8-4-2017

17-14 Reputation and Multilateral Punishment under Uncertainty

Aidin Hajikhameneh
Chapman University, hajikhameneh@chapman.edu

Jared Rubin
Chapman University, jrubin@chapman.edu

17-14 Reputation and Multilateral Punishment under Uncertainty

Reputation and Multilateral Punishment under Uncertainty*

Aidin Hajikhameneh† **Jared Rubin**‡

August 4, 2017

Abstract

Prinicpal-agent problems can reduce gains from exchange available in long distance trade. One solution to mitigate this problem is multilateral punishment, whereby groups of principals jointly punish cheating agents by giving them bad reputations. But how does such punishment work when there is uncertainty regarding whether an agent actually cheated or was just the victim of bad luck? And how might such uncertainty be mitigated—or exacerbated—by non-observable, pro-social behavioral characteristics? We address these questions by designing a simple modified trust game with uncertainty and the capacity for principals to employ multilateral punishment. We find that a modest amount of uncertainty *increases* overall welfare because principals are more willing to trust agents with bad reputations.

JEL classifications: *C91, C92, D02, D83, F10, N70*
Keywords: *Multilateral punishment, reputation, uncertainty, exchange, lab experiment, trust game*

*We thank Gary Charness, Ellen Green, Laurence Iannaccone, Erik Kimbrough, Michael McBride, Elena Pikulina, and participants in the Southwest Experimental and Behavioral Economics Workshop at UC Santa Barbara, and the Economic Science Association World Meeting at UC San Diego for excellent feedback. Experiments were programmed using z-Tree (Fischbacher, 2007). Some figures were created using the open-source statistical software R. All mistakes are our own.

†Institute for the Study of Religion, Economics and Society, Chapman University, 338 N. Glassell, Orange, CA 92866, USA, e-mail: hajikhameneh@chapman.edu

‡Argyros School of Business and Economics, Chapman University, 338 N. Glassell, Orange, CA 92866, USA, e-mail: jrubin@chapman.edu

1 Introduction

Long-distance impersonal exchange is fraught with opportunities for cheating. Since exchange is sequential, it is in the second player's interest to renege if their past conduct does not matter for their ability to conduct future exchange (Greif, 2000). One way societies solve this "fundamental problem of exchange," is to establish (relatively) impartial legal institutions which serve as a third-party enforcement mechanism and discourage cheating. Yet, impersonal exchange pre-dates widespread legal institutions by millennia, and even in the contemporary world, exchange happens all the time in the absence of formal legal enforcement (Ostrom, 1990, 2005).

The most common mechanism used to link past conduct to future reward is *reputation*. Bilateral reputation works to facilitate exchange when individuals have much to gain from maintaining the exchange relationship and there are few outside options that yield similar returns.[1] However, bilateral reputation is often not enough to sustain exchange. For instance, firms in most industries have little incentive to maintain their reputation with one individual customer, especially when cheating is highly profitable. When bilateral reputation mechanisms fail, societies often establish *multilateral* reputation-building institutions, whereby coalitions form to coordinate on punishment (i.e., loss of reputation) for breach of contract. For instance, the Amish shun offenders of church rules. This requires punishment by the entire community; such multilateral punishment breaks down quickly if group members fail to punish (Posner and Rasmusen, 1999).

In the contemporary West, technology has greatly facilitated the capacity of individuals and organizations to build reputations. This is precisely the objective of websites such as Yelp, TripAdvisor, and eBay (Resnick *et al.*, 2006). Historically, in the absence of technology that can quickly and cheaply disseminate information about individuals who have reneged on their obligations, societies established a wide range of private-order institutions that connect one's past conduct to future reward. For example, the Maghribi Traders' Coalition studied by Greif (1993) was a group

[1] A large literature shows how bilateral reputation, via repeat interaction, can sustain cooperation without third-party enforcement (e.g., Kreps and Wilson (1982); Kreps *et al.* (1982)). While this works under certain circumstances, we focus on situations where bilateral reputation is not effective, such as when groups are large and histories unknown (Ghosh and Ray, 1996; Kranton, 1996; Kranton and Minehart, 2001; Leeson, 2008).

of traders and agents who sent each other letters if they felt that a fellow trader cheated them. Since they were a closed group and merchants incentivized their agents by paying them a premium, the coalition worked for over a century to facilitate exchange. Clay (1997) and Okazaki (2005) study similarly-structured reputation-based institutions designed to facilitate trade between merchants in 19^{th} century Mexican California and pre-modern Japan, respectively. Bernstein (2001) provides a contemporary example from the cotton industry, where for over a century disputes have been settled in an extra-legal "private legal system" which depends on reputation-based sanctions to incentivize compliance with its trade rules. Bernstein (1992) shows that dispute-resolution in the diamond industry works using similar, reputation-based, extra-legal institutions.[2]

Yet, institutions which facilitate trade by permitting individuals to establish a reputation face two problems.[3] First, it must be in the incentive of the cheated parties to report they were cheated. In a world of Yelp, TripAdvisor, or eBay, this cost is minimal. However, in a pre-modern setting where communication over long distance was costly and time consuming, such costs could be prohibitive. Second, and perhaps more important, "cheating" is often not black and white and is often unverifiable. This problem arises due to asymmetric information. For example, consider a merchant-agent relationship. A merchant sends goods to an agent to be sold in a distant land and expects to receive a portion of the proceeds in return. The agent reports back to the merchant that the goods were damaged at sea and she was only able to receive half of the expected proceeds. The agent may be telling the truth, although of course she also has incentive to lie if the goods actually arrived in perfect condition. When the merchant receives the bad news and reputation-based institutions are present, will the merchant engage with the institution and provide the agent with a bad reputation? Under what circumstances will she do so? How do such considerations affect the incentive to engage in exchange relationships in the first place?

[2]The papers cited above highlight cases where institutions established individual-based reputation in order to facilitate exchange. Such institutions are the focus of the present paper. Another mechanism employed under different circumstances is group-based reputation, which has been studied theoretically and empirically in a variety of contexts. For some examples, see Milgrom *et al.* (1990); Besley and Coate (1995); Tirole (1996); Ghatak and Guinnane (1999); Greif *et al.* (1994); Greif (2002, 2004); Winfree and McCluskey (2005); Richardson (2005); Levin (2009); Boerner and Ritschl (2009).

[3]See Leeson (2008) for a nice overview of when multilateral punishment institutions fail. Ali *et al.* (2017) provide a theoretical model which gives conditions under which multilateral enforcement is renegotiation-proof.

We address these questions by designing a laboratory experiment in which subjects play a modified trust game (Berg *et al.*, 1995) in the presence of a multi-lateral punishment institution.[4] Previous experiments on cheating in trade reveal that the opportunity to cheat increases market segmentation and reduces gains from exchange (Cassar *et al.*, 2009), while information-sharing networks reduce cheating and increase efficiency (Cassar *et al.*, 2010; Kimbrough and Rubin, 2015). Likewise, reputation-building encourages trust and trustworthiness in repeated trust games among strangers with known reputations (Bohnet and Zeckhauser, 2004; Bohnet *et al.*, 2005; Bracht and Feltovich, 2009; Charness *et al.*, 2011).[5] Yet, to our knowledge the literature does not address how such reputation-building institutions work when there is uncertainty regarding whether the agent actually cheated or not. This is an important issue, because it is unclear whether the institutions work in the desired manner in the presence of uncertainty and, if they do, the degree to which uncertainty affects efficiency.

Our experiment is designed to address precisely these issues. In our experiment, principals and agents interact multiple times, although never with the same player twice. In each interaction, there are gains from exchange, and agents have the opportunity to cheat the principal. At the beginning of each interaction, the principal sees the agents reputation, which is established by past play. Reputation is established via an exogenously given institution that permits a cheated principal to provide the agent she is matched with a bad reputation (at cost) that they carry with them the rest of the game.

We employ a 3×2 experimental design where uncertainty and the cost of providing a bad reputation to the agent are varied. We vary the cost of giving a bad reputation between two

[4]Ho and Huffman (2017) provide an overview of the literature on trust and legal institutions. They note that the experimental literature generally finds that law and trust are substitutes when law is imposed exogenously (i.e., by the experimenter), but they are complements when laws are adopted endogenously.

[5]Second and third party enforcement is employed in various experiments to mitigate cheating in trust games. Smith and Wilson (2017) allow principals to retaliate after being cheated. Fehr and Rockenbach (2003) and Fehr and List (2004) let principals threaten agents with a fine prior to trade. This fine is imposed if agents cheat. Results of this research indicate that second party enforcement exacerbates cheating. Bohnet *et al.* (2001) incorporate a court system of varying strength into the trust game. The court system is a probabilistic function that catches and punishes cheating agents. They report that weak and strong court systems mitigate cheating compared to a medium strength court system. Fehr and Fischbacher (2004) employed a human third party enforcement who could punish non-cooperative behavior at cost in a dictator and prisoners' dilemma game. They find that human third party enforcement encourages cooperative behavior.

values: 0 and a non-trivial positive sum. We introduce uncertainty by letting nature destroy the principal's return with some positive probability–despite the fact that the agent acted honestly. The principal cannot verify whether the agent acted honestly but was unlucky or whether the agent acted dishonestly. The principal must then choose whether to give the agent a bad reputation in spite of this uncertainty. We vary the probability of nature destroying the principals return between 0, a modest positive amount (0.05), and a large amount (0.30).

Our most important finding is that a modest amount of uncertainty (i.e., the 0.05 treatment) *increases* total welfare relative to the case where there is no uncertainty or a high amount of uncertainty, *despite* the fact that uncertainty destroys resources.[6] The reason this result arises is that principals are more willing to engage with agents with bad reputations when there is a modest amount of uncertainty, presumably because agents may have a bad reputation despite acting honestly in the past. This is not a Pareto improving welfare gain, however; agents benefit at the expense of principals, some of whom end up being cheated because they trusted an agent who cheated in the past (and is more likely to continue cheating). On the other hand, a large amount of uncertainty destroys too many resources for the principal to be willing to engage in trade with an agent with a bad *or* good reputation, and the overall level of welfare is therefore lower.

The paper is structured as follow. In Section 2, we present a formal, yet simple, model to guide the reader's intuition regarding how subjects might act in a modified trust game with a multilateral punishment institution, costly reputation, and uncertainty. Section 3 lays out the experimental design, Section 4 reports the results, and Section 5 offers some concluding thoughts.

[6]This result is consistent with Cassar *et al.* (2009), who find that the overall volume of trade is greater when cheating is allowed than when it is not, because "extra-marginal" agents have the opportunity to engage in trade despite the fact that they end up cheating.

2 The Model

2.1 Setup

In this section, we present a model in the spirit of Greif (1993). We model a setting in which principals use agents to trade, gains from trade are possible, and agents have the opportunity to cheat the principal. We assume the existence of a multilateral punishment institution—like the one studied in Greif (1993), Bernstein (1992, 2001), and Clay (1997)—whereby principals can bestow a reputation, known to all other principals, on agents who they believe cheated them. Our model, however, adds two key features: i) there is uncertainty in the mapping from an agent's action to the outcome, and ii) multilateral punishment is costly to enact. We focus on these two features and generate testable predictions regarding how they affect players' actions.

Consider an economy consisting of two groups of infinitely-lived players: Principals (P) and Agents (A). Each are of measure 1. As in the experiment laid out in Section 3, principals and agents are randomly matched in each period and interact only once.

Within each period, there are three stages. In the first stage, P decides whether or not to *send* ($S \in \{0, 1\}$) goods to A. If P chooses not to send ($S = 0$), each player receives their opportunity cost payout $\Omega > 0$, entailing that total welfare is 2Ω, and the period ends. If P chooses $S = 1$, the total amount available to the players doubles from 2Ω to 4Ω, representing the gains made from the exchange relationship, and the game proceeds to the second stage.

In the second stage, A decides whether or not to *divide* ($D \in \{0, 1\}$) the proceeds between the two players. If A chooses not to divide ($D = 0$), A keeps the payout 4Ω, P receives 0, and the game proceeds to the third stage. In other words, $D = 0$ represents the agent "cheating" the principal. If A chooses to divide ($D = 1$), then A receives payout 2Ω. However, there is uncertainty regarding how much P receives. With probability $1 - \theta \in [0, 1]$, P receives 2Ω and the period ends. With probability θ, P receives 0 and the game proceeds to the third stage. The random element θ is interpreted as the unobservable randomness associated with long-distance exchange relationships (Greif, 1993). For instance, if one ships goods to an overseas agent, it is possible that some of the

goods are damaged in transit (i.e., with probability θ), while it is also possible that the agent is lying about the shape the goods are in and keeps the proceeds for himself. P cannot distinguish between A choosing $D = 0$ and A choosing $D = 1$ but receiving a bad shock.

The game reaches the third stage if and only if P receives 0 from the exchange relationship. In this stage, we assume the existence of a multi-lateral punishment institution in which principals can transmit reputation about agents to all other principals, and a "black list" is kept in perpetuity. We do not model the emergence of this institution or study whether it is self-enforcing. We simply assume it exists and can be employed by principals at a cost.

In the third stage, P can contribute to the multilateral punishment institution by choosing whether or not to *inform* ($I \in \{0, 1\}$) other P's about the agent at cost $C \geq 0$. If P chooses $I = 1$, that agent carries a bad *reputation*, $R \in \{0, 1\}$, for the remainder of its life. Prior to the first period, each A has a good reputation ($R = 1$). A maintains its good reputation until one P with whom it is matched chooses $I = 1$. In all remaining periods, A has a bad reputation ($R = 0$). P sees the reputation of the A it is matched with prior to its first stage decision to send. Figure 1 provides the game tree for each period of the game.

2.2 Utility and Player Types

We assume that principals attempt to maximize their lifetime payout, but also exhibit pro-social behavior (Frey and Meier, 2004; Tirole, 2006; Meier, 2006; Kőszegi, 2014). For the sake of the model, we define pro-social behavior as the willingness to take on a cost to warn other principals that the agent they are matched with caused them harm, even though this punishment does not affect their future payoff. That is, pro-social behavior is purely a behavioral feature. It is not strategic—indeed, it reduces one's monetary payout. Principals differ only in their pro-social behavior parameter, β_i, which is distributed throughout the population over some known cumulative distribution function $F(\cdot)$.[7]

[7]There are numerous behavioral features that we could have incorporated into the model in order to generate principals choosing to inform on agents in spite of the negative monetary payoff associated with doing so. These include negative behaviors aimed at agents such as betrayal (Bohnet and Zeckhauser, 2004; Hong and Bohnet, 2007;

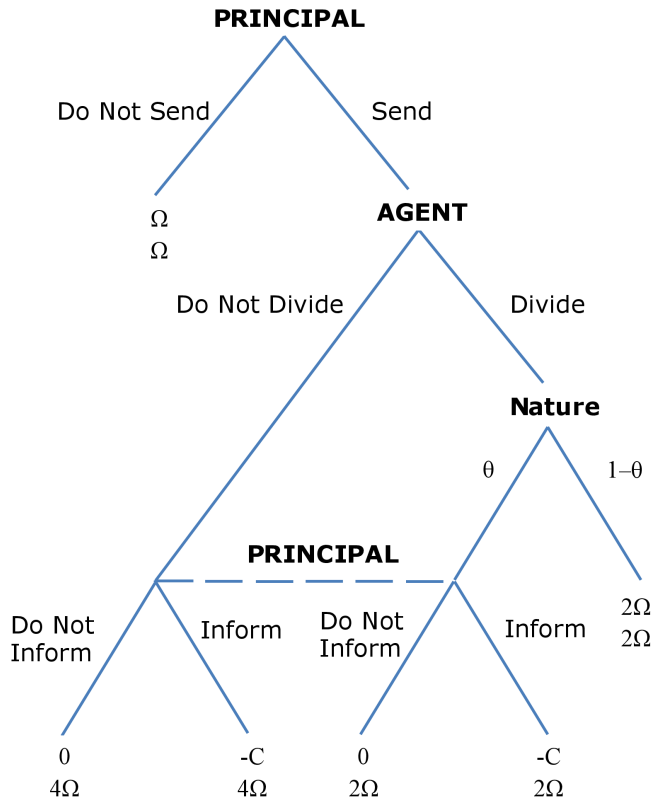

Figure 1: **Game Tree.** The top and bottom payoffs belong to the principal and agent, respectively.

