentive to misreport revenues). We model this as the taxpayer being subject to a higher audit probability if reported revenues are below third-party revenues. Specifically, the taxpayer maximizes:

$$EU = (1 - qp(\frac{\widehat{\pi} + \varepsilon}{\widehat{R}}))U(\pi - \tau\widehat{\pi}) + qp(\frac{\widehat{\pi} + \varepsilon}{\widehat{R}})U(\pi - \tau\pi - \theta\tau(\pi - \widehat{\pi}))$$

where

$$p = p(\frac{\widehat{\pi} + \varepsilon}{\widehat{R}}) \text{ if } \widehat{R} \geq R_T$$
$$p = p(\frac{\widehat{\pi} + \varepsilon}{\widehat{R}}) + D \text{ if } \widehat{R} < R_T$$

We first note that, as in Case 3, will never be optimal for firms to choose $\widehat{R} > R_T$ unless that was optimal in the absence of third-party reporting.

In this framework, firms may optimally choose a level of reported revenues that is below R_T, but they will never choose an amount that is "close" to R_T. The logic is as follows: if the firm is just below R_T, it can instead report R_T. There is a loss to the firm from doing this in the form of having to pay higher tax but a benefit because the audit probability decreases *discontinuously* by D. Therefore, there will be a range of reported revenues over which firms will not locate, instead bunching at R_T.

We can think of the policy notifications as changing R_T relative to firm priors. Specifically, we model the notifications as an effective increase in R_T. Since the notifications reference previous tax returns, firms can only respond by filing an amendment. If filing an amendment is costless, firms will reoptimize given the new level of R_T. However, if there is a fixed cost to filing an amendment, firms will compare their expected utility if they reoptimize given the new level of R_T to the expected utility of not responding to the notification at all. They will file an amendment only if the difference in the expected utility exceeds the fixed cost. In practice, these fixed costs might include finding old records or hiring an accountant.

3 Background and Empirical Predictions

3.1 Firm Taxation and Third-Party Information in Ecuador

3.1.1 Rates and Reporting Requirements

We now turn to our empirical setting. We examine taxpayer behavior in the context of the corporate income tax in Ecuador. All incorporated firms in Ecuador are required to file an annual profit tax return (Form F101). Firms calculate pre-tax profit as the difference between total revenues and total costs. They must distribute 15% of pre-tax profits among their employees and are then taxed at a flat rate of 25% on the remainder. The 25% rate is independent of firm size and had not changed in over 20 years up to and including the years that were affected by the intervention in this study.[4] The Ecuadorian fiscal year corresponds to the calendar year and firms file the annual corporate tax return (F101) the following April.

In addition, all firms are required to file a monthly value added tax (VAT) return (Form F104). In order to deduct input costs, this form must include a purchase annex which lists the amount purchased from each supplier along with the supplier tax ID.[5] A similar annex for sales to client firms has to be submitted by large firms with annual sales above $200,000 as well as Large Taxpaying Units, public sector firms, financial institutions, credit card companies, and firms requesting refunds.

3.1.2 Third-Party Information and Cross Checks

The SRI has several sources of information against which to cross check firm self-reports. Data from the sales and purchase annexes described above can be used to verify the accuracy of sales and purchases reported by firms' trading partners. For example, SRI can compare a firm's own reported sales to the sum of all purchases reported *from* that firm by other firms. SRI supplements this information with sales numbers from credit cards (credit card companies report total sales per month registered for a firm through credit card payments), exports and imports recorded by the Ecuadorian Customs, and

[4] Corporate tax rates were reduced to 24%, 23% and 22% in fiscal years 2011, 2012 and 2013, respectively.

[5] Firms withhold a percentage of their payments to suppliers, which they transfer to the tax authority. This withholding can be used by suppliers as a credit against their tax liability.

returns to financial investments recorded by financial institutions.

The ability of the SRI to perform large scale cross-checks is quite recent. For example, the annex data has only been collected since 2007 and was not initially used in any systematic way. Discrepancies between self-reported and third-party reported data were computed, but only for a relatively small number of firms (for example, in the process of auditing a large company). SRI started conducting large scale cross-checks of taxpayers in Ecuador in 2011.

3.2 Policy Intervention

Our results are based on a series of natural policy experiments in which the SRI notified selected firms about detected discrepancies between third-party calculated revenue and firms' self-reported revenues in previous corporate income tax returns. Starting in 2011, SRI staff computed revenue discrepancies for previously submitted returns as well as the potential effects of adjustments on tax revenues. Based on this information, SRI sent notifications of revenue discrepancies to selected firms. While the specific selection methodology to determine which firms would be notified is confidential to SRI, key factors included the magnitude of discrepancies and potential revenue adjustment.

We examine three rounds of notifications corresponding to detected discrepancies on 2008, 2009, and 2010 tax returns. We refer to these as the 2008, 2009, and 2010 rounds respectively. The notifications corresponding to the 2008 returns were sent in the summer of 2011; the notifications corresponding to the 2009 and 2010 returns were sent in the spring of 2012. Note that in all cases, notifications were sent for previously filed tax returns after all real behavioral decisions for the relevant returns had been made. Therefore, any changes we observe in response to the notifications are reporting rather than real responses. Firms are required to file the tax return online; notifications were sent by email to the address on record, which typically belongs to the general manager or accountant of the firm.

For the 2008 round, 3,136 firms were selected for notification. The email to firms in this group informed them that in a cross-check with third-party information, a discrepancy in their reported revenue for 2008 was found. Firms were asked to submit an amendment.

The original email (in Spanish) is available in the appendix and the relevant portion of the message is translated below:

"Dear Taxpayer [XXX]

After a careful data cross-check, the Tax Administration has estimated that the revenue that firm [XXX] received during the 2008 fiscal year is larger than the total revenue reported on its 2008 corporate income tax return ... Please submit an amendment within the next 10 business days."

The following year, the SRI selected 2,221 firms for the 2009 round and 2,636 firms for the 2010 round. The most important difference is that in these rounds, the notifications included the SRI estimate of the firm revenue based on third-party sources. The relevant portion of the message is translated below:

"Dear Taxpayer [XXX],

After a careful data cross-check, the Tax Administration has estimated that the total revenue that firm [XXX] received during the 20XX fiscal year is larger than the total revenue reported on its 20XX corporate income tax return:

Fiscal Year	Corporate Income Tax Line Item	Total Revenue Computed by SRI	Total Revenue Declared by Taxpayer
2010	699. Total Revenue	$255,300	$190,500

The Tax Administration urges you to submit an amendment."

We observe firms' initial filings as well as any subsequent amendments made to these returns. Note that firms are only legally required to respond to an SRI request when a written notice is delivered to them in person. This requires delivery by SRI staff and is thus very expensive. The email option was chosen due to resource constraints, although it was known that firms were not legally required to submit an amendment.[6]

3.3 Empirical Predictions

In our empirical analysis, we consider three sets of firms: the full sample, the notification sample, and the amended sample. The full sample consists of economically active formal firms, which we define (year by year) to be firms who filed a F101 and reported non-

[6]The SRI also conducted legally binding notifications. In these cases, if the firm did not submit an amendment, the SRI estimate of revenue was taken as correct and the firm was required to pay the resulting increase in tax liability. Anecdotal evidence suggests that similar evasion substitution as discussed below was also found among the sample of firms treated with legally binding letter notifications. We are in the process of obtaining data on these notifications.

zero revenues or costs or for whom there is non-zero third-party revenue information.[7] The notification sample consists of firms selected by SRI to be notified about a revenue discrepancy. The amended sample consists of firms that submitted an amended return after receiving the notification.

