University of Innsbruck

Truth, trust, and sanctions: On institutional selection in sender-receiver games

**Ronald Peeters, Marc Vorsatz,
Markus Walzl**

Working Papers in Economics and Statistics

2011-28

Research Platform
eeecon
Empirical and Experimental Economics

University of Innsbruck
http://eeecon.uibk.ac.at/

University of Innsbruck
Working Papers in Economics and Statistics

The series is jointly edited and published by

Department of Economics

Department of Public Finance

Department of Statistics

Contact Address:
University of Innsbruck
Department of Public Finance
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 7171
Fax: + 43 512 507 2970
e-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
http://eeecon.uibk.ac.at/wopec/

For a list of recent papers see the backpages of this paper.

Truth, trust, and sanctions: On institutional selection in sender-receiver games[*]

Ronald Peeters[†] Marc Vorsatz[‡] Markus Walzl[§]

October 27, 2011

Abstract

We conduct a laboratory experiment to investigate the impact of institutions and institutional choice on truth-telling and trust in sender-receiver games. We find that in an institution with sanctioning opportunities, receivers sanction predominantly after having trusted lies. Individuals who sanction are responsible for truth-telling beyond standard equilibrium predictions and are more likely to choose the sanctioning institution. Sanctioning and non-sanctioning institutions coexist if their choice is endogenous and the former shows a higher level of truth-telling but lower material payoffs. It is shown that our experimental findings are consistent with the equilibrium analysis of a logit agent quantal response equilibrium with two distinct groups of individuals: one consisting of subjects who perceive non-monetary lying costs as senders and non-monetary costs when being lied to as receivers and one consisting of payoff maximizers.

JEL Classification: A13, C72, Z13.

Keywords: Experiment; Sender-receiver games; Strategic information transmission; Institutional selection.

[*] We are grateful to Rene Saran, the editor, and two referees for very helpful remarks and suggestions. Financial support by Meteor, the Dutch Science Foundation and the Spanish Ministry of Education and Science (project ECO2009-07530 and the Ramón y Cajal program) is gratefully acknowledged.

[†] Corresponding author. Department of Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: r.peeters@maastrichtuniversity.nl.

[‡] Fundación de Estudios de Economía Applicada (FEDEA), Calle Jorge Juan 46, 28001 Madrid, Spain. Email: mvorsatz@fedea.es.

[§] Department of Economics, Innsbruck University, Universitaetsstr. 15, 6020 Innsbruck, Austria. Email: markus.walzl@uibk.ac.at.

1 Introduction

The strategic transmission of information is ubiquitous in economic interactions. As long as information is asymmetrically distributed among market participants, there is an incentive to strategically hide or release private information. For instance, a seller of a financial asset who is privately informed about future price changes may wish to talk a potential buyer into an early purchase if he expects the market price to fall while trying to postpone the transaction if he expects the price to rise (see e.g. Wang et al., 2010). Likewise, a seller of a commodity who has private information about its quality (as in the lemons-market in Akerlof, 1970) may wish to communicate a good quality (e.g. by offering a warranty) while seeking to conceal any information that reduces the potential buyer's willingness to pay.

A framework to study the strategic transmission of payoff relevant information is *sender-receiver games*. In their seminal contribution on information transmission between payoff maximizing individuals, Crawford and Sobel (1982) demonstrate that the more the preferences of the informed (the sender) and the uninformed agent (the receiver) are aligned, the more information is transmitted in sequential equilibrium. Evidence in favor of this main finding has been provided by Dickhaut et al. (1995) via laboratory experiments. Experiments by Cai and Wang (2006), however, show that senders over-communicate relative to standard equilibrium predictions; that is, on average, individuals reveal more private information than predicted by Crawford and Sobel (1982). Regarding the willingness to enforce truth-telling, Sánchez-Pagés and Vorsatz (2007) identify receivers who voluntary incur a cost in order to punish the sender after having trusted a lie and find that these subjects are responsible for the observed over-communication.[1] The objective of our study is to investigate how far this over-communication is driven by intrinsic motivations for truth-telling and how institutional selection (between a sanctioning and non-sanctioning institution) affects the composition of individuals and the performance of each institution.[2]

[1] Throughout the paper we will label a report by the sender that is equal to her private information as *truth* and a choice by the receiver that resembles a best response if the sender's message is truthful as *trust*. Hence, 'trust' labels the choice of a receiver who is categorized as a 'believer' in Crawford (2003). The labels *lie* and *distrust* are defined accordingly. Based on these labels, we will use *excessive truth-telling* as a synonym for information revelation beyond the predictions by Crawford and Sobel (1982).

[2] Already Arrow (1968, p.538) noted that "[o]ne of the character[i]stics of a successful economic system is that the relations of trust and confidence between principal and agent are sufficiently strong so that the agent will not cheat even though it may be "rational economic behavior" to do so." We are indebted to one of the referees for this quote.

In our theoretical analysis, we develop a logit agent quantal response equilibrium (logit-AQRE) model with individuals who experience non-monetary costs of lying that is able to account for the existing experimental evidence. In particular, the model predicts (i) the frequency of truthful revelations of private information to increase in the non-monetary costs from misreporting private information (*costs of lying*), (ii) the frequency of sanctioning after trusting a lie to increase in the non-monetary costs from being exposed to lies (*costs of being lied to*), and (iii) the expected utility to be independent of the expected truth-telling frequency if the individual does not perceive non-monetary costs – otherwise it is increasing in the truth-telling frequency (*anticipation of non-monetary costs*).

Regarding a deeper understanding of the motives behind truth-telling, the enforcement of truth-telling through costly sanctions, and the anticipation of these effects in institutional choice, two research questions emerge from these theoretical predictions for the experimental part of our study. (i) Is there a correlation between the two main ingredients of the model (costs of lying and costs of being lied to), i.e., do individuals who sanction lies also tell the truth excessively while others do not? (ii) Do individuals with different attitudes towards sanctioning also differ in their choices for an institutional environment, i.e. is there a self-selection of individuals with distinct costs of lying and being lied to into different institutional environments?

To analyze these questions, we conduct an experiment that consists of two institutions and two phases. The *sanction-free institution* corresponds to a simple constant-sum sender-receiver game with antagonistic payoffs that, in its reduced form, resembles matching pennies. The *sanctioning institution* extends the sanction-free institution by giving the receiver the option – after observing whether the sender told the truth or lied about her private information – to reduce the payoffs of both players to zero. In the first 60 rounds of the experiment (*random assignment phase*), we randomly assign subjects to the two institutions. This matching procedure allows for a within subjects analysis of truth-telling in the presence *and* absence of sanctioning opportunities, answering question (i). For an investigation of question (ii), we add a second phase of another 40 rounds (*selection phase*) where individuals can choose in each round which institution to join.

In the experiment, we observe excessive truth-telling and trust compared to the standard equilibrium predictions and identify *sanctioners* as individuals who predominantly sanction after having trusted a lie. We find that sanctioners are responsible for the excessive truth-telling in *both* institutions and phases, which implies that the two non-monetary costs are correlated. With respect to institutional choice in the selection phase, we observe that sanctioners choose the sanctioning institution as often as the sanction-free institution while the vast majority of the remaining subjects opts for the sanction-free institution. Hence, the two institutions typically coexist throughout the selection phase.[3] Since the sanctioning institution exhibits more truth-telling than the sanction-free institution and since we also observe sanctions throughout the selection phase, we can conclude that there are individuals (predominantly sanctioners) who deliberately choose an institution with lower material payoffs but a higher level of truth-telling.

We proceed as follows. In Section 2, we analyze the impact of non-monetary lying costs on truth-telling, trust, and sanctioning in a logit-AQRE. We present the experimental design and procedures in Section 3. Testable hypotheses are derived in Section 4. The experimental results are presented in Section 5 and the logit-AQRE estimates in Section 6. In Section 7, we relate our findings to the existing literature, and discuss possible implications. Formal proofs, the sensitivity analysis, and the experimental instructions are relegated to the Appendices.

2 Theoretical analysis

We consider the sender-receiver game depicted in Figure 1. There are two players: the *sender* and the *receiver*. The sender is either of type A or type B. The actual type is drawn by nature and the realization is only known by the sender. The players are informed that both types are equally likely. The receiver decides whether to take action A or action B. In case the action matches with the sender's type, the receiver gets a payoff of 5 ECU and leaves the sender with a payoff of 1 ECU.[4] Payoffs are reversed in case the

action A	action B
1 ; 5	5 ; 1

type A

action A	action B
5 ; 1	1 ; 5

type B

Figure 1: Sender-receiver game.

[3] This is anything but trivial: sanctions are necessarily inefficient as they only destroy payoff and therefore, it is always optimal for payoff maximizers to choose the sanction-free institution.

[4] ECU stands for Experimental Currency Unit, the currency used in the experiment.

action does not match with the sender's type.

Before the receiver chooses his action, but after the sender has learnt her type, the sender transmits one of the following two messages to the receiver: message A ("the type selected by nature is A") or message B ("the type selected by nature is B"). For simplicity, we say throughout that the sender tells the *truth* if her message is equal to her true type, otherwise we say she tells a *lie*. Similarly, we say that the receiver *trusts* if his action resembles a best response to the reported type of the sender, otherwise we say he *distrusts* the message. Hence, the combinations truth–trust and lie–distrust lead to a payoff of 5 ECU to the receiver and only 1 ECU to the sender and the combinations truth–distrust and lie–trust lead to the reversed payoffs.[5]

We consider this game in two different institutional settings: the *sanction-free institution* (SFI) and the *sanctioning institution* (SI). In the sanction-free institution, the sender-receiver game depicted in Figure 1 is played. In the sanctioning institution, the receiver has additionally the option to *sanction* after the game in Figure 1 has been played and he learned the real type of the sender. If the receiver sanctions, the payoffs of both players are reduced to zero, otherwise the payoffs remain unchanged.

One can easily show that under the standard assumption that individuals are selfish profit maximizers and fully rational, receivers never sanction and all sequential equilibria of the game in the sanction-free and the sanctioning institution are such that the sender tells the truth with probability one-half and the receiver trusts with probability one-half (see Crawford and Sobel, 1982). Hence, no information is transmitted as the receiver's prior belief about the true type is not affected by the sender's message.

One problem of deriving null hypotheses for the experiment on the basis of these standard assumptions is that the best response correspondences are discontinuous: if the sender tells the truth with more than fifty percent chance, the receiver should always trust; and, if the receiver trusts with more than fifty percent chance, the sender should always lie. Goeree and Holt (2001) demonstrate in experiments on symmetric and asymmetric games of matching pennies that such an extreme response is unlikely to be observed empirically. A better description of behavior is provided by probabilistic choice models such as the logit agent quantal response equilibrium (logit-AQRE) introduced by McKelvey and Palfrey

[5] Due to the symmetry of the game, we can abstain from conditioning strategies on the actually chosen table. In reduced form, the game resembles a two-by-two constant-sum game like matching pennies.

(1998). Unlike in the standard best response correspondence, in the logit-AQRE model strategies that yield lower payoffs are played with lower but positive probability. Applying the logit-AQRE to the game in Figure 1, one obtains that the sender tells the truth with probability

$$p = \frac{e^{\lambda E[u(\text{truth})]}}{e^{\lambda E[u(\text{truth})]} + e^{\lambda E[u(\text{lie})]}},$$

where $E[u(\text{truth})]$ and $E[u(\text{lie})]$ denote the expected utilities for the sender from telling the truth and lying, respectively. Similarly, the receiver trusts with probability

$$q = \frac{e^{\lambda E[v(\text{trust})]}}{e^{\lambda E[v(\text{trust})]} + e^{\lambda E[v(\text{distrust})]}},$$

where $E[v(\text{trust})]$ and $E[v(\text{distrust})]$ denote the expected utilities for the receiver from trusting and distrusting, respectively. The parameter $\lambda \in [0, \infty)$ captures the level of "rationality" of the agent. If $\lambda = 0$, individuals act totally at random and each action is played with equal chance. As λ increases, individuals get more and more rational, and in the limit –as λ converges to infinity– individuals become fully rational and play a best response. The logit-AQRE is thus a natural generalization of sequential equilibrium incorporating the possibility of boundedly rational behavior.

The equations for p and q hold for both the sanction-free and the sanctioning institutions. Expected utilities, however, may vary across the two institutions. This is because for any message $m \in \{\text{truth}, \text{lie}\}$ of the sender and action $a \in \{\text{trust}, \text{distrust}\}$ of the receiver, the receiver reduces the payoffs of both players with probability

$$r_{m,a} = \frac{e^{\lambda v(m,a,1)}}{e^{\lambda v(m,a,1)} + e^{\lambda v(m,a,0)}}$$

in the sanctioning institution. Here, $v(m, a, 0)$ denotes the utility of the receiver if the sender reports m, the receiver plays a after observing m, and, finally, the receiver decides not to sanction the sender after learning the history of the game (sanctioning $s \in \{0, 1\}$ takes a value of zero). So, since the terminal utilities in the sanctioning institution are affected by the occasional punishment by the receiver, expected utilities (and thereby the equilibrium probabilities p^* and q^*) may differ across institutions.