Principals receive utility from exhibiting pro-social behavior when they inform on an agent ($I = 1$) and their informing changes the agent's reputation from "good" to "bad". Principals do not receive utility from informing on an agent that already has a bad reputation, since the agent will maintain that reputation regardless of what the principal does. In total, principals receive utility from their monetary payoff in addition to their pro-social utility. Denoting the principal's discount factor by $\delta \in (0, 1)$, we write principal i's period t utility as:

$$U_{i,t}^{P} = \sum_{j=t}^{\infty} \delta^{j-t} \left[E\left[\pi_{i,j}^{P}\right] + \beta_i I_j R_j \right], \tag{1}$$

Bohnet *et al.*, 2008), blame (Gurdal *et al.*, 2013; Bartling *et al.*, 2014), spite (Pillutla and Murnighan, 1996; Cason *et al.*, 2002; Kimbrough and Reiss, 2012), and punishment (Fehr and Fischbacher, 2004; Fehr and List, 2004; Smith and Wilson, 2017); or positive features such as gift exchange (Fehr *et al.*, 1997, 2007; Charness and Kuhn, 2011; Rubin and Sheremeta, 2016) aimed at inducing cooperative behavior by agents.

where principal i's payoff in period t, $\pi_{i,t}^{P}$, is written:

$$\pi_{i,t}^{P} = (1 - S_t)\,\Omega + S_t\left(D_t\left[-\theta I_t C + (1-\theta)\,2\Omega\right] - \left[1 - D_t\right] I_t C\right). \tag{2}$$

Meanwhile, we assume that there are two types of agents: altruists and strategic. Altruistic agents, which comprise $\alpha \in [0,1]$ of the agent population, always choose $D = 1$ regardless of their reputation (Andreoni, 1990; Levine, 1998; Charness and Haruvy, 2002; Fehr and Schmidt, 2006; Carpenter *et al.*, 2008). Strategic agents, which comprise the remaining $1 - \alpha$ portion of the agent population, only choose $D = 1$ if doing so maximizes their expected lifetime payout. We write (strategic) agent i's period t utility as:

$$U_{i,t}^{A} = \sum_{j=t}^{\infty} \delta^{j-t} E\left[\pi_{i,j}^{A}\right], \tag{3}$$

where agent i's payoff in period t, $\pi_{i,t}^{A}$, is written:

$$\pi_{i,t}^{A} = (1 - S_t)\,\Omega + S_t\left(D_{i,t} 2\Omega + \left[1 - D_{i,t}\right] 4\Omega\right) = \left[1 + S_t\left(3 - 2D_{i,t}\right)\right]\Omega. \tag{4}$$

2.3 Solving the Model

Before we solve the model, note that P's choice in any one period does not affect its utility in future periods. We therefore solve for P's decision in a single period given the information that P has at the beginning of the period (i.e., A's reputation).

We begin by solving for periods in which A begins with a bad reputation ($R = 0$). In this case, P never informs in the third stage regardless of its spite parameter (i.e., $I = 0$), because informing does not affect the reputation of A. Moving to the second stage, altruistic agents by definition always divide ($D = 1$). Meanwhile, strategic agents never divide ($D = 0$), because choosing $D_t = 0$ gives them a greater period t payoff than choosing $D_t = 1$, while not affecting their future payoff (since they will maintain their bad reputation).

Finally, moving to the first stage, P chooses $S = 1$ if its expected payoff in (2) is greater when $S = 1$ than it is when $S = 0$. A little simplification entails that this occurs when:

$$E\left[D_t | R = 0\right] \geq \frac{1}{2\left(1-\theta\right)}. \tag{5}$$

Since $D = 1$ if and only if A is altruistic, $E\left[D_t | R = 0\right]$ can be re-written $pr\left(A \text{ is altruistic} | R = 0\right)$. In period t, an altruistic agent has a bad reputation if and only if it was hit with a shock (with probability θ) in the same period it was matched with a P that had a high enough β_i to inform.

From (1), it is straight-forward to see that P will only inform on an agent with reputation $R = 1$ if $\beta_i \geq C$. Hence, the probability of being matched with a P that will choose $I = 1$ when the agent has a good reputation is $1 - F\left(C\right)$. Consequently, the probability of an altruistic agent being informed upon in any one period is $\theta\left[1 - F\left(C\right)\right]$. Hence, if P's always choose to send A's with good reputations, which we shall assume going forward, the probability that an altruistic agent has a good reputation in period t is $p^G = \left(1 - \theta\left[1 - F\left(C\right)\right]\right)^t$, and the probability it has a bad reputation is $p^B = 1 - \left(1 - \theta\left[1 - F\left(C\right)\right]\right)^t$. Denoting q as the probability a strategic agent has a bad reputation in period t, we write:

$$pr\left(A \text{ is altruistic} | R = 0\right) = \frac{\alpha p^B}{\alpha p^B + \left(1-\alpha\right) q}, \tag{6}$$

We therefore formally write the equilibrium strategies for periods in which $R = 0$ as follows.

Proposition 1: For periods in which $R = 0$, P and A choose the following strategies in equilibrium:
Stage 1: P chooses $S = 1$ if and only if $\frac{\alpha p^B}{\alpha p^B + (1-\alpha) q} \geq \frac{1}{2(1-\theta)}$; otherwise P chooses $S = 0$;
Stage 2: A chooses $D = 1$ if altruistic, $D = 0$ if strategic;
Stage 3: P chooses $I = 0$.

In our model, the two primary testable points of departure from Greif (1993) are: i) the multilateral punishment institution is costly (C), and ii) there is uncertainty (θ) regarding whether the agent cheated or not.[8] Therefore, in the experiment, we vary C and θ. It is therefore useful to provide comparative statics predictions with respect to these two parameters.

[8] Greif (1993) also does not include behavioral elements α or β. Since we do not directly test these parameters in the experiment, we do not focus on them in the analysis.

Proposition 1 indicates that, conditional on $R = 0$, both A's decision in Stage 2 and P's decision in Stage 3 are invariant to C and θ. However, P's send decision in Stage 1 is a function of both C and θ. We start by considering comparative statics with respect to C. An increase in C incentivizes strategic A's with good reputations to choose $D = 0$, since the probability they will be informed on is lower. C also affects the probability that an A (either strategic or altruistic) with a good reputation gets a bad reputation by having bad luck. This probability is $\theta(1 - F(C))$ if A chooses $D = 1$ and it is $1 - F(C)$ if it chooses $D = 0$. In either case, this probability is decreasing in C. In Prediction 1, we show that the combination of these two effects entail that, conditional on seeing $R = 0$, at low values of C an increase in C increases the probability that P is matched with a strategic agent, who will choose $D = 0$ in Stage 2. Therefore, an increase in C disincentivizes P from choosing $S = 1$. However, at sufficiently high values of C, further increases in C have no effect on strategic A's decision-making (it chooses $D = 0$ regardless), but it does make it less likely that strategic A's will have a bad reputation. As a result, it is more likely that, conditional on seeing an A with a bad reputation, that A is an altruist, thereby incentivizing P to choose $S = 1$. The size of the parameter space over which P chooses $S = 1$ is therefore U-shaped in C.

Meanwhile, an increase in θ has two countervailing effects on P's decision to send. On the one hand, higher levels of θ decrease the probability of a "good" outcome for P even if it is matched with an altruist (since the good outcome happens with probability $1 - \theta$ if A chooses $D = 1$). On the other hand, an increase in θ also means that more altruists receive a bad reputation, since they are more likely to have been informed upon in the past. This entails that, conditional on seeing $R = 0$, the probability that the agent that P is matched with is an altruist is increasing in θ. This latter effect dominates the first effect at low values of θ. However, at some point θ becomes large enough that P's choose $S = 0$ regardless of the probability that the agent is altruistic; even altruists only get the "good" outcome for P with probability $1 - \theta$. This is the key result of our paper: a modest, though not too large, level of uncertainty (θ) *increases* overall welfare by encouraging P's to send when they are matched with an A with a bad reputation. This logic is formalized in the following prediction, which we proceed to test in the experiment. Proofs of all predictions are in

Appendix B.

Prediction 1: For periods in which $R = 0$, the size of the parameter space over which P chooses $S = 1$ is hump-shaped in θ and U-shaped in C, ceteris paribus.

Next, consider periods in which A begins with a good reputation ($R = 1$). If the period gets to the third stage, P will inform on A if and only if $\beta_i \geq C$. Moving to the second stage, altruistic agents (by definition) always divide ($D = 1$). Meanwhile, strategic agents choose divide ($D = 1$) if and only if the future benefit of (potentially) keeping a good reputation outweighs the cost of foregone payoff in the present.

Intuitively, strategic agents choose $D = 1$ if and only if θ is sufficiently small. When θ is too large, the probability that A will lose its reputation in spite of choosing $D = 1$ is large enough that it is not worth it for A to forego the increased payoff in the present period associated with choosing $D = 0$. Formally, in Appendix B we show that there is some $\Psi : \mathbb{R}^2 \to \mathbb{R}$ where a necessary condition for strategic agents to choose $D = 1$ is $\theta \leq \Psi(\delta, C)$. On the other hand, if θ is sufficiently small, P chooses $S = 1$ in periods in which A has a bad reputation. In this case, A has nothing to lose from choosing $D = 0$, since it will not be harmed by having a bad reputation in the future. Combined, this intuition indicates that there is a range of θ over which a strategic A with a good reputation chooses $D = 1$.

Finally, moving to the first stage, it is clear that P chooses $S = 1$ if it sees $R = 1$ and θ is in the range in which strategic A's choose $D = 1$ in the second stage (as long as θ is not too large so as to destroy P's return with high probability). Even at very low levels of θ, if there are enough altruists in the economy, P is better off choosing $S = 1$ despite the fact that strategic A's will not divide. On the other hand, if $\theta > \Psi(\delta, C)$, P may choose $S = 1$ if there are enough altruistic agents in the economy, but only if θ is not too large (formally, if $\theta \leq \theta^*$, where θ^* is defined in the Appendix). If θ is too large, there is too high of a probability of nature destroying P's return, and P therefore chooses $S = 0$. Formally, these equilibrium actions are written as follows:

Proposition 2: For periods in which $R = 1$, P and A choose the following strategies in equilibrium:

Stage 1: P chooses $S = 1$ if and only if $\theta \leq \theta^*$; otherwise P chooses $S = 0$;
Stage 2: A chooses $D = 1$ if altruistic; if strategic, A chooses $D = 1$ if and only if $\theta \in \left(1 - \frac{1}{2\alpha}, \Psi(\delta, C)\right]$, otherwise it chooses $D = 0$;
Stage 3: P chooses $I = 1$ if and only if $\beta_i \geq C$; otherwise P chooses $I = 0$.

Next, we analyze comparative statics with respect to C and θ for the strategic A's decision to divide in the second stage. First, note that there are no clean comparative statics with respect to θ. If $1 - \frac{1}{2\alpha} \leq 0$ and $\Psi(\delta, C) > 0$, then the parameter space over which strategic A's choose $D = 1$ is decreasing in θ. If $\Psi(\delta, C) < 1 - \frac{1}{2\alpha}$, there is no value of θ for which strategic A's choose $D = 1$. Finally, if $0 < 1 - \frac{1}{2\alpha} < \Psi(\delta, C)$, the parameter space over which strategic A's choose $D = 1$ is hump-shaped in θ. Moreover, even if we see agents choosing $D = 1$ in the experiment, we cannot infer that $\Psi(\delta, C) > 1 - \frac{1}{2\alpha}$, since it may be altruistic agents choosing $D = 1$. Since there are no clean comparative static predictions with respect to θ, we do not make a formal prediction.

However, the comparative static prediction with respect to C is clean. As C increases, the likelihood of being informed upon in Stage 3 decreases, since it is less likely that A will be matched with a P with a sufficiently high β_i. Hence, as C increases there is less incentive for strategic agents to choose $D = 1$; instead they prefer to take the larger payoff in the present period. We formalize this insight in Prediction 2.

Prediction 2: For periods in which $R = 1$, the size of the parameter space over which a strategic A chooses $D = 1$ is decreasing in C, ceteris paribus.

Finally, consider comparative statics with respect to θ and C for P's decision to send in the first stage. It follows directly from Proposition 2 that P is less likely to choose $S = 1$ as θ increases. This is due primarily to the fact that higher levels of θ increase the probability that P will receive 0 return regardless of whether she is matched with a strategic or altruistic A. Meanwhile, an increase in C makes choosing $S = 1$ less attractive to P for two reasons. First, it decreases the likelihood that A will choose $D = 1$ in Stage 2 (see Prediction 2). Second, an increase in C also decreases the likelihood that a strategic A will have been punished in the past, thereby increasing their probability of having a good reputation.[9] Combined, these two effects indicate that an increase in C incentivizes

[9]Note that an increase in C also increases the attractiveness of choosing $D = 0$ to the stategic A. However, in

principals to choose $S = 0$, conditional on seeing $R = 1$. This logic is formalized as follows.

Prediction 3: For periods in which $R = 1$, the size of the parameter space over which P chooses $S = 1$ is decreasing in θ and C, ceteris paribus.

3 Experimental Design and Procedures

To investigate the effect of multilateral punishments on trade behavior, we designed a modified trust game in which uncooperative behavior could be reported to all principals (i.e., investors). In the beginning of each session, subjects are randomly assigned as the principal or agent. In the instructions we employed neutral language; for example, the principal and agent were named "participant 1" and "participant 2", respectively.[10]

In each session, there are 12 principals and 12 agents. The matching scheme follows a perfect stranger matching protocol. That is, each principal plays the same game with a new agent in each period. This is to ensure that repeated relationships are not a reason that principals report agents. Each session lasts 12 periods.

As in the model, each period consists of three stages. These stages are depicted, like before, in Figure 1. In the first stage, principals can opt out of the trade (Do Not Send in Figure 1) and earn their reservation payoff of $\Omega > 0$. If the principal decides to engage in the trade (Send), the game proceeds to Stage 2. In this stage, agents have the option to split the return on investment (Divide) or pocket the entire investment (Do Not Divide). If they split the return on investment they receive payment 2Ω, and if they pocket the investment they receive payment 4Ω. Moreover, if they split the return, nature overrides their decision, preventing principals from receiving their share, with probability $\theta \in [0, 1]$.

If the agent decides to not divide the return, or if nature overrides the agent's decision to divide, the principal does not receive her share. The principal only observes the outcome and not the

this range, the strategic A chooses $D = 0$ anyways, meaning that the salient effect on its reputation is the fact that it is less likely to have been punished in the past.

