

Exchange in the Absence of Legal Enforcement: Reputation and Multilateral Punishment under Uncertainty*

Aidin Hajikhameneh
San Jose State University
aidin.hajikhameneh@sjsu.edu

Jared Rubin
Chapman University
jrubin@chapman.edu

Abstract

Principal-agent problems can reduce gains from exchange available in long distance trade. One solution historically used to mitigate such problems 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. Our experimental results indicate that a modest amount of uncertainty has little effect on overall welfare: while part of the surplus is destroyed by uncertainty, principals are also more willing to trust agents with bad reputations, thereby increasing the frequency of welfare-enhancing exchange.

JEL classifications: C91, C92, D02, D83, F10, N70

Keywords: multilateral punishment, reputation, uncertainty, exchange, lab experiment, trust game

*We thank Gary Charness, Raymond Fisman, Ellen Green, Laurence Iannaccone, Erik Kimbrough, Michael McBride, Elena Pikulina, two anonymous referees, and participants in the Southwest Experimental and Behavioral Economics Workshop at UC Santa Barbara, the Economic Science Association World Meeting at UC San Diego, and the Association for the Study of Religion, Economics, and Culture at Chapman University 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.

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; Johnson et al. 2002).

The most common mechanism used to link past conduct to future reward is *reputation*. Bilateral reputation refers to the reputation established between two parties in a repeated relationship, and it 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 when groups are large and histories unknown (Ghosh and Ray 1996; Kranton 1996; Kranton and Minehart 2001; Leeson 2008). 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.

One famous example of such an institution is the Maghribi Traders’ Coalition studied by Greif (1993). This group of traders and agents 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, a self-enforcing equilibrium emerged in which agents had incentive to act honestly and merchants had incentive to report agents who violated agreements. As a result, the multilateral punishment mechanism worked for over a century to facilitate exchange. Clay (1997) and Okazaki (2005) study similarly-structured reputation-based institutions that employed information sharing and self-enforcing punishments for rule-breakers in an attempt to facilitate exchange between merchants in 19th 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 employs an intricate information transmission mechanism (an active trade press and circulars that report the names of transactors who refuse to arbitrate) along with social 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 that arbitrate between bourse (i.e., trading club) members. An industry long dominated by Orthodox Jews (i.e., a “closed group”), information about arbitration decisions is quickly transmitted to

¹A large literature shows how bilateral reputation, via repeated interaction, can sustain cooperation without third-party enforcement (Kreps and Wilson 1982; Kreps et al. 1982).

all bourses in the world federation, and members can be suspended or expelled for failing to live up to their agreements.² Similar mechanisms are also used in religious groups. 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).³

Yet, institutions which facilitate trade by permitting individuals to establish a reputation face two problems.⁴ First, it must be in the incentive of the cheated parties to report they were cheated. For instance, in the diamond industry, bourse members had incentive to report cheaters in part because doing so is inexpensive; information on past transactions was centralized and transmitted inexpensively via fax (Bernstein 1992). However, in the pre-modern settings in many of the studies cited above, where communication over long distance was costly and time consuming, such costs could be prohibitive. For instance, Greif (1989) gives numerous examples of Maghribi merchants sending extensive letters to their fellow traders after they felt they had been cheated. Such actions were time consuming and expensive—long distance travel was not cheap—and it was thus not in the narrow monetary incentive of the trader to send such letters, as it was unlikely they would engage in bilateral relations with the agent in question in the future. Clay (1997) notes that in 19th-century Mexican California, information was expensive to disseminate and merchants had an incentive to free-ride off of others, but repeated relationships encouraged merchants to engage in truthful correspondence.

Second, and perhaps more important, “cheating” is often not black and white and is frequently unverifiable. This problem arises due to asymmetric information. For example, Greif (1989) notes that in the 11th century, a trip from Egypt to Sicily—two trade entrepôts of the time—could take anywhere from 13 to 50 days. Ships did not always reach their destination, goods were often damaged in transit (along the choppy waters of the Mediterranean), and ship captains often pilfered goods. Clay (1997) shows that similar information asymmetries plagued merchant-agent relations in 19th century Mexican California, where price variability was high due to irregular and uncertain supply, while agents operated in distant ports and were nearly impossible to monitor. McMillan and Woodruff (1999) cite a similar case in contemporary Vietnam, where imperfect monitoring is rampant. Under such circumstances, an agent might very well truthfully report back to the merchant that external circumstances (such as the goods being damaged at sea) reduced

²Richman (2017a) shows that recently-eroded trust among diamond merchants has limited their capacity to conduct reputation-based exchange. For much more on ethnic trading networks in the diamond industry, see Richman (2017b). For more on the role that ethnic homogeneity plays in sustaining such groups, see Landa (1981, 1994).

³The papers cited above highlight cases in which 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), and Kimbrough and Rubin (2015).

⁴See 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.

the return the agent was able to obtain. But this is ostensibly cheap talk: the agent may be telling the truth, but he also has incentive to lie if the goods actually arrived in perfect condition. This makes it difficult for merchants to tell whether their customers cheated them or genuinely could not make repayment due to financial difficulties. When the “cheated” party receives bad news from their partner and reputation-based institutions are present, will the former engage with the institution and provide the latter with a bad reputation? Under what circumstances will they do so? How do such considerations affect the incentive to engage in exchange relationships in the first place? Perhaps most importantly, will *pro-social* characteristics (e.g., altruism, reciprocity) affects decisions in such a setting?

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.⁵ 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; Bolton et al. 2004, 2005; Bracht and Feltovich 2009; Charness et al. 2011).⁶ Yet, to our knowledge the experimental literature does not address how such reputation-building institutions work when there are various degrees of uncertainty regarding whether the agent actually cheated or not.⁷ This is an important issue, because it is unclear whether such 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,

⁵Ho and Huffman (forthcoming) 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.

⁶Second 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) employ 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 increases social welfare by punishing non-cooperative behavior.

⁷There is theoretical work on uncertainty in trade relations and third-party enforcement. Kandori (1992) models multi-lateral reputation mechanisms in the presence of uncertainty about the reputation of others. This is different from the type of uncertainty of interest in the present study (i.e., uncertainty regarding the actions of others in the *present* period). Moreover, there are some experimental works on uncertainty in reputation in contexts different from the ones considered in the present study. Rice (2012) studied the effect of uncertainty on “reputation ratings” using a combination of trust and ultimatum games. However, her focus was not on the multilateral punishment mechanism. In the context of public goods game, Bornstein and Weisel (2010) and Ambrus and Greiner (2012) studied the effect of uncertainty of the level of investment in the public good in a finitely-repeated game setting. They found that high levels of uncertainty significantly reduce social welfare. Ghidoni (2017) presents a related experimental study on trust and uncertainty in the context of the funding of large projects with uncertain outcomes. He finds that the proposer of the project can generate trust by the hosts by disclosing information about the project.

the principal sees the agent’s 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. While this is an admittedly strict design—reputation is dichotomous (unlike, say, Yelp, where it is more continuous) and agents cannot regain good reputations after losing them—it provides a straight-forward baseline for studying multilateral reputation-based institutions that does not suffer from the confounds inherent in complex institutional design. Indeed, the model presented in Ali and Miller (2016) suggests that our strict design should result in relatively *little use* of the punishment institution, since permanent ostracism deters individuals from reporting cheaters. In this light, our results present a lower bound on the effectiveness of such institutions.

We employ a 4×2 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 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 principal’s return between 0, a modest positive amount (0.05), a moderate positive amount (0.10), and a large amount (0.30).

Our most important finding is that a modest amount of uncertainty (i.e., the 0.05 treatment) does not affect total welfare relative to the case where there is no uncertainty, *despite* the fact that resources are destroyed with positive probability. This result arises due to a trade-off faced by principals: on the one hand nature destroys some of the gains from exchange, on the other hand principals are more willing to engage with agents with bad reputations when there is a modest amount of uncertainty.⁸ The latter result presumably arises because agents may have a bad reputation despite acting honestly in the past. Uncertainty sufficiently muddies the information about agents to the point that some principals are willing to trust them (similar to what McMillan and Woodruff (1999) find empirically among Vietnamese merchants). On the other hand, a large amount of uncertainty (i.e., the 0.30 treatment) 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.⁹

The paper is structured as follow. Section 2 lays out the experimental design. In Section 3, we present

⁸This 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.

⁹This result parallels the one found in Falk and Kosfeld (2006), who study the impact of minimum reciprocity rules on trust. They find a similar trade-off to the one we find: at low levels of minimum reciprocity, such rules *decrease* trust and are counterproductive. However, higher minimum levels mitigate the loss of trust engendered by the rules.

some intuition to guide the reader regarding how we might expect subjects to act in a modified trust game with a multilateral punishment institution, costly reputation, and uncertainty. Section 4 reports the results, and Section 5 offers some concluding thoughts.

2 Experimental Design and Procedures

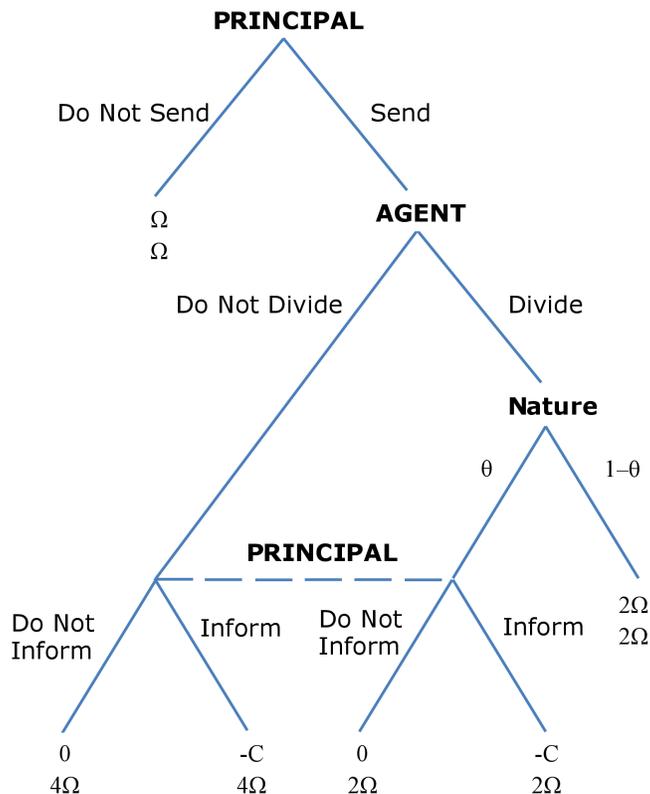
To investigate the effect of multilateral punishments on trade behavior, we designed a modified trust game in which uncooperative behavior could be reported to all principals (i.e., investors). In the beginning of each session, subjects were randomly assigned as a principal or agent. In the instructions we employed neutral language; for example, the principal and agent were named “participant 1” and “participant 2,” respectively (for the full instructions, see Appendix A). Further, to minimize confusion, instructions were read out loud in front of the subjects and also a quiz was administered at the end of the instructions (for details on the quiz, see Appendix A.2).

In each session, there were 12 principals and 12 agents. The matching scheme follows a perfect stranger matching protocol. That is, each principal played the same game with a new agent in each period. This was done to ensure that repeated relationships were not a reason that principals reported agents. Each session lasted 12 periods. We reserved the laboratory for thirty minutes longer than needed so that subjects would not know when the game ended. This was intended to prevent “period T ” behavior common in finitely-repeated lab experiments.

Each period consisted of three stages, which are depicted in Figure 1. In the first stage, principals could opt out of the trade (Do Not Send) and earn their reservation payoff of $\Omega > 0$. If the principal decided to engage in trade (Send), the game proceeded to Stage 2. In this stage, agents had 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 received payment 2Ω , and if they pocketed the investment they received payment 4Ω . If they split the return, nature overrode their decision, preventing principals from receiving their share, with probability $\theta \in [0, 1]$.

If the agent decided to not divide the return, or if nature overrode the agent’s decision to divide, the principal did not receive her share. The principal only observed the outcome and not the agent’s action. If either of these events occurred, the game proceeded to a third stage. In this stage, the principal had the option of informing all the prospective principals for a fee C (Inform). If the principal chose to inform other principals, that particular agent would have a bad reputation for the remainder of game. In the beginning of each period, principals observed whether the agents they were matched with had a “bad” or “good” reputation. We define reputation to be a binary signal in which the agent has a “good reputation” if and

Figure 1: Game Tree



Note: The top and bottom payoffs belong to the principal and agent, respectively.

only if no principal has ever informed on her in the past, and she has a “bad reputation” if at least one principal informed on her in the past.