The model predicts several patterns for revenue and cost discrepancies in the cross-section (full sample) of firms. If firms have full information about R_T, we would still expect to see some firms with $\widehat{R} > R_T$ (firms for whom the constraint is non-binding). $\widehat{R} < R_T$ would occur only if there are limits to enforcement. If firms do not have full information about R_T, this would result in additional deviations of $\widehat{R}$ from R_T, potentially in both directions. The model would also predict bunching at R_T.

The model also predicts that some firms may under-report costs, setting $\widehat{C} < C$. While we do not observe C, we can perform a stricter test for cost under-reporting by examining whether firms report $\widehat{C} < C_T$. This should provide a lower bound on the extent of cost under-reporting, since only a small share of firms in Ecuador are required to file sales annexes. If third party reporting is partial, $C_T < C$.[8]

In the context of the framework, we can think of the 2009 and 2010 notifications as updating firms about R_T. All firms in the notification sample have $\widehat{R} < R_T$ by definition, since SRI only sent notifications to firms that were under-reporting revenues relative to third-party estimates. If there are limits to enforcement and some costs of filing an amended return, we would expect to see some firms choose not to file an amendment. Among firms that do file an amendment, we should see bunching at R_T, as discussed in Section 2. We also expect to see firms offsetting the increase in reported revenue with an increase in reported costs. If the gradient of the audit probability with respect to the profit rate is fairly flat, as seems plausible for the case of Ecuador, these substitution effects will be large.

[7] It is possible that a firm is engaging in economic activity but reports these variables as zero and lacks any third-party information. We consider these firms to be essentially informal. In 2008, we have third-party information for the notification sample but not the full sample. We therefore define economically active firms for this year as firms with non-zero F101 revenues or costs. 77% of firms in the economically active sample in 2009-10 meet this criterion.

[8] Note that it could be the case that partner firms are misreporting their sales, on which the third-party cost variable is based. However, since the incentive is for firms to under-report revenues, this type of misreporting will typically result in a *lower* C_T.

4 Data and Methodology

To analyze firms' responses to the notifications, we combine several data sources from SRI. Information about declared revenues and costs for all original and amended tax returns is collected from the corporate income tax form F101. We observe values for all line items as well as the submission date for each version of the return.[9] This allows us to calculate the change in revenue and costs declared in the F101 between the pre- and post-intervention periods.

Revenue and cost discrepancies are calculated using third-party data. Specifically, third-party revenue is the sum of exports, bank interest, and the maximum of purchases reported by other firms and credit card purchases. We take the maximum because we do not have transaction level credit card purchases and therefore cannot determine the overlap between the final two categories. Third-party costs are the sum of imports and sales reported by other firms. Discrepancies are calculated as the firm report minus the third-party report, so that a positive revenue or cost discrepancy indicates a firm self-report that is higher than the third-party estimate.

Table 1 shows summary statistics for the full sample of economically active firms for fiscal years 2008-2010 (pooled). The mean declared annual revenue is $1.48 million, with a median of $39,000. Declared costs have a mean of $1.38 million and a median of $40,000. Correspondingly, there is a large range in tax liabilities. The mean is $22,000; however, the median is practically zero, and the standard deviation is over $600,000. There are large discrepancies between firms' own declared revenues and costs and the information the tax authority can estimate from third parties. The mean and median revenue discrepancies are positive, indicating that most firms declare more revenue than what the tax authority could obtain from third-party sources. In our framework, this is consistent with the third-party constraint not binding for some firms or with firms overestimating the extent of third-party reporting. In addition, it may be that the tax authority has signals of true revenue beyond the third-party reports.

The mean and median of cost discrepancies are also positive, indicating that most firms declare higher costs than what is available from third parties. This is less surprising, since information on costs is incomplete, and firms in general have an interest to report high costs. However, as we will discuss below, a substantial share of firms also report

[9] Firms can make amendments to F101 without presenting any additional documentation to SRI.

lower costs than the third-party estimate.

[Table 1 about here.]

Summary statistics for the policy experiment samples are shown in Table 2. The table displays information for the firms that SRI selected in each of the three notification rounds (corresponding to firms' 2008, 2009, and 2010 returns), as well as for firms that amended their returns after receiving the email. We include any amendments filed within a three month window after the notification.[10]

[Table 2 about here.]

The share of firms responding to the notifications by amending their returns was 19% in round 1 (for 2008 returns), 11% in round 2 (for 2009 returns) and 16% in round 3 (for 2010 returns). This measure is an underestimate of the actual response rate, since about 7% of messages bounced due to invalid email addresses, and some share of "non-bounced" emails may have been delivered to old addresses or may have been left unread. In addition, as discussed above, firms are not legally obligated to respond to the notifications. If there are limits to enforcement and costs of filing an amendment, we would expect to see a substantial share of firms choosing not to amend. The firms that revised their returns (amending firms) tend to be slightly smaller in terms of economic activity than the full notification sample. Note that in this sample, revenue discrepancies are negative since all firms that were notified reported less revenue than the SRI third-party estimate. Revenue discrepancies are large: the average report was lower than the third-party estimate by $307,000, $176,000 and $197,000 in 2008, 2009 and 2010 respectively. The revenue discrepancies for the amended sample were slightly smaller, consistent with their lower overall economic activity.

Given these large discrepancy amounts, revisions in response to the notifications have potentially large effects on firm tax liability and collected tax revenue. For example, if

[10] In 2009 and 2010, we observe the firm specific notification date. In 2008, as we discuss below, we impute the notification start date and we do not have firm specific notifications dates. We assume that the notifications in 2008 were made over a one month period following the start date, as in 2009 and 2010, and therefore consider amendments filed within a four month window of the start date in 2008 to be as consistent as possible with the later rounds. In practice, over half of the firms that amended their return responded in less than one month and our results are robust to the choice of the post notification window.

all firms in the SRI sample were to amend their 2009 and 2010 returns to match figures SRI provided in the notifications, reported aggregate pre-tax revenue could potentially increase by about $391 and $522 million for the 2009 and 2010 fiscal years, respectively. As a back-of-the-envelope calculation, this would increase total firms' tax liability by about over $194 million. Among the group of firms that amended their returns, it was expected that total tax liability would increase $39 million in 2009 and $76 million in 2010 if firms only adjusted revenues to match SRI-estimated revenues. As we show below, actual tax revenue increases were substantially smaller as a result of firms compensating for their revenue adjustments with offsetting cost adjustments.

[Figure 1 about here.]

Figure 1 plots amendment rates both for the notified sample (right hand Column) and for the rest of the economically active firms in the country (left hand Column) relative to the start of the notification period (indicated by zero on the x-axis). In the notified sample, we see a sharp increase in the amendment rate just following the commencement of the notifications. There is no such increase in the full sample. For the 2009 round, the amendment rate is essentially zero prior to the notifications. The pre-notification amendment rate is somewhat higher for the 2010 round, which is not surprising: the notifications are closer to the time of initial filing for this round, so the natural rate of amendment is higher than for previous years. However, we see no difference in the amendment rate between the notification and full samples until the notification start date.

For 2008, we do not know the notification start date. However, we can impute the date based on the timing of the stark spike in the amendment rate for the notification sample relative to what we observe in 2009 and 2010. Based on this discontinuity, we impute the start date for the 2008 round as August 11, 2011. In the subsequent analysis, we use the firm-specific notification dates for the notifications about the 2009 and 2010 tax declarations, and August 11, 2011 as the start date for notifications about 2008. The pre-notification values are defined by the last observed filing or amendment prior to these dates. The post-notification values are defined by the first observed amendment (if any) after the notification during the windows discussed above.