[10]For the full instructions, see Appendix A.

agent's action. If either of these events occur, the game proceeds to a third stage. In this stage, the principal has the option of informing all the prospective principals for a fee C (Inform). If the principal chooses to inform other principals, that particular agent will have a bad reputation for the remainder of game. In the beginning of each period, the principals observe whether the agents that they are matched with established a "bad" or "good" reputation. We define reputation to be a binary signal in which the agent has a "good reputation" if and only if no principal has ever informed on the agent in the past, and it has a "bad reputation" if at least one principal informed on the agent in the past.

The novel aspect of our experiment is that it provides insight into how uncertainty (θ) and costly punishment (C) affect the effectiveness of a multilateral punishment institution. To this end, we employ a 3×2 design in which treatments are based on variations in θ and C. Table 1 reports the treatments.

		Parameters			Procedure	
Uncertainty	Informing	θ	C	Ω	Sessions	Subjects per Session
No	Costless	0	0	10	3	24
No	Costly	0	5	10	3	24
Weak	Costless	0.05	0	10	3	24
Weak	Costly	0.05	5	10	3	24
Strong	Costless	0.3	0	10	3	24
Strong	Costly	0.3	5	10	3	24

Table 1: **Summary of experimental treatments**

The parameter space is designed to capture a possible "hump shaped" relationship between the principal's send decision and θ. Hence, we set θ equal to 0, 0.05, and 0.3. We label each level of θ as no nature, weak nature, or strong nature, respectively. We set C equal to 0 and 5 to test the hypothesized relationship between the principal's send decision and the informing cost. We label each level of C as costless informing and costly informing, respectively. The reservation payoff of both the principals and agents, Ω, is set to 10.

After subjects played the modified trust game, we conducted an incentivized risk elicitation task

to control for subjects' risk attitudes (Holt and Laury, 2002). In this task subjects have the option between two sets of paired lotteries. Lottery A offers a sequence of monetary prizes of \$1 or \$3, each with the constant probability of 0.5. Lottery B offers a sequence of monetary prizes of \$0.10 or \$4 in which the probability of wining the prize of \$4 is increasing in the sequence (for details, see Appendix A.2). Our measure of risk-aversion is the number of times that a subject chooses lottery B. After the risk elicitation task, in the last stage of the experiment, subjects filled out a questionnaire.

A total of 18 sessions, 3 sessions per treatment, were conducted in a laboratory at a medium-sized university in the United States. 24 subjects participated in each session, totaling 432 subjects. As noted before, upon arriving to the lab, subjects are randomly assigned as the principals or the agents. The currency used in the experiments was francs. Francs were converted to USD at rate of 20 francs to \$1. At the end of each session, the subjects' total francs were summed from all 12 periods and converted to USD. Then, for the total payment, earnings from the risk elicitation task and show up fee of \$7 were added to the earnings from the game. The range of the total payment was \$11 to \$28 and the average payment was \$18.77. Although each session lasted a maximum of 1 hour, subjects were recruited for 90 minutes.

4 Results

We begin by reporting results from the experiment as a whole. Since the theory provides different predictions depending on the reputation of the agent at the beginning of the period, we proceed to report results conditional on the agent's reputation.

4.1 Summary Statistics and Analysis

In this section, we examine principals' decision to send and agents' decision to divide independent of the agents' reputation. Table 2 reports the mean, standard error, and number of observations of the send and divide decisions across treatments. We also report the principals' decision to inform,

conditional on the agent having a good reputation.[11]

	Send			Divide			Inform		
Treatments	Mean	SE	N	Mean	SE	N	Mean	SE	N
No nature, costless informing	0.78	0.02	432	0.89	0.02	338	1.00	0.00	19
No nature, costly informing	0.64	0.02	432	0.81	0.02	276	0.50	0.08	40
Weak nature, costless informing	0.66	0.02	432	0.73	0.03	286	0.86	0.06	35
Weak nature, costly informing	0.68	0.02	432	0.72	0.03	292	0.48	0.08	42
Strong nature, costless informing	0.41	0.02	432	0.59	0.04	177	0.78	0.06	45
Strong nature, costly informing	0.44	0.02	432	0.56	0.04	190	0.31	0.05	91

Table 2: **Summary statistics.** In the Inform column, we condition Informing on the agent having a good reputation.

Conditional on informing being costless, the frequency that principals choose to send is decreasing in uncertainty (θ): 0.78 ($\theta = 0$) vs. 0.66 ($\theta = 0.05$) vs. 0.41 ($\theta = 0.30$).[12] However, conditional on information being costly, there is a slight hump-shape as θ increases, although the difference in means is not statistically significant between $\theta = 0$ and $\theta = 0.05$: 0.64 ($\theta = 0$) vs. 0.68 ($\theta = 0.05$) vs. 0.44 ($\theta = 0.30$).[13] On the other hand, conditional on θ, the cost of informing is related (and statistically significant) to the frequency principals choose to send only when $\theta = 0$ (0.78 vs 0.64, $p = 0.05$). Figure 2 depicts these patterns.

The average frequency which agents chose to divide, conditional on the game reaching Stage 2, is depicted in Figure 3. Conditional on informing cost equaling 0, the frequency that divide is chosen is decreasing in θ: 0.89 ($\theta = 0$) vs. 0.73 ($\theta = 0.05$) vs 0.59 ($\theta = 0.30$).[14] Likewise, conditional on information being costly, the frequency that divide is chosen is decreasing in θ: 0.81 ($\theta = 0$) vs. 0.72 ($\theta = 0.05$) vs 0.56 ($\theta = 0.30$).[15] As was the case for the send decision, the cost of informing is

[11] Informing on an agent that already has a bad reputation does not change the agent's reputation. We therefore focus on the more informative case where the decision to inform affects the agent's reputation.

[12] P-values for the various tests of differences are 0.05 ($\theta = 0$ vs. $\theta = 0.05$), 0.05 ($\theta = 0$ vs. $\theta = 0.30$), and 0.13 ($\theta = 0.05$ vs. $\theta = 0.30$). All p-values reported for comparisons in this paper are from two-sided Mann-Whitney-Wilcoxon rank sum tests at the session level.

[13] P-values for the various tests of differences are 0.51 ($\theta = 0$ vs. $\theta = 0.05$), 0.05 ($\theta = 0$ vs. $\theta = 0.30$), and 0.05 ($\theta = 0.05$ vs. $\theta = 0.30$).

[14] P-values for the various tests of differences are 0.05 ($\theta = 0$ vs. $\theta = 0.05$), 0.05 ($\theta = 0$ vs. $\theta = 0.30$), and 0.05 ($\theta = 0.05$ vs. $\theta = 0.30$).

[15] P-values for the various tests of differences are 0.05 ($\theta = 0$ vs. $\theta = 0.05$), 0.05 ($\theta = 0$ vs. $\theta = 0.30$), and 0.05 ($\theta = 0.05$ vs. $\theta = 0.30$).

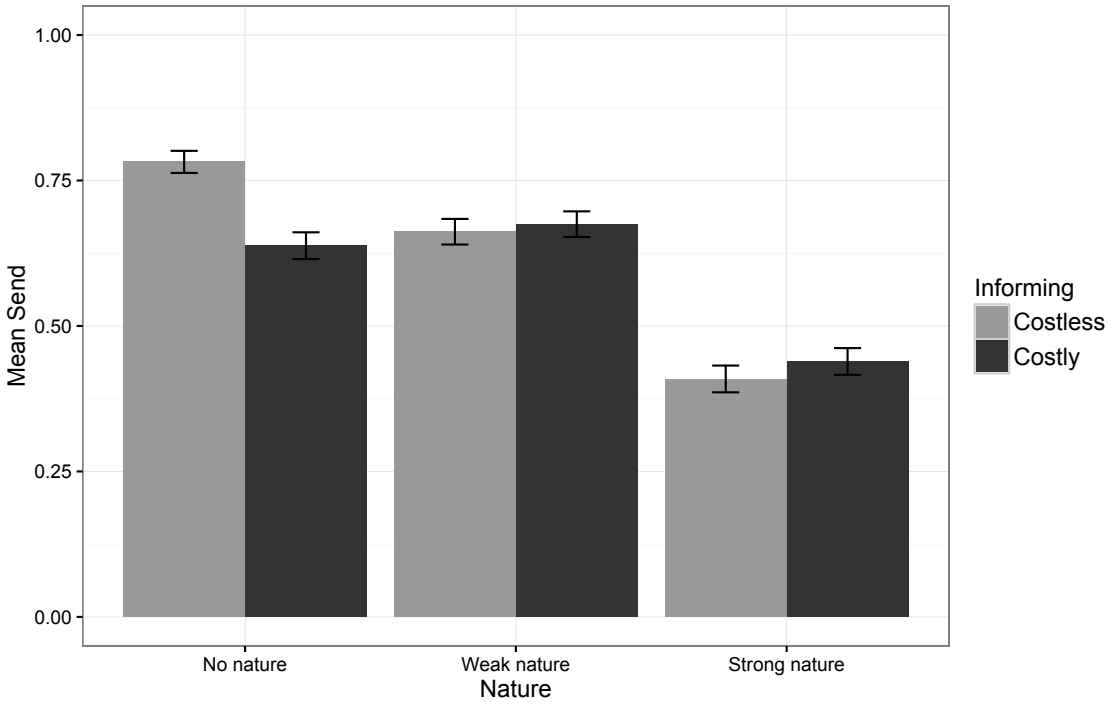

Figure 2: **Bar Plot of Mean Send by Treatment.** Error bars represent mean send plus or minus one standard error.

only significantly related to the divide frequency when $\theta = 0$ (0.89 vs. 0.81, $p = 0.05$).

Why do we observe these differences across treatments? Does reputation matter, as the model and the previous literature suggest, or does the mere imposition of uncertainty (θ) or informing cost (C) affect decisions across treatments independent of reputation? While the model does not directly speak to trends observed in the overall sample—its predictions are conditional on the agent's reputation—insight into the mechanisms driving cross-treatment differences can be gleaned from observing the send frequency across periods, as depicted in Figure 4. In period one, there is no statistically difference across *any* of the treatments except the one in which $\theta = 0$ and $C = 5$. Yet, while there is decay in the send frequency over time in all treatments, the decay rate varies by treatment. One obvious explanation for these treatment differences is that reputation matters, and principals matched with agents with good reputations send more frequently. Figure 5 supports this assertion. It reveals that the decay in sending frequency is closely and inversely associated with the fraction of agents who have a good reputation in that period.

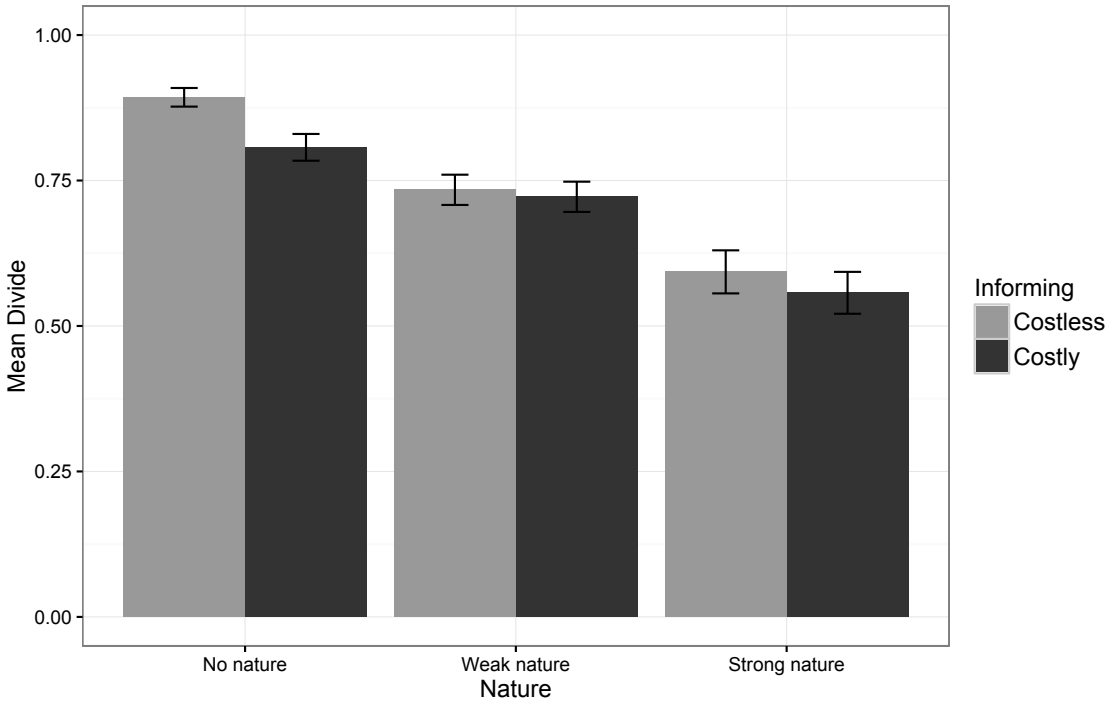

Figure 3: **Bar Plot of Mean Divide by Treatment.** Error bars represent mean divide plus or minus one standard error.

Finally, we test the effect of θ and C on the send and divide decisions in a more rigorous manner by estimating a random-effects model using the GLS estimator where the dependent variable is a send or divide dummy. In the first specification, we include treatment dummies. In the second specification, we include a reputation dummy, since Figures 4 and 5 suggest this may affect decision-making. In the third specification, we include controls for risk-aversion (obtained from the incentivized risk elicitation task (Holt and Laury, 2002)), gender, and path dependence (period dummy). We split the sample into treatments where informing is costless and when it is costly (Tables 3 and 5), and, in Tables 4 and 6, treatments where $\theta = 0$, 0.05, or 0.30.[16]

In Table 3, the omitted category is "weak nature" (i.e., $\theta = 0.05$). This permits us to test for a "hump-shape" with respect to θ, as predicted by Prediction 1 (in the case in which agents have a bad reputation). A few observations immediately jump out of Table 3. First, the agent's reputation is strongly and positively associated with the send frequency. This is not surprising, and it suggests, at least in part, that principals act in accordance with the incentives provided by

[16]We report regression results in which all sessions are included in one specification in Appendix Tables C2-C4.