The novel aspect of our experiment is that it provides insight into how uncertainty (θ) and costly punishment (C) augment the effectiveness of a multilateral punishment institution. To this end, we employ a 4×2 design in which treatments are based on variations in θ and C . Table 1 reports the treatments.

Table 1: Summary of Experimental Treatments

Uncertainty	Informing	Parameters			Procedure	
		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
Moderate	Costless	0.10	0	10	3	24
Moderate	Costly	0.10	5	10	3	24
Strong	Costless	0.30	0	10	3	24
Strong	Costly	0.30	5	10	3	24

The parameter space is designed to capture a possible “hump shaped” relationship between the principal’s send decision and θ (see Section 3). Hence, we set θ equal to 0, 0.05, 0.10, and 0.30. We chose these values because, if the relationship is indeed hump-shaped, the fraction of principals choosing send should be higher at a sufficiently modest value of θ relative to $\theta = 0$, while that fraction should be lower when θ is large (i.e., $\theta = 0.30$). We label each level of θ as no nature, weak nature, moderate nature, and strong nature, respectively. We set C equal to 0 and 5 to test how the principal’s send decision and the informing cost interact with uncertainty.¹⁰ We label each level of C as costless informing and costly informing, respectively. The reservation payoff of both the principals and agents, Ω , was 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 winning the prize of \$4 is increasing in the sequence (for details, see Appendix A.3). Our measure of risk aversion is the number of times 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 24 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 576 subjects. As noted before, upon arriving to the lab, subjects were randomly assigned as principals or agents. The currency used in the experiments was francs. Francs were converted to USD at a rate of 20 francs to \$1. At the end of each session, subjects’ total francs were summed from all 12 periods and converted to USD. 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 \$18.64.

3 Predictions

The experiment laid out in the previous section is motivated by the types of multilateral punishment institutions studied in Greif (1989, 1993), Bernstein (1992, 2001), Landa (1994), Clay (1997), Okazaki (2005), and Richman (2017b). Those papers describe historical and contemporary private-order institutions that provide solutions for long distance and (quasi-) impersonal exchange in situations where third-party, public institutions are either unavailable or lack enforcement mechanisms. The common thread connecting these institutions is that they consisted of closed groups who relied on *reputation* engendered by one’s past actions

¹⁰Note that $C = 5$ is a large cost—it is 25% of the best payout the principal could receive in a period. We chose such a high cost in order to observe how behavior changes when the cost of punishment is salient. Numerous experiments find that modest punishments crowd out trust (Fehr and Rockenbach 2003; Fehr and List 2004; Smith and Wilson 2017).

and communication of those actions to other members of the group.

In this paper we focus on two issues that affect how multilateral punishment institutions work: uncertainty regarding whether one party reneged on the agreement and costly institutional use. We are not the first to consider these issues. Indeed, the experiment laid out in Section 2 is in the spirit of Kandori (1992), who considers a game with the following three attributes: i) a label is attached to each agent (i.e., reputation); ii) before executing trade each agent observes her and her partner's label; iii) a player and her partner's actions and labels today determine their labels tomorrow. Yet, there is an important difference between the focus of our studies: Kandori (1992) considers trade under uncertainty regarding the *reputation* of others, similar to the theoretical literature on the problem of collusion in oligopoly (e.g., Green and Porter 1984; Abreu et al. 1986; Ali et al. 2017) and studies of institutions that mitigate such uncertainty (e.g., Milgrom et al. 1990; Bernstein 1992). Our experiment is closer in spirit to Greif (1989, 1993), Clay (1997), and McMillan and Woodruff (1999), who highlight uncertainty with respect to the action taken by one's partner in the present transaction (although Greif (1993) explicitly models a full information game).

While it is straight-forward enough to add trade uncertainty and costly institutional use to standard game-theoretic models, we are also interested in how uncertainty and costly institutions interact with *behavioral* features that often entail actions inconsistent with standard neo-classical economic theory. For instance, in trust games such as the one laid out here, it is common for participants to send significantly more than the Nash equilibrium amount of zero, although this behavior tends to decay over time (Burnham et al. 2000). In the context of our experiment, it is possible that some degree of altruism encourages agents to return (i.e., Divide) to the principal even if the agent has a bad reputation and can thus gain nothing in the future from doing so (Andreoni 1990; Levine 1998; Charness and Haruvy 2002; Fehr and Schmidt 2006; Carpenter et al. 2008). Likewise, it is possible that the principal punishes agents in the third stage, even though it is costly to do so and the principal will never interact again with the agent, for behavioral reasons such as a feeling of betrayal (Bohnet and Zeckhauser 2004; Hong and Bohnet 2007; 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), or punishment (Fehr and List 2004; Fehr and Fischbacher 2004; Smith and Wilson 2017).¹¹ Our experiment does not allow us to delineate between these various behavioral motives and we thus remain agnostic as to which are salient.

For the remainder of this section, we provide intuition for how subjects might act by considering the trade-offs they face in light of their own behavioral characteristics.¹² First, consider how principals act in any period if the game enters the third stage, meaning that either the agent cheated them (i.e., chose Do

¹¹On the contrary, there are also pro-social behavioral reasons why principals may punish agents, such as gift exchange (Fehr et al. 1997, 2007; Charness and Kuhn 2011; Rubin and Sheremeta 2016) aimed at inducing cooperative behavior by agents.

¹²In Appendix B, we provide a formal model with certain behavioral assumptions that provides similar results.

Not Divide) or the agent acted honestly but nature destroyed the principal's share of the surplus (with probability θ). Clearly, for purely economic reasons, the likelihood that a principal punishes should be decreasing in the cost of punishment, C . Meanwhile, if the agent already has a bad reputation, there will be little incentive—behavioral or economic—to choose a costly punishment, since such a punishment does not affect the agent's reputation. On the other hand, if the agent has a good reputation, the principal may decide to punish the agent, even if it is costly. When making this decision, a principal whose decision is affected by a sense of betrayal, blame, spite, or punishment might also take into account that there is some probability that the agent acted honestly (except in the $\theta = 0$ treatment). We therefore conjecture that the likelihood that principals punish (i.e. Inform) in the third stage, conditional on being matched with an agent with a good reputation, is decreasing in θ . This yields the following prediction:

Prediction 1: The proportion of principals choosing Inform in the third stage is decreasing in θ and C .

Next, consider agents' actions in the second stage. Agents face the following trade-off: choosing Do Not Divide gives them a greater payoff in the present period but may affect their reputation, and thus payoff, in future periods. Agents with bad reputations do not face the latter motivation, meaning that a money-maximizing agent with a bad reputation would always choose Do Not Divide. However, some agents with bad reputations may be motivated by altruism or reciprocity to choose Divide. We conjecture that such agents are more likely to be those that chose to Divide in previous periods but were hit with a bad stroke of luck by nature.¹³ This would indicate that, on average, agents with bad reputations are more likely to choose Divide as θ increases, since the proportion of such agents who choose Divide for behavioral reasons is greater. On the other hand, higher values of θ are associated with a greater likelihood that the agent's sacrifice of half their income will be for naught, since nature will take it away in any case. Which of these effects dominates is ambiguous and depends on the strength of the behavioral parameters and their distribution in the population. Since our experiment does not permit such tests, we do not provide a prediction regarding how θ affects the Divide decision of agents with bad reputations.

A similar logic can be applied to agents who enter the period with a good reputation, although these agents also consider the expected value of future payouts from maintaining a good reputation. This expected value is clearly decreasing in θ , since the probability that agents receive a bad reputation even if they choose Divide is increasing in θ (due to the higher likelihood that the principal receives nothing due to nature). Whether this monetary effect outweighs the behavioral effects noted above is an empirical question. Hence, regardless of the agent's reputation, the proportion of agents choosing Divide in the second stage is ambiguous

¹³Although it is possible that some agents who chose Not Divide in previous periods and received a bad reputation as a result will feel gratitude towards the principal who chose to Divide with them in the present period.

in θ .

Finally, we consider the decisions made by principals in the first stage. When principals are matched with agents who have bad reputations, they must consider the following: i) agents with bad reputations have no monetary incentive to choose Divide in the following round; ii) there may be some agents with behavioral motivation (such as altruism) to Divide but were hit with bad luck in previous periods; iii) nature might destroy the principal's share even if the agent chooses Divide. The second and third considerations are a function of θ . By definition, the likelihood that nature destroys the principal's share is increasing in θ . On the other hand, as θ increases, it is also more likely that the principal is matched with an agent who chose Divide in the past but was subject to bad luck. At sufficiently high values of θ the third consideration must takeover: regardless of what the agent does, the principal is likely to get 0 from choosing send and thus chooses Do Not Send (indeed, at $\theta > 0.5$ the expected value to the principal from choosing Send is lower than choosing Do Not Send *even if* the principal expects the agent to divide with probability 1). It is an empirical question as to which effect dominates at low levels of θ . Principals matched with agents with good reputations face a similar decision as those matched with bad reputations, although they are more likely to be matched with an agent who chooses to Divide. Combined, these insights yield the following prediction:

Prediction 2: The proportion of principals choosing Send in the first stage is either hump-shaped or decreasing in θ .

4 Results

4.1 Summary Statistics and Analysis

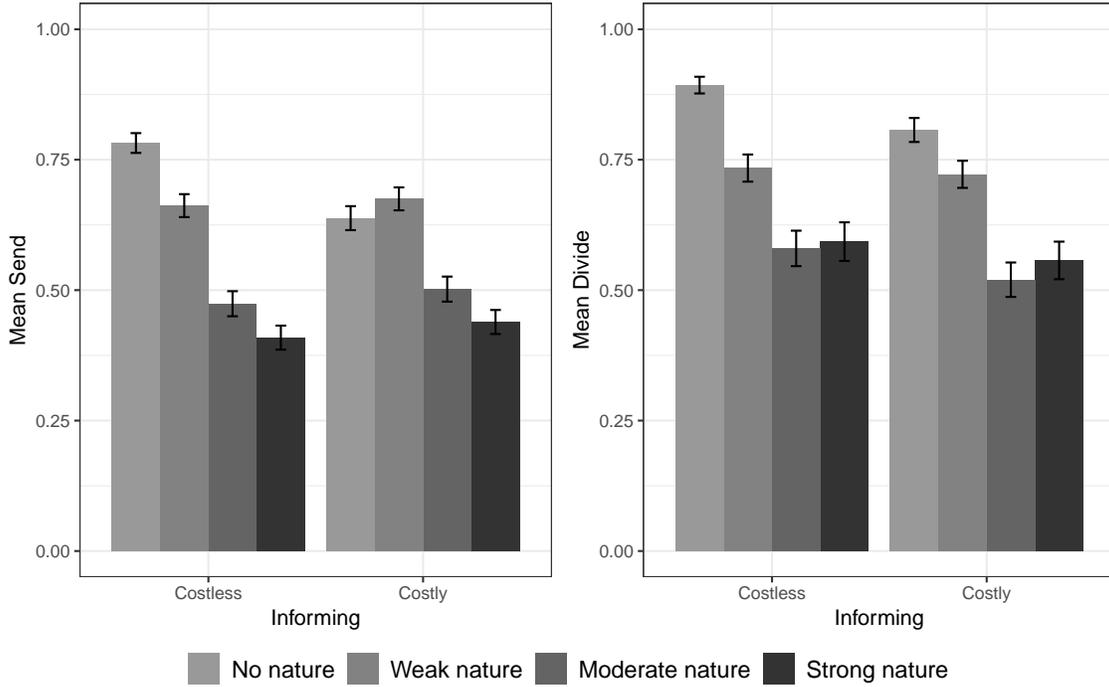
In this section, we examine principals' decision to send and agents' decision to divide unconditional on the agents' reputation. Figure 2 reports the mean amount sent in each treatment.¹⁴

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. 0.47 ($\theta = 0.10$) vs. 0.41 ($\theta = 0.30$).¹⁵ 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. 0.50

¹⁴Appendix Table D.1 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. Appendix Table D.4 reports the earnings by participants by treatment.

¹⁵P-values for the various tests of differences are 0.05 ($\theta = 0$ vs. $\theta = 0.05$), 0.05 ($\theta = 0.05$ vs. $\theta = 0.10$), and 0.51 ($\theta = 0.10$ vs. $\theta = 0.30$). All p-values reported for non-parametric comparisons in this paper are from two-sided Mann-Whitney-Wilcoxon rank sum tests at the session level.

Figure 2: Send and Divide Decisions by Treatment



($\theta = 0.10$) vs. 0.44 ($\theta = 0.30$).¹⁶ 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, $p = 0.05$).