To examine the response to the notification, we analyze firm revisions to the relevant tax returns. For this analysis, we do not explicitly use a control group. Instead, we use

the pre-amendment value as the counterfactual to calculate the treatment effect of the notification. This assumption is based on the fact that, as can be seen in Figure 1, the probability that non-notified firms spontaneously revise their returns for previous years after such a long period is quite low. For example, among the almost 100,000 firms that submitted a 2008 tax return and did not receive an email notification from SRI, only 222 (0.3%) amended their 2008 tax return between September 2011 and August 2012. Correspondingly, the comparison between the pre-notification and the amended return of firms that received a notification can provide a causal estimate of the detected discrepancy notification on firms' reporting decisions. In ongoing work, we are planning to confirm the validity of this identifying assumption by (a) comparing firms just above and below the notification threshold; (b) examining the types of amendments made by non-notified firms; and (c) comparing pre- and post- notification amendments for the notification sample.

The subsample that responds by revising their return is obviously self-selected. These firms may, for example, be more risk averse or may find it easier to make offsetting adjustments on their returns than the average firm in the notification sample. The magnitude of responses of these firms may thus differ from a counterfactual in which all firms in the notification groups were required to submit revised returns. Nevertheless, examining responses of this subgroup is still informative. First, it provides a valid test of the existence of evasion substitution. Second, as we show below, the specific ways in which firms made revisions provide strong evidence that observed adjustments are in fact changes in evasion rather than legitimate adjustments to previous reports.

5 Results

5.1 Evidence from the Full Sample of Firms

5.1.1 Revenue Discrepancies

Before analyzing the impacts of the discrepancy notifications, we examine the pattern of revenue and cost discrepancies in the full sample of active firms in 2009 and 2010, the two years for which we have data on third-party reported revenues and costs for all firms. Figure 2 Panel A plots the discrepancy between the logs of firms' declared revenue and of the revenue estimate based on information from third parties. We observe both cases in which third-party reported revenue is higher and cases where it is lower than the

self-reported amount. We also see a notable spike in the distribution at zero.

[Figure 2 about here.]

5.1.2 Cost Discrepancies

Figure 2 Panel B shows the discrepancies between the logs of firms' declared costs and of the cost estimate based on third-party information. As discussed above, one of the predictions of the model is that firms may report $\widehat{C} < C$. A strong test of this prediction is to look at whether firms ever report $\widehat{C} < C_T$, even though all else equal, firms pay higher taxes when declaring lower costs. We find that about 20% of do in fact declare lower costs than the third-party estimate. This finding would not be predicted by many alternative models of firm tax evasion but is consistent with our conceptual framework. More generally, this finding has important implications for effectiveness of the "self-enforcing" properties of the VAT. Self-enforcement is premised on the idea that transacting partners have strictly conflicting incentives: sellers should want to under-report the value of the transaction while buyers should wish to over-report it in order to maximize their cost deduction. However, if firms have incentives to under-report true costs, there may be opportunities for collusion and evasion in the VAT chain, particularly for firms for whom the third-party reported information on revenues is significantly below true revenue.

After having described discrepancies in the full sample, we next turn to the analysis of how the notified firms responded to the treatment.

5.2 Response to Notifications

5.2.1 Revenue Adjustments

In this section, we first examine how firms' reported revenues responded to the notifications. We then examine effects on costs and tax liabilities. We begin with the 2009 and 2010 rounds, in which firms were provided a specific value of the third-party reported revenue R_T. Recall that our conceptual framework predicts that among firms that amend their return in response to the notification, we should observe bunching at R_T. This is indeed what we observe in the data.

[Figure 3 about here.]

Panel A of Figure 3 shows the difference between the log of post-amendment revenue and third-party revenue for firms that file an amendment following receipt of the discrepancy notice. We observe a sharp spike around zero, indicating that firms are adjusting their revenues to match the provided estimate of R_T. 39% of firms in the 2009 round and 35% of firms in the 2010 round match exactly, setting $\widehat{R}' = R_T$.

[Figure 4 about here.]

Figure 4 Panel A plots the changes in declared revenue against the pre-treatment discrepancies. The figure shows clearly that the observed bunching around zero in Figure 3 is not limited to small firms: even firms with very large discrepancies matched the third-party amount closely in their amended returns for the years 2009 and 2010, in which they were informed about the precise amount of discrepancies.

Note that approximately 15% of amending firms filed an amendment but did not change reported revenues or any other major variables. One interpretation is that the fixed cost of filing a return with essentially no adjustments is very low. Under this interpretation, we can think of these firms as analogous to the non-amenders. Amenders that do adjust reported revenues tend to locate along the 45 degree line, indicating that they are matching R_T. The regression line and confidence interval confirm that there is a strong alignment of the amendment by firms that adjust their declared revenue and the third-party estimate.

We next turn to the 2008 round, in which firms were told that $R_T > \widehat{R}$, but were not given a specific estimate of R_T. The first graph in Panel A of Figure 3 shows revenue adjustments relative to R_T for 2008. As expected, we see substantially lower bunching around zero. We still see 6% of firms that match R_T exactly, suggesting that they are ultimately able to determine which components of revenue can be cross-checked. This is confirmed in Panel A of Figure 4: in the 2008 round, firms with larger revenue discrepancies tend to make larger revenue adjustments, but there is much higher variance than in 2009 or 2010.

5.2.2 Substitution Effects

We next show that even though firms were only notified about discrepancies in their revenues, they made substantial offsetting adjustments to reported costs. Panel B of

Figures 3 and 4 show that many firms closely match their change in reported revenue with a change in reported costs. Panel B of Figure 4 shows that this behavior is not restricted to small firms, but seems to hold along the entire distribution of changes in reported revenue, even when these revenue adjustments are very large. As a result, as displayed in Panel C of Figure 3, changes in final tax liability were close to zero for most firms.

Figure 5 shows the same results restricting the sample to amending firms that made some adjustment to their reported revenues. We see the same patterns even more sharply: firms tend to exactly match the third-party estimate of revenues in 2009 and 2010 but not in 2008, and firms essentially fully offset increases in declared revenues with increases in declared costs in all years.

[Figure 5 about here.]

Table 3 displays these findings in regression form. Panel A shows results for firms that submitted an amended declaration after receiving the email notification and Panel B for all firms that were intended to be notified. The coefficient on 'post' indicates the difference in declared amounts before the intervention compared to the end of the post-intervention period. All results are highly significant in both logs and levels.

[Table 3 about here.]

Panel A shows that amending firms on average increased their reported revenue by about $86,000 and their costs by about $80,000, resulting in an average change in tax liability of $1,900. The results in logs indicate about a 30-fold increase in revenue and costs, and about 7-fold increase in tax liability. So while the increase in dollar amounts is relatively small, the increase in percentage terms is not that trivial, reflecting the fact that some of the treated firms had very low pre-treatment tax declared. We see similar patterns in the full notification sample as well. These effects are literally attenuated by the amendment rate, since non-amender firms in the notification sample had zero adjustments by definition. However, even in this sample, the observed effects are highly statistically significant.

Overall, our results show both strong evidence of evasion substitution and misreporting by firms. While the notifications did cause firms to adjust their reported revenues to

match third-party estimates, they offset much of this adjustment in increases in reported costs. These results cannot be driven by real behavioral responses, since all decisions for the relevant tax returns were made well prior to the notifications. The pattern of behavioral responses provides strong evidence of misreporting by firms, both before and after the notifications.

6 Conclusion

This paper analyzes multitasking and evasion substitution in the context of tax enforcement. We present a simple conceptual framework to think about firm tax evasion decisions and demonstrate conditions under which an increase in monitoring ability on one margin through improved third-party reporting will lead to offsetting adjustments on other margins.

We first show that descriptive evidence from the universe of firms provides empirical support for our model. We observe bunching of reported revenues and costs at the third-party amount. In addition, we demonstrate that firms sometimes under-report rather than over-report costs, even though over-reporting should - all else equal - reduce tax liability. This finding is consistent with our framework but would not be predicted under many alternative models, and it has important implications beyond the context of this study. If firms have incentives to under-report costs, self-enforcement in the VAT chain may be undermined: reporting of input costs is what leads to the paper trail for suppliers sales, which is at the core of the self-enforcing mechanism in the VAT.