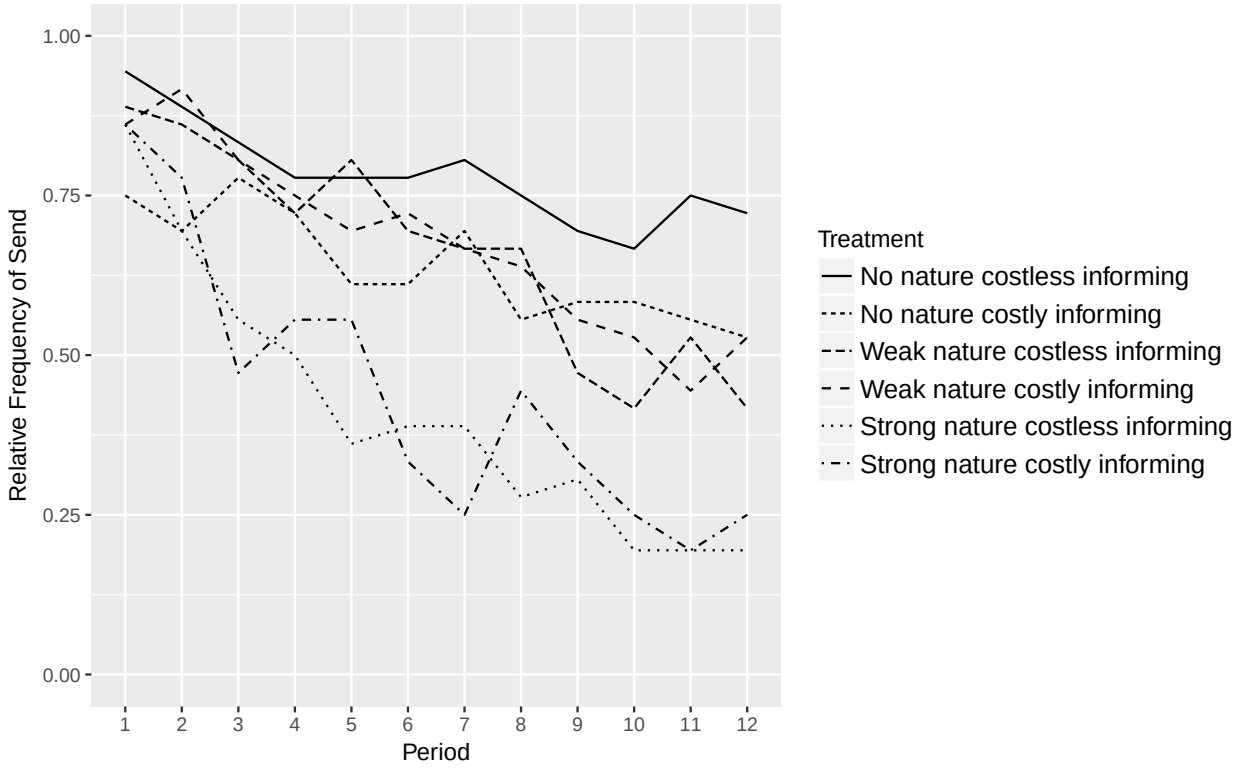

Figure 4: **Relative Frequency of Send by Treatment and Period.**

the institution. Focusing on the most robust specifications presented in Columns (3) and (6), it appears that θ is statistically unrelated to the send decision when informing is costless ($C = 0$) but is "hump-shaped" when informing is costly ($C = 5$). Indeed, these results indicate that principals are more likely to send in the $\theta = 0.05$ treatment than in the $\theta = 0$ or $\theta = 0.30$ treatments. In other words, when informing is costly, a modest amount of uncertainty can *increase overall welfare*, since total welfare is determined primarily by the principal's send decision.[17] The model suggests that this result arises from principals being more willing to send to agents with *bad* reputations when uncertainty is modest relative to when there is no uncertainty. We test whether this is the case in the following two sections.

In Table 4, the omitted category is "costless informing" (i.e. $C = 0$). The informing cost is negatively and significantly related to the frequency sent only when there is no nature (i.e., $\theta = 0$). The results related to the principal's send decision are summarized in Result 1.

[17]Total welfare is also affected by the random element θ, since the principal's payoff is reduced with probability θ if the agent chooses to divide.

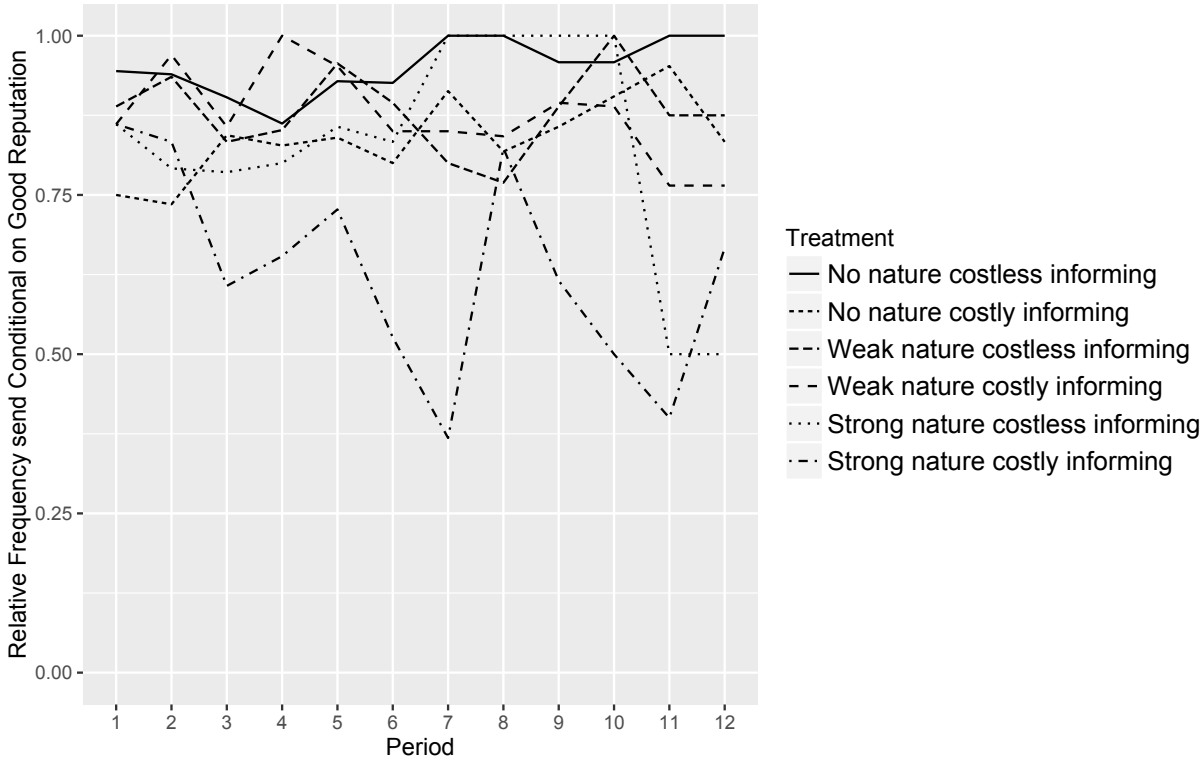

Figure 5: **Relative Frequency of Send Conditional on Being Matched with an Agent with a Good Reputation by Treatment and Period.**

Result 1: The frequency that principals send is hump-shaped in θ when informing is costly, unrelated to θ when informing is costless, and decreasing in C when $\theta = 0$.

The results presented in Table 5 suggest that agents are less likely to divide as θ increases, although some of the estimates are noisy. Meanwhile, Table 6 indicates that the informing cost is negatively and significantly correlated to the frequency divide when there is no nature (i.e., $\theta = 0$) or weak nature (i.e., $\theta = 0.05$). The results related to the agent's divide decision are summarized in Result 2.

Result 2: The frequency that agents divide is decreasing in θ, and it is decreasing in C when $\theta = 0$ and $\theta = 0.05$.

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable = Send (0/1)					
	Costless Informing			Costly Informing		
No Nature	0.120**	-0.016	-0.013	-0.037	-0.085***	-0.078***
$\theta = 0$	(0.057)	(0.058)	(0.054)	(0.053)	(0.030)	(0.025)
Strong Nature	-0.252***	-0.101	-0.110	-0.236***	-0.196***	-0.193***
$\theta = 0.3$	(0.093)	(0.087)	(0.083)	(0.060)	(0.036)	(0.033)
Reputation		0.573***	0.549***		0.581***	0.556***
		(0.040)	(0.056)		(0.045)	(0.054)
Risk			0.008			0.007
			(0.013)			(0.008)
Female			0.029			-0.025
			(0.045)			(0.043)
Period			-0.007			-0.009
			(0.005)			(0.006)
Intercept	0.662***	0.361***	0.371***	0.676***	0.311***	0.371***
	(0.043)	(0.053)	(0.108)	(0.037)	(0.039)	(0.074)
Observations	1296	1296	1296	1296	1296	1296
R^2	0.102	0.383	0.387	0.044	0.353	0.358

Standard errors clustered by session in parentheses. * p<0.1, ** p<0.05, *** p<0.01.
Weak Nature ($\theta = 0.05$) is the omitted category.

Table 3: **Regression Analysis of the Decision to Send**.

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable = Send (0/1)					
	No Nature		Weak Nature		Strong Nature	
Costly Informing	-0.144***	-0.094**	0.014	-0.037	0.030	-0.103
$C = 5$	(0.055)	(0.039)	(0.059)	(0.054)	(0.098)	(0.087)
Reputation		0.704***		0.486***		0.454***
		(0.042)		(0.045)		(0.060)
Risk		0.005*		0.002		0.015
		(0.003)		(0.018)		(0.015)
Female		-0.082**		0.026		0.057
		(0.038)		(0.060)		(0.040)
Period		0.006		-0.012***		-0.023***
		(0.004)		(0.004)		(0.004)
Intercept	0.782***	0.223***	0.662***	0.467***	0.410***	0.349***
	(0.038)	(0.055)	(0.045)	(0.139)	(0.085)	(0.104)
Observations	864	864	864	864	864	864
R^2	0.025	0.472	0.000	0.287	0.000	0.289

Standard errors clustered by session in parentheses. * p<0.1, ** p<0.05, *** p<0.01.
Costless Informing ($C = 0$) is the omitted category.

Table 4: **Regression Analysis of the Decision to Send.**

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable = Divide (0/1)					
	Costless Informing			Costly Informing		
No Nature	0.147***	0.072*	0.082**	0.098**	0.064**	0.080**
$\theta = 0$	(0.048)	(0.043)	(0.041)	(0.041)	(0.029)	(0.033)
Strong Nature	-0.147*	-0.089	-0.107	-0.148***	-0.164***	-0.196***
$\theta = 0.3$	(0.079)	(0.086)	(0.087)	(0.038)	(0.043)	(0.054)
Reputation		0.357***	0.301***		0.270***	0.190***
		(0.037)	(0.044)		(0.053)	(0.060)
Risk			0.017**			-0.018
			(0.008)			(0.019)
Female			0.048			0.089
			(0.038)			(0.061)
Period			-0.013***			-0.019***
			(0.004)			(0.003)
Intercept	0.693***	0.460***	0.484***	0.648***	0.448***	0.646***
	(0.043)	(0.059)	(0.089)	(0.012)	(0.043)	(0.096)
Observations	801	801	801	758	758	758
R^2	0.078	0.277	0.265	0.045	0.179	0.144

Standard errors clustered by session in parentheses. * p<0.1, ** p<0.05, *** p<0.01.
Weak Nature ($\theta = 0.05$) is the omitted category.

Table 5: **Regression Analysis of the Decision to Divide**.

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable = Divide (0/1)					
	No Nature		Weak Nature		Strong Nature	
Costly Informing	-0.090*	-0.092***	-0.042	-0.072*	-0.041	-0.136
$C = 5$	(0.046)	(0.026)	(0.047)	(0.039)	(0.077)	(0.090)
Reputation		0.301***		0.355***		0.148***
		(0.066)		(0.048)		(0.040)
Risk		-0.016		0.001		0.024
		(0.017)		(0.010)		(0.024)
Female		0.066		0.011		0.147
		(0.045)		(0.059)		(0.096)
Period		-0.008**		-0.016***		-0.024***
		(0.004)		(0.004)		(0.008)
Intercept	0.836***	0.672***	0.692***	0.540***	0.547***	0.422***
	(0.023)	(0.125)	(0.045)	(0.107)	(0.068)	(0.116)
Observations	614	614	578	578	367	367
R^2	0.014	0.161	0.000	0.268	0.001	0.092

Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.
Costless Informing ($C = 0$) is the omitted category.

Table 6: **Regression Analysis of the Decision to Divide.**

4.2 Effect of Reputation

In this section, we present results of principals' send and agents' divide decisions conditional on the agent's reputation at the beginning of the period. These results provide a direct test of the theory laid out in Section 2.

4.2.1 Bad Reputation

We begin by summarizing the send and divide decisions conditional on agents beginning the period with a bad reputation. Prediction 1 indicates that the send frequency is U-shaped in C and is hump-shaped in θ. Table 7, which summarizes the decisions of the principals and agents, provides some preliminary evidence in support of these predictions. Conditional on C, the send frequency appears hump-shaped in θ. Meanwhile, conditional on θ, the send frequency is decreasing in C. Although we cannot test for the hypothesized U-shape with only two values of C, the fact that the send frequency is decreasing at $C = 0$ is consistent with a U-shape. This is easily seen in Figure 6, which depicts the means and standard errors of the conditional send decision in each treatment.[18]

	Send|bad reputation			Divide|bad reputation		
Treatments	Mean	SE	N	Mean	SE	N
No nature, costless informing	0.25	0.04	102	0.32	0.10	25
No nature, costly informing	0.17	0.03	125	0.38	0.11	21
Weak nature, costless informing	0.42	0.03	205	0.38	0.05	86
Weak nature, costly informing	0.33	0.04	161	0.19	0.05	53
Strong nature, costless informing	0.26	0.02	319	0.40	0.05	83
Strong nature, costly informing	0.15	0.03	191	0.28	0.08	29

Table 7: **Send and Divide Conditional on Bad Reputation**

We test Prediction 1 more rigorously by estimating a random-effects model using the GLS estimator where the dependent variable is a send or divide dummy. These specifications are similar to those reported in the previous section, with the exception that we no longer control for reputation (since we confine observations only to those periods in which an agent has a bad reputation). Tables

[18] Appendix Figure C1 depicts the relative frequency of send conditional on bad reputation by treatment and period. In period 1, no agent has a bad reputation. Hence, the starting point is period 2.

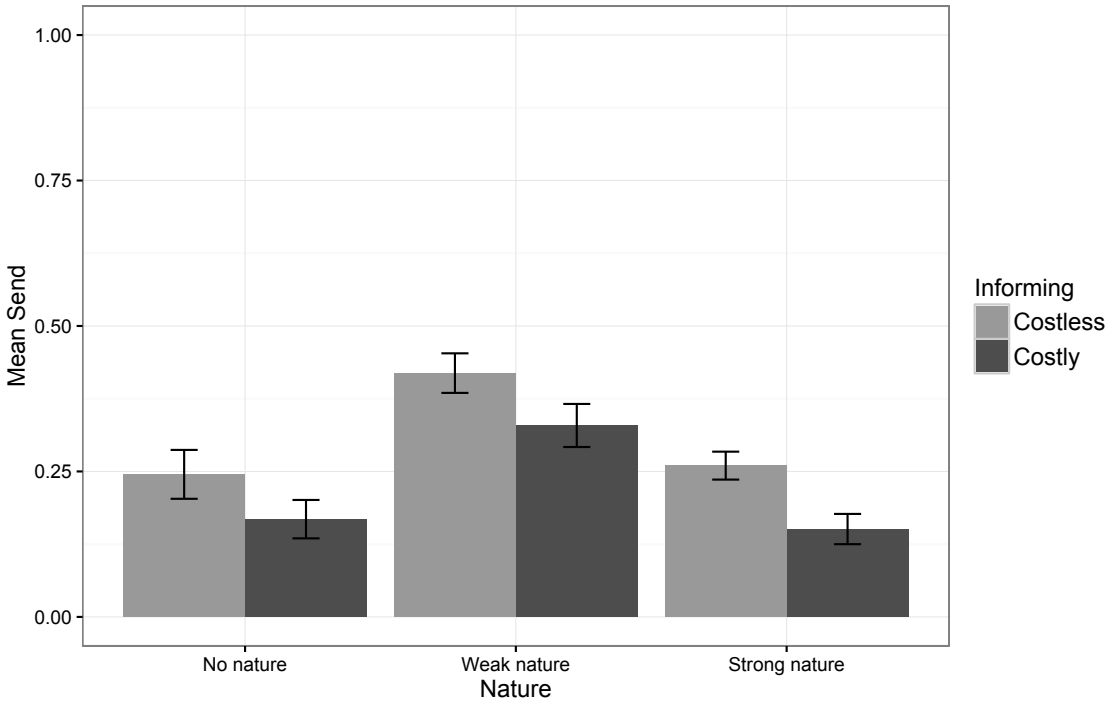

Figure 6: **Bar Plot of Mean Send Conditional on Bad Reputation by Treatment.** Error bars represent mean send plus or minus one standard error.