Sánchez-Pagés and Vorsatz (2007) have shown that the logit-AQRE is unable to explain the experimental data of constant-sum sender-receiver games when there are only two states.[6] In particular, the model is unable to incorporate the following two experimental findings. (i) According to the logit-AQRE, the sanctioning probability "only" depends on the terminal utilities. Since the histories truth–distrust and lie–trust lead to the same payoff for both players, the sanctioning rate should hence be the same for both histories. Yet, experiments have shown that the sanctioning rate is significantly greater when the receiver has trusted a lie than when the receiver has distrusted the truth.[7] (ii) The logit-AQRE predicts perfectly randomized truth-telling and trust for all λ, yet experimental data establishes that the sender tells the truth and the receiver trusts in more than half of the cases in both institutions.

To account for the existing experimental evidence, we analyze in this sequel a logit-AQRE with players who experience non-monetary costs of lying and being lied to. Although it is not possible to derive a closed form solution, equilibrium comparative statics can be used to derive testable hypotheses. Let us assume that the utility function u of the sender is $u(m, a, s) = \pi_S(m, a, s) - c \cdot \mathbf{I}(m = \text{lie})$ with the sender's message $m \in \{\text{truth}, \text{lie}\}$, the receiver's action $a \in \{\text{trust}, \text{distrust}\}$ and the receiver's sanctioning decision $s \in \{0, 1\}$. If sanctions are not available, we set $s = 0$. The term $\pi_S(m, a, s)$ corresponds to the monetary payoff for the sender and $\mathbf{I}(m = \text{lie})$ is an indicator function that takes the value 1 if the sender lies and 0 otherwise. Hence, $c \geq 0$ measures the lying costs of the sender (see Kartik, 2009; Hurkens and Kartik, 2009; Sánchez-Pagés and Vorsatz, 2009). Similarly, the utility function v of the receiver is assumed to be $v(m, a, s) = \pi_R(m, a, s) - d \cdot \mathbf{I}(m = \text{lie}, s = 0)$ with the receiver's monetary payoff $\pi_R(m, a, s)$ and indicator function $\mathbf{I}(m = \text{lie}, s = 0)$. The indicator function takes the value 1 if the sender lies to the receiver and the lie remains unsanctioned. The parameter $d \geq 0$ measures the costs of being lied to.[8]

[6] See, Cai and Wang (2006) for an application of the logit-AQRE with payoff maximizing players when there are more than two states.

[7] Further evidence is provided by Xiao (2010) who allows for third party punishments in a sender-receiver game and finds that most third parties (19 out of 27) punish only false messages.

[8] $v(m, a, s)$ can be interpreted as the utility function of a (negatively) reciprocal individual, who considers lies as unkind but reduces (or, in our specification, nullifies) his suffering from being exposed to an unkind act through retaliation. For a more detailed discussion of the relation between this model and other models of non-standard preferences see Subsection 7.2. Note that our specification of the receiver's utility highlights the interplay between the costs of being lied to and the payoff consequences of a sanction. For instance, receivers may well sanction after lie–trust but not after lie–distrust as already observed in Sánchez-Pagés and Vorsatz (2007). Furthermore, our model predicts receivers to sanction more after truth–distrust (without lie, but lower payoff) than after lie–distrust (with lie, but with higher payoff) if and only if the cost of being exposed to a lie d is less than the payoff difference between these two histories.

Proposition 1. (Sanction-free Institution)
In the unique logit-AQRE of the sanction-free institution, (i) p^* and q^* are independent of d, (ii) $c = 0$ or $\lambda = 0$ implies that $p^* = q^* = \frac{1}{2}$, and (iii) if $\lambda > 0$, both p^* and q^* are strictly increasing in c.

Proof. See Appendix A.

Proposition 1 shows that in the sanction-free institution, the equilibrium probabilities are independent of d and strictly increasing in c whenever $\lambda > 0$. Next, we turn our attention to the sanctioning institution. In particular, we obtain that if $c = d = 0$, the sender tells the truth with probability one-half and the receiver trusts with probability one-half in this institution as well. A strictly positive d, however, induces the receiver to sanction more often after history lie–trust than after truth–distrust and the equilibrium probabilities of truth and trust are again strictly increasing in c.

Proposition 2. (Sanctioning Institution)
In the unique logit-AQRE of the sanctioning institution, (i) $c = d = 0$ or $\lambda = 0$ imply that $p^* = q^* = \frac{1}{2}$, $r^*_{truth,trust} = r^*_{lie,distrust}$, and $r^*_{lie,trust} = r^*_{truth,distrust}$, (ii) $r^*_{lie,trust} > r^*_{truth,distrust}$ if and only if $d > 0$ and $\lambda > 0$, and (iii) if $\lambda > 0$, both p^* and q^* are strictly increasing in c.

Proof. See Appendix A.

Propositions 1 and 2 provide the following insights: (a) in the absence of non-monetary costs, i.e. if $c = d = 0$, the sender tells the truth with probability one-half and the receiver trusts with probability one-half in both institutions and there is no difference between sanctioning rates after histories that lead to the same payoff distribution; (b) if $c = 0$ and $d > 0$, then the receiver sanctions more often after lie–trust than after truth–distrust and, at the same time, both players behave in the sanction-free institution as if they were payoff maximizers; (c) if $c > 0$ and $d = 0$, then there is no difference between sanctioning rates after histories that lead to the same payoff distribution and there is excessive truth-telling and excessive trust in the sanction-free institution; and, (d) if $c > 0$ and $d > 0$, then the receiver sanctions more often after lie–trust than after truth–distrust and there is excessive truth-telling and trust in the sanction-free institution.

3 Experimental design and procedures

The experiment was conducted with the help of the z-Tree toolbox (Fischbacher, 2007) in the experimental computer laboratory at Maastricht University. All students of the Faculty of Economics and Business Administration were invited via email to register for the experiment. In total, we had 8 sessions with 20 subjects per session. Subjects received written and context-free instructions (see Appendix C) that they could study at their own pace. Clarifying questions were dealt with privately. Before the experiment started, every subject had to answer some control questions correctly.

The experiment consists of two phases. The first phase is referred to as *random assignment phase* (RAP) and it lasts 60 rounds. In each round, the 20 subjects are randomly divided in such a way that 6 subjects are assigned to the sanction-free institution and 14 to the sanctioning institution.[9] Next, subjects within the same institution are randomly matched into pairs. Within each pair, one subject is randomly chosen to be the sender, the other subject is the receiver. After all subjects are informed about the institution they are assigned to and their role, the respective game is played.

The second phase of a session, the *selection phase* (SP), lasts 40 rounds. At the beginning of each round, subjects decide in which institution to play. After all subjects have made their decisions, subjects within the same institution are randomly matched into pairs. In case of an odd number of subjects in an institution, one randomly chosen subject in each institution stays unmatched and receives a fixed payoff of 3 ECU. In each pair, one subject is randomly chosen to be the sender, the other subject is the receiver. After all subjects are informed about their role, the respective game is played.[10]

After each round, subjects were informed about all decisions taken within the respective pair, the resulting payoffs, and the individual cumulative payoffs. Subjects were never given any feedback on the identity of the players they were matched to.[11] Subjects were paid privately in cash immediately after the experiment. The payoffs gathered throughout the

[9] We chose this imbalanced assignment in order to increase the probability that receivers have the opportunity to sanction after different histories.

[10] The particular sequence of the random assignment phase before the selection phase serves two goals. First, it is guaranteed that subjects acquire some experience with both institutions prior to any selection opportunities. Second, it facilitates a proper type identification based on sanctioning behavior.

[11] Observe that subjects do not have incentives to coordinate on a particular action as preferences are completely antagonistic and payoffs are constant-sum. Moreover, we preserve the anonymity of the matchings. Hence, supergame effects can be excluded.

session were transferred into euros at an exchange rate of 0.05; that is, one ECU was worth 5 Eurocents.[12] The average payment was € 16.59 (including 3 euros for showing up). A session lasted 105 minutes on average.

4 Hypotheses

In our experiment, subjects take decisions with respect to truth-telling, trust, sanctioning, and institutional choice. Decisions regarding truth-telling, trust, and institutional choice can be driven by the subject's preference *and* her expectation over actions of the other player. In contrast, the receiver decides upon sanctioning under complete information (observing all preceding actions and the sender's type) and at the final stage of the game. Hence, sanctioning decisions do not depend on (unobserved) beliefs over actions and are therefore a more direct expression of preferences. For example, individuals who sanction after having trusted a lie but not after having distrusted the truth reveal a preference for truth-telling (*i.e.*, costs of being lied to) as their willingness to costly reduce the payoff of the sender depends on the particular message being sent.

Our main methodological approach will therefore be to analyze whether the decisions towards truth-telling and institutional selection depend on the sanctioning behavior; in this way, we separate beliefs from preferences to the largest extent possible. At the same time, this approach will also allow us to address several important questions: Is it true that those who reveal a preference for truth-telling in the role of the receiver are responsible for the excessive truth-telling in the presence and absence of sanctioning opportunities found in earlier studies? If yes, we would be able to conclude that *preferences for truth-telling of the senders* (*i.e.*, lying aversion) are likely to play a crucial role why more information than predicted by the standard sequential equilibrium is contained in the messages. Also, is it true that somebody who reveals a preference for truth-telling as a sender (as identified in the former step) opts more often for the sanctioning institution than somebody without such preference? If yes, individuals self sort into different environments according to their preferences with possibly important consequences for the functioning of these institutions in terms of economic efficiency and information revelation.

[12] We decided to pay subjects according to their cumulative payoffs because a lottery (one round is randomly determined for payment) would provide subjects with a device to randomize over actions. Paying for the sum of the payoffs, on the other hand, implies that subjects have to randomize explicitly.

Given this structure, we have to identify first the subjects with non-standard preferences as receivers. This is done as follows. For each subject, the sanctioning decisions after the history lie–trust in the random assignment phase are assumed to be independent Bernoulli trials with success probability 0.5. Using the actual data, we can then test the null hypothesis that the success probability is smaller than or equal to 0.5 against the alternative hypothesis that the success probability is greater than 0.5. The degree of confidence with which this test is rejected is finally used to classify subjects. If the one-sided p-value of this test is smaller than or equal to 0.20, then a subject is assigned to the group of *sanctioners*. All subjects that are not classified as sanctioners, are classified as *others*.[13,14]

Our first hypothesis regards the sanctioning behavior of these groups.

Hypothesis 1. (Sanctioning)
Sanctioners punish significantly more often after lie–trust than after truth–distrust. The others sanction as often after truth–distrust as after lie–trust.

Under Hypothesis 1, we are able to conclude from Proposition 2 that only the sanctioners have a strictly positive d. Given our main objective of identifying subjects with different preferences towards truth-telling, we aim to show in the next step of our analysis that only the sanctioners suffer if they lie. Since it follows from Proposition 1 that senders with a strictly positive lying cost tell the truth excessively in the sanction-free institution, we have to establish that the sanctioners tell the truth in more than half of the cases in this institution while the others do so with probability one-half.[15] Hypothesis 2 is slightly more general as it requires that all excessive truth-telling found at the population level in both institutions and phases to be caused by the sanctioners.

Hypothesis 2. (Truth-telling)
Sanctioners tell the truth excessively, the others with probability one-half.

[13] A separate analysis of individuals who sanction after lie–trust *and* after truth–distrust (i.e., sanctioning contingent on the payoff distribution) and of individuals who only sanction after lie–trust but not after truth–distrust (i.e., sanctioning contingent on a specific history) is impossible due to the small number of observations of the history truth–distrust.

[14] It is shown in Appendix B that our results are robust for a wide range of p-values. To be more concrete, we consider eight different p-values between 0.05 and 0.40 plus an extreme classification in which everybody who punishes at least once is classified as a sanctioner. Observe that a smaller p-value implies that a subject is less likely to be classified as a sanctioner; thus, the group of sanctioners increases as the p-value is relaxed.

[15] If there is (anticipated) type heterogeneity, it is a best response for the others to lie excessively in such a way that no information is transmitted on the aggregate. This contradicts the existing experimental results on over-communication.

Under Hypotheses 1 and 2, the others behave on the aggregate as if they were payoff maximizers. However, it is not necessarily true that the sanctioners face a strictly positive lying cost; after all, they could simply believe that the receivers distrust very often in the sanction-free institution so that telling the truth frequently becomes a rational choice even in the absence of any intrinsic motives to do so. The decision which institution to join during the selection phase provides additional evidence whether subjects truly care about truth-telling. In particular, the expected utility of a subject who is equally likely to be the sender and the receiver is given by the expression

$$(1 - \underbrace{\textstyle\sum_{h \in H} \sigma_h \cdot r_h}_{\text{prob. of sanctions}}) \cdot 3 - \underbrace{(1-p) \cdot (\tfrac{c}{2} + \tfrac{d}{2}\left[q(1 - r_{\text{lie,trust}}) + (1-q)(1 - r_{\text{lie,distrust}})\right])}_{\text{expected loss due to lies}} .$$

In this equation, H denotes the set of all possible histories of the sanction-free institution, σ_h is the probability with which h is played given p and q, and r_h is the likelihood that the receiver sanctions after observing h (if applicable).