We next test for evasion substitution directly by exploiting a natural policy experiment in Ecuador in which the tax authority notified selected firms about large discrepancies between declared revenue and revenue calculated from third-party sources. Firms respond strongly to the notifications, amending their returns with higher reported revenues and often exactly matching the amount indicated by the tax authority. However, firms offset much of this increase through an increase in reported costs, leaving their profits and corresponding corporate tax liabilities nearly unchanged. This is true even when adjustments are in the tens or even hundreds of thousands of dollars. As a result, total tax collection from amending firms was an order of magnitude smaller than if firms had not made offsetting cost adjustments. Critically, the notifications referenced previously filed tax returns, so observed responses clearly indicate misreporting by firms rather than

real behavioral responses.

These results demonstrate the importance of taking possible substitution effects into account when examining the effectiveness of tax enforcement measures. They also indicate the limits of third-party reporting in increasing tax revenues when third-party information is partial. The collection of third-party information may have different levels of effectiveness depending on the amount of information already available on other margins: as in the O-ring theory of economic development (Kremer, 1993), the weakest link may play a preponderant role for tax collection.

Finally, we note that third-party reporting may still have beneficial effects even if firm-level evasion substitution is high if there are spillovers across margins and spillovers across firms. An increase in third-party reporting provides a low-cost way to monitor certain types of transactions. This could then allow the tax authority to focus its costly auditing activities on non-third-party reported margins, thereby decreasing the ability to evade on these weaker links. In addition, third-party information may increase compliance among a firm's trading partners. For example, as the web of third-party reported transactions increases, firms increasingly have less discretion in under-reporting revenue. In response, firms might increase their request of receipts from their suppliers, in order to be able to deduct more legitimate costs. This type of spillover effect could generate a paper trail for the revenue of their suppliers, consistent with the findings of Pomeranz (2011). Such spillover effects along the supply chain may be of particular importance if firms otherwise have incentives to under-report costs. Examining the implications of within-firm and cross-firm spillover effects in the design of optimal information collection and enforcement policies is an important avenue for future research.

References

Allingham, Michael G. and Agnar Sandmo, "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1972, *1* (3-4), 323–338.

Alm, James, John Deskins, and Michael McKee, "Do Individuals Comply on Income Not Reported by Their Employer?," *Public Finance Review*, 2009, *37* (2), 120–141.

Aparicio, Gabriela, Paul Carrillo, and M Shahe Emran, "Taxes, Prisons, and CFOs: The Effects of Increased Punishment on Corporate Tax Compliance in Ecuador," 2011. Working Paper. 2011-02.

Engstrom, Per and Patrik Hesselius, "The Information Theory: Theory and Application," 2007. Working Paper 17, Institute for Labour Market Policy Evaluation 2007.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler, "Testing Enforcement Strategies in the Field: Legal Threat, Moral Appeal and Social Information," *Journal of the European Economic Association, Forthcoming.*

Gadenne, Lucie and Monica Singhal, "Decentralization in Developing Economies," 2013.

Holmstrom, Bengt and Paul Milgrom, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 1991, *7*, 24.

Klepper, Steven and Nagin Daniel, "The Anatomy of Tax Evasion," *Journal of Law, Economics, and Organization*, 1989, *5* (1), 1–24.

Kleven, Henrik J., Claus T. Kreiner, and Emmanuel Saez, "Why can Modern Governments Tax so Much? An Agency Model of Firms as Fiscal Intermediaries," 2009. NBER Working Papers 15218, National Bureau of Economic Research 2009.

_ , Martin Knudsen, Claus T. Kreiner, Soren L. Pedersen, and Emmanuel Saez, "Un-willing or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark," *Econometrica*, 2010, *79* (3), 651–692.

Kopczuk, Wojciech, "Polish Business Flat Tax and its Effect on Reported Incomes: a Pareto Improving Tax Reform?," 2012. Columbia University Working Paper.

Kremer, Michael, "The O-ring Theory of Economic Development," *The Quarterly Journal of Economics*, 1993, *108* (3), 551–575.

Kumler, Todd, Eric Verhoogen, and Judith A. Fras, "Enlisting Employees in Improving Payroll-Tax Compliance: Evidence from Mexico," NBER Working Papers 19385, National Bureau of Economic Research, Inc August 2013.

Pomeranz, Dina, "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax," 2011. Working Paper.

Slemrod, Joel, "Does It Matter Who Writes the Check to the Government? The Economics

of Tax Remittance," *National Tax Journal*, 2008, *61* (2), 75–251.

— **and Shlomo Yitzhaki**, "Tax Avoidance, Evasion, and Administration," *Handbook of Public Economics*, 2002, *3*, 1423–1470.

Yang, Dean, "Can Enforcement Backfire? Crime Displacement in the Context of Customs Reform in the Philippines," *Review of Economics and Statistics*, 2008, *90* (1), 1–14.

Yitzhaki, Shlomo, "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1974, *3* (2), 201–202.

7 Tables and Figures

Table 1: Descriptive Statistics, Active Firms, 2008-2010

Revenue ($000)	1,479.3 (40,756.8) [39.4]
Costs ($000)	1,384.8 (40,166) [39.9]
Tax liability ($000)	22.0 (604.0) [0.0]
Own reported revenue - third-party reported revenue ($000)	587.1 (29,200.0) [5.9]
Own reported costs - third-party reported costs ($000)	926.2 (35,844.7) [30.1]
Observations	207,874

Notes: Group means are reported along with standard deviations in parentheses and medians in brackets. All monetary figures in USD.

Table 2: Descriptive Statistics By Year, Notified and Amending Firms

	Fiscal Year		
	2008	**2009**	**2010**
Notified Firms			
Revenue ($000)	1,541.5	624.1	591.8
	(6,117.7)	(1,259.8)	(1,236.9)
	[136.3]	[120.4]	[102.7]
Costs ($000)	1,438.7	583.3	552.0
	(5,397.9)	(1,170.3)	(1,152.8)
	[129.2]	[113.1]	[95.6]
Tax liability ($000)	25.6	10.5	9.9
	(170.6)	(31.0)	(27.6)
	[0.9]	[0.7]	[0.6]
Own reported revenue - SRI estimated revenue ($000)	-306.6	-176.1	-197.0
	(1,740.9)	(312.3)	(388.3)
	[-41.7]	[-63.7]	[-66.3]
Observations	3,136	2,221	2,636
Number of firms with invalid email addresses	NA	159	163
Amending Firms			
Revenue ($000)	1,401.8	370.0	416.7
	(4367.1)	(943.9)	(958.8)
	[179.0]	[55.8]	[34.1]
Costs ($000)	1331.1	351.6	397.4
	(4172.1)	(886.7)	(923.4)
	[162.6]	[62.2]	[32.6]
Tax liability ($000)	18.1	5.1	5.1
	(69.7)	(16.8)	(13.3)
	[1.4]	[0.1]	[0.0]
Own reported revenue - SRI estimated revenue ($000)	-216.6	-150.9	-173.1
	(805.7)	(298.8)	(332.9)
	[-40.2]	[-54.5]	[-62.7]
Observations	596	249	421
Start of notification period	August 11, 2011	March 26, 2012	March 26, 2012
End of notification period		April 20, 2012	April 20, 2012