8 and 9 report the results. Table 8 shows strong support for Prediction 1. The regression results indicate that principals send significantly less frequently upon seeing a bad reputation—between 13.6 and 17.0 percentage points less frequently—in the $\theta = 0$ and $\theta = 0.3$ treatments relative to the $\theta = 0.05$ treatment. In other words, the relationship between the send frequency and θ, conditional on the agent having a bad reputation, is hump-shaped.

The results reported in Table 9 are also consistent with Prediction 1, although they are more noisily estimated. In all regressions, the point estimates indicate that principals send between 7.8 and 10.8 percentage points less frequently when informing is costly relative to when it is not, conditional on being matched with an agent with a bad reputation. However, these estimates are statistically significant only when $\theta = 0$. These findings are summarized in Result 3.

Result 3: Conditional on being matched with an agent with a bad reputation, the principals' send frequency is hump-shaped in θ and decreasing (although noisily estimated) in C.

Next, we turn to the agents' divide decision. The model predicts that a set fraction of agents

	(1)	(2)	(3)	(4)
	Dependent Variable = Send (0/1)			
	Costless Informing		Costly Informing	
No Nature	-0.124*	-0.136*	-0.155***	-0.152***
$\theta = 0$	(0.071)	(0.075)	(0.042)	(0.050)
Strong Nature	-0.140*	-0.170**	-0.166**	-0.166**
$\theta = 0.3$	(0.085)	(0.083)	(0.078)	(0.082)
Risk		0.005		-0.003
		(0.023)		(0.017)
Female		0.057		0.047
		(0.048)		(0.065)
Period		-0.026***		-0.021***
		(0.005)		(0.006)
Intercept	0.400***	0.578***	0.324***	0.478***
	(0.055)	(0.158)	(0.036)	(0.108)
Observations	626	626	477	477
R^2	0.026	0.055	0.038	0.060

Standard errors clustered by session in parentheses.
* p<0.1, ** p<0.05, *** p<0.01.
Weak Nature ($\theta = 0.05$) is the omitted category.

Table 8: **Regression Analysis of the Effect of Bad Reputation on the Decision to Send.**

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable = Send (0/1)					
	No Nature		Weak Nature		Strong Nature	
Costly Informing	-0.108**	-0.102*	-0.078	-0.093	-0.102	-0.087
$C = 5$	(0.051)	(0.053)	(0.068)	(0.076)	(0.097)	(0.093)
Risk		0.003		-0.029		0.018
		(0.018)		(0.031)		(0.017)
Female		0.005		0.002		0.106*
		(0.070)		(0.062)		(0.056)
Period		-0.020*		-0.026***		-0.025***
		(0.010)		(0.010)		(0.003)
Intercept	0.279***	0.425***	0.401***	0.740***	0.260***	0.314***
	(0.046)	(0.131)	(0.057)	(0.221)	(0.066)	(0.083)
Observations	227	227	366	366	510	510
R^2	0.009	0.018	0.008	0.045	0.016	0.058

Standard errors clustered by session in parentheses.
* p<0.1, ** p<0.05, *** p<0.01.
Costless Informing ($C = 0$) is the omitted category.

Table 9: **Regression Analysis of the Effect of Bad Reputation on the Decision to Send.**

(denoted "altruists" in the model) will divide conditional on having a bad reputation, and this fraction should not be a function of θ or C. We find mixed evidence in support of this prediction. The results reported in Table 10 indicate that agents with bad reputations choose to divide more in the $\theta = 0$ treatment when informing is costly. Table 11 similarly reports, in column (4), that agents with a bad reputation are less likely to divide when informing is costly and $\theta = 0.05$. We have no obvious explanation for these results, although we note that the sample is small due to the fact that principals do not frequently send conditional on seeing a bad reputation. Across all treatments, only 43% of agents had a bad reputation in any one given period, and only 27% of principals chose to send upon being matched with an agent with a bad reputation.

	(1)	(2)	(3)	(4)
	Dependent Variable = Divide (0/1)			
	Costless Informing		Costly Informing	
No Nature	0.036	0.033	0.172***	0.223***
$\theta = 0$	(0.122)	(0.139)	(0.027)	(0.054)
Strong Nature	-0.003	-0.010	0.022	0.011
$\theta = 0.3$	(0.134)	(0.117)	(0.113)	(0.134)
Risk		0.034		-0.036
		(0.022)		(0.033)
Female		0.225***		0.076
		(0.082)		(0.108)
Period		-0.014		-0.024**
		(0.011)		(0.010)
Intercept	0.396***	0.261*	0.213***	0.504***
	(0.097)	(0.147)	(0.026)	(0.111)
Observations	194	194	103	103
R^2	0.002	0.029	0.025	0.065

Standard errors clustered by session in parentheses.
* p<0.1, ** p<0.05, *** p<0.01.
Weak Nature ($\theta = 0.05$) is the omitted category.

Table 10: **Regression Analysis of the Effect of Bad Reputation on the Decision to Divide.**

4.2.2 Good Reputation

We conclude this section by analyzing the periods in which agents had a good reputation. Predictions 2 and 3 indicate that, conditional on the agent having a good reputation, the fraction of

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable = Divide (0/1)					
	No Nature		Weak Nature		Strong Nature	
Costly Informing	-0.048	0.077	-0.171	-0.213**	-0.147	-0.156
$C = 5$	(0.080)	(0.117)	(0.104)	(0.099)	(0.138)	(0.156)
Risk		-0.097***		0.025*		0.054**
		(0.021)		(0.013)		(0.025)
Female		0.455***		0.130		0.016
		(0.056)		(0.111)		(0.077)
Period		0.005		-0.030**		-0.009
		(0.025)		(0.015)		(0.011)
Intercept	0.442***	0.553*	0.398***	0.484***	0.397***	0.238***
	(0.078)	(0.332)	(0.100)	(0.169)	(0.078)	(0.080)
Observations	46	46	139	139	112	112
R^2	0.004	0.330	0.042	0.069	0.012	0.075

Standard errors clustered by session in parentheses. * p<0.1, ** p<0.05, *** p<0.01.
Costless Informing ($C = 0$) is the omitted category.

Table 11: **Regression Analysis of the Effect of Bad Reputation on the Decision to Divide.**

principals sending should be decreasing in θ and C, and the fraction of agents dividing should be decreasing in C. Table 12 reports the average frequency that principals chose send and agents chose divide, conditional on the agent having a good reputation, and Figure 7 reports this information graphically.[19] A glance at the means provides modest support for Predictions 2 and 3. Conditional on informing being costless, both the send and divide frequencies are indeed decreasing in θ. However, when informing is costly, both the send and divide frequencies appear somewhat hump-shaped in θ. Further, while both the send and divide frequencies are decreasing in C, conditional on θ, this effect is small at $\theta = 0.05$.

In order to provide statistical support for Predictions 2 and 3, we conduct regression analyses of the principals' decision to send and the agents' decision to divide, conditional on agents having a good reputation. These results, reported in Tables 13-16, estimate the same models as Tables 8-11, restricting the sample to periods in which agents have a good reputation.

The regressions reported in Table 13 provide some support for Prediction 3. The point estimates in Column (2) indicate that when informing is costless, the fraction of principals choosing send is

[19]Figure 5 depicts the relative frequency of send by treatment and period.

Treatments	Send|good reputation			Divide|good reputation		
	Mean	SE	N	Mean	SE	N
No nature, costless informing	0.95	0.01	330	0.94	0.01	313
No nature, costly informing	0.83	0.02	307	0.84	0.02	255
Weak nature, costless informing	0.88	0.02	227	0.89	0.02	200
Weak nature, costly informing	0.88	0.02	271	0.84	0.02	239
Strong nature, costless informing	0.83	0.04	113	0.77	0.04	94
Strong nature, costly informing	0.67	0.03	241	0.61	0.04	161

Table 12: **Send and Divide Conditional on Good Reputation**

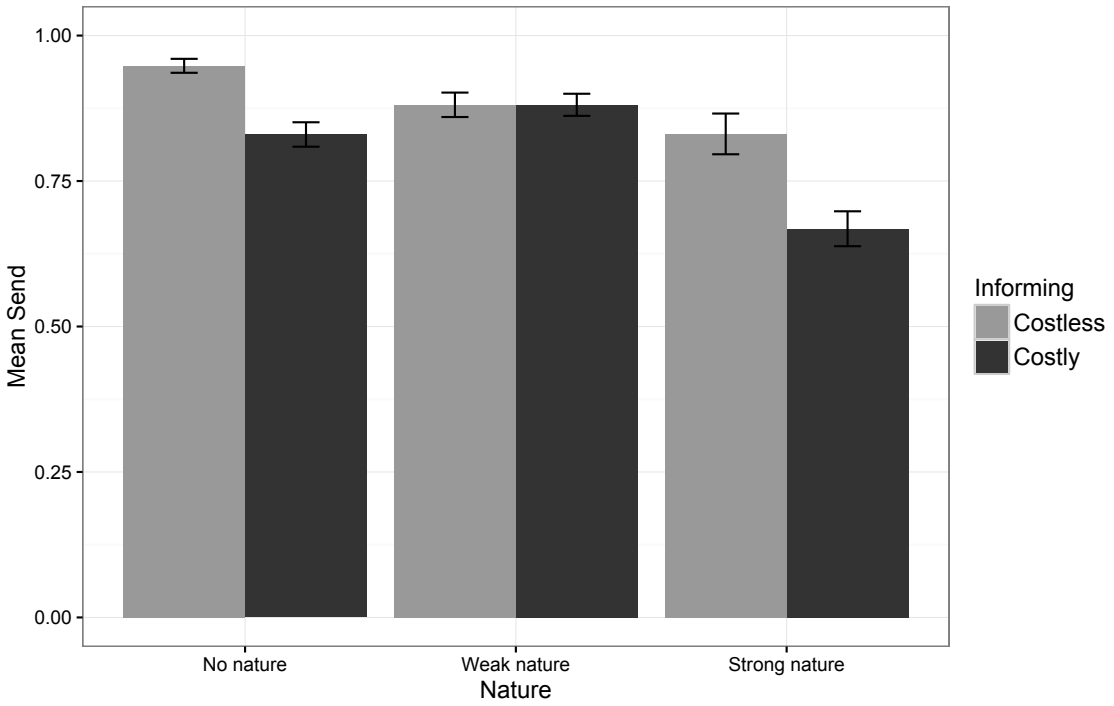

Figure 7: **Bar Plot of Mean Send Conditional on Good Reputation by Treatment.** Error bars represent mean send plus or minus one standard error.

decreasing in θ. However, these results are not statistically significant, and we do not interpret them as such. Column (4) indicates that when informing is costly, principals send less often when uncertainty is high ($\theta = 0.3$), but there is no statistically significant difference in the send frequency between the $\theta = 0$ and $\theta = 0.05$ treatments.

Meanwhile, the results reported in Table 14 indicate that the fraction of principals choosing send is decreasing in C when $\theta = 0$. However, the effect of C on the principals' send decision is

	(1)	(2)	(3)	(4)
	Dependent Variable = Send (0/1)			
	Costless Informing		Costly Informing	
No Nature	0.064	0.060	-0.050	-0.035
$\theta = 0$	(0.056)	(0.051)	(0.051)	(0.048)
Strong Nature	-0.056	-0.057	-0.214***	-0.206***
$\theta = 0.3$	(0.094)	(0.090)	(0.033)	(0.034)
Risk		0.010*		0.015
		(0.006)		(0.010)
Female		0.033		-0.068
		(0.050)		(0.042)
Period		0.001		-0.005
		(0.004)		(0.007)
Intercept	0.885***	0.823***	0.887***	0.889***
	(0.055)	(0.076)	(0.011)	(0.074)
Observations	670	670	819	819
R^2	0.023	0.026	0.048	0.066

Standard errors clustered by session in parentheses.
* p<0.1, ** p<0.05, *** p<0.01.
Weak Nature ($\theta = 0.05$) is the omitted category.

Table 13: **Regression Analysis of the Effect of Good Reputation on the Decision to Send.**

practically 0 when $\theta = 0.05$, and it is statistically insignificant when $\theta = 0.3$. These findings are summarized in Result 4.

Result 4: Conditional on being matched with an agent with a good reputation, the principals' send frequency is statistically unrelated to θ when informing is costless but it is weakly decreasing in θ when informing is costly. Meanwhile, the principals' send frequency is decreasing in C when $\theta = 0$, but it is statistically unrelated to C when $\theta = 0.05$ or $\theta = 0.3$.

Finally, we turn to the agents' divide decision. Prediction 2 indicated that the fraction of agents dividing is decreasing in C. The results reported in Table 16 largely support this conjecture. While, columns (2) and (4) of Table 15 reveal that the frequency with which agents choose to divide is decreasing in θ, with all point estimates being marginally significant, columns (2), (4), and (6) in Table 16 indicate that the frequency with which agents choose to divide is decreasing in C, with all point estimates being marginally significant. These findings are summarized in Result 5.

Result 5: Conditional on having a good reputation at the beginning of the period, the frequency

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable = Send (0/1)					
	No Nature		Weak Nature		Strong Nature	
Costly Informing	-0.112**	-0.092*	0.003	0.009	-0.157*	-0.117
$C = 5$	(0.053)	(0.054)	(0.058)	(0.048)	(0.089)	(0.097)
Risk		0.006		0.023*		0.012
		(0.004)		(0.014)		(0.014)
Female		-0.115***		0.042		0.027
		(0.030)		(0.065)		(0.067)
Period		0.010**		-0.007*		-0.023**
		(0.004)		(0.004)		(0.009)
Intercept	0.949***	0.923***	0.885***	0.804***	0.830***	0.843***
	(0.012)	(0.048)	(0.057)	(0.105)	(0.084)	(0.139)
Observations	637	637	498	498	354	354
R^2	0.035	0.086	0.000	0.020	0.028	0.056

Standard errors clustered by session in parentheses. * p<0.1, ** p<0.05, *** p<0.01.
Costless Informing ($C = 0$) is the omitted category.

Table 14: **Regression Analysis of the Effect of Good Reputation on the Decision to Send.**

with which agents choose to divide is decreasing in θ and C (with marginal statistical significance).

	(1)	(2)	(3)	(4)
	Dependent Variable = Divide (0/1)			
	Costless Informing		Costly Informing	
No Nature	0.054	0.081*	0.039	0.060
$\theta = 0$	(0.051)	(0.048)	(0.042)	(0.045)
Strong Nature	-0.127	-0.155	-0.165***	-0.208***
$\theta = 0.3$	(0.100)	(0.098)	(0.028)	(0.033)
Risk		0.002		-0.004
		(0.013)		(0.019)
Female		0.009		0.119
		(0.064)		(0.075)
Period		-0.018***		-0.021***
		(0.004)		(0.004)
Intercept	0.805***	0.871***	0.708***	0.765***
	(0.044)	(0.064)	(0.009)	(0.099)
Observations	607	607	655	655
R^2	0.038	0.028	0.057	0.036

Standard errors clustered by session in parentheses.
* p<0.1, ** p<0.05, *** p<0.01.
Weak Nature ($\theta = 0.05$) is the omitted category.