We next turn to the average frequency which agents chose to divide, conditional on the game reaching Stage 2 (see Figure 2). 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. 0.58 ($\theta = 0.10$) vs. 0.59 ($\theta = 0.30$).¹⁷ 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. 0.52 ($\theta = 0.10$) vs. 0.56 ($\theta = 0.30$).¹⁸ As was the case for the send decision, the cost of informing is only significantly related to the divide frequency when $\theta = 0$ (0.89 vs. 0.81, $p = 0.05$).

Regardless of informing cost, the mean of francs earned per period by principals is decreasing in uncertainty (θ): 16.15 ($\theta = 0$) vs. 12.45 ($\theta = 0.05$) vs. 10.25 ($\theta = 0.10$) vs. 9.18 ($\theta = 0.30$) when informing is costless¹⁹ and 13.65 ($\theta = 0$) vs. 12.45 ($\theta = 0.05$) vs. 9.40 ($\theta = 0.10$) vs. 8.72 ($\theta = 0.30$) when informing is costly (see Figure 3).²⁰ On the other hand, the mean francs earned per period by agents is hump-shape as

¹⁶P-values for the various tests of differences are 0.51 ($\theta = 0$ vs. $\theta = 0.05$), 0.05 ($\theta = 0.05$ vs. $\theta = 0.10$), and 0.51 ($\theta = 0.10$ vs. $\theta = 0.30$).

¹⁷P-values for the various tests of differences are 0.05 ($\theta = 0$ vs. $\theta = 0.05$), 0.28 ($\theta = 0.05$ vs. $\theta = 0.10$), and 0.83 ($\theta = 0.10$ vs. $\theta = 0.30$).

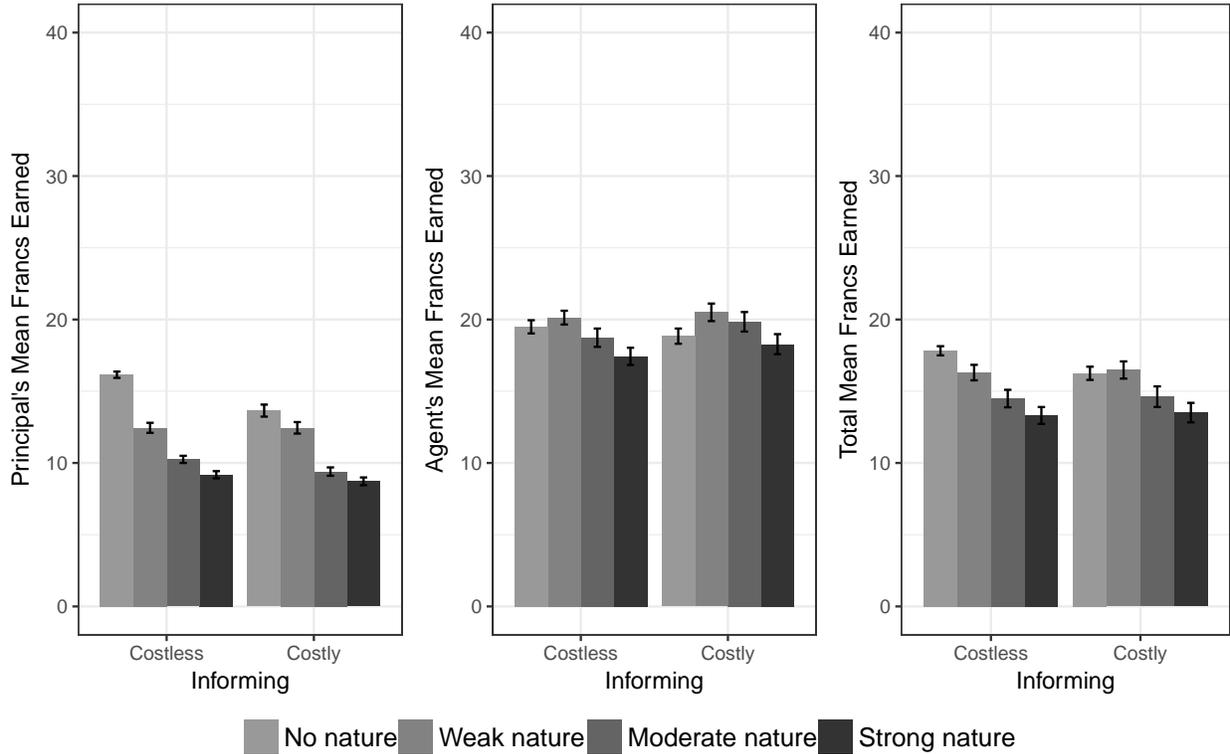
¹⁸P-values for the various tests of differences are 0.05 ($\theta = 0$ vs. $\theta = 0.05$), 0.05 ($\theta = 0.05$ vs. $\theta = 0.10$), and 0.83 ($\theta = 0.10$ vs. $\theta = 0.30$).

¹⁹P-values for the various tests of differences are 0.05 ($\theta = 0$ vs. $\theta = 0.05$), 0.05 ($\theta = 0.05$ vs. $\theta = 0.10$), and 0.05 ($\theta = 0.10$ vs. $\theta = 0.30$).

²⁰P-values for the various tests of differences are 0.28 ($\theta = 0$ vs. $\theta = 0.05$), 0.05 ($\theta = 0.05$ vs. $\theta = 0.10$), and 0.51 ($\theta = 0.10$ vs. $\theta = 0.30$).

uncertainty (θ) increases: 19.49 ($\theta = 0$) vs. 20.13 ($\theta = 0.05$) vs. 18.73 ($\theta = 0.10$) vs. 17.43 ($\theta = 0.30$) when informing is costless²¹ and 18.84 ($\theta = 0$) vs. 20.50 ($\theta = 0.05$) vs. 19.84 ($\theta = 0.10$) vs. 18.28 ($\theta = 0.30$) when informing is costly.²²

Figure 3: Francs Earned per Period by Role and Treatment



These results suggest the possibility, laid out in Section 3, that at low levels of uncertainty (i.e., $\theta = 0.05$) principals give agents the benefit of the doubt and send to them even if their reputation is bad. This could entail greater average payouts for agents and lower average payouts for principals, if the latter end up giving “too much” benefit of the doubt to the former. Moreover, these effects could cancel each other out with respect to *total welfare*, as is suggested in the right-most figure in Figure 3. We test this possibility in Section 4.2, which breaks down actions conditional on the agent’s reputation at the beginning of the period.

Indeed, our key results relate to *total welfare*, the combined amount earned by principals and agents. When informing is costless, there is total welfare is decreasing in θ (17.82 vs. 16.30 vs. 14.49 vs. 13.31, $p = 0.05$ for all differences except for $\theta = 0.10$ vs. $\theta = 0.30$, $p = 0.28$). However, when informing is costly—the historical case of interest—there is no difference in the $\theta = 0$ and $\theta = 0.05$ treatments (16.25 vs. 16.48,

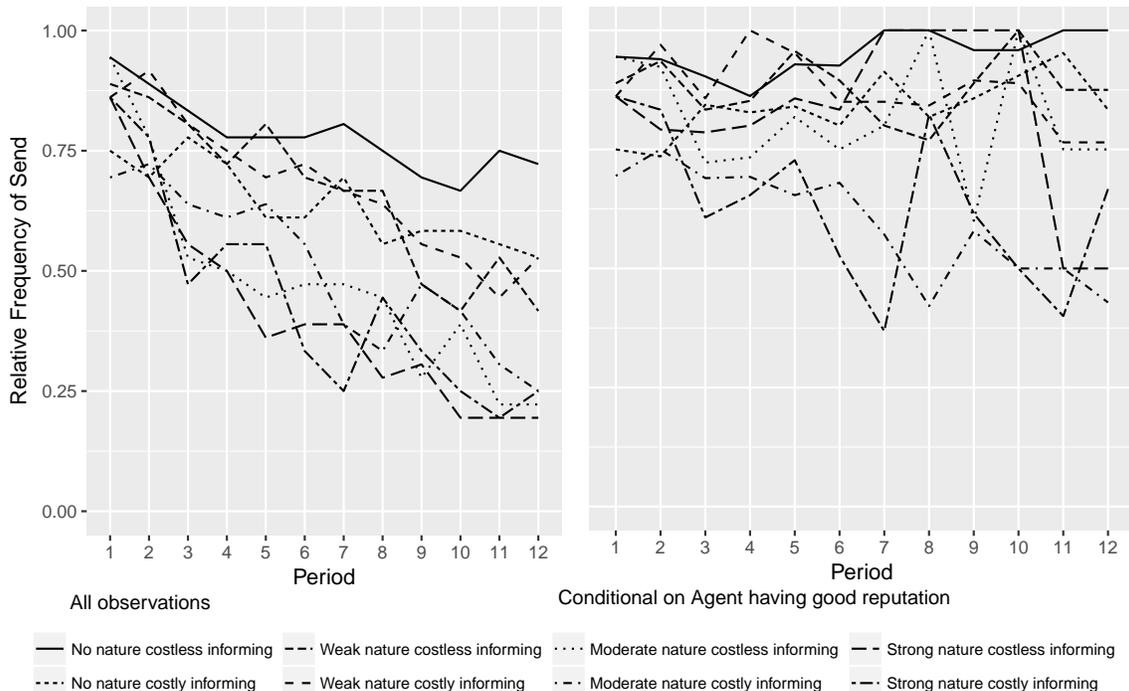
²¹P-values for the various tests of differences are 0.38 ($\theta = 0$ vs. $\theta = 0.05$), 0.18 ($\theta = 0.05$ vs. $\theta = 0.10$), and 0.28 ($\theta = 0.10$ vs. $\theta = 0.30$).

²²P-values for the various tests of differences are 0.13 ($\theta = 0$ vs. $\theta = 0.05$), 0.26 ($\theta = 0.05$ vs. $\theta = 0.10$), and 0.28 ($\theta = 0.10$ vs. $\theta = 0.30$).

$p = 0.51$). This is despite the fact that nature destroys half of the welfare in 5% of cases in the $\theta = 0.05$ treatment. We will attempt to provide an explanation for this result in the remainder of the paper.

The interpretation put forth in Section 3 suggests that our results may depend on the *reputation* of the agent at the beginning of the period. Hence, we now ask whether reputation did indeed matter in our experiment, or did the mere imposition of uncertainty (θ) or informing cost (C) affect decisions across treatments independent of reputation?

Figure 4: Relative Frequency of Send, by Treatment and Period



For figure in color, see Appendix Figure D.1.

Some preliminary insight into the mechanisms driving cross-treatment differences can be gleaned from observing the send frequency across periods, as depicted in the left panel of Figure 4. In period one, there is no statistically difference across any of the treatments except for $\{\theta = 0, C = 5\}$ and $\{\theta = 0.10, C = 5\}$. 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. The right panel of Figure 4 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.²³ These results suggest the possibility—also suggested in Section 3—that the patterns associated with the send and divide decisions depend on the agent’s reputation. We

²³In Appendix C, we present a regression analysis to further uncover the determinants of the send and divide decisions. We relegate these results to an appendix because our primary insights relate to the send and divide decisions *conditional* on reputation.

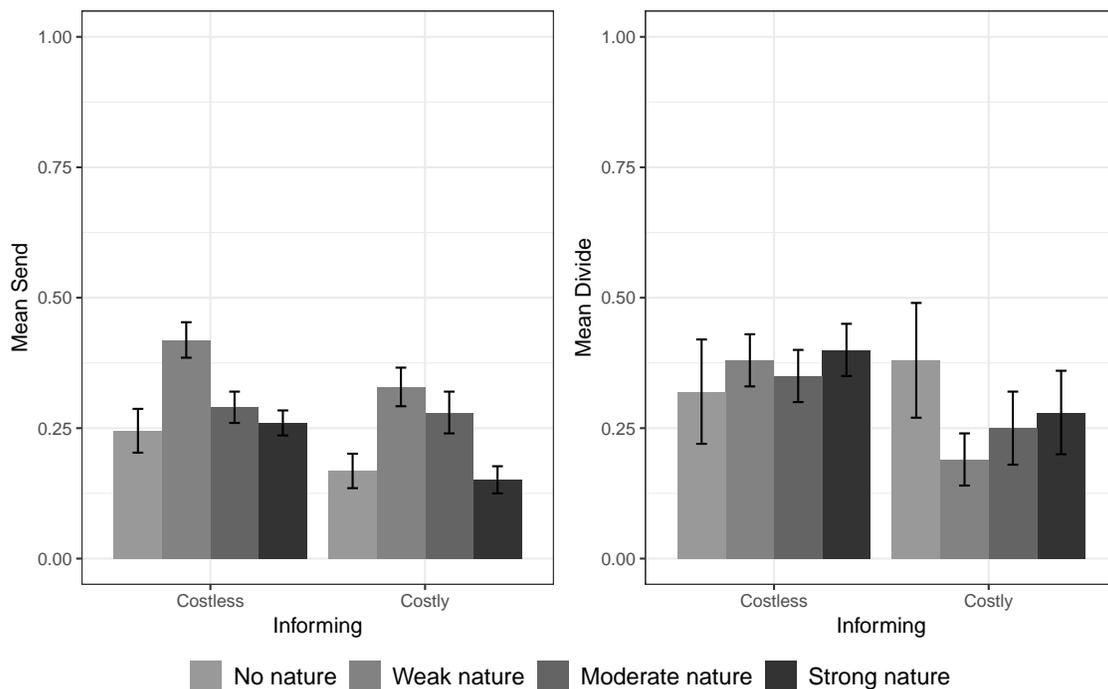
therefore turn to results which condition on the agent’s reputation for the remainder of the analysis.