One sees that there is no reason for payoff maximizers to select the sanctioning institution beyond some degree of experimentation. As discussed in Section 2, there is some sanctioning for all histories in the sanctioning institution and therefore, the expected monetary payoff is necessarily lower in this institution. This is different for an individual with non-standard preferences. The expected utility is now increasing in p so that if the level of truth-telling is higher in the sanctioning than in the sanction-free institution *and* if the individual cares sufficiently about truth-telling, then the expected utility is higher in the sanctioning institution.[16] Our third hypothesis therefore states implicitly that the sanctioners anticipate higher levels of truth-telling in the sanctioning institution.

Hypothesis 3. (Institutional selection)
Sanctioners choose the sanctioning institution significantly more often than the others.

Under Hypotheses 1–3, it is possible to divide the experimental population into two subgroups: one group that cares sufficiently about truth-telling and anticipates a higher level of truth-telling in the sanctioning institution and another group that can be modeled as

[16] Observe that $\sum_{h \in H} \sigma_h \cdot r_h$ can be rewritten as $p \cdot q \cdot r_{\text{truth,trust}} + p \cdot (1-q) \cdot r_{\text{truth,distrust}} + (1-p) \cdot q \cdot r_{\text{lie,trust}} + (1-p) \cdot (1-q) \cdot r_{\text{truth,trust}}$. By Proposition 2, $r^*_{\text{lie,trust}} > r^*_{\text{truth,distrust}}$ whenever $d > 0$. Also, we know from former experiments that $r_{\text{truth,trust}} = r_{\text{lie,distrust}} \approx 0$ and that $q > 0.5$ in both institutions. Hence, the probability of sanctioning is decreasing in p for the parameter range we are interested in; in particular, the derivative with respect to p reduces to $(1-q) \cdot r_{\text{truth,distrust}} - q \cdot r_{\text{lie,trust}} < 0$.

payoff maximizers. To obtain this insight we did not have to take into account whether or not the receiver trusts. So, this decision is not directly related to intrinsic preferences towards truth-telling. Nevertheless, these preferences may matter indirectly if the receiver tends to best reply to the expected behavior of the sender. In particular, it follows from Propositions 1 and 2 that the excessive truth-telling for the sanctioners triggers excessive trust. Payoff maximizers are again expected to randomize perfectly.

Hypothesis 4. (Trust)
Sanctioners trust excessively, the others with probability one-half.

5 Results

This section is divided into four parts. We analyze first the sanctioning behavior throughout the experiment. In particular, we classify subjects by means of their sanctioning behavior in the random assignment phase as indicated in the former section. This allows us to study truth-telling, institutional selection, and trust separately for individuals with possibly different preferences regarding truth-telling.

In our statistical analysis, we proceed as follows. First, we calculate for each session the overall percentage of the variables of interest (truth-telling, trust, sanctioning for each history, and institutional selection). This results in eight truly independent observations (one per session). Next, we apply one-sided Wilcoxon signed-rank tests to these observations to evaluate our hypotheses.[17] For the tests on excessive truth-telling/trust the experimental data is paired with the logit-AQRE equilibrium prediction when players are payoff maximizers.

5.1 Sanctioning

Figure 2 illustrates the development of the sanctioning rates after the histories lie–trust and truth–distrust, the two histories when the receiver gets the low payoff, over rounds (clustered per 5 rounds). Sanctioning after truth–trust and lie–distrust, the two histories when the receiver gets the high payoff, only took place once for each history. Therefore, we ignore these histories from now on.

[17] As our hypotheses on truth-telling and trust for the others are not directional, tests should actually be two-sided. Yet, the p-values in Tables 4 and 7 show that the conclusions remain the same.

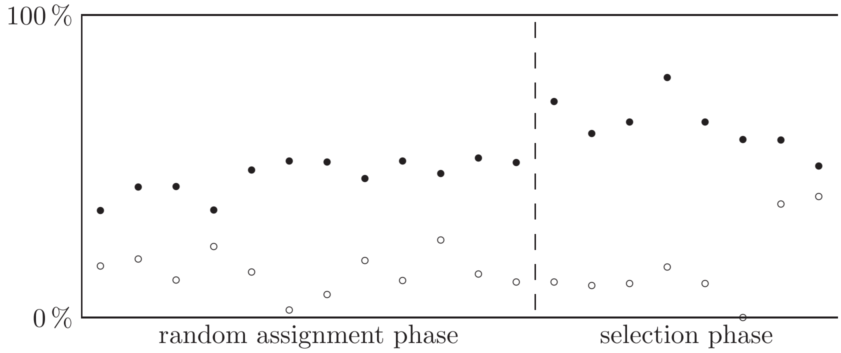

Figure 2: Sanctioning rates after truth–distrust (circles) and lie–trust (bullets) over rounds (5-round averages).

First of all, the occurrence of sanctioning is in sharp contrast with the prediction of the sequential equilibrium with payoff maximizing players. Moreover, Figure 2 suggests that there is more sanctioning after lie–trust than after truth–distrust contradicting the logit-AQRE with payoff maximizing players.[18] We also see that the transition to the selection phase increases sanctioning after lie–trust but not after truth–distrust. These observations are confirmed by the test results displayed in Table 1.

	lie–trust		truth–distrust
RAP	47 %	[0.0072]	15 %
	[0.0150]		[0.3363]
SP	64 %	[0.0173]	14 %

Table 1: Average sanctioning rates for the whole population. Between brackets, we display the p-values of the one-sided Wilcoxon signed-rank tests on the difference in sanctioning between the two histories and the two phases.

Result 1 (Sanctioning–I). *The sanctioning rate after lie–trust is higher than after truth–distrust. Institutional selection increases the sanctioning rate after lie–trust but not after truth–distrust.*

Table 2 summarizes the average sanctioning rates after the histories lie–trust and truth–distrust in the two phases for the two subgroups. Based on the procedure introduced in the former section, 53 out of the 160 participants are classified as sanctioners.

[18] The difference between the two trends narrows down at the end of the experiment. Note, however, that the sanctioning rate after the history truth–distrust over the last ten rounds is only based on 14 observations.

	lie–trust		truth–distrust
RAP	92 %	[0.0059]	20 %
	[0.3892]		[0.0262]
SP	91 %	[0.0137]	10 %

	lie–trust		truth–distrust
RAP	15 %	[0.3099]	14 %
	[0.0344]		[0.3371]
SP	31 %	[0.0865]	17 %

Table 2: Average sanctioning rates for the sanctioners (left panel) and the others (right panel). Between brackets, we display the p-values of the one-sided Wilcoxon signed-rank tests on the difference in sanctioning between the two histories and the two phases.

Most importantly, the sanctioners punish more often after lie–trust than after truth–distrust even though both histories lead to the same payoff distribution. The sanctioning behavior of the others, however, does not differ between these two histories. Secondly, the sanctioners punish more often than the others after lie–trust throughout the selection phase ($p = 0.0072$), but their sanctioning behavior does not differ from that of the others after truth–distrust in either phase ($p = 0.2419$ for RAP and $p = 0.2858$ for SP). Finally, comparing the sanctioning behavior across phases, it can be observed that the sanctioners punish less often after history truth–distrust in the selection phase than in the random assignment phase and that the others punish more often after history lie–trust in the selection phase than in the random assignment phase.

Result 2 (**Sanctioning-II**). *The sanctioners punish more often after lie–trust than after truth–distrust in both phases, the others do so only in the selection phase. The sanctioning rate after truth–distrust does not differ between subgroups.*

Result 2 and Propositon 1 show together that the sanctioners have a positive d in both phases. For the others, this is only the case for the selection phase. Hence, our data largely supports Hypothesis 1.

5.2 Truth-telling

Figure 3 displays the development of the average truth-telling rate during the sessions. It can be seen that subjects tend to tell the truth excessively in both institutions and both phases. The excessive truth-telling seems most prominent in the sanctioning institution during the selection phase, however the difference between the two institutions is most visible in the random assignment phase over the first twenty rounds. Finally, the opportunity to select an institution does not seem to impact the level of truth-telling.

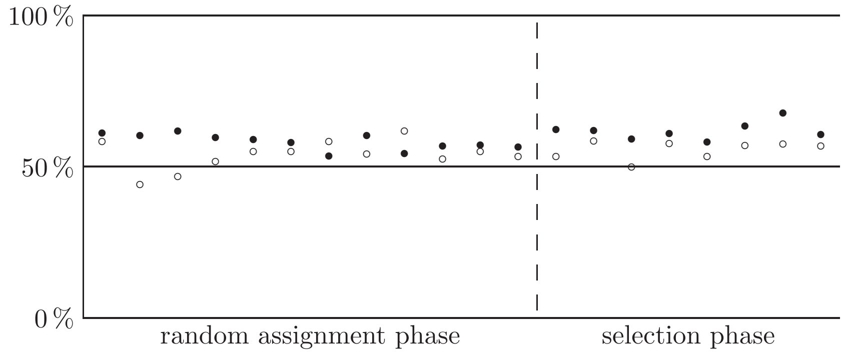

Figure 3: Truth-telling in the sanction-free institution (circles) and in the sanctioning institution (bullets) over rounds (5-round averages).

Table 3 summarizes the average truth–telling rates in both institutions and phases and the test results for excessive truth-telling. In addition, it displays the test results for the difference in truth-telling rates between the two institutions for each of the two phases and between the two phases for each of the two institutions. Except for the sanctioning institution during the selection phase, we find significant excessive truth-telling. The lack of significance, however, seems to be driven by one session in which the sanctioning institution has only been chosen for a few times. To see this, note that in the session in question, the average truth-telling level of 30 % throughout the selection phase is based upon 48 messages, while the number of messages in the other seven sessions ranges from 160 to 320 (with resulting truth-telling levels between 55 % to 60 % and even 85 % in one session). Accordingly, the p-value changes from 0.0708 to 0.0196 once the session in question is left out of the analysis.

	SFI		SI
RAP	54 % (0.0209)	[0.0517]	58 % (0.0105)
	[0.2643]		[0.4168]
SP	55 % (0.0105)	[0.1355]	62 % (0.0708)

Table 3: Average truth-telling rates for the overall population. In parenthesis, we display the p-values of the one-sided Wilcoxon signed-rank tests for excessive truth-telling. In brackets, we display the p-values of the one-sided Wilcoxon signed-rank tests on the difference in truth-telling between the two institutions and the two phases.

The data also reveals that during the random assignment phase, the truth-telling rate in the sanctioning institution is actually greater than the one in the sanction-free institution.

However, the differences do not turn out to be significant. Finally, for both institutions, the transition to the selection phase does not lead to a significant change of the truth-telling rate.

Result 3 (Truth-telling–I). *Subjects tell the truth excessively in both institutions throughout both phases. Institutional selection has no significant effect on truth-telling.*

For an interpretation of these results, we continue with a comparison of the two subgroups. Table 4 provides the relevant numbers on subgroup-averages and test results.

	SFI		SI
RAP	64 %	[0.0465]	73 %
	(0.0059)		(0.0059)
	[0.5000]		[0.0912]
SP	63 %	[0.0807]	78 %
	(0.0178)		(0.0178)

	SFI		SI
RAP	49 %	[0.1039]	51 %
	(0.2201)		(0.3365)
	[0.0395]		[0.0211]
SP	53 %	[0.0058]	42 %
	(0.1629)		(0.0608)

Table 4: Average truth-telling rates for the sanctioners (left panel) and the others (right panel). In parenthesis, we display the p-values of the one-sided Wilcoxon signed-rank tests for excessive truth-telling. In brackets, we display the p-values of the one-sided Wilcoxon signed-rank tests on the difference in truth-telling between the two institutions and the two phases.

The sanctioners tell the truth excessively in both institutions throughout both phases. Excessive truth-telling among the others is nowhere found to be significant. In fact, the data indicates that the others even lie excessively in the sanctioning institution during the selection phase. Moreover, the sanctioners tell the truth more often than the others on all occasions. Since a pairwise comparison of the respective entries in Table 4 yields the matrix of p-values $\begin{bmatrix} 0.0059 & 0.0058 \\ 0.0618 & 0.0059 \end{bmatrix}$, the only instance where this difference is not significant at the 5 % confidence level is the sanction-free institution during the selection phase. The sanctioners also tend to tell the truth more often when there are sanctioning opportunities. For the random assignment phase this effect is significant, and so it is for the selection phase if the first session is not taken into account ($p = 0.0196$). Once the institution is subject to choice, the others are found to lie more in the sanctioning institution than in the sanction-free institution. A similar result is not found for the random assignment phase. Institutional selection does not affect the truth-telling behavior of the sanctioners. On the other hand, the transition to the selection phase causes the others to tell the truth more in the sanction-free and less in the sanctioning institution.

Result 4 (**Truth-telling–II**). *Sanctioners tell the truth more often than the others. Only sanctioners tell the truth excessively, and they do so in both institutions and phases.*

Result 4 indicates that the sanctioners are responsible for the excessive truth-telling in the presence *and* absence of sanctioning opportunities, which is in line with what has been predicted by Hypothesis 2. Hence, we can conclude that the others behave on the aggregate as if they are payoff maximizers. Also, the sanctioners must be assumed to have a strictly positive lying cost c if one wants to explain their behavior with the proposed logit-AQRE.

5.3 Institutional selection

Table 5 presents the relevant data on institutional selection.

	SFI	SI
all	70 %	30 %
sanctioners	49 %	51 %
others	80 %	20 %

Table 5: Institutional selection.