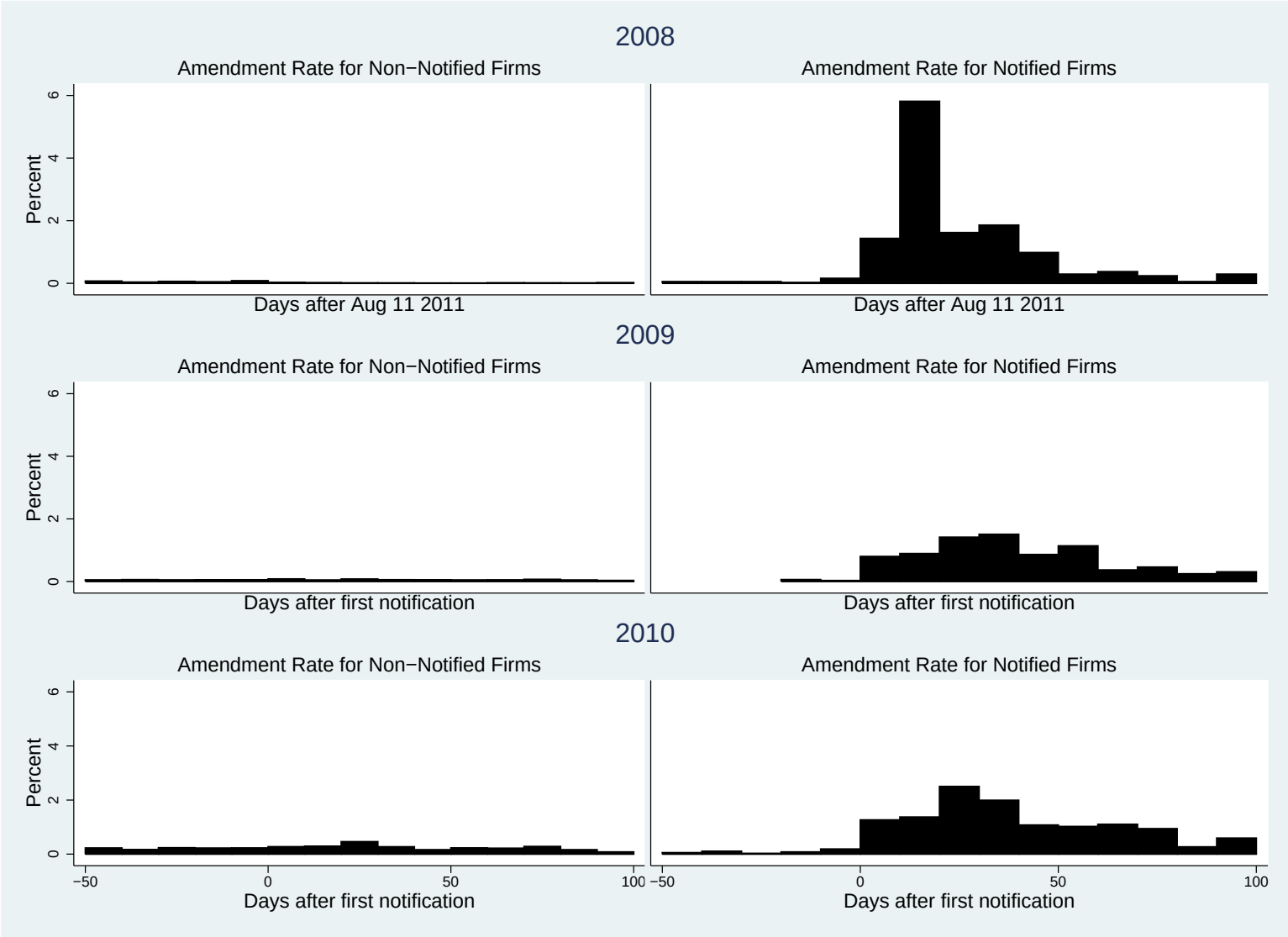
Notes: Group means are reported, along with standard deviations in parentheses and medians in brackets. Notified firms are defined as those to whom the SRI sent an email notification (including those for whom the email bounced back). Amending firms are defined as those who filed an amended return in our post-notification window; see text for details. All monetary figures in USD.

Table 3: Treatment Effects, 2008-2010

	(1) Revenue	(2) Costs	(3) Tax Liability	(4) Log Revenue	(5) Log Costs	(6) Log Tax Liability
Panel A: Amending Firms						
Post	86,203***	80,155***	1,857***	3.448***	3.405***	1.939***
	(15,346)	(15,037)	(222)	(0.204)	(0.205)	(0.125)
Constant	8.71e+05***	8.28e+05***	11,237***	8.595***	8.604***	4.815***
	(7,673)	(7,518)	(111)	(0.102)	(0.103)	(0.062)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.996	0.995	0.996	0.721	0.713	0.828
Observations	2,532	2,532	2,532	2,532	2,532	2,532
Number of firms	1,266	1,266	1,266	1,266	1,266	1,266
Panel B: Notified Firms						
Post	13,653***	12,695***	294***	0.546***	0.539***	0.307***
	(2,363)	(2,310)	(35)	(0.036)	(0.037)	(0.022)
Constant	9.73e+05***	9.09e+05***	16,224***	9.012***	9.016***	5.201***
	(1,181)	(1,155)	(17)	(0.018)	(0.018)	(0.011)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.998	0.998	0.998	0.916	0.915	0.924
Observations	15,986	15,986	15,986	15,986	15,986	15,986
Number of firms	7,993	7,993	7,993	7,993	7,993	7,993

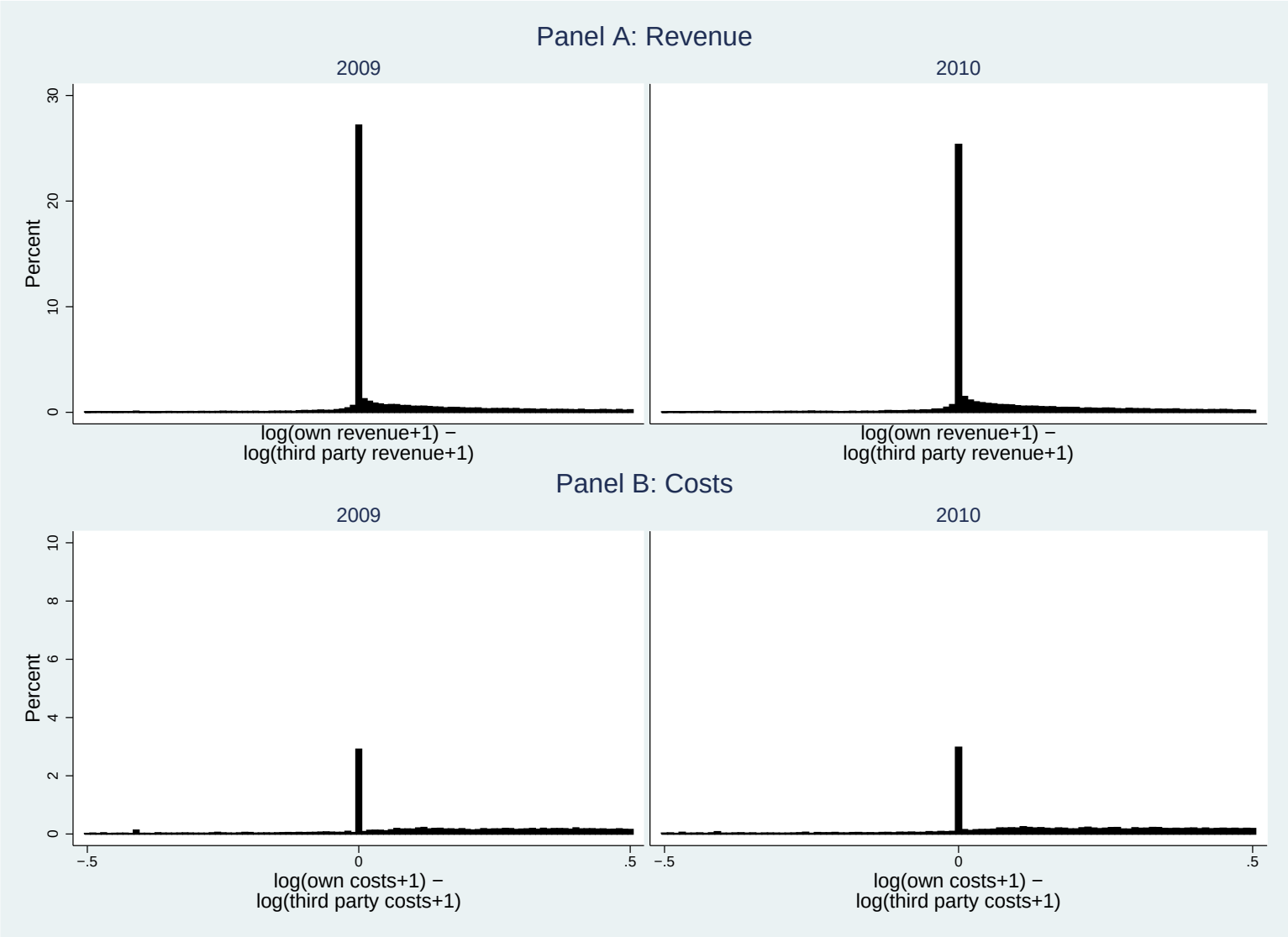
Notes: In Columns (1)-(3), dependent variables are in levels. In Columns (4)-(6), dependent variables are in *log + 1* to avoid exclusion of zero values. Panel A shows the sample of amending firms. Panel B shows the sample of notified firms. Firms are pooled across 2008-2010. All monetary figures in USD. Standard errors clustered by firm in parentheses. Level of significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Figure 1: Amendment Rates