4.2 Effect of Reputation

4.2.1 Bad 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. We begin by summarizing the send and divide decisions conditional on agents beginning the period with a bad reputation. Prediction 2 suggests that the send frequency should either be hump-shaped or decreasing in θ . Figure 5, which summarizes the decisions of the principals and agents, provides some preliminary evidence in support of this prediction.²⁴ Conditional on C , the send frequency appears hump-shaped in θ . Meanwhile, conditional on θ , the send frequency is decreasing in C .²⁵

Figure 5: Send and Divide Decision by Treatment, Conditional on Agent Having Bad Reputation



We test the relationship between uncertainty and subjects’ decisions more rigorously by estimating a random-effects model using the GLS estimator where the dependent variable is a send or divide dummy.²⁶ In the first specification (columns 1 and 4), we include treatment dummies. In the second specification

²⁴The summary statistics of the send and divide decisions, conditional on agents having a bad reputation, are available in Appendix Table D.2.

²⁵Appendix Figure D.2 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.

²⁶In all the specifications reported in the body of the paper we have clustered the standard errors at the session level. As a robustness check, in Tables D.5-D.7, we replicated all tables in the body of the paper using the bootstrap method. All the main findings are robust with respect to the clustering method.

(columns 2 and 5), we include controls for risk-aversion obtained from the incentivized risk elicitation task (for the results of this task, see Appendix E), gender,²⁷ and path dependence (period dummy). In the third specification (columns 3 and 6), we include interactions between the uncertainty treatment dummies and the costly informing treatment dummy. In all specifications, the omitted uncertainty dummy is “weak nature.”

The results from our most robust specification (column 3) indicate that principals send significantly less frequently upon seeing a bad reputation— 14.1, 13.9, and 16.9 percentage points less frequently—in the $\theta = 0$, $\theta = 0.10$, and $\theta = 0.30$ treatments relative to the $\theta = 0.05$ treatment when informing is costless. When informing is costly, there is no significant difference in the frequency of send between $\theta = 0.05$ and $\theta = 0.10$ ($p = 0.57$)²⁸ while principals send significantly less frequently in the $\theta = 0$ and $\theta = 0.30$ treatments relative to the $\theta = 0.05$ treatment ($p = 0.00$ and 0.04 , respectively). In other words, the relationship between the send frequency and θ , conditional on the agent having a bad reputation, is *hump-shaped*.

Moreover, principals send 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 differences are statistically significant only when $\theta = 0$ (p – value = 0.08). These findings are summarized in Result 1.

Result 1: 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 results reported in Table 2 indicate that agents with bad reputations choose to divide more in the $\theta = 0$ treatment relative to the $\theta = 0.05$, $\theta = 0.10$, and the $\theta = 0.30$ treatments when informing is costly ($p = 0.00$, 0.07 , and 0.05 , respectively). Meanwhile, when informing is costless, there are no cross-treatment differences in the proportion of agents with bad reputations who divide. 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 45% of agents had a bad reputation in any one given period, and only 28% of principals chose to send upon being matched with an agent with a bad reputation.

Combined, these results suggest that principals indeed do give too much “benefit of the doubt” to agents with a bad reputation in the presence of low levels of uncertainty (i.e., $\theta = 0.05$ or 0.10). Regardless of whether informing is costly or not, principals send much more frequently to agents with bad reputation when $\theta = 0.05$ than they do when $\theta = 0$. In the presence of costly punishment, this is also true when $\theta = 0.10$. This occurs *despite* the fact that bad-reputation agents do not divide (significantly) more in the $\theta = 0.05$ or $\theta = 0.10$ treatments; indeed, when informing is costly, bad-reputation agents divide *less* in

²⁷In the questionnaire, we collected information regarding the gender and major of the subjects. The summary statistics of the gender and major are available in Appendix Table F.1.

²⁸All p-values reported for parametric comparisons in this paper are from two-sided Wald tests.

Table 2: Regression Analysis of the Decision to Send and Divide Conditional on Bad Reputation

Dependent Variable:	Send (0/1)			Divide (0/1)		
	(1)	(2)	(3)	(4)	(5)	(6)
No Nature	-0.143***	-0.145***	-0.141*	0.099	0.117	0.032
$\theta = 0.00$	(0.041)	(0.046)	(0.073)	(0.069)	(0.081)	(0.133)
Moderate Nature	-0.076*	-0.086*	-0.139**	-0.017	-0.058	-0.105
$\theta = 0.10$	(0.045)	(0.046)	(0.071)	(0.075)	(0.079)	(0.108)
Strong Nature	-0.153***	-0.165***	-0.169**	-0.001	-0.022	-0.023
$\theta = 0.30$	(0.056)	(0.057)	(0.081)	(0.092)	(0.091)	(0.122)
Costly Informing	-0.067*	-0.062*	-0.093	-0.132**	-0.139**	-0.197**
	(0.036)	(0.037)	(0.066)	(0.058)	(0.059)	(0.099)
No Nature \times Costly Informing			-0.007			0.177
			(0.086)			(0.140)
Moderate Nature \times Costly Informing			0.116			0.124
			(0.081)			(0.146)
Strong Nature \times Costly Informing			0.007			-0.009
			(0.114)			(0.172)
Risk		0.000	0.001		-0.001	-0.001
		(0.012)	(0.012)		(0.016)	(0.016)
Female		0.031	0.035		0.227***	0.232***
		(0.038)	(0.038)		(0.057)	(0.056)
Period		-0.026***	-0.026***		-0.021***	-0.021***
		(0.003)	(0.003)		(0.006)	(0.006)
Intercept	0.397***	0.594***	0.604***	0.381***	0.435***	0.454***
	(0.040)	(0.081)	(0.095)	(0.072)	(0.105)	(0.127)
Observations	1548	1548	1548	426	426	426
R^2	0.04	0.03	0.04	0.04	0.10	0.11

Standard errors clustered by session in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Moderate Nature ($\theta = 0.05$) and Costless Informing ($C = 0$) are the omitted categories.

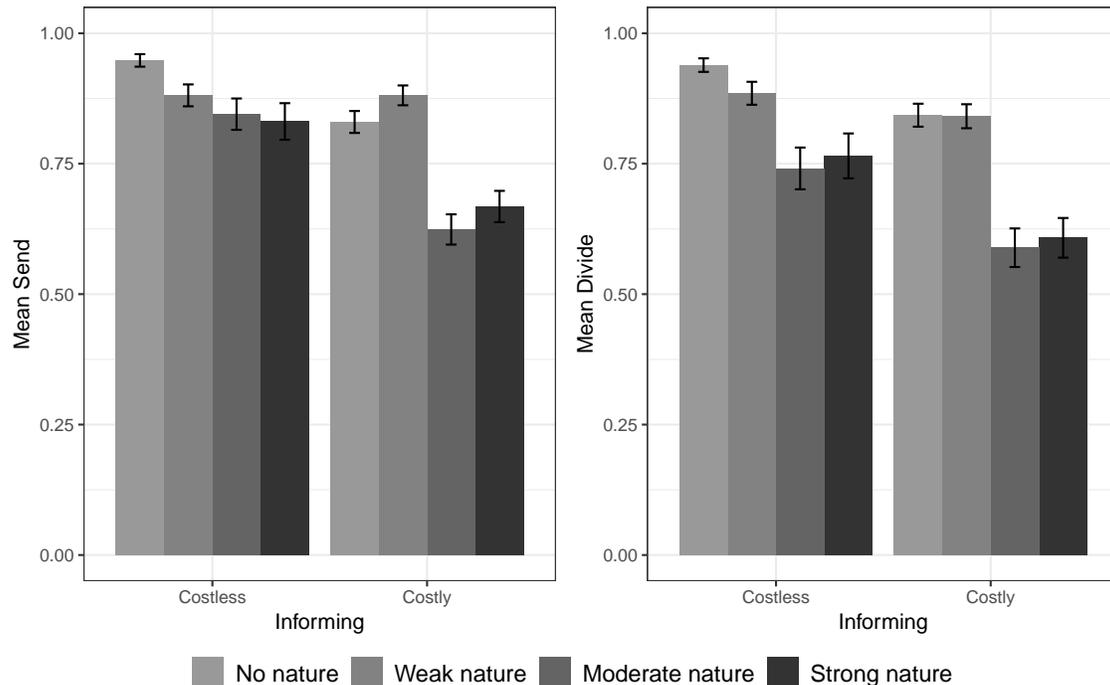
these treatments. These results thus provide insight into the pattern of welfare noted in Figure 3: welfare is decreasing in θ for principals but is hump-shaped in θ for agents.

4.2.2 Good Reputation

We conclude this section by analyzing the periods in which agents had a good reputation. Figure 6 shows the average frequency that principals chose send and agents chose divide, conditional on the agent having a good reputation.²⁹ Conditional on informing being costless, both the send and divide frequencies are decreasing in θ . However, when informing is costly, both the send and divide frequencies appear somewhat hump-shaped in θ . Further, both the send and divide frequencies are decreasing in C , conditional on θ (although this effect is small at $\theta = 0.05$).

²⁹Appendix Table D.3 reports the corresponding summary statistics for the send and divide decisions conditional on having a good reputation.

Figure 6: Send and Divide Decision by Treatment, Conditional on Agent Having Good Reputation



In order to provide statistical support for these insights, 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 Table 3, estimate the same model as Table 2, restricting the sample to periods in which agents have a good reputation.

The point estimates in Column 3 indicate that when informing is costless, the fraction of principals choosing send is decreasing in θ . However, these results are not statistically significant, and we do not interpret them as such. Column 3 also indicates that when informing is costly, principals send less often when uncertainty is moderate ($\theta = 0.10$) or high ($\theta = 0.30$), but there is no statistically significant difference in the send frequency between the $\theta = 0$ and $\theta = 0.05$ treatments ($p = 0.36$). Meanwhile, the results reported in Column 3 indicate that the fraction of principals choosing send is decreasing in C when $\theta = 0$, $\theta = 0.10$, or $\theta = 0.30$ ($p = 0.02, 0.00$, and 0.09 , respectively). However, the effect of C on the principals’ send decision is practically 0 and insignificant when $\theta = 0.05$. These findings are summarized in Result 2.

Result 2: Conditional on being matched with an agent with a good reputation, the principals’ send frequency is weakly decreasing in θ and C , although these trends are noisily estimated.