In more than two-thirds of the cases, individuals have selected the sanction-free institution. However, the sanctioners have chosen the sanctioning institution in more than half of the cases, whereas the others only selected this institution in one out of five cases. The difference between the subgroups is significant ($p = 0.0059$). Figure 4 indicates that the subjects' aggregate behavior towards institutional selection is rather stable throughout the selection phase.

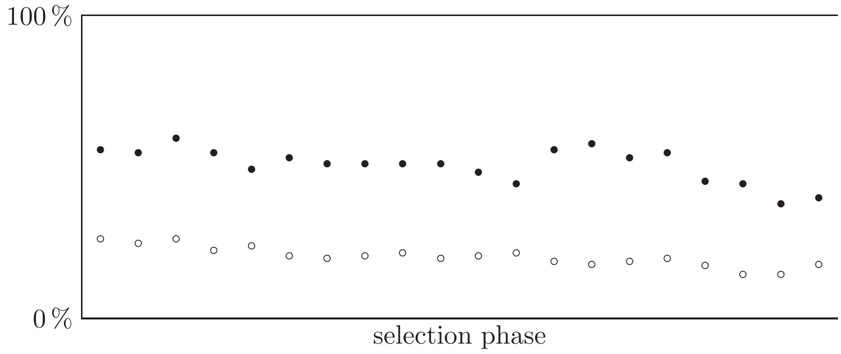

Figure 4: Choice ratio of the sanctioning institution during the selection phase for the sanctioners (bullets) and the others (circles) over rounds (2-round averages).

To investigate the connection between aggregate behavior and individual institutional choice, we test our data on institutional selection for the hypothesis that individuals randomize over institutions with probabilities as depicted in Table 5 (i.e., sanctioners (others) randomize in each round in such a way that they end up in the sanction-free institution in 49 % (80 %) of the cases). Figure 5 presents the cumulative distributions of switching frequencies for the sanctioners and others based on the randomization hypothesis and experimental data.

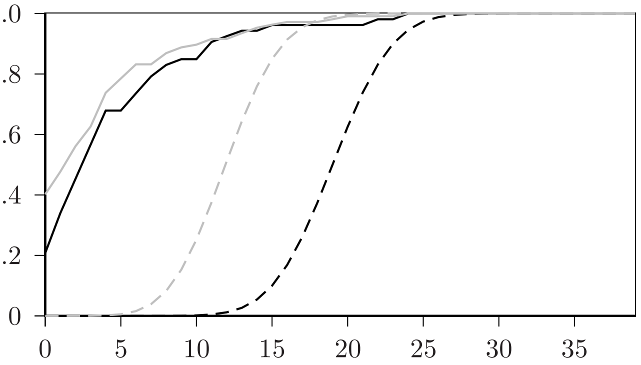

Figure 5: Cumulative distributions of switching frequencies among the sanctioners (black curves) and the others (gray curves). The dashed curves refer to the theoretical prediction for randomization, the continuous curves refer to the data coming from the experiment.

The figure indicates that both types switch less often than predicted by the randomization hypothesis. In fact, the sanctioners (the others) switch with probability 0.1180 (0.0911). Also, the randomization hypothesis can be rejected for both subgroups (sanctioners: $p = 0.0058$, others: $p = 0.0058$). Hence, institutional choice is not random, individuals rather tend to stick to “their” institution.

Result 5 (**Institutional selection**). *Both institutions co-exist. Sanctioners choose the sanctioning institution more often than the others.*

Since the data on institutional selection is as predicted by Hypothesis 3, we can conclude that the sanctioners anticipate a higher level of truth-telling in the sanctioning than in the sanction-free institution throughout the selection phase. This interpretation is also supported by the parameter estimation of the logit-AQRE presented in Section 6, where it is shown that the expected utility for the sanctioners (the others) is higher (lower) in the sanctioning than in the sanction-free institution. Finally, note that the average per round payoff in the sanctioning institution is 16.67 % lower than in the sanction-free institution in the random assignment phase (2.50 ECU versus 3.00 ECU) and 19.45 % lower in the selection phase (2.42 ECU versus 3.00 ECU). Thus, the sanctioners willingly forego a monetary payoff in order to participate in an institution with a higher level of truth-telling. Indeed, sanctioners (others) earned on average 13.06 (13.86) Euro not taking into account the show-up fee. This difference is significant at $p = 0.0085$.

5.4 Trust

Figure 6 displays the development of the average trust rate during the sessions.

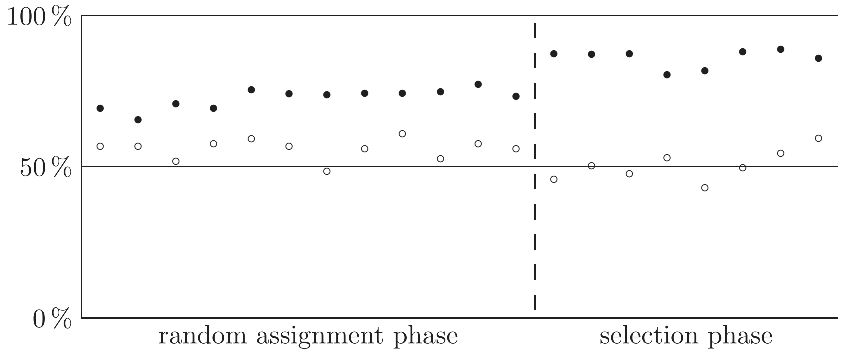

Figure 6: Trust in the sanction-free institution (circles) and in the sanctioning institution (bullets) over rounds (5-round averages).

In the sanctioning institution, receivers trust excessively and there seems to be more trust when the institution is an element of choice. In the sanction-free institution, subjects only seem to trust excessively when randomly assigned to an institution. Table 6 summarizes the average trust rates in both institutions and phases and the test results for excessive trust. It also displays the test results for the difference in trust rates between the two institutions for each of the two phases and between the two phases for each of the two institutions.

In the random assignment phase, we find excessive trust in both institutions, whereas in the selection phase, there is only excessive trust in the sanctioning institution. Moreover, trust rates are higher in the sanctioning institution for both phases. Finally, the transition to the selection phase causes a significant change in trust rates. For the sanctioning institution trust increases, but for the sanction-free institution it decreases.

Result 6 (**Trust–I**). *Subjects trust excessively in both institutions during the random assignment phase and in the sanctioning institution during the selection phase. The presence*

	SFI		SI
RAP	56 %	[0.0004]	73 %
	(0.0071)		(0.0072)
	[0.0072]		[0.0072]
SP	50 %	[0.0004]	86 %
	(0.4721)		(0.0072)

Table 6: Average trust rates for the overall population. In parenthesis, we display the p-values of the one-sided Wilcoxon signed-rank tests for excessive trust. In brackets, we display the p-values of the one-sided Wilcoxon signed-rank tests on the difference in trust between the two institutions and the two phases.

of sanctions enhances trust. Institutional selection increases trust in the sanctioning and reduces trust in the sanction-free institution.

We now compare behavior with respect to trust for the two subgroups. Table 7 provides the relevant numbers on subgroup-averages and test results.

	SFI		SI
RAP	53 %	[0.0058]	87 %
	(0.1462)		(0.0058)
	[0.1313]		[0.2641]
SP	46 %	[0.0058]	92 %
	(0.2201)		(0.0059)

	SFI		SI
RAP	57 %	[0.0126]	66 %
	(0.0342)		(0.0058)
	[0.0611]		[0.0059]
SP	52 %	[0.0059]	78 %
	(0.1999)		(0.0059)

Table 7: Average trust rates for the sanctioners (left panel) and the others (right panel). In parenthesis, we display the p-values of the one-sided Wilcoxon signed-rank tests for excessive trust. In brackets, we display the p-values of the one-sided Wilcoxon signed-rank tests on the difference in trust between the two institutions and the two phases.

In the sanctioning institution, both subgroups trust excessively throughout both phases. In the sanction-free institution, excessive trust is only found for the others during the random assignment phase. If one compares the two subgroups, one finds that the sanctioners trust more than the others in the sanctioning institution during the random assignment phase (a pairwise comparison of the entries in Table 7 yields the following matrix of p-values $\begin{bmatrix} 0.2419 & 0.0059 \\ 0.1632 & 0.1459 \end{bmatrix}$). Moreover, we find that the presence of sanctioning opportunities triggers more trust among both types throughout both phases. Finally, the transition to the selection phase enhances trust of the others in the sanctioning institution.

Result 7 (Trust–II). *The others trust excessively in the sanction-free institution throughout the random assignment phase and both subgroups trust excessively in the sanctioning* *institution. The sanctioners trust more than the others in the sanctioning institution during the random assignment phase. Institutional selection generates more trust by the others in the sanctioning institution.*

Hypothesis 4 is only partly supported by Result 7. While there is excessive trust at the aggregate level in the random assignment phase in both institutions (as implied by a positive c and/or a positive d in the logit-AQRE), we do not find that this result is triggered by the sanctioners alone. In fact, only the others trust excessively in the sanction-free institution during the random assignment phase. This finding supports the interpretation that this group expects the sender to tell the truth frequently in this institution, perhaps because they are aware of the presence of subjects with non-standard preferences towards truth-telling. Also, both groups trust very often in the sanctioning institution. Hence, they take into account that the sender tends to tell the truth, either because of the fear of being punished or because of non-standard preferences.

6 Model estimates

We now present the parameter estimates of the logit-AQRE for the whole population and the two subgroups. The maximum likelihood estimation procedure applied is the standard one: we calculate the equilibrium probabilities of the logit-AQRE numerically on a fine grid and evaluate the objective function at these equilibrium values. In the sanction-free institution, the log-likelihood function to be maximized is equal to

$$L(\lambda, c) = \textstyle\sum_{k \in K} n_k \, \ln(\rho_k^*),$$

where $K = \{\text{truth}, \text{lie}, \text{trust}, \text{distrust}\}$ denotes the union of the strategy sets of the sender and the receiver, n_k indicates how often k has been chosen in the experiment, and ρ_k^* is the equilibrium probability of k given λ and c.

Observe that the data of the sender and the receiver is simultaneously used to determine the value of the log-likelihood function, yet the probability of truth-telling of the sender depends on her own lying cost while the probability that the receiver trusts depends on the expected lying cost of the sender. Hence, our estimations will not provide us with the actual lying cost of the representative individual of the considered group but with an average of the actual lying cost (from the data of the sender) and the expected lying cost of the sender (from the data of the receiver). It is important to keep this limited interpretability of the estimates in mind.

The log-likelihood function in the sanctioning institution is

$$L(\lambda, c, d) = \sum_{k\in K} n_k \ln(\rho^*_k),$$

where the set K contains now additionally the sanctioning decisions of the receiver.

	Sanctioners	Others	Population
Sanction–free			
λ	0.11 [0.02,0.28]	1.27×10^5 [$0.95, 1.70 \times 10^5$]	0.76 [0.25,2.25]
c	9.81 [2.30,31.05]	0.53 [0.30,0.80]	0.66 [0.36,1.12]
Exp. Util.	*0.38*	*2.74*	*2.52*
Sanctioning			
λ	0.96 [0.85,1.00]	1.83 [1.60,2.15]	1.52 [1.10,1.55]
c	0.25 [-0.05,0.70]	1.02 [0.80,1.25]	0.56 [0.40,0.75]
d	3.06 [2.75,3.40]	0.06 [-0.10,0.20]	0.91 [0.80,1.05]
Exp. Util.	*2.23*	*1.93*	*2.07*

Table 8: Logit-AQRE estimation results for the random assignment phase. In brackets, we display the 95 % confidence interval (obtained via bootstrapping with 1 000 repetitions using 70 % of the actual data).

Table 8 presents our estimation results for the whole population and the two subgroups for the random assignment phase. We bootstrap standard errors to determine the accuracy of the estimates. In particular, we re-estimate the parameters 1 000 times for random samples that consist of 70 % of the actual data. This provides us with a distribution of estimates for which we calculate the 95 % confidence interval (via standardizing the empirical cdf). We also calculate the expected utility of the representative individuals, which is the crucial value for deciding whether to join the sanction-free or the sanctioning institution during the selection phase. When calculating these expectations it is taken into account that each subject plays the game in the role of the sender and the role of the receiver with equal probability and that the player in the other role is drawn from the whole population. Also, since Proposition 1 shows that it is impossible to get an estimate of d in the sanction-free institution, we have to assume that it is constant across institutions.

With respect to the sanction-free institution, we find that the sanctioners have a substantial c. The disutility parameter for the others, on the other hand, is rather small. The very large standard error for λ for the others is perhaps surprising, but it has an easy explanation: λ is undefined if the probabilities of truth-telling and trust are equal to 0.5, so the log-likelihood function is extremely flat in λ for the observed truth-telling and trust rates (49 % and 57 %).