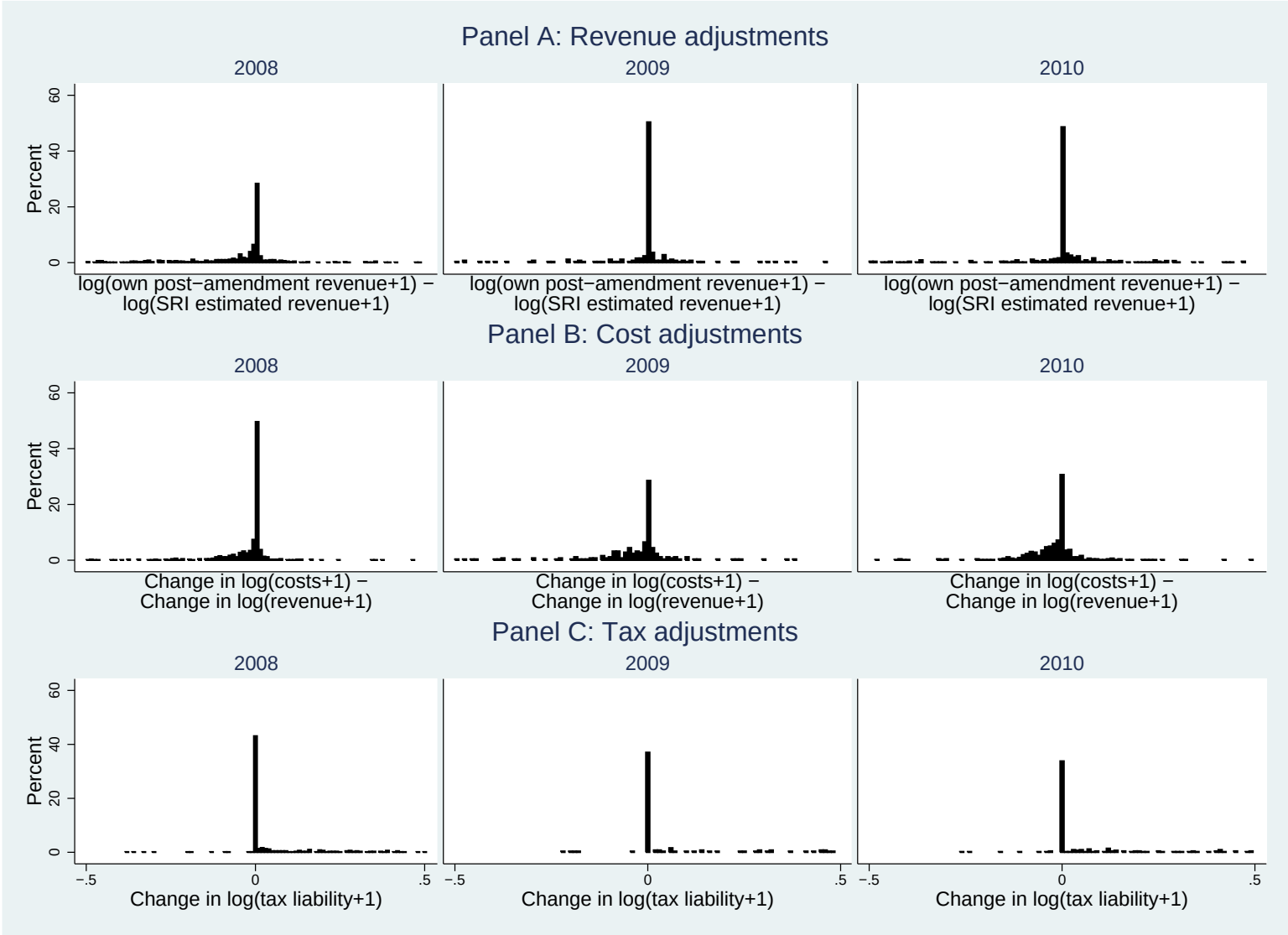
Notes: The left Column plots amendment rates for the universe of non-notified firms, the right Column for notified firms before and after the start of the intervention.