In conjunction with Result 1, Result 2 sheds light on the primary finding of this paper: in the presence of a costly reputation institution, a small amount of uncertainty (i.e., $\theta = 0.05$) *does not affect total welfare*. As

Table 3: Regression Analysis of the Decision to Send and Divide Conditional on Good Reputation

Dependent Variable:	Send (0/1)			Divide (0/1)		
	(1)	(2)	(3)	(4)	(5)	(6)
No Nature	0.006	0.015	0.073	0.045	0.067**	0.083*
$\theta = 0.00$	(0.045)	(0.041)	(0.051)	(0.031)	(0.031)	(0.047)
Moderate Nature	-0.150***	-0.155***	-0.039	-0.192***	-0.217***	-0.228***
$\theta = 0.10$	(0.045)	(0.043)	(0.052)	(0.033)	(0.039)	(0.055)
Strong Nature	-0.138***	-0.139***	-0.058	-0.150***	-0.181***	-0.161*
$\theta = 0.30$	(0.053)	(0.053)	(0.095)	(0.048)	(0.046)	(0.093)
Costly Informing	-0.119***	-0.107***	0.011	-0.106***	-0.092***	-0.079*
	(0.033)	(0.032)	(0.051)	(0.029)	(0.029)	(0.042)
No Nature \times Costly Informing			-0.115*			-0.032
			(0.068)			(0.065)
Moderate Nature \times Costly Informing			-0.217***			0.018
			(0.058)			(0.076)
Strong Nature \times Costly Informing			-0.154			-0.039
			(0.099)			(0.098)
Risk		0.017***	0.016***		-0.003	-0.003
		(0.007)	(0.006)		(0.010)	(0.010)
Female		-0.001	-0.007		0.057	0.058
		(0.028)	(0.028)		(0.045)	(0.046)
Period		-0.007	-0.007		-0.020***	-0.020***
		(0.005)	(0.005)		(0.003)	(0.003)
Intercept	0.947***	0.906***	0.851***	0.814***	0.885***	0.879***
	(0.045)	(0.069)	(0.068)	(0.030)	(0.058)	(0.057)
Observations	1908	1908	1908	1555	1555	1555
R^2	0.12	0.14	0.16	0.01	0.01	0.01

Standard errors clustered by session in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Moderate Nature ($\theta = 0.05$) and Costless Informing ($C = 0$) are the omitted categories.

uncertainty increases, principals are less willing to engage in exchange with agents with good reputations, presumably because nature may reduce their payout even if the agent acts honestly. Meanwhile, principals are more willing to trust agents with bad reputations—giving them too much “benefit of the doubt”—which results in greater welfare for agents and less welfare for principals. Combined—in our experiment, at least—these two effects seem to cancel each other out, resulting in a similar level of total welfare in the $\theta = 0$ and $\theta = 0.05$ treatments.

Finally, we turn to the agents’ divide decision. Columns 4-6 of Table 3 reveal that the frequency with which agents choose to divide is decreasing in θ , with all point estimates being marginally significant. Column 6 indicates that the frequency with which agents choose to divide is decreasing in C , with some point estimates being marginally significant. These findings are summarized in Result 3.

Result 3: Conditional on having a good reputation at the beginning of the period, the frequency with which

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

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 has little effect on overall welfare, especially when informing is costly, because modest uncertainty affects willingness to trade in two countervailing ways. On the one hand, higher levels of uncertainty entail that principals are less likely to engage in exchange with agents with good reputations, presumably because there is some probability that the exchange will turn out bad for the principal even if the agent acts honestly. On the other hand, when there is modest uncertainty principals are *more* likely to engage in exchange with agents with bad reputations, presumably because there is some chance that the agent is “pro-social” and was hit with a stroke of bad luck in a previous period.

Our results have implications for the functioning of reputation-based institutions. They suggest the possibility that these institutions not only function when a modest amount of uncertainty is present—they may function about as well in terms of total welfare. Importantly, the type of uncertainty we are concerned with here is with respect to the actions of parties in the *present transaction*, not uncertainty about their past actions (as in Milgrom et al. (1990), Kandori (1992), and Ali and Miller (2016)). Such uncertainty is inherent to all types of exchange relationships, both in historical settings where information is costly to transmit (Greif 1993; Clay 1997; Okazaki 2005) and modern settings where information asymmetries make verification difficult (McMillan and Woodruff 1999). Our experiment suggests the possibility that pro-social behavioral characteristics can mitigate some of the disincentives to exchange and thus can augment the effectiveness of reputation-based institutions. This is consistent with McMillan and Woodruff (1999), who find that Vietnamese merchants often continue to interact with those who have been known to renege in the past, knowing that there is a decent probability that those individuals acted honestly but were hit with a stroke of bad luck.

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 (or, at least, not increase). 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

- Abreu, Dilip, Pearce, David, and Stacchetti, Ennio. 1986. "Optimal Cartel Equilibria with Imperfect Monitoring." 39 *Journal of Economic Theory* 251–269.
- Ali, S. Nageeb and Miller, David A. 2016. "Ostracism and Forgiveness." 106 *American Economic Review* 2329–48.
- Ali, S. Nageeb, Miller, David A., and Yang, David Yilin. 2017. "Renegotiation-Proof Multilateral Enforcement." *Working paper*.
- Ambrus, Attila and Greiner, Ben. 2012. "Imperfect Public Monitoring with Costly Punishment: An Experimental Study." 102 *American Economic Review* 3317–32.
- Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-glow Giving." 100 *Economic Journal* 464–477.
- Bartling, Björn, Engl, Florian, and Weber, Roberto A. 2014. "Does Willful Ignorance Deflect Punishment?—An Experimental Study." 70 *European Economic Review* 512–524.
- Berg, Joyce, Dickhaut, John, and McCabe, Kevin. 1995. "Trust, Reciprocity, and Social History." 10 *Games and Economic Behavior* 122–142.
- Bernstein, Lisa. 1992. "Opting out of the Legal System: Extralegal Contractual Relations in the Diamond Industry." 21 *Journal of Legal Studies* 115–157.
- Bernstein, Lisa. 2001. "Private Commercial Law in the Cotton Industry: Creating Cooperation through Rules, Norms, and Institutions." 99 *Michigan Law Review* 1724–90.
- Besley, Timothy and Coate, Stephen. 1995. "Group Lending, Repayment Incentives and Social Collateral." 46 *Journal of Development Economics* 1–18.
- Boerner, Lars and Ritschl, Albrecht. 2009. "The Economic History of Sovereignty: Communal Responsibility, the Extended Family, and the Firm." 165 *Journal of Institutional and Theoretical Economics* 99–112.
- Bohnet, Iris, Frey, Bruno S., and Huck, Steffen. 2001. "More Order with Less Law: On Contract Enforcement, Trust, and Crowding." 95 *American Political Science Review* 131–144.
- Bohnet, Iris, Greig, Fiona, Herrmann, Benedikt, and Zeckhauser, Richard. 2008. "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States." 98 *American Economic Review* 294–310.
- Bohnet, Iris, Harmgart, Heike, Tyran, Jean-Robert, et al. 2005. "Learning Trust." 3 *Journal of the European Economic Association* 322–329.

- Bohnet, Iris and Zeckhauser, Richard. 2004. "Trust, Risk and Betrayal." 55 *Journal of Economic Behavior & Organization* 467–484.
- Bolton, Gary E., Katok, Elena, and Ockenfels, Axel. 2004. "How Effective are Electronic Reputation Mechanisms? An Experimental Investigation." 50 *Management Science* 1587–1602.
- Bolton, Gary E., Katok, Elena, and Ockenfels, Axel. 2005. "Cooperation among Strangers with Limited Information about Reputation." 89 *Journal of Public Economics* 1457–1468.
- Bornstein, Gary and Weisel, Ori. 2010. "Punishment, Cooperation, and Cheater Detection in Noisy Social Exchange." 1 *Games* 18–33.
- Bracht, Juergen and Feltovich, Nick. 2009. "Whatever You Say, Your Reputation Precedes You: Observation and Cheap Talk in the Trust Game." 93 *Journal of Public Economics* 1036–1044.
- Burnham, Terence, McCabe, Kevin, and Smith, Vernon L. 2000. "Friend-or-Foe Intentionality Priming in an Extensive Form Trust Game." 43 *Journal of Economic Behavior & Organization* 57–73.
- Carpenter, Jeffrey, Connolly, Cristina, and Myers, Caitlin Knowles. 2008. "Altruistic Behavior in a Representative Dictator Experiment." 11 *Experimental Economics* 282–298.
- Cason, Timothy N., Saijo, Tatsuyoshi, and Yamato, Takehiko. 2002. "Voluntary Participation and Spite in Public Good Provision Experiments: An International Comparison." 5 *Experimental Economics* 133–153.
- Cassar, Alessandra, Friedman, Daniel, and Schneider, Patricia Higinio. 2009. "Cheating in Markets: A Laboratory Experiment." 72 *Journal of Economic Behavior & Organization* 240–259.
- Cassar, Alessandra, Friedman, Daniel, and Schneider, Patricia Higinio. 2010. "A Laboratory Investigation of Networked Markets." 120 *Economic Journal* 919–943.
- Charness, Gary, Du, Ninghua, and Yang, Chun-Lei. 2011. "Trust and Trustworthiness Reputations in An Investment Game." 72 *Games and Economic Behavior* 361–375.
- Charness, Gary and Haruvy, Ernan. 2002. "Altruism, Equity, and Reciprocity in a Gift-Exchange Experiment: An Encompassing Approach." 40 *Games and Economic Behavior* 203–231.
- Charness, Gary and Kuhn, Peter. 2011. "Lab Labor: What can Labor Economists Learn from the Lab?" 4 *Handbook of Labor Economics* 229–330.
- Clay, Karen. 1997. "Trade without Law: Private-order Institutions in Mexican California." 13 *Journal of Law, Economics, and Organization* 202–231.
- Falk, Armin and Kosfeld, Michael. 2006. "The Hidden Costs of Control." 96 *American Economic Review* 1611–1630.
- Fehr, Ernst and Fischbacher, Urs. 2004. "Third-party Punishment and Social Norms." 25 *Evolution and Human Behavior* 63–87.
- Fehr, Ernst, Gächter, Simon, and Kirchsteiger, Georg. 1997. "Reciprocity as a Contract Enforcement Device: Experimental Evidence." *Econometrica* 833–860.
- Fehr, Ernst, Klein, Alexander, and Schmidt, Klaus M. 2007. "Fairness and Contract Design." 75 *Econometrica* 121–154.
- Fehr, Ernst and List, John A. 2004. "The Hidden Costs and Returns of Incentives—Trust and Trustworthiness among CEOs." 2 *Journal of the European Economic Association* 743–771.
- Fehr, Ernst and Rockenbach, Bettina. 2003. "Detrimental Effects of Sanctions on Human Altruism." 422 *Nature* 137–140.

- Fehr, Ernst and Schmidt, Klaus M. 2006. "The Economics of Fairness, Reciprocity and Altruism—Experimental Evidence and New Theories." 1 *Handbook of the Economics of Giving, Altruism and Reciprocity* 615–691.
- Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." 10 *Experimental Economics* 171–178.
- Ghatak, Maitreesh and Guinnane, Timothy W. 1999. "The Economics of Lending with Joint Liability: Theory and Practice." 60 *Journal of Development Economics* 195–228.
- Ghidoni, Riccardo. 2017. "Mistrust and Opposition to Large-Scale Projects: An Experiment on the Role of Uncertainty." *Working Paper*.
- Ghosh, Parikshit and Ray, Debraj. 1996. "Cooperation in Community Interaction without Information Flows." 63 *Review of Economic Studies* 491–519.
- Green, Edward J. and Porter, Robert H. 1984. "Noncooperative Collusion under Imperfect Price Information." *Econometrica* 87–100.
- Greif, Avner. 1989. "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders." 49 *Journal of Economic History* 857–882.
- Greif, Avner. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." 83 *American Economic Review* 525–548.
- Greif, Avner. 2000. "The Fundamental Problem of Exchange: A Research Agenda in Historical Institutional Analysis." 4 *European Review of Economic History* 251–284.
- Greif, Avner. 2002. "Institutions and Impersonal Exchange: From Communal to Individual Responsibility." 158 *Journal of Institutional and Theoretical Economics* 168–204.
- Greif, Avner. 2004. "Impersonal Exchange without Impartial Law: The Community Responsibility System." 5 *Chicago Journal of International Law* 109–138.
- Greif, Avner, Milgrom, Paul, and Weingast, Barry R. 1994. "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild." 102 *Journal of Political Economy* 745–776.
- Gurdal, Mehmet Y., Miller, Joshua B., and Rustichini, Aldo. 2013. "Why Blame?" 121 *Journal of Political Economy* 1205–1247.
- Ho, Benjamin and Huffman, David. forthcoming. "Trust and the Law." In: *Handbook of Behavioral Economics and the Law*.
- Holt, Charles A. and Laury, Susan K. 2002. "Risk Aversion and Incentive Effects." 92 *American Economic Review* 1644–1655.
- Hong, Kessely and Bohnet, Iris. 2007. "Status and Distrust: The Relevance of Inequality and Betrayal Aversion." 28 *Journal of Economic Psychology* 197–213.
- Johnson, Simon, McMillan, John, and Woodruff, Christopher. 2002. "Courts and Relational Contracts." 18 *Journal of Law, Economics, and Organization* 221–277.
- Kandori, Michihiro. 1992. "Social Norms and Community Enforcement." 59 *Review of Economic Studies* 63–80.
- Kimbrough, Erik O. and Reiss, J Philipp. 2012. "Measuring the Distribution of Spitefulness." 7 *PloS One* e41812.
- Kimbrough, Erik O. and Rubin, Jared. 2015. "Sustaining Group Reputation." 31 *Journal of Law, Economics, and Organization* 599–628.