The results for the sanctioning institution show that only the sanctioners suffer significantly when being lied to. This is not surprising given that the sanctioners punish predominantly after history lie–trust while the others punish equally often after truth–distrust and lie–trust. Since the others do not tell the truth excessively in this institution either, the positive c must again be caused by their excessive trust. Most importantly, the average bootstrapped c for the sanctioners is not significantly different from zero. The excessive truth-telling of the sanctioners in this institution is hence not driven by lying costs, it rather seems that the fear of being punished when lying is the main underlying reason for their behavior (even though this group has at the same time a high lying cost in the absence of sanctioning opportunities).[19]

Finally, we also see that the expected utility for the sanctioners is higher in the sanctioning than in the sanction-free institution, which is consistent with our experimental finding that these subjects choose the sanctioning institution frequently. The expected utility for the others, on the other hand, is considerably higher in the sanction-free than in the sanctioning institution. This suggests that (i) the others should select the sanction-free institution more often than the sanctioning institution and (ii) the sanction-free institution should be selected more often by the others than by the sanctioners. We have found both results to hold true in our statistical analysis.

To complete our econometric analysis, we finally present in Table 9 the estimation results for the selection phase. We have seen in our statistical analysis that the aggregate group behavior does not change much between the two phases. This suggests that individuals do not switch types when moving from the random assignment to the selection phase. Consequently, we should still obtain a substantial c (d) for the sanctioners in the

[19] Note that λ and c are not comparable across institutions because the estimated c is independent of d in the sanction-free institution but a function of the actual and the expected cost the receiver faces when not sanctioning a lie in the sanctioning institution. And, since c is not comparable across institutions, neither is λ.

sanction-free (sanctioning) institution while the corresponding values for the others should still be rather small. Also, the expected utility for the others should still be higher in the sanction-free than in the sanctioning institution whereas it should still be worthwhile for the sanctioners to opt for the sanctioning institution.

	Sanctioners	Others	Population
Sanction–free			
λ	0.02	1.27×10^5	0.02
	[0.01,0.09]	$[0.01, 1.70 \times 10^5]$	[0.01,0.42]
c	51.92	0.66	10.70
	[5.50,81.50]	[-0.30,4.00]	[0.60,31.8]
Exp. Util.	*-7.17*	*2.56*	*0.13*
Sanctioning			
λ	1.04	2.52	1.68
	[0.9,1.30]	[1.83,3.46]	[1.30,2.25]
c	0.59	1.00	0.46
	[0.00,1.30]	[0.50,1.52]	[0.20,0.80]
d	2.96	0.69	1.27
	[2.50,3.50]	[0.51,0.84]	[1.15,1.65]
Exp. Util.	*2.04*	*1.66*	*1.94*

Table 9: Logit-AQRE estimation results for the selection phase. In brackets, we display the 95 % confidence interval (obtained via bootstrapping with 1 000 repetitions using 70 % of the actual data).

Table 9 supports our interpretation that individuals do not switch types during the experiment. The estimated parameters in the sanctioning institution during the selection phase are for all groups close to those obtained for the random assignment phase. The only considerable change across phases is in the sanction-free institution for the sanctioners: their estimated c increases from 9.81 in the random assignment to 51.92 in the selection phase. So, if anything, the sanctioners suffer more from lying during the selection phase, giving them even more incentives to opt for the sanctioning institution.

7 Concluding discussion

In this study, we proposed a logit-AQRE model with individuals who experience a disutility from lying and being lied to. This model is able to describe the central observations of our laboratory experiment: (i) excessive truth-telling and trust, (ii) history-dependent sanctions, and (iii) the persistent choice of the sanctioning institution. In this concluding section, we relate our findings to the existing theoretical and empirical literature, and identify implications for the modeling and the economic impact of strategic information transmission.

7.1 Bounded rationality and non-standard preferences

There are two possible reasons why a logit-AQRE with payoff maximizing individuals could fail to explain aggregate behavior in our experiment. First, there could be a lack of sophistication in the belief-formation process. For instance, Cai and Wang (2006) and Kawagoe and Takizawa (2009) demonstrate the descriptive power of a behavioral type analysis as discussed in Crawford (2003) or Ellingsen and Östling (2010) in sender-receiver games of varying degrees of conflict. In such a behavioral type analysis, players are assumed to be either "mortal" or "sophisticated". Mortal agents of level 0 stick to a prescribed strategy such as always telling the truth and trust, mortal agents of level 1 play a best response to an opponent who is of level 0 etc.; this iteratively defines level k. Sophisticated agents play strategically and best-respond to the actual distribution of sophisticated and mortal players. While the presence of mortal agents who are "programmed" to tell the truth could certainly explain excessive truth-telling, sanctioning and choosing the sanctioning institution would not be an option for those individuals as long as they maximize their own payoff.[20] Moreover, the propensity of sanctioners to choose the sanctioning institution demonstrates that sanctioners are not just annoyed from having trusted a lie or tell the truth more frequently because of different beliefs about receiver behavior; they rather anticipate a higher level of truth-telling (due to the existence of sanctions) and regard this as a sufficient compensation for lower aggregate material payoffs in this institution.

Once it is acknowledged that mistakes in the belief-formation process are not causing the observed phenomena, we are left with the second potential explanation which is that information transmission is driven by non-standard preferences. This view has already been suggested by Gneezy (2005) and Sutter (2009) who emphasize that deception as observed in the lab crucially depends on the individual *and* social consequences of a lie (or telling the truth). Specifically, Gneezy (2005) shows that the average person in his sample prefers not

[20] Our results imply that lie and trust are the best replies for a payoff maximizer to the observed aggregate behavior in all phases and institutions; and so is choosing the sanction-free institution. Hence, the observed truth-telling is not only excessive relative to the logit-AQRE with payoff maximizing individuals but also relative to the best response to actual behavior. In contrast, trust is excessive relative to logit-AQRE with payoff maximizing individuals but *not* relative to the best response.

to lie if this increases her payoff a little but reduces the receiver's payoff a great deal, i.e. if truth-telling is believed to enhance the total surplus and thereby efficiency.[21] However, as our sanction-free institution is a constant-sum game, efficiency concerns cannot explain the behavior of sanctioners in our experiment. In particular, efficiency oriented individuals would never sanction or choose the sanctioning institution as it generates a lower surplus than the sanction-free institution in our experiment nor would they tell the truth excessively in the sanction-free institution.

While efficiency concerns are therefore unfit to explain our central findings, other models of distributional preferences (e.g., inequity aversion or maximin-preferences)[22] are able to explain certain aspects such as the persistence of sanctioning and the choice of a sanctioning institution. But, as long as preferences only depend on the eventual payoff distribution and do not account for truth-telling, histories such as lie–trust and truth–distrust cannot induce different sanctioning behavior, nor can a sender be motivated to tell the truth excessively. An introduction of costs of lying and being lied to closes this gap, as we have demonstrated.

However, the disutility from lying and being lied to is not necessarily triggered by the truthfulness of the sender's message. The sender's cost of lying could also depend on the sender's beliefs about the receiver's actions and/or the sender's expectations about how much the receiver blames her for a bad outcome. For example, the sender may expect the receiver to trust in more than half of the cases and therefore tell the truth out of altruistic motives or because she would feel guilty if the receiver trusted a lie as discussed in Charness and Dufwenberg (2011).[23] Similarly, the receiver's cost from being lied to could be based

[21] Sutter (2009) corroborates these findings and additionally demonstrates the importance of 'sophisticated deception' by telling the truth to receivers who are expected not to trust. The importance of the consequence of a lie is also highlighted in Wang et al. (2010) who demonstrate that pupil dilation of senders depends on the payoff consequences for the receiver. Hurkens and Kartik (2009) have found that Gneezy's data is also consistent with the hypothesis that subjects are of one of the following two behavioral types: either an individual does not lie at all or she lies whenever the outcome obtained from lying is preferred to the outcome obtained from telling the truth.

[22] For example, utility functions as proposed by Levine (1998), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), or Charness and Rabin (2002). For a recent overview of models of distributional preferences see Sobel (2005).

[23] As the type elicitation based on the random assignment phase can hardly be combined with a belief elicitation and as the constant-sum characteristic of the basic game excludes a Pareto improving information transmission that is crucial in Charness and Dufwenberg (2011), our design does not allow to disentangle per se and belief/intention-dependent costs of a lie. In our corresponding working paper (METEOR Research Memorandum 07/034, Maastricht University), we provide a sequential equilibrium analysis of per se and belief-dependent preferences for truth-telling and demonstrate that both models

on (negative) reciprocity (as specified e.g. in Rabin, 1993, Dufwenberg and Kirchsteiger, 2004, and Falk and Fischbacher, 2006). If the receiver believes that the sender expects him to trust, he may consider a lie as an unkind act because the sender keeps the large payoff for himself while the receiver may feel "entitled" to the large payoff as he is taking action. Observe, however, that an explanation of history-dependent sanctions based on negative reciprocity requires that a receiver only considers a low payoff the result of an unkind act if it was the result of trusting a lie. The same payoff (distribution) is also generated by distrusting a truthful message. In this case, however, we do not observe significant sanctioning. Hence, we need to assume that the kindness of an action does not only depend on the corresponding payoff but also upon whether the payoff has been generated by trusting a lie or by distrusting the truth. The costs of being lied to as introduced in our model can therefore be regarded as a reduced form reciprocity model that explicitly acknowledges whether a certain payoff configuration has been generated by a lie (and is therefore considered unkind and sanctioned) or truth-telling (and is therefore not sanctioned).

In any case, the existence of sanctioners in our experimental society suggests that truth-telling is more frequent or easier to implement (and a less severe obstacle to economic performance) in real-life situations than indicated by models with rational payoff-maximizing agents. In particular, details of institutional design (such as opportunities for costly sanctions) that are irrelevant in these models have a systematic impact on individual behavior and aggregate institutional performance.

7.2 Self selection

The self-selection of individuals into different institutions has been addressed by several laboratory studies. Feld and Tyran (2002) and Alm et al. (1999) allow individuals to vote on the enforcement of a tax that finances a public good and analyze the impact of voting on tax evasion. Typically, voters do not support the enforcement of penalties on tax evasion (with a negative impact on tax compliance). In Decker et al. (2003), Botelho et al. (2005), Guillen et al. (2006), Kroll et al. (2007), Ertan et al. (2009), and Sutter et al. (2010) participants can vote for different sanctioning or reward opportunities in a public good game. In these studies, endogenous institutional choice typically enhances contribution

yield similar predictions for our experimental set-up.

levels.

More closely related to our set-up with self-selection rather than voting is the paper by Gürerk et al. (2006). In their experiment, individuals repeatedly have to select a sanctioning or a sanction-free institution (as in our selection phase) while playing a public good game. At the beginning of the experiment, most individuals choose the sanction-free institution where contribution levels break down in early rounds (resulting in low payoffs). Subsequently these individuals migrate to the sanctioning institution where a few participants who are willing to sanction free-riders enforce high contribution levels (resulting in high payoffs). Ultimately, no individuals interact in the sanction-free institution and high contribution levels are sustained in the sanctioning institution until the end of the experiment. As argued by Henrich (2006), Gürerk et al. (2006)'s experiment is a neat example of competition between groups or institutions that is assumed to be at the heart of social learning processes that can be made responsible for the establishment of social norms in large scale societies (see e.g. Henrich and Boyd, 2001, Friedman and Singh, 2009, or Herold, 2010). While some (e.g., negatively reciprocal) individuals have a propensity to sanction low contributions to the public good, prefer the sanctioning institution already at the beginning of the experiment, and establish high contribution levels therein, other (e.g., profit maximizing) individuals seek to adopt the practices of the payoff-superior institution and therefore vacate the sanction-free institution over time.

In our experiment, self-selection yields a rather stable co-existence of institutions; unlike in Gürerk et al. (2006) where individuals finally coordinate on one of the environments. Our theoretical model allows for such a co-existence. Individuals with sufficiently pronounced costs of lying and being lied to have a propensity to sanction lies, prefer the sanctioning institution, and establish higher levels of truth-telling therein, while payoff maximizers see no reason to migrate into the payoff-inferior sanctioning institution. As individuals are randomly assigned a role in each round and the game without sanctions is constant-sum, establishing truth-telling does not generate higher payoffs in the sanctioning institution in our experiment whereas establishing high contributions did so in the sanctioning institution in Gürerk et al. (2006). As a consequence, sanctioners may stay in the sanctioning institution but do not create an environment that eventually attracts the others. This creates "sub-societies" with distinct economic performance (i.e., aggregate payoffs) and communication modes (i.e., levels of truth-telling). While the (payoff superior) sanction-free institution is fairly described by a population of profit maximizing individuals, the persistence of the (payoff inferior) sanctioning institution requires a more complex modeling of individual preferences (as suggested in our model with costs of lying and being lied to).

References

1. Alm J, G McClelland and W Schulze (1999). Changing the social norm of tax compliance by voting. Kyklos 52 (2): 141-171.

2. Akerlof G (1970). The market for 'lemons': Quality uncertainty and the market mechanism. Quarterly Journal of Economics 84 (3): 488-500.

3. Arrow K (1968). The economics of moral hazard: Further comment. American Economic Review 58 (3): 537-539.

4. Bolton P and A Ockenfels (2000). ERC: A theory of equity, reciprocity, and competition. American Economic Review 90 (1): 166-193.

5. Botelho A, G Harrison, L Costa Pinto and E Rutstroem (2005). Social norms and social choice. University of Central Florida, working paper.

6. Cai H and J Wang (2006). Overcommunication in strategic information transmission games. Games and Economic Behavior 56 (1): 7-36.

7. Charness G and M Dufwenberg (2011). Participation. American Economic Review 101 (4): 1213-1239.

8. Charness G and M Rabin (2002). Understanding social preferences with simple tests. Quarterly Journal of Economics 117 (3): 817-869.