Figure 2: Discrepancies



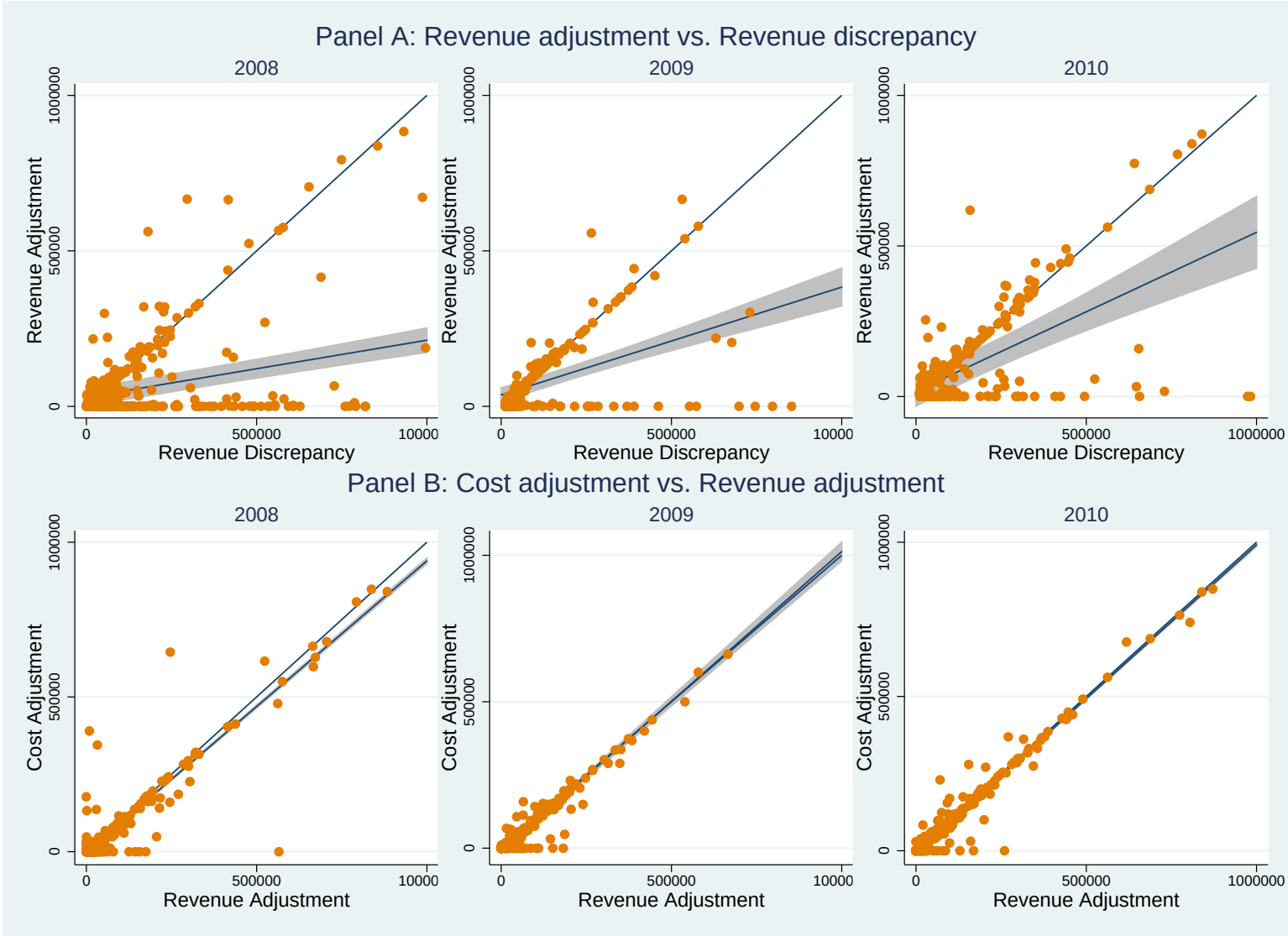
Notes: Panel A presents histograms of revenue discrepancies among the universe of active firms in the pre-notification period. Panel B does the same for cost discrepancies. In both panels, a *log + 1* transformation is applied before calculating discrepancies. Bins are of size 0.01. The top and bottom 1% of the sample are omitted when calculating bin heights for computational reasons.

Figure 3: Adjustments



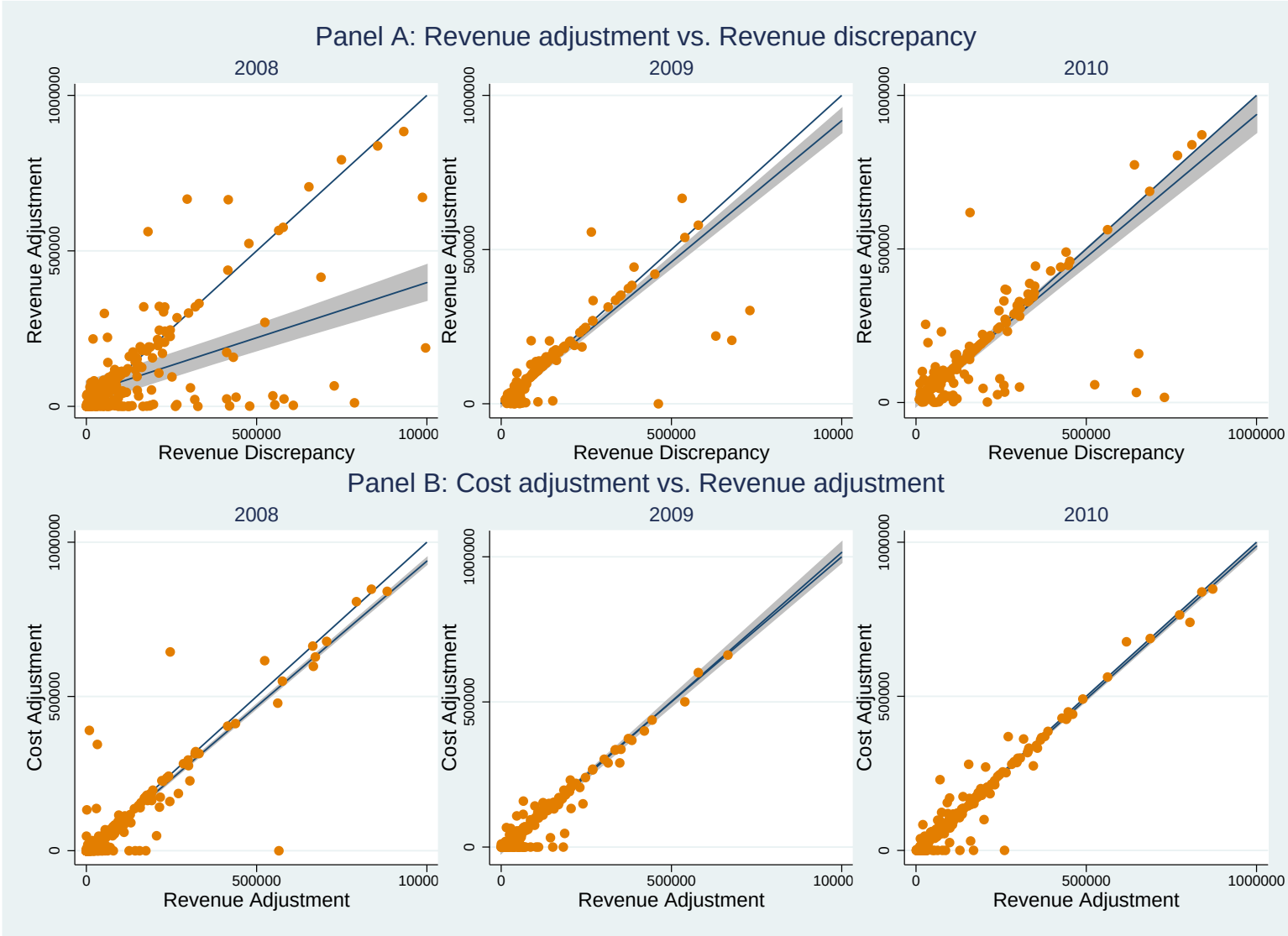
Notes: Panel A presents histograms of log(own revenue) minus log(SRI estimated revenue) for the sample of amending firms. Panel B does the same for cost adjustments minus revenue adjustments. Panel C shows the same for the tax adjustments. In all three panels, a *log + 1* transformation is applied before calculating discrepancies and adjustments. Bins are of size 0.01, and the top and bottom 1% of the sample are omitted when calculating bin heights for computational reasons.

Figure 4: Matching



Notes: Panel A plots revenue adjustments on the y-axis against pre-notification revenue discrepancies on the x-axis by year for the sample of amending firms. Panel B shows the same for cost adjustments on the y-axis and revenue adjustments on the x-axis. Column 1 shows results for the 2008 declarations, Column 2 for 2009, and Column 3 for 2010. Also shown is a 45-degree line in blue, a fitted line in green, and a 95% confidence interval for the fitted line in grey. The range shown is zero to a million, which comprises over 90% of the sample.

Figure 5: Matching Among Adjusters



Notes: Panel A plots revenue adjustments on the y-axis against pre-notification revenue discrepancies on the x-axis by year for the sample of amending firms that actually adjusted revenues. Panel B shows the same with cost adjustments on the y-axis and revenue adjustments on the x-axis. Column 1 shows the adjustments for 2008, Column 2 for 2009, and Column 3 for 2010. Also shown is a 45-degree line in blue, a fitted line in green, and a 95% confidence interval for the fitted line in grey. The range shown is zero to a million, which comprises over 90% of the sample.