- Kranton, Rachel E. 1996. "Reciprocal Exchange: A Self-sustaining System." *American Economic Review* 830–851.
- Kranton, Rachel E. and Minehart, Deborah F. 2001. "A Theory of Buyer-Seller Networks." 91 *American Economic Review* 485–508.
- Kreps, David M., Milgrom, Paul, Roberts, John, and Wilson, Robert. 1982. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." 27 *Journal of Economic Theory* 245–252.
- Kreps, David M. and Wilson, Robert. 1982. "Reputation and Imperfect Information." 27 *Journal of Economic Theory* 253–279.
- Landa, Janet T. 1981. "A Theory of the Ethnically Homogeneous Middleman Group: An Institutional Alternative to Contract Law." 10 *Journal of Legal Studies* 349–362.
- Landa, Janet T. 1994. *Trust, Ethnicity, and Identity: Beyond the New Institutional Economics of Ethnic Trading Networks, Contract Law, and Gift-Exchange*. University of Michigan Press.
- Leeson, Peter T. 2008. "Social Distance and Self-enforcing Exchange." 37 *Journal of Legal Studies* 161–188.
- Levin, Jonathan. 2009. "The Dynamics of Collective Reputation." 9 *The BE Journal of Theoretical Economics*.
- Levine, David K. 1998. "Modeling Altruism and Spitefulness in Experiments." 1 *Review of Economic Dynamics* 593–622.
- McMillan, John and Woodruff, Christopher. 1999. "Dispute Prevention without Courts in Vietnam." 15 *Journal of Law, Economics, and Organization* 637–658.
- Milgrom, Paul R., North, Douglass C., and Weingast, Barry R. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs." 2 *Economics & Politics* 1–23.
- Okazaki, Tetsuji. 2005. "The Role of the Merchant Coalition in Pre-modern Japanese Economic Development: An Historical Institutional Analysis." 42 *Explorations in Economic History* 184–201.
- Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, Cambridge University Press.
- Ostrom, Elinor. 2005. *Understanding Institutional Diversity*. Princeton University Press: Princeton, NJ.
- Pillutla, Madan M. and Murnighan, J. Keith. 1996. "Unfairness, Anger, and Spite: Emotional Rejections of Ultimatum Offers." 68 *Organizational Behavior and Human Decision Processes* 208–224.
- Posner, Richard A. and Rasmusen, Eric B. 1999. "Creating and Enforcing Norms, with Special Reference to Sanctions." 19 *International Review of Law and Economics* 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.
- Rice, Sarah C. 2012. "Reputation and Uncertainty in Online Markets: An Experimental Study." 23 *Information Systems Research* 436–452.
- Richardson, Gary. 2005. "Craft Guilds and Christianity in Late-Medieval England: A Rational-Choice Analysis." 17 *Rationality and Society* 139–189.
- Richman, Barak D. 2017a. "An Autopsy of Cooperation: Diamond Dealers and the Limits of Trust-Based Exchange." 9 *Journal of Legal Analysis* 247–283.
- Richman, Barak D. 2017b. *Stateless Commerce: The Diamond Network and the Persistence of Relational Exchange*. Harvard University Press.

- Rubin, Jared and Sheremeta, Roman. 2016. "Principal-Agent Settings with Random Shocks." 62 *Management Science* 985–999.
- Smith, Vernon L. and Wilson, Bart J. 2017. "Sentiments, Conduct, and Trust in the Laboratory." 34 *Social Philosophy and Policy*.
- Tirole, Jean. 1996. "A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality)." 63 *Review of Economic Studies* 1–22.
- Winfrey, Jason A. and McCluskey, Jill J. 2005. "Collective Reputation and Quality." 87 *American Journal of Agricultural Economics* 206–213.

Appendices

A Experiment Instructions

This Appendix 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 various nature treatments is the probability that nature overrides Participant 2's decision. Here we only report the strong nature, costly narration (i.e., $\theta = 0.30$ and $C = 5$). The instructions of the other treatments are available upon request.

A.1 Instructions

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 20 francs to 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.

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.

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.

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.

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.

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.

Table A1: Summary of Earnings

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

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 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 to 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.

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

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 earnings 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 Quiz

1. What happens if Participant 1 chooses “not send” in the first stage?
 - (a) The period ends and each participant receives 20 francs
 - (b) The period ends and each participant receives 10 francs
 - (c) Participant 2 has to decide to split or not split in stage 2
 - (d) Participant 1 may decide to report or not report in stage 3

2. When Participant 1 observes that the Participant 2 he/she is matched with has never been reported, it means:
 - (a) Participant 2 has never chosen to not split
 - (b) Other Participant 1s have never reported the Participant 2 in previous rounds
 - (c) Participant 2 has never chosen to split
 - (d) Other Participant 1s must have reported the Participant 2 at least once in previous rounds

3. When Participant 1 observes that the Participant 2 he/she is matched with has been reported, it means:
 - (a) Participant 2 must have chosen not split in all previous rounds
 - (b) Other Participant 1s have never reported the Participant 2 in previous rounds
 - (c) Participant 2 has never chosen to split
 - (d) Other Participant 1s must have reported the Participant 2 at least once in previous rounds

4. When Participant 2 chooses split:
 - (a) The period ends and each participant receives 20 francs
 - (b) Participant 1 may receive 0 francs
 - (c) Participant 1 may report
 - (d) All of the above

5. When Participant 2 chooses not split:
 - (a) Participant 2 receives 20 francs
 - (b) Other Participant 1s will definitely see this Participant 2 as reported in future periods
 - (c) Participant 1 may or may not report
 - (d) Participant 1 receives 10 francs

A.3 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.

		<u>Options</u>	
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 A Simple Model

B.1 Setup

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 2, principals and agents are randomly matched in each period and interact only once. In each period, they interact in the same manner as described in the experiment, which is summarized in the game tree in Figure 1.

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. 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.

B.2 Utility and Player Types

We assume that principals attempt to maximize their lifetime payout, but also exhibit some willingness to punish (WTP) when they receive a bad outcome. For the sake of the model, we define WTP as the willingness to take on a cost to change the reputation of the agent with whom they are matched, even though this punishment does not affect their future monetary payoff. Principals differ only in their WTP parameter, β_i , which is distributed throughout the population over some known cumulative distribution function $F(\cdot)$.³⁰

Principals receive utility from punishing agents 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 the utility they receive from punishing. Denoting the principal's discount factor by $\delta \in (0, 1)$, we write principal i 's period t utility as:

$$U_{i,t}^P = \sum_{j=t}^{\infty} \delta^{j-t} [E[\pi_{i,j}^P] + \beta_i I_j R_j], \quad (\text{B.1})$$

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

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

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. 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:

$$U_{i,t}^A = \sum_{j=t}^{\infty} \delta^{j-t} E[\pi_{i,j}^A], \quad (\text{B.3})$$

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

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

³⁰We do not assume that principals act on their *beliefs* about an agent's true action, which (if it were modeled) would be a function of θ . Incorporating such beliefs into the model would not affect comparative statics results, since beliefs move in the same direction as uncertainty, which we model as the mechanism generating comparative statics with respect to θ .

B.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 WTP 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 (B.2) is greater when $S = 1$ than it is when $S = 0$. A little simplification entails that this occurs when:

$$E[D_t|R=0] \geq \frac{1}{2(1-\theta)}. \quad (\text{B.5})$$

Since $D = 1$ if and only if A is altruistic, $E[D_t|R=0]$ can be re-written $pr(A \text{ is altruistic}|R=0)$. 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 (B.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 $1 - F(C)$. Consequently, the probability of an altruistic agent being informed upon in any one period is $\theta[1 - F(C)]$. Hence, if P 's always choose to send to A 's with good reputations, which we shall assume for the moment, the probability that an altruistic agent has a good reputation in period t is $p^G = (1 - \theta[1 - F(C)])^t$, and the probability it has a bad reputation is $p^B = 1 - p^G$. Denoting q as the probability a strategic agent has a bad reputation in period t (for which we solve below), we write:

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

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 $I = 0$.

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 focus here, as in the paper, on comparative statics with respect to θ . 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 former 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* the size of the parameter space over which P 's choose to send when they are matched with an A with a bad reputation. This logic is formalized in the following prediction.

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 , ceteris paribus.

Proof.

We know from the proof of Proposition 2 (see ahead) 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:

$$\frac{\alpha \left[1 - (1 - \theta [1 - F(C)])^t \right]}{\alpha \left[1 - (1 - \theta [1 - F(C)])^t \right] + (1 - \alpha) \left(1 - [F(C)]^t \right)} \geq \frac{1}{2(1 - \theta)}. \quad (\text{B.7})$$

Rearranging this inequality yields:

$$(1 - 2\theta) \left[1 - (1 - \theta [1 - F(C)])^t \right] \geq \left(\frac{1 - \alpha}{\alpha} \right) \left(1 - [F(C)]^t \right). \quad (\text{B.8})$$

It is clear that (B.8) does not hold at $\theta = 0$ or $\theta \geq 0.5$.³¹ To prove the prediction with respect to θ , we show that the LHS of (B.8), 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 :

$$\frac{\partial \Lambda}{\partial \theta} = -2 \left[1 - (1 - \theta [1 - F(C)])^t \right] + t(1 - 2\theta) [1 - F(C)] (1 - \theta [1 - F(C)])^{t-1}. \quad (\text{B.9})$$

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 $\theta < \frac{1}{2}$.

³¹Formally, it is possible that (B.8) 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 (B.8) is satisfied for some value of θ .

Solving for $\frac{\partial^2 \Lambda}{\partial \theta^2}$ gives:

$$\frac{\partial^2 \Lambda}{\partial \theta^2} = -4t [1 - F(C)] (1 - \theta [1 - F(C)])^{t-1} - t(t-1)(1-2\theta)[1 - F(C)]^2 (1 - \theta [1 - F(C)])^{t-2}, \quad (\text{B.10})$$

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

■

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 $D = 0$. We show below that there is some $\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ 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, and dropping the assumption that P always sends to an agent with a good reputation, 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 below). 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 $\theta \in (1 - \frac{1}{2\alpha}, \Psi(\delta, C)]$, otherwise it chooses $D = 0$;

Stage 3: P chooses $I = 1$ if and only if $\beta_i \geq C$; otherwise P chooses $I = 0$.

Proof.

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

Stage 2: The Stage 2 action of altruist A 's is $D_{i,t} = 1$ by definition. We therefore proceed by considering the Stage 2 actions of strategic A 's. Strategic A 's choose $D_{i,t} = 1$ iff $E[U_{i,t}^A(D_{i,t} = 1) | R_t = 1] \geq$

$E [U_{i,t}^A (D_{i,t} = 0) | R_t = 1]$, where:

$$E [U_{i,t}^A (D_{i,t} = 1) | R_t = 1] = 2\Omega + \delta\theta(1 - F(C))V^B + \delta[1 - \theta(1 - F(C))]V^G, \quad (\text{B.11})$$

$$E [U_{i,t}^A (D_{i,t} = 0) | R_t = 1] = 4\Omega + \delta(1 - F(C))V^B + \delta F(C)V^G, \quad (\text{B.12})$$

where V^B and V^G are A 's continuation payoff from having a bad and good reputation, respectively. In equations (B.11) and (B.12), V^G is recursive. Hence, the utilities can be re-written:

$$U_{i,t}^A [D_{i,t} = 1] = \frac{2\Omega + \delta\theta(1 - F(C))V^B}{1 - \delta[1 - \theta(1 - F(C))]}, \quad (\text{B.13})$$

$$U_{i,t}^A [D_{i,t} = 0] = \frac{4\Omega + \delta(1 - F(C))V^B}{1 - \delta F(C)}. \quad (\text{B.14})$$

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:

$$\begin{aligned} \frac{2\Omega + \delta\theta(1 - F(C))\frac{4\Omega}{1-\delta}}{1 - \delta[1 - \theta(1 - F(C))]} &\geq \frac{4\Omega + \delta(1 - F(C))\frac{4\Omega}{1-\delta}}{1 - \delta F(C)}, \\ &\Rightarrow 1 \geq 2, \end{aligned} \quad (\text{B.15})$$

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:

$$\frac{2\Omega + \delta\theta(1 - F(C))\frac{\Omega}{1-\delta}}{1 - \delta[1 - \theta(1 - F(C))]} \geq \frac{4\Omega + \delta(1 - F(C))\frac{\Omega}{1-\delta}}{1 - \delta F(C)}, \quad (\text{B.16})$$

$$\Rightarrow \theta \leq \frac{1}{3} \left(1 - \frac{2(1-\delta)}{\delta[1 - F(C)]} \right). \quad (\text{B.17})$$

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

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

$$E [D_{i,t} | R_t = 1] \geq \frac{1}{2(1-\theta)}. \quad (\text{B.18})$$

Since the RHS of (B.18) is greater than 1 when $\theta > \frac{1}{2}$, $S_t = 0$ whenever $\theta > \frac{1}{2}$. Moreover, from above, if $\theta \in \left(1 - \frac{1}{2\alpha}, \Psi(\delta, C) \right]$, then both types of agents choose $D_{i,t} = 1$. This entails that $E [D_{i,t} | R_t = 1] = 1$, and thus P chooses $S_t = 1$ if $\theta \leq \frac{1}{2}$, which must be true when $\theta \in \left(1 - \frac{1}{2\alpha}, \Psi(\delta, C) \right]$. Hence, P chooses $S_t = 1$ if $\theta \in \left(1 - \frac{1}{2\alpha}, \Psi(\delta, C) \right]$.