Table 15: **Regression Analysis of the Effect of Good Reputation on the Decision to Divide.**

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable = Divide (0/1)					
	No Nature		Weak Nature		Strong Nature	
Costly Informing	-0.110**	-0.115***	-0.092**	-0.068	-0.135	-0.141
$C = 5$	(0.050)	(0.035)	(0.046)	(0.047)	(0.096)	(0.099)
Risk		-0.012		-0.001		0.006
		(0.014)		(0.018)		(0.037)
Female		0.034		-0.025		0.213*
		(0.060)		(0.078)		(0.121)
Period		-0.012***		-0.015***		-0.035***
		(0.004)		(0.005)		(0.007)
Intercept	0.858***	0.991***	0.806***	0.902***	0.681***	0.639***
	(0.027)	(0.081)	(0.045)	(0.083)	(0.092)	(0.156)
Observations	568	568	439	439	255	255
R^2	0.024	0.020	0.004	0.001	0.025	0.035

Standard errors clustered by session in parentheses. * p<0.1, ** p<0.05, *** p<0.01.
Costless Informing ($C = 0$) is the omitted category.

Table 16: **Regression Analysis of the Effect of Good Reputation on the Decision to Divide.**

In sum, the results presented in Section 4 confirm the model's key predictions, even if not all of the predictions are confirmed at conventional levels of statistical precision. We find that the principals' send decision is hump-shaped with respect to θ when informing is costly. This entails that welfare is *improved* when there is a modest amount of uncertainty relative to when there is zero or a high level of uncertainty. The results presented in Section 4.2.1 confirm the model's insight that these outcomes are primarily driven by principals trusting agents with a *bad reputation* to a greater degree when there is a modest amount of uncertainty.

5 Conclusion

In this paper, we conduct a modified trust game in the presence of uncertainty and a multilateral punishment institution. We find that a modest amount of uncertainty regarding whether the agent actually cheated the principal or not can increase the overall level of trade, since principals are more willing to trust agents who have a bad reputation, knowing that their reputation may have been the result of bad luck rather than bad actions.

Our results have numerous implications for the functioning of reputation-based institutions. First, they suggest that these institutions not only function when a modest amount of uncertainty is present—they function *better* in terms of total welfare. Second, we find that the cost of giving a cheating agent a bad reputation matters only in the absence of uncertainty. Hence, it is possible that pro-social behavioral characteristics encourage people to use such institutions even when doing so is costly and expending resources to provide reputation is a public good. Finally, our results provide insight into the behavioral characteristics that allow institutions to operate in the manner in which they are designed. In our experiment, trade flourished in all of the treatments (albeit to varying degrees) despite an absence of economic incentive. We propose that the interaction between pro-social behavioral features and institutional design encourages the regular presence of trade.

Our results raise numerous questions related to institutional design and formation. First, the multilateral punishment institution in our experiment was exogenously imposed. When such institutions arise endogenously—as they do in most real world situations—does modest uncertainty arise conterminously as a *feature*, and not a bug, of the institution? Or, is it simply that multilateral punishment institutions are more likely to arise when there naturally exists some modest level of uncertainty (due to the state of monitoring or information technology)? Second, how will exogenous events that change the level of uncertainty in principal-agent relations affect the functioning of multilateral punishment institutions? For instance, if information or communication technologies improve, it may be possible for merchants to tell with higher probability whether or not their agents are lying to them regarding the quality of the goods they receive at the docks. Intuitively, such technological improvements should increase the level of trade, since the incentive to cheat is lower. Our results suggest, however, that under some conditions the overall level of trade may decrease. These open questions are testable empirically and experimentally, and finding their answers is an important task for those interested in understanding why and how economic institutions work.

References

Ali, N., Miller, D. and Yang, D. (2017). Renegotiation-proof multilateral enforcement. *Working paper*.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, **100** (401), 464–477.

Bartling, B., Engl, F. and Weber, R. A. (2014). Does willful ignorance deflect punishment?–an experimental study. *European Economic Review*, **70**, 512–524.

Berg, J., Dickhaut, J. and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, **10** (1), 122–142.

Bernstein, L. (1992). Opting out of the legal system: Extralegal contractual relations in the diamond industry. *The Journal of Legal Studies*, **21** (1), 115–157.

— (2001). Private commercial law in the cotton industry: Creating cooperation through rules, norms, and institutions. *Michigan Law Review*, **99** (7), 1724–90.

Besley, T. and Coate, S. (1995). Group lending, repayment incentives and social collateral. *Journal of development economics*, **46** (1), 1–18.

Boerner, L. and Ritschl, A. (2009). The economic history of sovereignty: communal responsibility, the extended family, and the firm. *Journal of Institutional and Theoretical Economics JITE*, **165** (1), 99–112.

Bohnet, I., Frey, B. S. and Huck, S. (2001). More order with less law: On contract enforcement, trust, and crowding. *American Political Science Review*, **95** (1), 131–144.

—, Greig, F., Herrmann, B. and Zeckhauser, R. (2008). Betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey, and the united states. *American Economic Review*, pp. 294–310.

—, Harmgart, H., Tyran, J.-R. *et al.* (2005). Learning trust. *Journal of the European Economic Association*, **3** (2-3), 322–329.

— and Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, **55** (4), 467–484.

Bracht, J. and Feltovich, N. (2009). Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game. *Journal of Public Economics*, **93** (9), 1036–1044.

Carpenter, J., Connolly, C. and Myers, C. K. (2008). Altruistic behavior in a representative dictator experiment. *Experimental Economics*, **11** (3), 282–298.

Cason, T. N., Saijo, T. and Yamato, T. (2002). Voluntary participation and spite in public good provision experiments: an international comparison. *Experimental Economics*, **5** (2), 133–153.

Cassar, A., Friedman, D. and Schneider, P. H. (2009). Cheating in markets: A laboratory experiment. *Journal of Economic Behavior & Organization*, **72** (1), 240–259.

—, — and — (2010). A laboratory investigation of networked markets. *The Economic Journal*, **120** (547), 919–943.

Charness, G., Du, N. and Yang, C.-L. (2011). Trust and trustworthiness reputations in an investment game. *Games and Economic Behavior*, **72** (2), 361–375.

— and Haruvy, E. (2002). Altruism, equity, and reciprocity in a gift-exchange experiment: an encompassing approach. *Games and Economic Behavior*, **40** (2), 203–231.

— and Kuhn, P. (2011). Lab labor: What can labor economists learn from the lab? *Handbook of labor economics*, **4**, 229–330.

Clay, K. (1997). Trade without law: private-order institutions in mexican california. *The Journal of Law, Economics, and Organization*, **13** (1), 202–231.

Fehr, E. and Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, **25** (2), 63–87.

—, Gächter, S. and Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica: journal of the Econometric Society*, pp. 833–860.

—, Klein, A. and Schmidt, K. M. (2007). Fairness and contract design. *Econometrica*, **75** (1), 121–154.

— and List, J. A. (2004). The hidden costs and returns of incentives—trust and trustworthiness among ceos. *Journal of the European Economic Association*, **2** (5), 743–771.

— and Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature*, **422** (6928), 137–140.

— and Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, **1**, 615–691.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, **10** (2), 171–178.

Frey, B. S. and Meier, S. (2004). Social comparisons and pro-social behavior: Testing" conditional cooperation" in a field experiment. *The American Economic Review*, **94** (5), 1717–1722.

Ghatak, M. and Guinnane, T. W. (1999). The economics of lending with joint liability: theory and practice. *Journal of development economics*, **60** (1), 195–228.

Ghosh, P. and Ray, D. (1996). Cooperation in community interaction without information flows. *The Review of Economic Studies*, **63** (3), 491–519.

Greif, A. (1993). Contract enforceability and economic institutions in early trade: The maghribi traders' coalition. *The American economic review*, pp. 525–548.

— (2000). The fundamental problem of exchange: a research agenda in historical institutional analysis. *European Review of Economic History*, **4** (03), 251–284.

— (2002). Institutions and impersonal exchange: from communal to individual responsibility. *Journal of Institutional and Theoretical Economics JITE*, **158** (1), 168–204.

— (2004). Impersonal exchange without impartial law: the community responsibility system. *Chi. J. Int'l L.*, **5**, 109.

—, Milgrom, P. and Weingast, B. R. (1994). Coordination, commitment, and enforcement: The case of the merchant guild. *Journal of political economy*, **102** (4), 745–776.

Gurdal, M. Y., Miller, J. B. and Rustichini, A. (2013). Why blame? *Journal of Political Economy*, **121** (6), 1205–1247.

Ho, B. and Huffman, D. (2017). Trust and the law. *In: Handbook of Behavioral Economics and the Law. Forthcoming.*

Holt, C. and Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review*, **92** (5), 1644–1655.

Hong, K. and Bohnet, I. (2007). Status and distrust: The relevance of inequality and betrayal aversion. *Journal of Economic Psychology*, **28** (2), 197–213.

Kimbrough, E. O. and Reiss, J. P. (2012). Measuring the distribution of spitefulness. *PloS one*, **7** (8), e41812.

— and Rubin, J. (2015). Sustaining group reputation. *The Journal of Law, Economics, and Organization*, **31** (3), 599–628.

Kőszegi, B. (2014). Behavioral contract theory. *Journal of Economic Literature*, **52** (4), 1075–1118.

Kranton, R. E. (1996). Reciprocal exchange: a self-sustaining system. *The American Economic Review*, pp. 830–851.

— and Minehart, D. F. (2001). A theory of buyer-seller networks. *The American Economic Review*, **91** (3), 485–508.

Kreps, D. M., Milgrom, P., Roberts, J. and Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, **27** (2), 245–252.

— and Wilson, R. (1982). Reputation and imperfect information. *Journal of Economic Theory*, **27**, 253–279.

Leeson, P. T. (2008). Social distance and self-enforcing exchange. *The Journal of Legal Studies*, **37** (1), 161–188.

Levin, J. (2009). The dynamics of collective reputation. *The BE Journal of Theoretical Economics*, **9** (1).

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of economic dynamics*, **1** (3), 593–622.

Meier, S. (2006). A survey of economic theories and field evidence on pro-social behavior.

Milgrom, P. R., North, D. C. *et al.* (1990). The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. *Economics & Politics*, **2** (1), 1–23.

Okazaki, T. (2005). The role of the merchant coalition in pre-modern japanese economic development: an historical institutional analysis. *Explorations in Economic History*, **42** (2), 184–201.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge, Cambridge University Press.

— (2005). *Understanding institutional diversity*, vol. 241. Princeton University Press Princeton, NJ.

Pillutla, M. M. and Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational behavior and human decision processes*, **68** (3), 208–224.

Posner, R. A. and Rasmusen, E. B. (1999). Creating and enforcing norms, with special reference to sanctions. *International Review of law and economics*, **19** (3), 369–382.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Resnick, P., Zeckhauser, R., Swanson, J. and Lockwood, K. (2006). The value of reputation on ebay: A controlled experiment. *Experimental economics*, **9** (2), 79–101.

Richardson, G. (2005). Craft guilds and christianity in late-medieval england: A rational-choice analysis. *Rationality and society*, **17** (2), 139–189.

Rubin, J. and Sheremeta, R. (2016). Principal–agent settings with random shocks. *Management Science*, **62** (4), 985–999.

Smith, V. L. and Wilson, B. J. (2017). Sentiments, conduct, and trust in the laboratory. *Social Philosophy and Policy. 34(1)*.

Tirole, J. (1996). A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *The Review of Economic Studies*, **63** (1), 1–22.

— (2006). Incentives and prosocial behavior. *The American economic review*, **96** (5), 1652–1678.

Winfree, J. A. and McCluskey, J. J. (2005). Collective reputation and quality. *american Journal of agricultural Economics*, **87** (1), 206–213.

Appendices

A Experiment Instructions

This part presents the instructions that were given to the subjects. Subjects randomly assumed their roles (that is, Participant 1 and Participant 2) in the beginning of the experiment. The only difference between the strong and weak nature treatments is the probability with which nature overrides the participant 2's decision. Therefore, here we only give the strong nature narration. The instructions of the other treatments are available upon request.

A.1 Instruction

This is an experiment in the economics of strategic decision-making. Various research agencies have provided funds for this research. The instructions are simple. If you follow them closely and make the appropriate decisions, you can earn an appreciable amount of money.

The currency used in the experiment is francs. Francs will be converted to U.S. Dollars at a rate of $\underline{20}$ francs to $\underline{1}$ dollar. You have already received a $7.00 participation fee. Your earnings from the experiment will be incorporated into your participation fee. At the end of today's experiment, you will be paid in private and in cash.

It is very important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid. We expect and appreciate your cooperation.

A.1.1 Your role assignment

The experiment consists of numerous decision-making periods. During each period, you will be randomly and anonymously placed into a group that consists of two participants: participant 1 and participant 2. At the beginning of the first period, you will be randomly assigned to be either as participant 1 or participant

2. You will remain in the same role assignment throughout the entire experiment. So, if you are assigned to be participant 2, then you will stay as participant 2 throughout the entire experiment. Each consecutive period you will be randomly re-grouped with another participant of opposite assignment. So, if you are participant 2, each period you will be randomly re-grouped with another participant 1. You will never be re-grouped with a participant you have been grouped with in a previous period.

A.1.2 Stage 1

Each period will proceed in three stages. Both participant 1 and participant 2 begin each period with 10 francs, meaning that there are 20 total francs in the group. In Stage 1, participant 1 will choose whether to send or not send his/her francs to participant 2.

If participant 1 chooses to not send his/her francs, each participant keeps the 10 francs they began the period with and the period ends.

If participant 1 chooses to send his/her francs, the total number of francs in the group will double to 40, and the experiment will proceed to stage 2.

Before making the decision, whether to send, participant 1 will be made aware of whether participant 2 has been reported in a previous period. We will describe how one is reported when we describe stage 3.

A.1.3 Stage 2

If participant 1 chooses to send in Stage 1, the computer will display to participant 2 that participant 1 chose to send. The total amount of francs is then doubled to 40. Then, in Stage 2, participant 2 chooses whether to split or not split the francs with participant 1.

If participant 2 chooses to split the francs, with a probability of 70% (determined by a random number generator), both participants receive 20 francs and the period ends. With a probability of 30%, participant 2 will receive 20 francs and participant 1 will receive 0 francs. The experiment will then proceed to Stage 3.

If participant 2 chooses to not split the francs, participant 2 will receive 40 francs and participant 1 will receive 0 francs. The experiment will then proceed to Stage 3.

A.1.4 Stage 3

If the period reaches stage 3, participant 1 will not know whether participant 2 chose to not split (which automatically sends the game to stage 3), or whether participant 2 chose to split (which sends the game to stage 3 with a probability of 30%).

In the case that participant 2 had chosen to not split, he/she receives 40 francs. On the other hand, if participant 2 had chosen to split, then he/she will get 20 francs.

In stage 3, participant 1 will choose to report or not report participant 2.

If participant 1 chooses to not report, the stage ends, and participant 1 will receive 0 francs.

If participant 1 chooses report, he/she will lose 5 francs, meaning that participant 1 will receive -5 francs. However, in each of the remaining periods, participant 2 will be noted as having been reported to the participant 1 that he/she is matched with. In other words, prior to participant 1 making their decision in Stage 1, they will see whether the participant 2 they are matched with has been reported by at least one of his/her group members in the past.