Appendix

Appendix A1: Policy Intervention Message: Year 2008

SERVICIO DE RENTAS INTERNAS

DEPARTAMENTO DE GESTIÓN TRIBUTARIA

Quito, 5 de septiembre del 2011

Señor (a) xxxxxx

Gerente General de xxxxx

El Art. 67 del Código Tributario y el segundo artículo de la Ley de Creación del Servicio de Rentas Internas otorgan a esta Administración Tributaria la facultad para efectuar la determinación, recaudación y control de los tributos internos del Estado.

Esta Administración Tributaria, luego de revisar las bases de datos con las que cuenta, *ha identificado valores atribuibles a ingresos de la sociedad a la que usted representa superiores al monto registrado en la declaración de impuesto a la renta correspondiente al ejercicio fiscal 2008*.

De conformidad a lo establecido por los artículos 89 del Código Tributario y 101 de la Ley de Régimen Tributario, las declaraciones de impuestos efectuadas por los sujetos pasivos tienen el carácter de definitivas y vinculantes, por lo que hacen responsable al declarante y, en su caso, al contador que firme la declaración, por la exactitud y veracidad de los datos que contenga; sin embargo el sujeto pasivo, a petición expresa del Servicio de Rentas Internas podrá, dentro de los seis años siguientes a la fecha de presentación de la declaración original, rectificar en una declaración sustitutiva, *los rubros requeridos por la Administración Tributaria*.

El Art. 19 de la Ley de Régimen Tributario Interno y el artículo 37 de su reglamento, establecen que todas las sociedades están obligadas a llevar contabilidad y declarar el impuesto en base a los resultados que arroje la misma. Adicionalmente los libros contables tienen que estar debidamente respaldados por los correspondientes comprobantes de venta y demás documentos pertinentes, documentación toda que puede ser requerida en

cualquier momento por la Administración Tributaria para fines de control.

En atención a los antecedentes y a las normas legales citadas, esta Administración le solicita presente la declaración sustitutiva correspondiente al impuesto a la renta del ejercicio fiscal 2008 vía Internet, dentro de los diez (10) días hábiles posteriores a la presente comunicación.

Adicionalmente le recordamos que en la declaración del impuesto a la renta del año 2008, debe registrar el valor del anticipo calculado de impuesto a la renta con cargo al ejercicio fiscal 2009, de conformidad al artículo 41 de la Ley de Régimen Tributario Interno.

A la vez se le informa que de ser el caso, el sujeto pasivo, deberá calcular el impuesto, interés y multa a pagar considerando los pagos previos efectuados, conforme la normativa tributaria vigente respecto a la imputación al pago.

Finalmente, se advierte al sujeto pasivo que la Administración Tributaria se reserva el derecho de verificar oportunamente la información contenida en las declaraciones de impuestos, que en el caso de que el sujeto activo ejerza su facultad determinadora procederá a cobrar un recargo del veinte por ciento (20%) calculado en base al impuesto determinado, y que en caso de comprobar la existencia de actos de ocultación o falsedad, por los que se haya dejado de pagar en todo o en parte los tributos debidos, en provecho propio o de un tercero, tales hechos se considerarán defraudación fiscal, conforme lo señala el artículo 342 del Código Tributario y cuyas sanciones se especifican en el Libro Cuarto del mismo cuerpo legal que se refiere al Ilícito Tributario.

En caso de requerir mayor información sobre la presente comunicación puede acercarse a las oficinas del Departamento de Gestión Tributaria, ubicadas a nivel nacional.

El envío de este correo es automático, por favor no lo responda.

Atentamente,

Servicio de Rentas Internas

Nota: Ahora es más fácil cumplir con sus obligaciones tributarias, utilizando nuestro servicio gratuito de declaraciones y anexos por internet, que le permitirá presentar ágilmente la información. Obtenga su clave de seguridad y el programa en cualquiera de las oficinas del Servicio de Rentas Internas a nivel nacional.

Appendix A2: POLICY INTERVENTION MESSAGE: YEARS 2009 - 2010

Declare correctamente su Impuesto a la Renta

SRi ...le hace bien al país!

SERVICIO DE RENTAS INTERNAS

DEPARTAMENTO DE GESTIÓN TRIBUTARIA

Quito, a viernes, 20 de abril de 2012

Señor (a) xx

Representante Legal de xx

El Art. 67 del Código Tributario y el segundo artículo de la Ley de Creación del Servicio de Rentas Internas otorgan a esta Administración Tributaria la facultad para efectuar la determinación, recaudación y control de los tributos internos del Estado.

El Servicio de Rentas Internas, ha realizado el cruce especial de información donde se verifican los valores declarados en el rubro Ventas Gravadas y No Gravadas. Así, luego de revisar las bases de datos con las que cuenta, ha detectado valores atribuibles a la sociedad a la que usted representa, diferentes a los montos registrados en la declaración de impuesto a la renta correspondiente al ejercicio fiscal 20XX, según se puede observar en el siguiente detalle:

Año Fiscal	Casillero de la Declaración de Impuesto a la Renta	Valor calculado por la Administración Tributaria	Valor declarado por el contribuyente
2010	699 - TOTAL INGRESOS	777.499,10	719.153,50

De conformidad a lo establecido por los artículos 89 del Código Tributario y 101 de la Ley de Régimen Tributario, las declaraciones de impuestos efectuadas por los su-

jetos pasivos tienen el carácter de definitivas y vinculantes, por lo que hacen responsables al declarante y al contador que firmen la declaración, por la exactitud y veracidad de los datos que contenga la misma; sin embargo el sujeto pasivo, a petición expresa del Servicio de Rentas Internas podrá, dentro de los seis años siguientes a la fecha de presentación de la declaracin original, rectificar en una declaración sustitutiva, los rubros requeridos por la Administración Tributaria.

El Art. 19 de la Ley de Régimen Tributario Interno y el artículo 37 de su reglamento, establecen que todas las sociedades están obligadas a llevar contabilidad y declarar el impuesto en base a los resultados que arroje la misma. Adicionalmente los libros contables tienen que estar debidamente respaldados por los correspondientes comprobantes de venta y demás documentos pertinentes, documentación toda que puede ser requerida por la Administración Tributaria para fines de control.

En atención a los antecedentes y a las normas legales citadas, esta Administración le apremia a presentar la declaración sustitutiva correspondiente al impuesto a la renta del ejercicio fiscal 2010 vía Internet.

Adicionalmente se le recuerda que en la declaración del impuesto a la renta del año 2010 debe registrar el valor del anticipo calculado de impuesto a la renta con cargo al ejercicio fiscal 2011, de conformidad al artículo 41 de la Ley de Régimen Tributario Interno.

De ser el caso, el sujeto pasivo deberá calcular el impuesto, interés y multa a pagar, considerando los pagos previos efectuados, conforme la normativa tributaria vigente respecto a la imputación al pago.

Finalmente, se informa al sujeto pasivo que la Administración Tributaria se reserva el derecho de verificar oportunamente la información contenida en las declaraciones de impuestos, y que en el caso de que el sujeto activo ejerza su facultad determinadora procederá cobrar un recargo del veinte por ciento (20%) calculado en base al impuesto determinado; así como también, que en caso de comprobar la existencia de actos de ocultación o falsedad, por los que se haya dejado de pagar en todo o en parte los tributos debidos, en provecho propio o de un tercero, tales hechos se considerarán defraudación fiscal, conforme lo señala el artículo 342 del Código Tributario y cuyas sanciones se especifican en el Libro Cuarto del mismo cuerpo legal que se refiere al Ilícito Tributario.

La asesora que se requiera para el cumplimiento de obligaciones tributarias, la puede obtener en todas las oficinas del Servicio de Rentas Internas a nivel nacional o a travs de nuestra página web (www.sri.gob.ec).

Atentamente,

Servicio de Rentas Internas