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

$p = (1 - \theta [1 - F(C)])^t$, while the probability of a strategic agent having a good reputation in period t is $[F(C)]^t$. Hence,

$$pr(A \text{ is altruist} \mid R = 1) = \frac{\alpha (1 - \theta [1 - F(C)])^t}{\alpha (1 - \theta [1 - F(C)])^t + (1 - \alpha) [F(C)]^t}, \quad (\text{B.19})$$

and P chooses $S_t = 1$ iff,

$$(1 - 2\theta) (1 - \theta [1 - F(C)])^t \geq \left(\frac{1 - \alpha}{\alpha} \right) [F(C)]^t, \quad (\text{B.20})$$

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

Finally, consider $\theta \in (\Psi(\delta, C), \frac{1}{2})$. In this range strategic A 's choose $D_{i,t} = 0$, meaning that P chooses $S_t = 1$ if (B.20) holds. It is straightforward to see that the LHS of (B.20) is decreasing in θ (as long as $\theta < \frac{1}{2}$). Denote $\bar{\theta}$ as the value of θ that sets (B.20) to equality. If $\bar{\theta} > \Psi(\delta, C)$, then $S_t = 1$ when $\theta \in [\Psi(\delta, C), \bar{\theta}]$. Combining this with the fact that $S_t = 1$ if $\theta \leq \Psi(\delta, C)$, we have the more general formulation that $S_t = 1$ if $\theta \leq \theta^*$, where $\theta^* = \max\{\bar{\theta}, \Psi(\delta, C)\}$.

■

Next, we analyze comparative statics with respect to θ for the strategic A 's decision to divide in the second stage. 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.

Finally, consider comparative statics with respect to θ 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 . This logic is formalized as follows.

Prediction 2: For periods in which $R = 1$, the size of the parameter space over which P chooses $S = 1$ is decreasing in θ , ceteris paribus.

Proof.

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.

■

C Regression Analyses of the Entire Sample

In this Appendix, 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 for various levels uncertainty. In the second specification, we include a reputation dummy, since Figure 4 suggests this may affect decision-making. In the third specification, we include controls for risk-aversion (obtained from the incentivized risk elicitation task), gender, and path dependence (period dummy).

In Table C.1, the omitted category is “weak nature” (i.e., $\theta = 0.05$). A few observations immediately jump out of Table C.1. 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 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 $\theta = 0$, $\theta = 0.10$, or $\theta = 0.30$ treatments ($p = 0.00$, 0.00 , and 0.00 , respectively). In other words, when informing is costly, a modest amount of uncertainty can *increase* the amount sent.

Result C1: 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 $\theta = 0$ or $\theta = 0.10$.

The results presented in columns 4-6 of Table C.1 suggest that agents are less likely to divide as θ increases, although some of the estimates are noisy. Meanwhile, the informing cost is negatively and significantly correlated to the frequency divide when there is no nature ($\theta = 0$) or moderate nature ($\theta = 0.10$). The results related to the agent’s divide decision are summarized in Result C2.

Result C2: The frequency that agents divide is noisily decreasing in θ , and it is decreasing in C when $\theta = 0$ or $\theta = 0.10$.

Table C.1: Regression Analysis of the Decision to Send and Divide

Dependent Variable:	Send (0/1)			Divide (0/1)		
	(1)	(2)	(3)	(4)	(5)	(6)
No Nature	0.042	-0.039	-0.002	0.122***	0.096***	0.107**
$\theta = 0.00$	(0.046)	(0.031)	(0.051)	(0.031)	(0.030)	(0.050)
Moderate Nature	-0.181***	-0.136***	-0.091*	-0.147***	-0.163***	-0.157***
$\theta = 0.10$	(0.031)	(0.035)	(0.052)	(0.031)	(0.035)	(0.056)
Strong Nature	-0.244***	-0.157***	-0.116	-0.148***	-0.163***	-0.134
$\theta = 0.30$	(0.055)	(0.046)	(0.083)	(0.043)	(0.049)	(0.084)
Costly Informing	-0.018	-0.098***	-0.037	-0.055**	-0.088***	-0.066
	(0.034)	(0.028)	(0.049)	(0.026)	(0.026)	(0.046)
No Nature \times Costly Informing			-0.074			-0.022
			(0.058)			(0.059)
Moderate Nature \times Costly Informing			-0.088			-0.012
			(0.064)			(0.065)
Strong Nature \times Costly Informing			-0.083			-0.059
			(0.090)			(0.097)
Good Reputation		0.508***	0.508***		0.196***	0.198***
		(0.036)	(0.037)		(0.035)	(0.035)
Risk		0.010	0.010		-0.003	-0.003
		(0.007)	(0.007)		(0.009)	(0.009)
Female		0.006	0.004		0.089**	0.091**
		(0.027)	(0.027)		(0.038)	(0.038)
Period		-0.012***	-0.012***		-0.019***	-0.019***
		(0.004)	(0.004)		(0.003)	(0.003)
Intercept	0.678***	0.458***	0.429***	0.698***	0.650***	0.636***
	(0.034)	(0.068)	(0.073)	(0.029)	(0.068)	(0.074)
Observations	3456	3456	3456	1981	1981	1981
R^2	0.23	0.28	0.28	0.10	0.23	0.23

Standard errors clustered by session in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Moderate Nature ($\theta = 0.05$) and Costless Informing ($C = 0$) are the omitted categories.

D Additional Tables and Figures

Figure D.1: Relative Frequency of Send by Treatment and Period

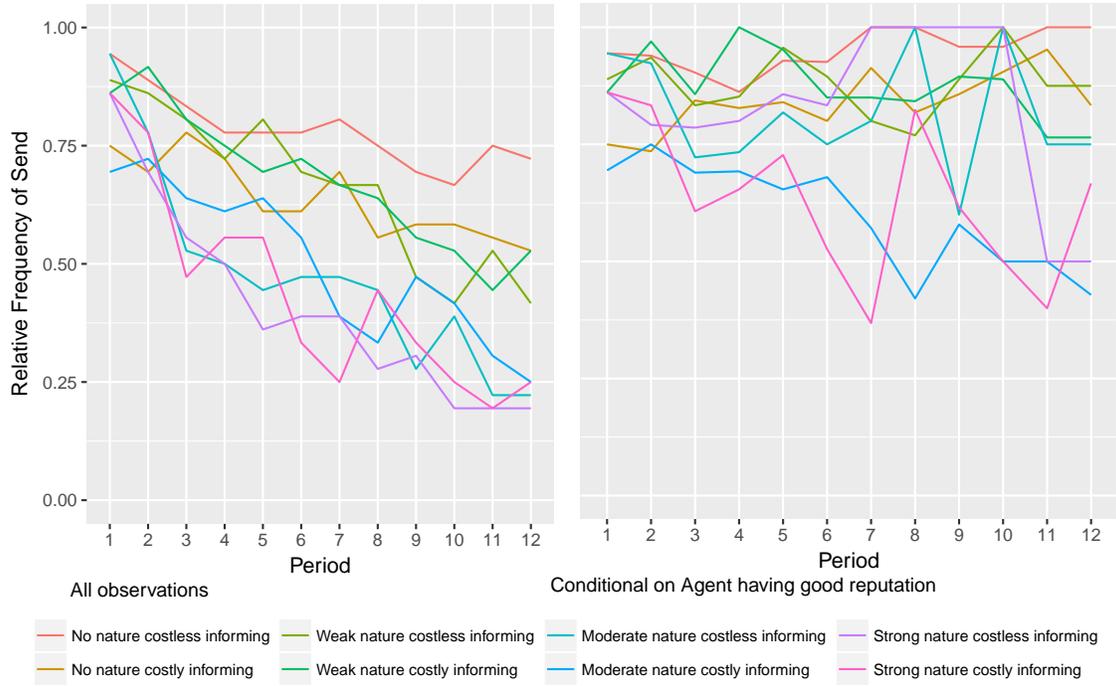


Figure D.2: Relative Frequency of Send Conditional on Bad Reputation by Treatment and Period

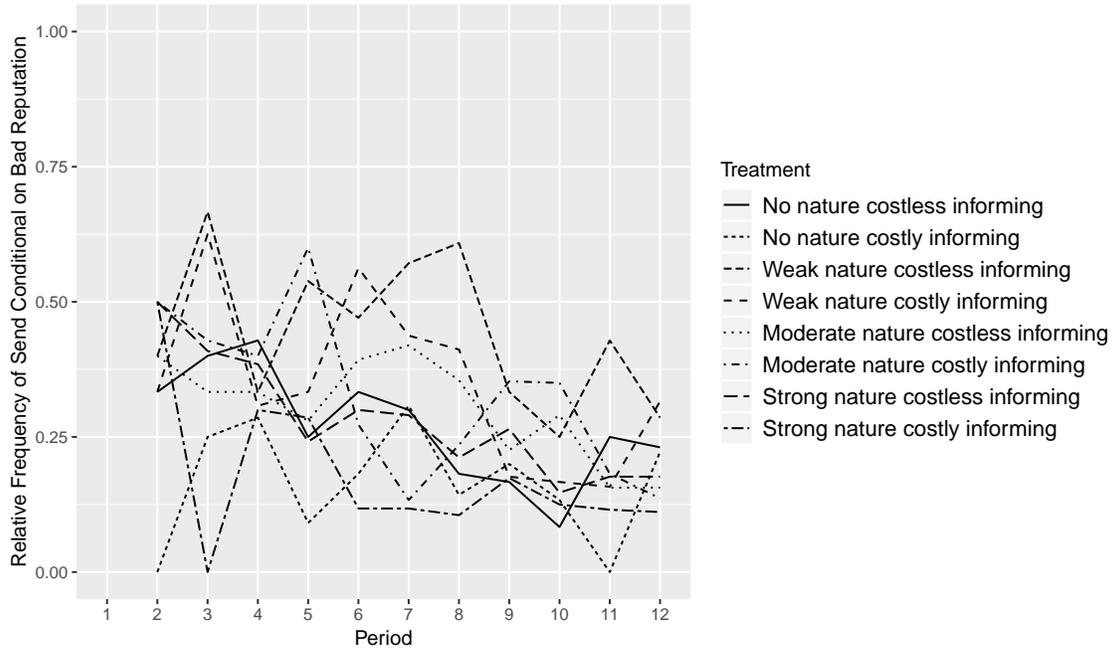


Table D.1: Summary Statistics: Send, Divide, and Inform

Treatments	Send			Divide			Inform		
	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
Moderate nature, costless informing	0.47	0.02	432	0.58	0.03	205	0.82	0.06	39
Moderate nature, costly informing	0.50	0.02	432	0.52	0.03	217	0.28	0.05	82
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

Note: In the Inform column, we condition Informing on the agent having a good reputation.