A.1.5 Earnings of participants 1 and 2

In each period, the earnings of both participants depend on the send/not send decision made by participant 1 in Stage 1, the split/not split decision made by participant 2 in Stage 2, and the report/not report decision made by participant 1 in Stage 3. The earnings are summarized in table A1:

A.1.6 Examples

Example 1

Assume the following scenario. In the first stage, participant 1 observes that the participant 2 he/she is matched with has never been reported in a previous period. After observing this information, participant 1 chooses to not send. The period ends and participant 1's earnings are 10 francs and participant 2's earnings are 10 francs.

Example 2

Assume the following scenario. In the first stage, participant 1 observes that the participant 2 he/she

Choices	Earnings	
	Participant 1	Participant 2
Stage 1		
Participant 1 chooses to not send	10	10
Stage 2		
Participant 2 chooses split and the random number generator determines that the francs are split	20	20
Stage 3		
Participant 2 chooses to split but the random number generator determines that participant 1 does not get the francs. Participant 1 chooses to not report	0	20
Participant 2 chooses to split but the random number generator determines that participant 1 does not get the francs. Participant 1 chooses to report	-5	20
Participant 2 chooses to not split and participant 1 chooses to not report	0	40
Participant 2 chooses to not split and participant 1 chooses to report	-5	40

Table A1: **Summary of earnings**

is matched with has been reported in a previous period. After observing this information, participant 1 chooses to send. Then, in Stage 2, participant 2 chooses to split. The random number generator determines (with a probability of 70%) that the francs are split. The period ends and participant 1's earnings are 20 francs and participant 2's earnings are 20 francs.

Example 3

Assume the following scenario. In the first stage, participant 1 observes that the participant 2 he/she is matched with has not been reported. After observing this information, participant 1 chooses to send. Then, in Stage 2, participant 2 chooses to split. The random number generator determines (with a probability of 30%) that participant 1 does not get the francs. Participant 1 observes that the period has reached Stage

3, but does not know if it is because participant 2 chose not split or whether participant 2 chose to split but the random number generator chose that participant 1 does not get the francs. In Stage 3, Participant 1 then chooses to report. The period ends and participant 1's earnings are -5 francs and participant 2's earnings are 20 francs.

Example 4

Assume the following scenario. In the first stage, participant 1 observes that the participant 2 he/she is matched with has been reported in a previous period. After observing this information, participant 1 chooses to send. Then, in Stage 2, participant 2 chooses to not split. Participant 1 observes that the period has reached Stage 3, but does not know if it is because participant 2 chose to not split or whether participant 2 chose to split but the random number generator chose to not split the francs. In Stage 3, Participant 1 then chooses to not report. The period ends and participant 1's earnings are 0 francs and participant 2's earnings are 40 francs.

A.1.7 End of the period

At the end of each period, the computer will calculate individual earnings and display to both participants the following information:

- Whether participant 1 chose to send or not send in Stage 1
- Whether participant 1 received the francs (when applicable)
- Whether participant 1 chose to report or not report in Stage 3 (when applicable)
- Your earnings

A.1.8 Important notes

Each period, you will be randomly and anonymously placed into a group which consists of two participants: participant 1 and participant 2. At the beginning of the first period you will be randomly assigned to be either participant 1 or participant 2. You will remain in the same role assignment throughout the entire experiment. So, if you are assigned to be participant 2, then you will stay as participant 2 throughout

the entire experiment. Each consecutive period you will be randomly re-grouped with a participant of the opposite assignment, and you will never be re-grouped with a participant you have been grouped with in a previous period. So, if you are participant 2, each period you will be randomly re-grouped with a participant 1 with whom you have not been grouped with in the past.

Each period proceeds in three stages. In Stage 1, participant 1 will choose to send or not send francs to participant 2. Prior to making this decision, participant 1 will observe whether the participant 2 he/she is matched with has been reported or not reported in a previous period. If participant 1 chooses to not send, the period ends and each participant earns 10 francs. If participant 1 chooses to send, the total amount of francs is doubled and the game proceeds to Stage 2. In Stage 2, participant 2 chooses to split or not split. If participant 2 chooses to split, then with a probability of 70% the total amount of francs are split, the period ends, and each participant earns 20 francs; with a probability of 30%, the francs are not split and the experiment proceeds to Stage 3. Likewise, if participant 2 chooses to not split, the experiment will proceed to Stage 3. In Stage 3, participant 1 chooses to report or not report. If participant 1 chooses to report, the period ends and participant 1 earns 5 francs, participant 2 earns 20 francs in case that he/she chose to split in Stage 2 and 40 francs in the case that he/she chose to not split in Stage 2. However, participant 2's group members in future periods will see that he/she has been reported. If participant 1 chooses to not report, the period ends, participant 1 earns 0 francs, and participant 2 earns 20 francs in the case that he/she chose to split in Stage 2 and 40 francs in the case that he/she chose to not split in Stage 2. After the period ends, your earning will be displayed on the screen, but participant 2's decision to split or not split will not be displayed.

Remember you have already received a $7.00 participation fee. In the experiment, depending on a period, you may receive either positive or negative earnings. At the end of the experiment we will sum your earnings from each period and convert them to a U.S. dollar payment at a rate of 20 francs to 1 dollar. If the earnings are negative, we will subtract them from your participation fee. If the earnings are positive, we will add them to your participation fee.

A.2 Risk preference elicitation instructions

In the questions that follow, you are going to be asked to make ten decisions. Each decision will be between Option A and Option B. One of the ten choices you make will be randomly selected to determine your earnings for this part of the experiment.

Options		
A	B	Your Choice
\$1 or \$3 each with probability 1/2	\$0.1 with probability 9/10 or \$4 with probability 1/10	A or B
\$1 or \$3 each with probability 1/2	\$0.1 with probability 8/10 or \$4 with probability 2/10	A or B
\$1 or \$3 each with probability 1/2	\$0.1 with probability 7/10 or \$4 with probability 3/10	A or B
\$1 or \$3 each with probability 1/2	\$0.1 with probability 6/10 or \$4 with probability 4/10	A or B
\$1 or \$3 each with probability 1/2	\$0.1 with probability 5/10 or \$4 with probability 5/10	A or B
\$1 or \$3 each with probability 1/2	\$0.1 with probability 4/10 or \$4 with probability 6/10	A or B
\$1 or \$3 each with probability 1/2	\$0.1 with probability 3/10 or \$4 with probability 7/10	A or B
\$1 or \$3 each with probability 1/2	\$0.1 with probability 2/10 or \$4 with probability 8/10	A or B
\$1 or \$3 each with probability 1/2	\$0.1 with probability 1/10 or \$4 with probability 9/10	A or B
\$1 or \$3 each with probability 1/2	\$0.1 with probability 0/10 or \$4 with probability 10/10	A or B

B Proofs

In this section we provide proofs of the three predictions as well as Proposition 2. Proposition 1 follows directly from the intuition laid out in the paper and therefore does not necessitate a formal proof. We begin with the proof of Proposition 2, since this proof is needed for the proofs of the three predictions.

B.1 Proof of Proposition 2

B.1.1 Stage 3

The Stage 3 action of P follows directly from (1).

B.1.2 Stage 2

The Stage 2 action of altruist A's, $D_{i,t} = 1$, is true by definition. We therefore proceed by considering the Stage 2 actions of strategic A's.

Strategic A's choose $D_{i.t} = 1$ iff $E\left[U_{i,t}^{A}\left(D_{i,t}=1\right)|R_t=1\right] \geq E\left[U_{i,t}^{A}\left(D_{i.t}=0\right)|R_t=1\right]$, where:

$$E\left[U_{i,t}^{A}\left(D_{i,t}=1\right)|R_t=1\right] = 2\Omega + \delta\theta\left(1-F\left(C\right)\right)V^{B} + \delta\left[1-\theta\left(1-F\left(C\right)\right)\right]V^{G}, \tag{B1}$$

$$E\left[U_{i,t}^{A}\left(D_{i.t}=0\right)|R_t=1\right] = 4\Omega + \delta\left(1-F\left(C\right)\right)V^{B} + \delta F\left(C\right)V^{G}, \tag{B2}$$

where V^B and V^G are A's continuation payoff from having a bad and good reputation, respectively. In equations (B1) and (B2), V^G is recursive. Hence, the utilities can be re-written:

$$U_{i,t}^{A}\left[D_{i,t}=1\right] = \frac{2\Omega + \delta\theta\left(1-F\left(C\right)\right)V^{B}}{1-\delta\left[1-\theta\left(1-F\left(C\right)\right)\right]}, \tag{B3}$$

$$U_{i,t}^{A}\left[D_{i.t}=0\right] = \frac{4\Omega + \delta\left(1-F\left(C\right)\right)V^{B}}{1-\delta F\left(C\right)}. \tag{B4}$$

The value of V^B depends on P's Stage 1 choice when $R_t = 0$. If P chooses $S_t = 0$, then $V^B = \frac{\Omega}{1-\delta}$. Meanwhile, if P chooses $S_t = 1$, then $V^B = \frac{4\Omega}{1-\delta}$. First consider the case where P chooses $S_t = 1$ when $R_t = 0$. In this case, A chooses $D_{i,t} = 1$ iff:

$$\frac{2\Omega + \delta\theta\left(1 - F\left(C\right)\right)\frac{4\Omega}{1-\delta}}{1 - \delta\left[1 - \theta\left(1 - F\left(C\right)\right)\right]} \geq \frac{4\Omega + \delta\left(1 - F\left(C\right)\right)\frac{4\Omega}{1-\delta}}{1 - \delta F\left(C\right)}, \tag{B5}$$

$$\Rightarrow 1 \geq 2,$$

which is obviously not true. Hence, if P chooses $S_t = 1$ when $R_t = 0$, A chooses $D_{i,t} = 0$ when $R_t = 1$. Turning to the case in which P chooses $S_t = 0$, A chooses $D_{i,t} = 1$ iff:

$$\frac{2\Omega + \delta\theta\left(1 - F\left(C\right)\right)\frac{\Omega}{1-\delta}}{1 - \delta\left[1 - \theta\left(1 - F\left(C\right)\right)\right]} \geq \frac{4\Omega + \delta\left(1 - F\left(C\right)\right)\frac{\Omega}{1-\delta}}{1 - \delta F\left(C\right)}, \tag{B6}$$

$$\Rightarrow \theta \leq \frac{1}{3}\left(1 - \frac{2\left(1-\delta\right)}{\delta\left[1 - F\left(C\right)\right]}\right). \tag{B7}$$

We therefore define $\Psi\left(\delta, C\right) \equiv \frac{1}{3}\left(1 - \frac{2(1-\delta)}{\delta[1-F(C)]}\right)$, where it follows that $\Psi\left(\delta, C\right) > 0$ if and only if $F\left(C\right) < \frac{3\delta-2}{\delta}$. Hence, A chooses $D_{i,t} = 1$ iff P chooses $S_t = 0$ when $R_t = 0$ and $\theta \leq \Psi\left(\delta, C\right)$. We must therefore find the condition under which P chooses $S_t = 0$ when $R_t = 0$.

To find this condition, note that an equilibrium exists in which A chooses $D_{i,t} = 1$ (conditional on $R_t = 1$) and P chooses $S_t = 0$ (conditional on the A they are matched with having $R_t = 0$) if and only if $\frac{\alpha p^B}{\alpha p^B + (1-\alpha)q} < \frac{1}{2(1-\theta)}$. In this case, $q = 1 - \left(1 - \theta\left[1 - F\left(C\right)\right]\right)^t$, since the strategic A chooses $D_{i,t} = 1$ when $R_t = 1$. It follows that $\frac{\alpha p^B}{\alpha p^B + (1-\alpha)q} = \alpha$, and thus the equilibrium exists iff $\alpha < \frac{1}{2(1-\theta)}$. Rearranging terms, this gives $\theta > 1 - \frac{1}{2\alpha}$. Hence, an equilibrium exists in which A chooses $D_{i,t} = 1$ iff $\theta \in \left(1 - \frac{1}{2\alpha}, \Psi\left(\delta, C\right)\right]$.

B.1.3 Stage 1

Turning to the first stage, P chooses $S_t = 1$ if and only if $E\left[U_t^P\left(S_t = 1\right)|R_t = 1\right] \geq E\left[U_t^P\left(S_t = 0\right)|R_t = 1\right]$, which can be re-written as

$$E\left[D_{i,t}|R_t = 1\right] \geq \frac{1}{2\left(1-\theta\right)}. \tag{B8}$$

First, since the RHS of (B8) is greater than 1 when $\theta > \frac{1}{2}$, $S_t = 0$ whenever $\theta > \frac{1}{2}$. Moreover, from above, if $\theta \in \left(1-\frac{1}{2\alpha}, \Psi\left(\delta, C\right)\right]$, then both types of agents choose $D_{i,t} = 1$. This entails that $E\left[D_{i,t}|R_t = 1\right] = 1$, and thus P chooses $S_t = 1$ if $\theta \leq \frac{1}{2}$, which must be true when $\theta \in \left(1-\frac{1}{2\alpha}, \Psi\left(\delta, C\right)\right]$. Hence, P chooses $S_t = 1$ if $\theta \in \left(1-\frac{1}{2\alpha}, \Psi\left(\delta, C\right)\right]$.

Next, consider $\theta \leq 1-\frac{1}{2\alpha}$. In this case, strategic A's choose $D_{i,t} = 0$, meaning that $E\left[D_{i,t}|R_t = 1\right] = pr\left(A \text{ is altruist } | \ R = 1\right)$. The probability of an altruistic agent having a good reputation in period t is $p = \left(1-\theta\left[1-F\left(C\right)\right]\right)^t$, while the probability of a strategic agent having a good reputation in period t is $\left[F\left(C\right)\right]^t$. Hence,

$$pr\left(A \text{ is altruist } | \ R = 1\right) = \frac{\alpha\left(1-\theta\left[1-F\left(C\right)\right]\right)^t}{\alpha\left(1-\theta\left[1-F\left(C\right)\right]\right)^t + \left(1-\alpha\right)\left[F\left(C\right)\right]^t}, \tag{B9}$$

and P chooses $S_t = 1$ iff,

$$\left(1-2\theta\right)\left(1-\theta\left[1-F\left(C\right)\right]\right)^t \geq \left(\frac{1-\alpha}{\alpha}\right)\left[F\left(C\right)\right]^t, \tag{B10}$$

which must be true, since $\theta \leq 1-\frac{1}{2\alpha} \Leftrightarrow 1-2\theta \geq \frac{1-\alpha}{\alpha}$ and $1-\theta\left[1-F\left(C\right)\right] \geq F\left(C\right)$. Hence, P chooses $S_t = 1$ if $\theta \leq 1-\frac{1}{2\alpha}$, and more generally, P chooses $S_t = 1$ if $\theta \leq \Psi\left(\delta, C\right)$.

Finally, consider $\theta \in \left(\Psi\left(\delta, C\right), \frac{1}{2}\right)$. In this range strategic A's choose $D_{i,t} = 0$, meaning that P chooses $S_t = 1$ if (B10) holds. It is straightforward to see that the LHS of (B10) is decreasing in θ (as long as $\theta < \frac{1}{2}$). Denote $\overline{\theta}$ as the value of θ that sets (B10) to equality. If $\overline{\theta} > \Psi\left(\delta, C\right)$, then $S_t = 1$ when $\theta \in \left[\Psi\left(\delta, C\right), \overline{\theta}\right]$. Combining this with the fact that $S_t = 1$ if $\theta \leq \Psi\left(\delta, C\right)$, we have the more general formulation that $S_t = 1$ if $\theta \leq \theta^*$, where $\theta^* = \max\left\{\overline{\theta}, \Psi\left(\delta, C\right)\right\}$.