9. Crawford V (2003). Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. American Economic Review 93 (1): 133-149.

10. Crawford V and J Sobel (1982). Strategic information transmission. Econometrica 50 (6): 1431-1451.

11. Decker T, A Stiehler and M Strobel (2003). A comparison of punishment rules in repeated public good games. Journal of Conflict Resolution 47 (6): 751-772.

12. Dufwenberg M and G Kirchsteiger (2004). A theory of sequential reciprocity. Games and Economic Behavior 47 (2): 268-298.

13. Dickhaut J, K McCabe and A Mukherji (1995). An experimental study on strategic information transmission. Economic Theory 6 (3): 389-403.

14. Ellingsen T and R Östling (2010). When does communication improve coordination? American Economic Review 100 (4): 1695-1724.

15. Ertan A, T Page and L Putterman (2009). Who to punish? Individual decisions and majority rule in mitigating the free-rider problem. European Economic Review 53 (5): 495-511.

16. Falk A and U Fischbacher (2006). A theory of reciprocity. Games and Economic Behavior 54 (2): 293-315.

17. Fehr E and K Schmidt (1999). A theory of fairness, competition, and cooperation. Quarterly Journal of Economics 114 (3): 817-868.

18. Feld L and JR Tyran (2002). Tax evasion and voting: An experimental analysis. Kyklos 55 (2): 197-221.

19. Fischbacher U (2007). zTree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10 (2): 171-178.

20. Friedman D and N Singh (2009). Equilibrium vengeance. Games and Economic Behavior 66 (2): 813-829.

21. Gneezy U (2005). Deception: The role of consequences. American Economic Review 95 (1): 384-394.

22. Goeree J and C Holt (2001). Ten little treasures of game theory, and ten intuitive contradictions. American Economic Review 90 (5): 1402-1422.

23. Guillen P, C Schwieren and G Staffiero (2006). Why feed the Leviathan? Public Choice 130 (1-2): 115-128.

24. Gürerk Ö, B Irlenbusch and B Rockenbach (2006). The competitive advantage of sanctioning institutions. Science 312 (5770): 108-111.

25. Henrich J and R Boyd (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. Journal of Theoretical Biology 208 (1): 79-89.

26. Henrich J (2006). Cooperation, punishment, and the evolution of human institutions. Science 312 (5770): 60-61.

27. Herold F (2010). Carrot or stick: The evolution of reciprocal preferences in a haystack model. American Economic Review, forthcoming.

28. Hurkens S and Kartik N (2009). Would I lie to you? On social preferences and lying aversion. Experimental Economics 12 (2): 180-192.

29. Kartik N (2009). Strategic communication with lying costs. Review of Economic Studies 76 (4): 1359-1395.

30. Kawagoe T and H Takizawa (2009). Equilibrium refinement vs. level-k analysis: An experimental study of cheap-talk games with private information. Games and Economic Behavior 66 (1): 238-255.

31. Kroll S, T Cherry and J Shogren (2007). Voting, punishment, and public goods. Economic Inquiry 45 (3): 557-570.

32. Levine D (1998). Modeling altruism and spitefulness in experiments. Review of Economic Dynamics 1 (3): 593-622.

33. McKelvey R and T Palfrey (1998). Quantal response equilibria for extensive form games. Experimental Economics 1 (1): 9-41.

34. Rabin M (1993). Incorporating fairness into game theory and economics. American Economic Review 83 (5): 1281-1302.

35. Sánchez-Pagés S and M Vorsatz (2007). An experimental study of truth-telling in a sender-receiver game. Games and Economic Behavior 61 (1): 86-112.

36. Sánchez-Pagés S and M Vorsatz (2009). Enjoy the silence: an experiment on truth-telling. Experimental Economics 12 (2): 220-241.

37. Sobel J (2005). Interdependent preferences and reciprocity. Journal of Economic Literature 43 (2): 392-436.

38. Sutter M (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. Economic Journal 119 (534): 47-60.

39. Sutter M, S Haigner and M Kocher (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. Review of Economic Studies 77 (4): 1540-1566.

40. Wang J, M Spezio and C Camerer (2010). Pinocchio's pupil: Using eyetracking and pupil dilation to understand truth-telling and deception in sender-receiver games. American Economic Review 100 (3): 984-1007.

41. Xiao E (2010). Profit-seeking punishment corrupts norm obedience. Mimeo.

A Proofs

Proof of Proposition 1

Fix $\lambda \in [0, \infty)$. For the sender,

$$p = \frac{e^{\lambda E[u(\text{truth})]}}{e^{\lambda E[u(\text{truth})]} + e^{\lambda E[u(\text{lie})]}} = \frac{1}{1 + e^{\lambda(E[u(\text{lie})] - E[u(\text{truth})])}},$$

together with $E[u(\text{truth})] = q + 5(1-q) = 5 - 4q$, and $E[u(\text{lie})] = 5q + (1-q) - c = 1 + 4q - c$ implies that

$$p = \frac{1}{1 + e^{\lambda(8q-4-c)}}.$$

For the receiver,

$$q = \frac{e^{\lambda E[v(\text{trust})]}}{e^{\lambda E[v(\text{trust})]} + e^{\lambda E[v(\text{distrust})]}} = \frac{1}{1 + e^{\lambda(E[u(\text{distrust})] - E[u(\text{trust})])}},$$

together with $E[v(\text{trust})] = 5p + (1-p)(1-d)$ and $E[v(\text{distrust})] = p + (1-p)(5-d)$ implies that

$$q = \frac{1}{1 + e^{\lambda(4-8p)}}.$$

Hence, p and q and thereby equilibrium probabilities p^* and q^* are independent of d *(Part (i))*.

If $c = 0$ or if $\lambda = 0$, then $p^* = q^* = \frac{1}{2}$ *(Part (ii))*.

To show that $\lambda > 0$ implies that p^* and q^* are strictly increasing in c *(Part (iii))*, fix $\lambda > 0$. We have that

$$\frac{\partial p}{\partial c} = -\frac{\lambda(8\frac{\partial q}{\partial c} - 1)\, e^{\lambda(8q-4-c)}}{(1 + e^{\lambda(8q-4-c)})^2} \tag{1}$$

and

$$\frac{\partial q}{\partial c} = \frac{8\lambda \frac{\partial p}{\partial c}\, e^{\lambda(4-8p)}}{(1 + e^{\lambda(4-8p)})^2}. \tag{2}$$

Suppose first, to the contrary, that $\frac{\partial p}{\partial c} = 0$. Then, $\frac{\partial q}{\partial c} = \frac{1}{8}$ by equation (1) and $\frac{\partial q}{\partial c} = 0$ by equation (2). So, this cannot be. Suppose next, again to the contrary, that $\frac{\partial p}{\partial c} < 0$. Then, $\frac{\partial q}{\partial c} < 0$ by equation (2) and $\frac{\partial q}{\partial c} > \frac{1}{8}$ by equation (1). Since this is again a contradiction, we can conclude that $\frac{\partial p}{\partial c} > 0$. So, it follows from equation (2) that $\frac{\partial q}{\partial c} > 0$. Finally, uniqueness is obtained from $\frac{\partial p}{\partial q} < 0$ and $\frac{\partial q}{\partial p} > 0$.

Proof of Proposition 2

Fix $\lambda \in [0, \infty)$. It is easy to see that $r^*_{\text{truth,trust}} = 1/(1 + e^{5\lambda})$, $r^*_{\text{truth,distrust}} = 1/(1 + e^{\lambda})$, $r^*_{\text{lie,trust}} = 1/(1 + e^{\lambda(1-d)})$, and $r^*_{\text{lie,distrust}} = 1/(1 + e^{\lambda(5-d)})$. Hence, $\lambda = 0$ or $d = 0$ implies $r^*_{\text{truth,trust}} = r^*_{\text{lie,distrust}}$ and $r^*_{\text{lie,trust}} = r^*_{\text{truth,distrust}}$. Moreover, $r^*_{\text{lie,trust}} > r^*_{\text{truth,distrust}}$ if and only if $\lambda > 0$ and $d > 0$. Using the sanctioning probabilities, we get

$$p = \frac{1}{1 + e^{\lambda f}} \quad \text{and} \quad q = \frac{1}{1 + e^{\lambda g}},$$

with

$$f \equiv E[u(\text{lie})] - E[u(\text{truth})] = \tfrac{5q}{1+e^{-\lambda(1-d)}} + \tfrac{(1-q)}{1+e^{-\lambda(5-d)}} - c - \tfrac{q}{1+e^{-5\lambda}} - \tfrac{5(1-q)}{1+e^{-\lambda}}$$

and

$$g \equiv E[u(\text{distrust})] - E[u(\text{trust})] = \tfrac{p}{1+e^{-\lambda}} + \tfrac{(1-p)(5-d)}{1+e^{-\lambda(5-d)}} - \tfrac{5p}{1+e^{-5\lambda}} - \tfrac{(1-p)(1-d)}{1+e^{-\lambda(1-d)}}.$$

If $\lambda = 0$, then $p^* = q^* = 0.5$.

Next, suppose that $\lambda > 0$. If $c = d = 0$, then $f = (1-2q)(1/(1+e^{-5\lambda}) + (10q-5)/(1+e^{-\lambda}))$ and $g = (2p-1)(1/(1+e^{-\lambda}) + (5-10p)/(1+e^{-5\lambda})$. One can then easily verify that $p^* = q^* = 0.5$. To show that p^* and q^* are strictly increasing in c given $\lambda > 0$ note that

$$\frac{\partial p}{\partial c} = -\frac{\lambda \frac{\partial f}{\partial c} e^{\lambda f}}{(1 + e^{\lambda f})^2} \quad \text{and} \quad \frac{\partial q}{\partial c} = -\frac{\lambda \frac{\partial g}{\partial c} e^{\lambda g}}{(1 + e^{\lambda g})^2}, \tag{3}$$

where

$$\tfrac{\partial f}{\partial c} = \tfrac{\partial q}{\partial c}\left(\tfrac{5}{1+e^{-\lambda(1-d)}} - \tfrac{1}{1+e^{-\lambda(5-d)}} - \tfrac{1}{1+e^{-5\lambda}} + \tfrac{5}{1+e^{-\lambda}}\right) - 1$$

and

$$\tfrac{\partial g}{\partial c} = \tfrac{\partial p}{\partial c}\left(\tfrac{1}{1+e^{-\lambda}} - \tfrac{5-d}{1+e^{-\lambda(5-d)}} - \tfrac{5}{1+e^{-5\lambda}} + \tfrac{1-d}{1+e^{-\lambda(1-d)}}\right).$$

Defining $a = \tfrac{5}{1+e^{-\lambda(1-d)}} - \tfrac{1}{1+e^{-\lambda(5-d)}} - \tfrac{1}{1+e^{-5\lambda}} + \tfrac{5}{1+e^{-\lambda}} > 0$[24] and $b = \tfrac{1}{1+e^{-\lambda}} - \tfrac{5-d}{1+e^{-\lambda(5-d)}} - \tfrac{5}{1+e^{-5\lambda}} + \tfrac{1-d}{1+e^{-\lambda(1-d)}} < 0$[25], one can rewrite equations (3) as

$$\frac{\partial p}{\partial c} = -\frac{\lambda(a\frac{\partial q}{\partial c} - 1)e^{\lambda f}}{(1 + e^{\lambda f})^2} \tag{4}$$

[24] This follows from $5/(1+e^{-\lambda}) \geq 5/2$, $1/(1+e^{-5\lambda}) \leq 1$, $5/(1+e^{-\lambda(1-d)}) \geq 0$ and $1/(1+e^{-\lambda(5-d)}) \leq 1$.

[25] Note that $(5-d)/(1+e^{-\lambda(5-d)}) > (1-d)/(1+e^{-\lambda(1-d)})$ and $5/(1+e^{-5\lambda}) > (1-d)/(1+e^{-\lambda})$.

and

$$\frac{\partial q}{\partial c} = -\frac{\lambda \frac{\partial p}{\partial c} b e^{\lambda g}}{(1+e^{\lambda g})^2}. \tag{5}$$

Suppose first that $\frac{\partial p}{\partial c} = 0$. Then, $\frac{\partial q}{\partial c} = 1/a > 0$ by equation (4) and $\frac{\partial q}{\partial c} = 0$ by equation (5). So, this cannot be. Suppose next, again to the contrary, that $\frac{\partial p}{\partial c} < 0$. Then, $\frac{\partial q}{\partial c} > 1/a > 0$ by equation (4) and $\frac{\partial q}{\partial c} < 0$ by equation (5) because $b < 0$. Consequently, it has to be that $\frac{\partial p}{\partial c} > 0$. So, it follows from $b < 0$ that $\frac{\partial q}{\partial c}$ as well. Finally, uniqueness is obtained from $\frac{\partial p}{\partial q} < 0$ and $\frac{\partial q}{\partial p} > 0$.

B Robustness Analysis

In this section, we show that our results are robust with respect to the type elicitation procedure. 70 out of 160 subjects never sanction after history lie–trust, so these subjects should always form part of the group of the others. Similarly, the 36 subjects who always sanction after this history should likely be classified as sanctioners.[26] Hence, the main task of a robustness analysis is to vary the assignment of the remaining 54 subjects.