Table D.2: Send and Divide Conditional on Bad Reputation

Treatments	Send bad reputation			Divide bad reputation		
	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
Moderate nature, costless informing	0.29	0.03	290	0.35	0.05	85
Moderate nature, costly informing	0.28	0.04	155	0.25	0.07	44
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 D.3: Send and Divide Conditional on Good Reputation

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
Moderate nature, costless informing	0.85	0.03	142	0.74	0.04	120
Moderate nature, costly informing	0.62	0.03	277	0.59	0.04	173
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 D.4: Summary Statistics: Francs Earned per Period by Role and Treatment

Treatments	Principal			Agent			Total		
	Mean	SE	N	Mean	SE	N	Mean	SE	N
No nature, costless informing	16.15	0.22	36	19.49	0.46	36	17.82	0.32	72
No nature, costly informing	13.65	0.42	36	18.84	0.53	36	16.25	0.46	72
Weak nature, costless informing	12.45	0.35	36	20.13	0.48	36	16.30	0.54	72
Weak nature, costly informing	12.45	0.40	36	20.50	0.61	36	16.48	0.60	72
Moderate nature, costless informing	10.25	0.25	36	18.73	0.64	36	14.49	0.61	72
Moderate nature, costly informing	9.40	0.29	36	19.84	0.68	36	14.62	0.72	72
Strong nature, costless informing	9.18	0.25	36	17.43	0.60	36	13.31	0.59	72
Strong nature, costly informing	8.72	0.27	36	18.28	0.70	36	13.51	0.68	72

Table D.5: Regression Analysis of the Decision to Send and Divide, Bootstrapped Standard Errors

Dependent Variable:	Send (0/1)			Divide (0/1)		
	(1)	(2)	(3)	(4)	(5)	(6)
No Nature	0.042	-0.039	-0.002	0.122**	0.096**	0.107*
$\theta = 0.00$	(0.036)	(0.035)	(0.043)	(0.055)	(0.048)	(0.055)
Moderate Nature	-0.181***	-0.136***	-0.091*	-0.147**	-0.163***	-0.157**
$\theta = 0.10$	(0.035)	(0.036)	(0.054)	(0.060)	(0.059)	(0.075)
Strong Nature	-0.244***	-0.157***	-0.116*	-0.148***	-0.163***	-0.134**
$\theta = 0.30$	(0.033)	(0.035)	(0.068)	(0.056)	(0.058)	(0.064)
Costly Informing	-0.018	-0.098***	-0.037	-0.055	-0.088**	-0.066
	(0.026)	(0.023)	(0.050)	(0.043)	(0.043)	(0.073)
No Nature \times Costly Informing			-0.074			-0.022
			(0.067)			(0.090)
Moderate Nature \times Costly Informing			-0.088			-0.012
			(0.074)			(0.122)
Strong Nature \times Costly Informing			-0.083			-0.059
			(0.080)			(0.104)
Good Reputation		0.508***	0.508***		0.196***	0.198***
		(0.021)	(0.025)		(0.039)	(0.044)
Risk		0.010*	0.010*		-0.003	-0.003
		(0.006)	(0.006)		(0.009)	(0.010)
Female		0.006	0.004		0.089**	0.091**
		(0.026)	(0.026)		(0.040)	(0.036)
Period		-0.012***	-0.012***		-0.019***	-0.019***
		(0.002)	(0.003)		(0.003)	(0.003)
Intercept	0.678***	0.458***	0.429***	0.698***	0.650***	0.636***
	(0.029)	(0.049)	(0.064)	(0.042)	(0.082)	(0.071)
Observations	3456	3456	3456	1981	1981	1981
R^2	0.23	0.28	0.28	0.10	0.23	0.23

Table D.6: Regression Analysis of the Effect of Bad Reputation on the Decision to Send and Divide, Bootstrapped Standard Errors

Dependent Variable:	Send (0/1)			Divide (0/1)		
	(1)	(2)	(3)	(4)	(5)	(6)
No Nature	-0.143***	-0.145**	-0.141*	0.099	0.117	0.032
$\theta = 0.00$	(0.046)	(0.057)	(0.085)	(0.116)	(0.107)	(0.166)
Moderate Nature	-0.076*	-0.086	-0.139***	-0.017	-0.058	-0.105
$\theta = 0.10$	(0.044)	(0.057)	(0.054)	(0.081)	(0.085)	(0.102)
Strong Nature	-0.153***	-0.165***	-0.169***	-0.001	-0.022	-0.023
$\theta = 0.30$	(0.042)	(0.050)	(0.062)	(0.095)	(0.090)	(0.123)
Costly Informing	-0.067*	-0.062*	-0.093	-0.132**	-0.139**	-0.197
	(0.036)	(0.034)	(0.071)	(0.064)	(0.065)	(0.123)
No Nature \times Costly Informing			-0.007			0.177
			(0.106)			(0.194)
Moderate Nature \times Costly Informing			0.116*			0.124
			(0.068)			(0.143)
Strong Nature \times Costly Informing			0.007			-0.009
			(0.098)			(0.159)
Risk		0.000	0.001		-0.001	-0.001
		(0.008)	(0.008)		(0.013)	(0.013)
Female		0.031	0.035		0.227***	0.232***
		(0.040)	(0.040)		(0.063)	(0.057)
Period		-0.026***	-0.026***		-0.021***	-0.021***
		(0.003)	(0.004)		(0.007)	(0.007)
Intercept	0.397***	0.594***	0.604***	0.381***	0.435***	0.454***
	(0.040)	(0.068)	(0.075)	(0.065)	(0.102)	(0.131)
Observations	1548	1548	1548	426	426	426
R^2	0.04	0.03	0.04	0.04	0.10	0.10

Standard errors clustered by session in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Moderate Nature ($\theta = 0.05$) and Costless Informing ($C = 0$) are the omitted categories.

Table D.7: Regression Analysis of the Effect of Good Reputation on the Decision to Send and Divide, Bootstrapped Standard Errors

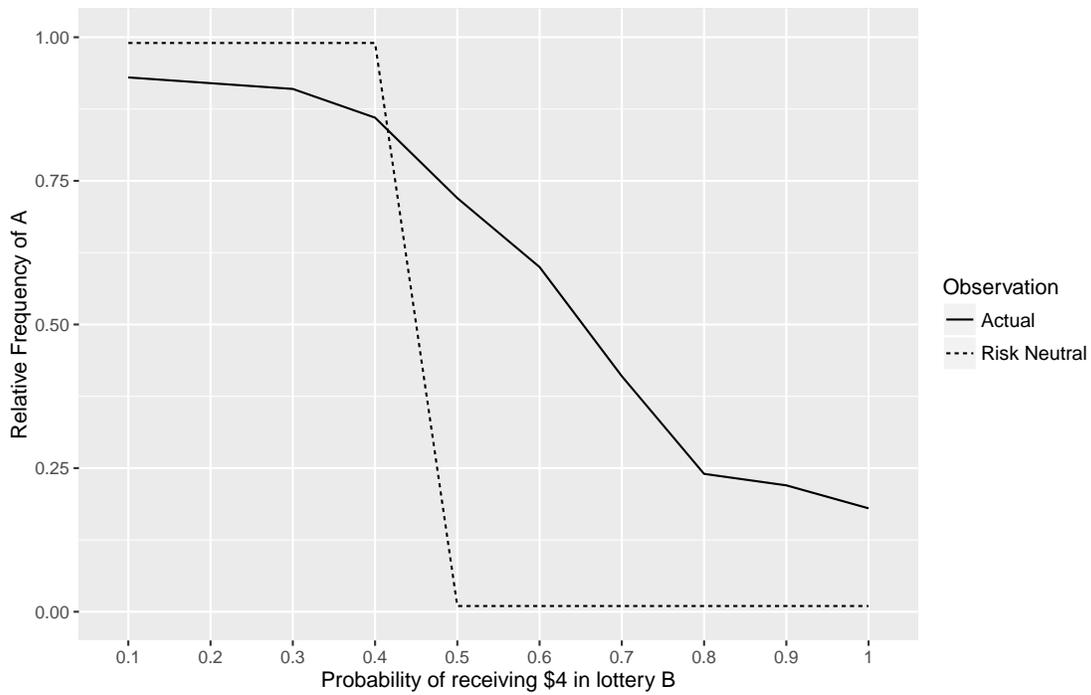
Dependent Variable:	Send (0/1)			Divide (0/1)		
	(1)	(2)	(3)	(4)	(5)	(6)
No Nature	0.006	0.015	0.073	0.045	0.067	0.083
$\theta = 0.00$	(0.037)	(0.035)	(0.054)	(0.043)	(0.053)	(0.068)
Moderate Nature	-0.150***	-0.155***	-0.039	-0.192***	-0.217***	-0.228***
$\theta = 0.10$	(0.047)	(0.037)	(0.057)	(0.048)	(0.047)	(0.079)
Strong Nature	-0.138***	-0.139***	-0.058	-0.150***	-0.181***	-0.161**
$\theta = 0.30$	(0.045)	(0.047)	(0.066)	(0.056)	(0.062)	(0.078)
Costly Informing	-0.119***	-0.107***	0.011	-0.106**	-0.092**	-0.079
	(0.024)	(0.030)	(0.063)	(0.044)	(0.042)	(0.073)
No Nature \times Costly Informing			-0.115			-0.032
			(0.071)			(0.108)
Moderate Nature \times Costly Informing			-0.217***			0.018
			(0.073)			(0.107)
Strong Nature \times Costly Informing			-0.154			-0.039
			(0.097)			(0.116)
Risk		0.017**	0.016**		-0.003	-0.003
		(0.007)	(0.007)		(0.010)	(0.010)
Female		-0.001	-0.007		0.057	0.058
		(0.031)	(0.028)		(0.048)	(0.040)
Period		-0.007**	-0.007**		-0.020***	-0.020***
		(0.003)	(0.003)		(0.002)	(0.003)
Intercept	0.947***	0.906***	0.851***	0.814***	0.885***	0.879***
	(0.033)	(0.055)	(0.065)	(0.043)	(0.064)	(0.065)
Observations	1908	1908	1908	1555	1555	1555
R^2	0.12	0.14	0.16	0.01	0.01	0.01

Standard errors clustered by session in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Moderate Nature ($\theta = 0.05$) and Costless Informing ($C = 0$) are the omitted categories.

E Risk Preference Elicitation Results

Following Holt and Laury (2002), we use a sequence of 10 paired lotteries to elicit risk preferences. According to the expected utility theorem, risk-neutral subjects should choose the lottery A the first 4 times in the sequence and then switch to the lottery B. Figure E.1 depicts the results of the risk preference elicitation task. The vertical axis represents the relative frequency of the choice of lottery A and the horizontal axis represents the probability of receiving the highest payoff of lottery B (i.e., \$4). In the same figure, the dashed line represents the expected utility theorem prediction and the solid line represents the actual choice of subjects.

Figure E.1: The Distribution of Choices in the Risk Preference Elicitation Task



F Demographic Analysis

In the questionnaire, we collected information regarding subjects' gender and major. We define an EBF major dummy which takes a value of 1 when subjects' major is Economics, Business, or Finance and 0 otherwise. Table F.1 reports the mean, standard error, and number of observations of female and EBF major across treatments.

Table F.1: Gender and EBF Major, Summary Statistics

Treatments	Female			EBF Major		
	Mean	SE	N	Mean	SE	N
No nature, costless informing	0.53	0.06	72	0.35	0.06	72
No nature, costly informing	0.54	0.06	72	0.33	0.06	72
Weak nature, costless informing	0.47	0.06	72	0.29	0.05	72
Weak nature, costly informing	0.54	0.06	72	0.22	0.05	72
Moderate nature, costless informing	0.67	0.06	72	0.31	0.05	72
Moderate nature, costly informing	0.60	0.06	72	0.29	0.05	72
Strong nature, costless informing	0.54	0.06	72	0.25	0.05	72
Strong nature, costly informing	0.68	0.06	72	0.31	0.05	72

On average, we expect to have equal number of male and female subjects per treatment. Using t-tests, we investigated whether mean number of females is significantly different from 0.5 across treatments. In two treatments, moderate nature costly informing ($\theta = 0.10$ and $C = 5$) and strong nature costly informing ($\theta = 0.30$ and $C = 5$), mean number of females is significantly more than 0.5. However, according to our regression analyses, the only significant correlation between being female and a variable of interest occurs in Table 2 in which female agents with the bad reputation tend to split more often than their male counterparts.

As for the effect of being Economics, Business, or Finance major, we reran all the regressions reported in this paper controlling for the EBF major. Our findings are robust with respect to the inclusion of the EBF major and in most of these regressions the EBF coefficient is practically 0 and insignificant. Only in the case of send conditional on good reputation, subjects who were majoring in Economics, Business, or Finance tend to send 8 percentage point less than non-EBF majors ($p - value = 0.05$). This might be due to a rudimentary backward induction analysis that yields Do Not Send as the optimal action of a principal regardless of an agent's reputation.

Further, we also ran a regression to see whether non-EBF major subjects sent less than EBF majors in period t when they inform and incur the cost of 5 francs in period $t - 1$. We find no significant correlation between EBF major and the decision to send in period t when a principal incurred the cost of informing in period $t - 1$.