B.2 Proof of Prediction 1

We know from the proof of Proposition 2 that, in equilibrium, if P chooses $S = 1$ when $R = 0$, strategic A's will choose $D = 0$ when $R = 1$. Therefore, the probability that strategic A's have a bad reputation in period t is $q = 1 - [F(C)]^t$. Hence, P chooses $S = 1$ iff:

$$\frac{\alpha\left[1-(1-\theta[1-F(C)])^t\right]}{\alpha\left[1-(1-\theta[1-F(C)])^t\right]+(1-\alpha)\left(1-[F(C)]^t\right)} \geq \frac{1}{2(1-\theta)}. \tag{B11}$$

Rearranging this inequality yields:

$$(1-2\theta)\left[1-(1-\theta[1-F(C)])^t\right] \geq \left(\frac{1-\alpha}{\alpha}\right)\left(1-[F(C)]^t\right). \tag{B12}$$

It is clear that (B12) does not hold at $\theta = 0$ or $\theta \geq 0.5$.[20] To prove the prediction with respect to θ, we show that the LHS of (B12), denoted by Λ, is hump-shaped in θ for $\theta \leq \frac{1}{2}$. First, we solve for the first derivative of Λ with respect to θ:

$$\frac{\partial\Lambda}{\partial\theta} = -2\left[1-(1-\theta[1-F(C)])^t\right] + t(1-2\theta)[1-F(C)](1-\theta[1-F(C)])^{t-1}. \tag{B13}$$

At $\theta = 0$, $\frac{\partial\Lambda}{\partial\theta} = t[1-F(C)] \geq 0$, while at $\theta = \frac{1}{2}$, $\frac{\partial\Lambda}{\partial\theta} = -2\left[1-(0.5[1+F(C)])^t\right] \leq 0$. In other words, Λ is increasing in θ at $\Lambda = 0$ and decreasing in θ at $\Lambda \geq \frac{1}{2}$. All that remains to show is that $\frac{\partial^2\Lambda}{\partial\theta^2} < 0$ for $\theta < \frac{1}{2}$. Solving for $\frac{\partial^2\Lambda}{\partial\theta^2}$ gives:

$$\frac{\partial^2\Lambda}{\partial\theta^2} = -4t[1-F(C)](1-\theta[1-F(C)])^{t-1} - t(t-1)(1-2\theta)[1-F(C)]^2(1-\theta[1-F(C)])^{t-2}, \tag{B14}$$

which is clearly less than 0 when $\theta \leq \frac{1}{2}$.

[20] Formally, it is possible that (B12) is not satisfied for any value of θ. This would entail that P never chooses $S = 1$ when $R = 0$, and thus the parameter space over which P chooses $S = 1$ is invariant in θ. However, since we observe many instances in the experiment of principals choosing $S = 1$ after seeing $R = 0$, we focus on the part of the parameter space where (B12) is satisfied for some value of θ.

Turning to comparative statics with respect to C, rearranging (B12) and taking the derivative with respect to C yields that the parameters space in which P chooses $S=1$ is decreasing in C if:

$$-t\theta F'(C)(1-2\theta)(1-\theta[1-F(C)])^{t-1}+tF'(C)\left(\frac{1-\alpha}{\alpha}\right)[F(C)]^{t-1}<0. \tag{B15}$$

Since t and $F'(C)$ are positive, rearranging this inequality yields the condition:

$$F(C)<\frac{(1-\theta)\left[\frac{\alpha\theta(1-2\theta)}{1-\alpha}\right]^{\frac{1}{t-1}}}{1-\theta\left[\frac{\alpha\theta(1-2\theta)}{1-\alpha}\right]^{\frac{1}{t-1}}}. \tag{B16}$$

In other words, there is some $\overline{F}=\min\left\{1,\frac{(1-\theta)\left[\frac{\alpha\theta(1-2\theta)}{1-\alpha}\right]^{\frac{1}{t-1}}}{1-\theta\left[\frac{\alpha\theta(1-2\theta)}{1-\alpha}\right]^{\frac{1}{t-1}}}\right\}$, for which the parameter set over which $S_t=1$ is decreasing in C if $F(C)\leq\overline{F}$ and increasing in C otherwise. That is, the parameter set over which $S_t=1$ is U-shaped in C.

B.3 Proof of Prediction 2

It is straight-forward to verify that $\frac{\partial\Psi(\delta,C)}{\partial C}<0$. Hence, the parameter space over which strategic A's choose $D_{i,t}$ conditional on $R_{i,t}=1$ is decreasing in C.

B.4 Proof of Prediction 3

It follows directly from the fact that P chooses $S=1$ if and only if $\theta\leq\theta^*$ that the size of the parameter space over which P chooses $S=1$ is decreasing in θ, ceteris paribus.

With respect to C, recall that $\theta^*=\max\left\{\overline{\theta},\Psi(\delta,C)\right\}$. It is straight-forward to verify that $\frac{\partial\Psi(\delta,C)}{\partial C}<0$. Meanwhile, the implicit function theorem yields:

$$\frac{\partial\overline{\theta}}{\partial C}=-\frac{tF'(C)\,\overline{\theta}\left(1-2\overline{\theta}\right)\left(1-\overline{\theta}[1-F(C)]\right)^{t-1}-tF'(C)\left(\frac{1-\alpha}{\alpha}\right)[F(C)]^{t-1}}{-2\left(1-\overline{\theta}[1-F(C)]\right)^{t}-t\left(1-2\overline{\theta}\right)[1-F(C)]\left(1-\overline{\theta}[1-F(C)]\right)^{t-1}}. \tag{B17}$$

It follows that $\frac{\partial \overline{\theta}}{\partial C}$ is the same sign as the numerator, meaning that $\frac{\partial \overline{\theta}}{\partial C} \leq 0$ if:

$$\overline{\theta}\left(1-2\overline{\theta}\right)\left(1-\overline{\theta}\left[1-F\left(C\right)\right]\right)^{t-1} \leq \left(\frac{1-\alpha}{\alpha}\right)\left[F\left(C\right)\right]^{t-1}. \tag{B18}$$

Plugging $\frac{1-2\overline{\theta}}{\frac{1-\alpha}{\alpha}} = \left[\frac{F(C)}{1-\overline{\theta}[1-F(C)]}\right]^{t}$ into (B18) reveals that $\frac{\partial \overline{\theta}}{\partial C} \leq 0$ if:

$$\overline{\theta}\left[\frac{F\left(C\right)}{1-\overline{\theta}\left[1-F\left(C\right)\right]}\right] \leq 1, \tag{B19}$$

which is obviously true. Therefore, $\frac{\partial \Psi(\delta,C)}{\partial C} < 0$ and $\frac{\partial \overline{\theta}}{\partial C} \leq 0$, and thus $\frac{\partial \theta^{*}}{\partial C} \leq 0$. It follows that the size of the parameter space over which P chooses $S = 1$ is decreasing in C, ceteris paribus.

C Additional Figures and Robustness Checks

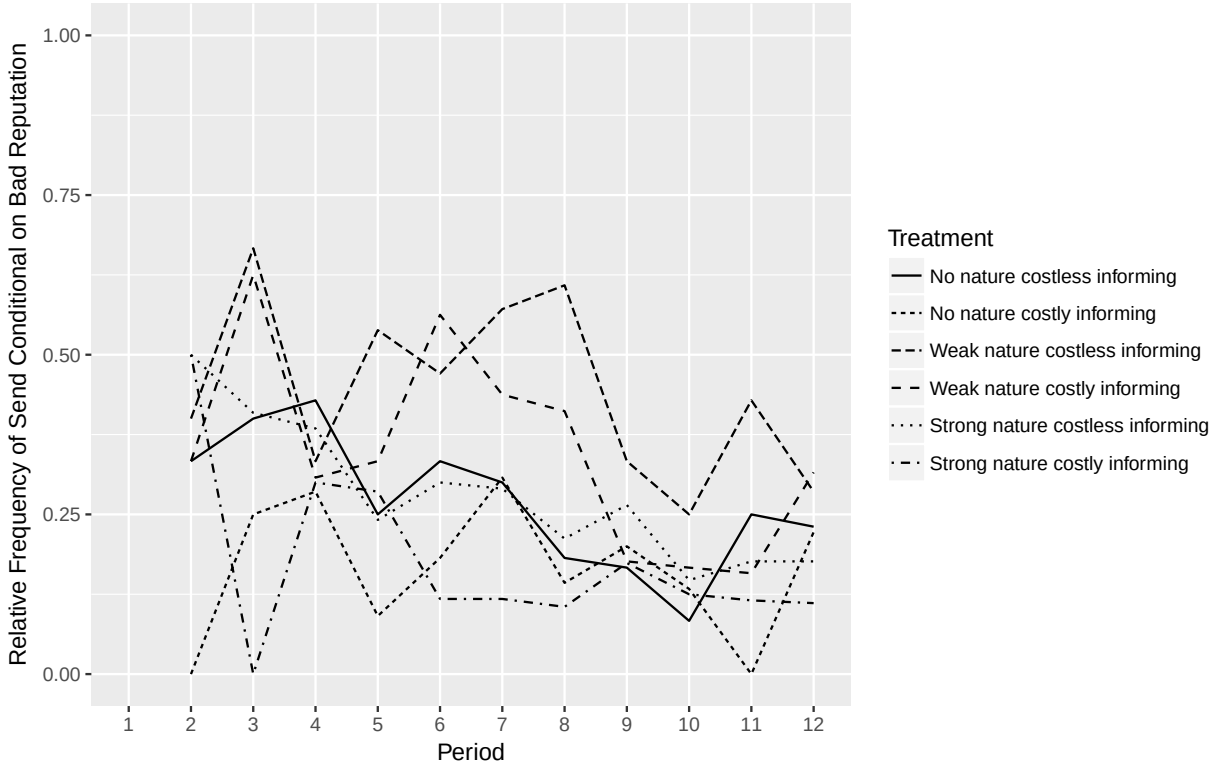

Figure C1: **Relative Frequency of Send Conditional on Bad Reputation by Treatment and Period.**

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable = Send (0/1)			**Dependent Variable = Divide (0/1)**		
No nature, costly Informing	-0.144***	-0.113***	-0.110***	-0.090**	-0.091***	-0.087**
	(0.052)	(0.033)	(0.030)	(0.043)	(0.030)	(0.035)
Weak nature, costless Informing	-0.120**	0.017	0.012	-0.146***	-0.079*	-0.094**
	(0.055)	(0.055)	(0.053)	(0.047)	(0.045)	(0.047)
Weak nature, costly Informing	-0.106**	-0.028	-0.030	-0.188***	-0.148***	-0.165***
	(0.051)	(0.031)	(0.030)	(0.024)	(0.017)	(0.025)
Strong nature, costless Informing	-0.373***	-0.083	-0.093	-0.294***	-0.175**	-0.216***
	(0.087)	(0.075)	(0.077)	(0.068)	(0.074)	(0.075)
Strong nature, costly Informing	-0.343***	-0.224***	-0.226***	-0.336***	-0.315***	-0.354***
	(0.059)	(0.036)	(0.036)	(0.040)	(0.043)	(0.049)
Reputation		0.577***	0.552***		0.321***	0.252***
		(0.029)	(0.038)		(0.030)	(0.034)
Risk			0.008			0.002
			(0.007)			(0.010)
Female			0.002			0.075**
			(0.030)			(0.037)
Period			-0.008**			-0.016***
			(0.004)			(0.003)
Intercept	0.782***	0.341***	0.380***	0.838***	0.561***	0.670***
	(0.036)	(0.035)	(0.068)	(0.021)	(0.032)	(0.060)
Observations	2592	2592	2592	1559	1559	1559

Standard errors clustered by session in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table C1: **Regression Analysis of the Send and Divide Decision**

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable = Send (0/1)			**Dependent Variable = Divide (0/1)**		
No nature, costly informing	-0.105**	-0.112**	-0.104*	-0.034	-0.033	-0.029
	(0.049)	(0.051)	(0.058)	(0.073)	(0.084)	(0.094)
Weak nature, costless informing	0.127*	0.130*	0.138*	-0.030	-0.036	-0.027
	(0.069)	(0.070)	(0.074)	(0.119)	(0.126)	(0.132)
Weak nature, costly informing	0.049	0.050	0.046	-0.203***	-0.217**	-0.225**
	(0.056)	(0.064)	(0.065)	(0.077)	(0.086)	(0.095)
Strong nature, costless informing	-0.014	-0.020	-0.031	-0.030	-0.024	-0.043
	(0.076)	(0.073)	(0.076)	(0.114)	(0.112)	(0.122)
Strong nature, costly informing	-0.116	-0.123	-0.118	-0.201	-0.232*	-0.240
	(0.080)	(0.081)	(0.086)	(0.127)	(0.136)	(0.152)
Risk		0.001	0.001		0.005	0.008
		(0.013)	(0.013)		(0.018)	(0.019)
Female		0.057	0.051		0.176**	0.180**
		(0.039)	(0.038)		(0.072)	(0.070)
Period			-0.024***			-0.017**
			(0.004)			(0.008)
Intercept	0.274***	0.242***	0.441***	0.424***	0.319***	0.435***
	(0.044)	(0.069)	(0.088)	(0.072)	(0.111)	(0.137)
Observations	1103	1103	1103	297	297	297

Standard errors clustered by session in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table C2: **Regression Analysis of the Effect of Bad Reputation on the Decision to Send and Divide**

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable = Send (0/1)			**Dependent Variable = Divide (0/1)**		
No nature, costly informing	-0.112**	-0.102**	-0.103**	-0.110**	-0.105**	-0.110**
	(0.050)	(0.047)	(0.047)	(0.047)	(0.047)	(0.048)
Weak nature, costless informing	-0.065	-0.065	-0.069	-0.054	-0.051	-0.080*
	(0.054)	(0.054)	(0.054)	(0.050)	(0.053)	(0.048)
Weak nature, costly informing	-0.062***	-0.060***	-0.061***	-0.147***	-0.149***	-0.162***
	(0.016)	(0.011)	(0.011)	(0.027)	(0.029)	(0.025)
Strong nature, costless informing	-0.120	-0.113	-0.120	-0.181**	-0.177**	-0.238***
	(0.077)	(0.079)	(0.082)	(0.092)	(0.090)	(0.088)
Strong nature, costly informing	-0.275***	-0.268***	-0.270***	-0.314***	-0.326***	-0.360***
	(0.032)	(0.032)	(0.034)	(0.036)	(0.039)	(0.038)
Risk		0.013**	0.013**		-0.002	-0.003
		(0.006)	(0.006)		(0.012)	(0.012)
Female		-0.020	-0.020		0.058	0.062
		(0.034)	(0.034)		(0.048)	(0.051)
Period			-0.002			-0.019***
			(0.004)			(0.003)
Intercept	0.949***	0.905***	0.919***	0.858***	0.833***	0.948***
	(0.012)	(0.038)	(0.050)	(0.025)	(0.067)	(0.065)
Observations	1489	1489	1489	1262	1262	1262

Standard errors clustered by session in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table C3: **Regression Analysis of the Effect of Good Reputation on the Decision to Send and Divide**