In the main part of the paper, a subject is classified as a sanctioner if the degree of confidence with which we reject the hypothesis that this subject punishes the sender after the history lie–trust with a probability of at most one half during the random assignment phase is less than 0.20. We now analyze how the results change if the threshold p-value of this test is varied between 5 % and 40 % (in five-percent steps). Thereby, the condition is relaxed as the p-value increases and more subjects are classified as sanctioners. Figure 7 shows that the percentage of sanctioners increases from about 25 % for $p = 0.05$ to roughly 40 % for $p = 0.40$.

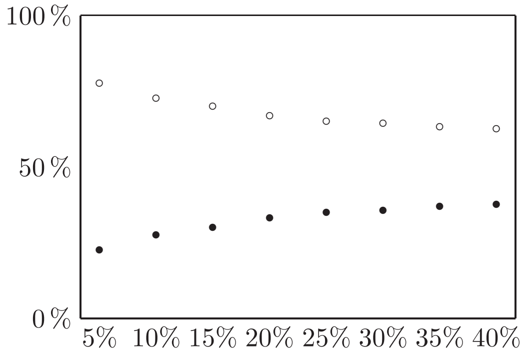

Figure 7: Percentages of sanctioners (bullets) and others (circles) in the experimental population for the eight different classification thresholds.

Results 1, 3, and 6 only contain observations on the entire population and are therefore unaffected by the type elicitation procedure. Result 2 establishes that the punishment rate of the sanctioners (others) is greater after history lie–trust than after history truth–distrust in both phases (in the selection phase). The estimations of the logit-AQRE supported this insight because the model parameter d turned out to be far bigger for the sanctioners.

Figure 8 analyzes how this finding changes with the employed p-value. The sanctioning

[26] It may happen that some of these subjects happen to play the history lie–trust only a few times, so that we cannot be very confident that they really sanction beyond a degree of experimentation.

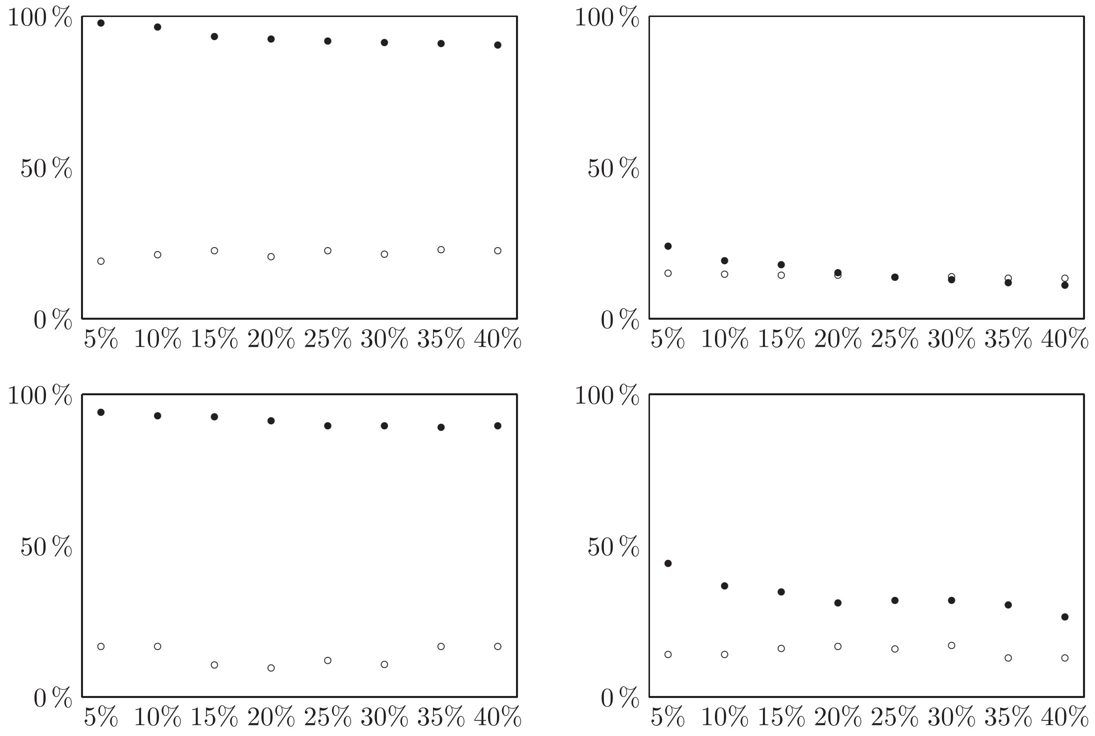

Figure 8: Sanctioning rates after truth–distrust (circles) and lie–trust (bullets) for the sanctioners (left column) and the others (right column) throughout the random assignment phase (top row) and the selection phase (bottom row) for the eight different classification thresholds.

rate after truth–distrust during the random assignment phase is almost identical for both subgroups independently of the assignment so that the size of the parameter d is entirely determined by the sanctioning rate after history lie–trust. We see that the sanctioning rate after lie–trust is slightly decreasing in the p-value for both subgroups, which is very intuitive: the marginal subjects added to the group of sanctioners and taken from the group of others when increasing the p-value punish less often than the subjects who already belong to the group of sanctioners but more than those who form part of the group of the others. The figure also shows the big persistent difference in behavior of the two subgroups. The difference in the sanctioning rate between the two histories is about 75 % for the sanctioners in both phases, whereas for the others, there is no difference in the random assignment phase and a small difference of about 10 % – 20 % (depending on the p-value) in the selection phase. Consequently, Result 2 is very robust with respect to the assignment procedure.

Result 4 states that only the sanctioners tell the truth excessively and that they do so in both phases. One main implication of this result in terms of the logit-AQRE is that only the sanctioners have a significantly positive c in the sanction-free institution.

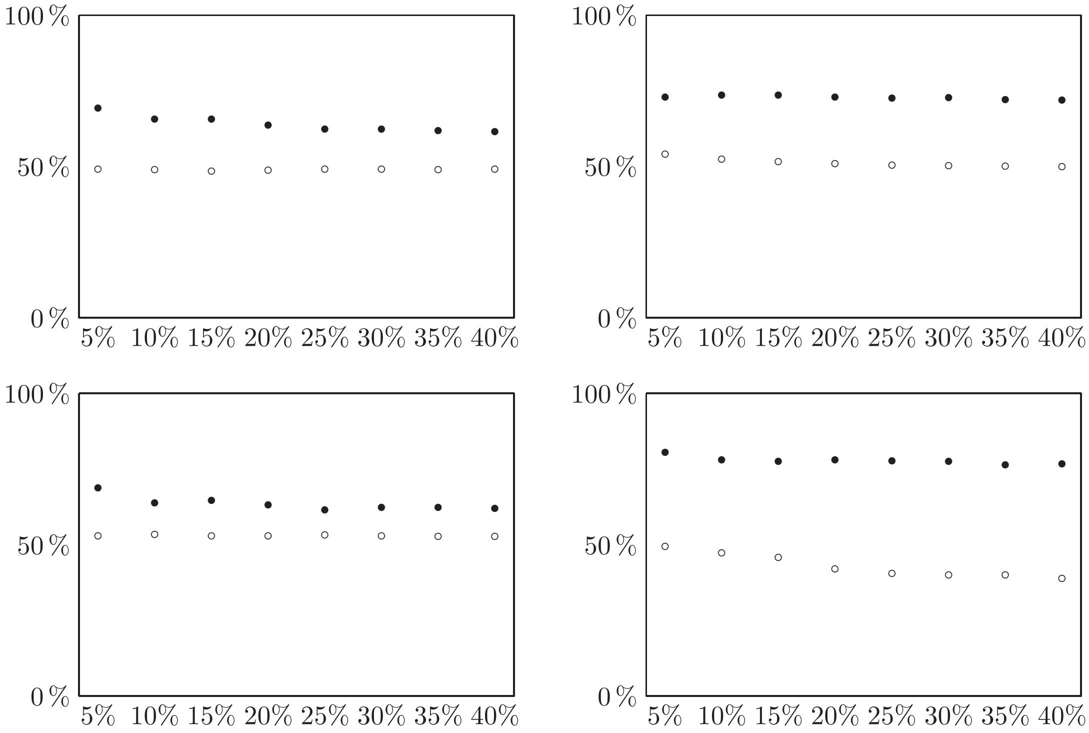

Figure 9: Truth-telling rates for the sanctioners (bullets) and the others (circles) in the sanction-free institution (left column) and the sanctioning institution (right column) during the random assignment phase (top row) and the selection phase (bottom row) for the eight different classification thresholds.

Figure 9 demonstrates the robustness of this finding. Independently of the assignment procedure, only the sanctioners tell the truth excessively in both the sanction–free and the sanctioning institution. The others even tend to lie excessively in the sanctioning institution for high p-values (that is, if many subjects are classified as sanctioners). Consequently, one of our main findings – the group of sanctioners is responsible for the over-communication phenomenon on the aggregate level – holds true for all eight assignment procedures.

Since Result 2 and 4 establish together that the others should be modeled as payoff maximizers and that the sanctioners display non-standard preferences towards truth-telling, the main question raised in this paper (and affirmatively answered by Result 5) is whether the subjects anticipate the different performance of the two institutions in terms of truth-telling and payoff, and self-select accordingly. Figure 10 shows that this result is robust as well. In fact, the sanctioners always select the sanctioning institution in more than 50 % of the cases, while the others do so only in about 20 % of the cases.

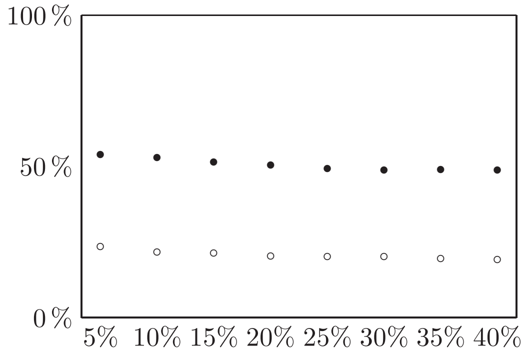

Figure 10: Institutional selection for the sanctioners (bullets) and the others (circles) for the eight different classification thresholds.

Result 7 is the last result in the paper analyzing subgroup behavior. It states that (a) in the sanction-free institution, only the others trust excessively and that they do so only during the random assignment phase, (b) both subgroups trust excessively in the sanctioning institution, and (c) the sanctioners trust more than the others in the sanctioning institution.

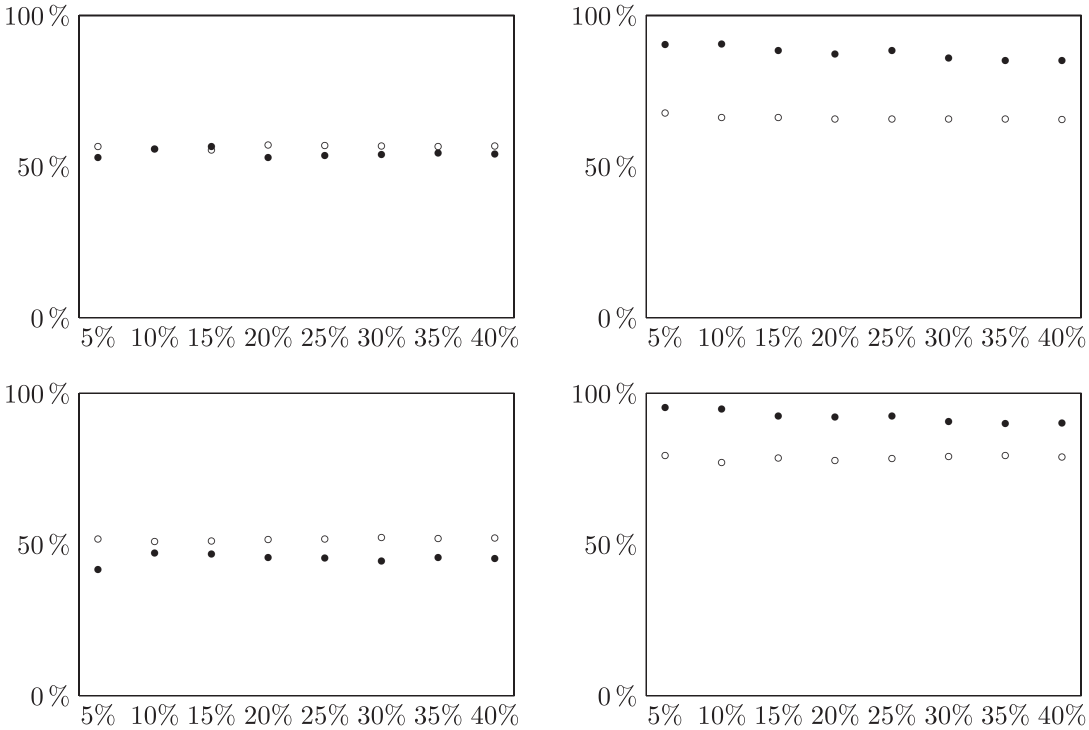

Figure 11: Trust rates for the sanctioners (bullets) and the others (circles) in the sanction-free institution (left column) and the sanctioning institution (right column) during the random assignment phase (top row) and the selection phase (bottom row) for the eight different classification thresholds.

The upper left panel of Figure 11 reveals that we are actually not able to claim that only the others trust excessively in the sanction–free institution during the random assignment phase. It rather seems that both subgroups trust with a probability slightly above 50 %, but depending on the actual assignment one group or the other trusts more. However, the two results mentioned in point (b) and (c) are very robust.

Our discussion has shown so far that the results presented in the paper hold true if the p-value is varied between 0.05 and 0.40. But did we consider a sufficient number of different p-values? In the end, even at a p-value of 0.40 there are still some of the 54 doubtful subjects who sanction sometimes but not always left in the group of the others. To be completely sure that the results are robust, we present in Table 10 the choice probabilities of the most extreme assignment procedure left; that is, the assignment procedure when the group of sanctioners consists of all subjects who sanction at least once after history lie–trust and the group of the others consists of all subjects who never sanction after this history.

	Sanctioners		Others	
	RAP	SP	RAP	SP
Sanctioning				
History lie–trust	93 %	78 %	0 %	7 %
History truth–distrust	27 %	20 %	17 %	4 %
Truth-telling				
SFI	58 %	59 %	49 %	52 %
SI	67 %	67 %	48 %	42 %
Choice for SI		53 %		20 %
Trust				
SFI	57 %	50 %	54 %	51 %
SI	79 %	89 %	65 %	75 %

Table 10: Choice probabilities when the group of sanctioners consists of all subjects who punish at least once after history lie–trust.

Again, none of our main results change. In particular, (a) the difference in the punishment rate after history lie–trust and truth–distrust is far greater for the sanctioners, (b) the sanctioners are responsible for the excessive truth-telling in both institutions, and (c) the sanctioners opt for the sanctioning institution far more often than the others. This demonstrates an independence of our results on the details of the classification procedure.[27]

[27] The results of the formal statistical analysis underlying the robustness analysis is available upon request from the authors.

C Instructions

Welcome

Dear participant,

thank you for taking part in this experiment! It will last about 2 hours. You will be compensated according to your performance during the experiment. In order to ensure that the experiment takes place in an optimal setting, we would like to ask you to follow the general rules during the whole experiment:

- please do not communicate with your fellow students!
- please do not forget to switch off your mobile phone!
- read the instructions carefully. If something is not well explained or any question turns up now or at any time later in the experiment, then ask one of the experimenters. Do, however, not ask out loud, but raise your hand! We will clarify questions privately.
- you may take notes on this instruction sheet if you wish.
- after the experiment, please remain seated till we paid you off.
- if you do not obey the rules, the data becomes useless for us. Therefore we will have to exclude you from this experiment and you will not receive any compensation.

Your decisions are anonymous. None of your fellow students nor anybody else will ever learn them from us.

Environment 1

The central situation of the experiment is the situation depicted in Figure 12 with the following underlying story.

A	*B*
1 ; 5	5 ; 1

Table *A*

A	*B*
5 ; 1	1 ; 5

Table *B*

Figure 12: Central situation of the experiment

There are two players, a *sender* and a *receiver*. In the beginning, the computer randomly selects one of the payoff tables A and B, each with equal probability. Only the sender will be (correctly) informed which table has been selected. Next, the sender transmits either the message *"Table A has been selected"* or the message *"Table B has been selected"* to the receiver. Please, observe that the sender can transmit whatever message he prefers. After observing the sender's message, the receiver decides whether to take *action A* (that is to select column A) or to take *action B* (that is to select column B). The interpretation of the actions is that the receiver says either *I believe the actual payoff table is A* or *I believe the actual payoff table is B*. The payoffs to the sender and the receiver, which are given by the numbers in the corresponding cell, depend only on the table actually chosen by the computer and the action selected by the receiver. The first number in the cell corresponds to the payoff of the sender, the second number to the payoff of the receiver. In short, if the receiver's action matches with the actual table she receives 5 ECU (Experimental Currency Units) and the sender 1 ECU. Otherwise, payoffs are the opposite. For example, if the computer chooses table A, the tells the receiver that table A has been selected, and the receiver takes action A, then the sender gets 1 ECU and the receiver 5 ECU.

Environment 2

The second environment extends the first environment. After receiving feedback on the table chosen by the computer and the decisions of the sender and the receiver, the receiver has to make a final decision. She has to decide whether to *accept* the payoffs for both participants or whether to *reduce* the payoff of both participants to zero.

Matching

This experimental session consists of 100 rounds. In total, 20 subjects participate in this experiment. In every of the first 60 rounds, the computer assigns you randomly to one of the two environments. With 70 % probability you will be assigned to the second environment. Next, you are randomly matched with another participant from the same environment to form a pair. In each pair, one participant is randomly chosen to be the sender, and one to be the receiver. This process is random. Your profile may change every round with respect to three variables: the environment you are assigned to (1 or 2), the participant you are matched with (some subject from the same environment), and the role you have (sender or receiver). The matching is anonymous, so you will never learn with whom you formed a pair. After every round you receive a complete feedback of the decisions of both players, the payoffs from the round, and your accumulate payoff.

In the second phase of the experiment, the last 40 rounds, you can decide whether you want to be in environment 1 or in environment 2. This decision is taken every round anew. Given your decision for the current round, you are again randomly matched with another participant from the same environment to form a pair. In each pair, one participant is randomly chosen to be the sender, and one to be the receiver. Observe that if an odd number of participants choose an environment it becomes impossible to divide all players into pairs. In this situation, the participant that stays single does not have to make decisions and gets a fixed payoff of 3 ECU. The matching is anonymous, so you will never learn with whom you formed a pair. After every round you receive a complete feedback of the decisions of both players, the payoffs from the round, and your accumulate payoff.

Payment

The points that you accumulate in course of the experiment will determine your payment. The exchange rate ECU/Euros is such that every ECU in the experiments is equal to 5 Eurocents.

Closing

At the end of the experiment, we would like to ask you to complete a short on-screen questionnaire. But, before we start, we would like to ask you to answer the control questions on the bottom of this page. Once ready, please raise your hand, and one of the experimenters will check your answers. The software will be started as soon as *all* answers have been checked. So, please, be patient.

Thank you again and good luck with the experiment! And, please, make your decisions carefully—your reward depends on your performance during the experiment.

Control questions

Please, answer the following questions! One of the experimenters will go round, check the answers and discuss any problems.

Please fill in your subject id: ______________________________

Statement	**True**	**False**
In the 43th round of the experiment, I will be able to select my favorite environment.		
If I am playing the role of sender this round, I can be sure to be playing the role as receiver next round.		
I never know whom of the other participants I am matched with.		
As a sender I can be sure that the receiver regards my message as credible.		
In the second environment, before making the decision of whether or not to reduce the payoffs of both participants, I am informed about the selected table and the payoffs resulting from my choice as a receiver.		
My decisions in the first phase do not influence my payoffs.		

University of Innsbruck - Working Papers in Economics and Statistics
Recent Papers can be accessed on the following webpage:

http://eeecon.uibk.ac.at/wopec/

2011-28 **Ronald Peeters, Marc Vorsatz, Markus Walzl:** Truth, trust, and sanctions: On institutional selection in sender-receiver games *forthcoming in Scandinavian Journal of Economics*

2011-27 **Haoran He, Peter Martinsson, Matthias Sutter:** Group Decision Making Under Risk: An Experiment with Student Couples *forthcoming in Economics Letters*

2011-26 **Andreas Exenberger, Andreas Pondorfer:** Rain, temperature and agricultural production: The impact of climate change in Sub-Sahara Africa, 1961-2009

2011-25 **Nikolaus Umlauf, Georg Mayr, Jakob Messner, Achim Zeileis:** Why Does It Always Rain on Me? A Spatio-Temporal Analysis of Precipitation in Austria

2011-24 **Matthias Bank, Alexander Kupfer, Rupert Sendlhofer:** Performance-sensitive government bonds - A new proposal for sustainable sovereign debt management

2011-23 **Gerhard Reitschuler, Rupert Sendlhofer:** Fiscal policy, trigger points and interest rates: Additional evidence from the U.S.

2011-22 **Bettina Grün, Ioannis Kosmidis, Achim Zeileis:** Extended beta regression in R: Shaken, stirred, mixed, and partitioned

2011-21 **Hannah Frick, Carolin Strobl, Friedrich Leisch, Achim Zeileis:** Flexible Rasch mixture models with package psychomix

2011-20 **Thomas Grubinger, Achim Zeileis, Karl-Peter Pfeiffer:** evtree: Evolutionary learning of globally optimal classification and regression trees in R

2011-19 **Wolfgang Rinnergschwentner, Gottfried Tappeiner, Janette Walde:** Multivariate stochastic volatility via wishart processes - A continuation

2011-18 **Jan Verbesselt, Achim Zeileis, Martin Herold:** Near Real-Time Disturbance Detection in Terrestrial Ecosystems Using Satellite Image Time Series: Drought Detection in Somalia

2011-17 **Stefan Borsky, Andrea Leiter, Michael Pfaffermayr:** Does going green pay off? The effect of an international environmental agreement on tropical timber trade

2011-16 **Pavlo Blavatskyy:** Stronger Utility

2011-15 **Anita Gantner, Wolfgang Höchtl, Rupert Sausgruber:** The pivotal mechanism revisited: Some evidence on group manipulation

2011-14 **David J. Cooper, Matthias Sutter:** Role selection and team performance

2011-13 **Wolfgang Höchtl, Rupert Sausgruber, Jean-Robert Tyran:** Inequality aversion and voting on redistribution

2011-12 **Thomas Windberger, Achim Zeileis:** Structural breaks in inflation dynamics within the European Monetary Union

2011-11 **Loukas Balafoutas, Adrian Beck, Rudolf Kerschbamer, Matthias Sutter:** What drives taxi drivers? A field experiment on fraud in a market for credence goods

2011-10 **Stefan Borsky, Paul A. Raschky:** A spatial econometric analysis of compliance with an international environmental agreement on open access resources

2011-09 **Edgar C. Merkle, Achim Zeileis:** Generalized measurement invariance tests with application to factor analysis

2011-08 **Michael Kirchler, Jürgen Huber, Thomas Stöckl:** Thar she bursts - reducing confusion reduces bubbles *modified version forthcoming in American Economic Review*

2011-07 **Ernst Fehr, Daniela Rützler, Matthias Sutter:** The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence

2011-06 **Octavio Fernández-Amador, Martin Gächter, Martin Larch, Georg Peter:** Monetary policy and its impact on stock market liquidity: Evidence from the euro zone

2011-05 **Martin Gächter, Peter Schwazer, Engelbert Theurl:** Entry and exit of physicians in a two-tiered public/private health care system

2011-04 **Loukas Balafoutas, Rudolf Kerschbamer, Matthias Sutter:** Distributional preferences and competitive behavior *forthcoming in Journal of Economic Behavior and Organization*

2011-03 **Francesco Feri, Alessandro Innocenti, Paolo Pin:** Psychological pressure in competitive environments: Evidence from a randomized natural experiment: Comment

2011-02 **Christian Kleiber, Achim Zeileis:** Reproducible Econometric Simulations

2011-01 **Carolin Strobl, Julia Kopf, Achim Zeileis:** A new method for detecting differential item functioning in the Rasch model

2010-29 **Matthias Sutter, Martin G. Kocher, Daniela Rützler and Stefan T. Trautmann:** Impatience and uncertainty: Experimental decisions predict adolescents' field behavior

2010-28 **Peter Martinsson, Katarina Nordblom, Daniela Rützler and Matthias Sutter:** Social preferences during childhood and the role of gender and age - An experiment in Austria and Sweden *Revised version forthcoming in Economics Letters*

2010-27 **Francesco Feri and Anita Gantner:** Baragining or searching for a better price? - An experimental study. *Revised version accepted for publication in Games and Economic Behavior*

2010-26 **Loukas Balafoutas, Martin G. Kocher, Louis Putterman and Matthias Sutter:** Equality, equity and incentives: An experiment

2010-25 **Jesús Crespo-Cuaresma and Octavio Fernández Amador:** Business cycle convergence in EMU: A second look at the second moment

2010-24 **Lorenz Goette, David Huffman, Stephan Meier and Matthias Sutter:** Group membership, competition and altruistic versus antisocial punishment: Evidence from randomly assigned army groups *Revised version forthcoming in Management Science*

2010-23 **Martin Gächter and Engelbert Theurl:** Health status convergence at the local level: Empirical evidence from Austria *(revised Version March 2011)*

2010-22 **Jesús Crespo-Cuaresma and Octavio Fernández Amador:** Business cycle convergence in the EMU: A first look at the second moment

2010-21 **Octavio Fernández-Amador, Josef Baumgartner and Jesús Crespo-Cuaresma:** Milking the prices: The role of asymmetries in the price transmission mechanism for milk products in Austria

2010-20 **Fredrik Carlsson, Haoran He, Peter Martinsson, Ping Qin and Matthias Sutter:** Household decision making in rural China: Using experiments to estimate the influences of spouses

2010-19 **Wolfgang Brunauer, Stefan Lang and Nikolaus Umlauf:** Modeling house prices using multilevel structured additive regression

2010-18 **Martin Gächter and Engelbert Theurl:** Socioeconomic environment and mortality: A two-level decomposition by sex and cause of death

2010-17 **Boris Maciejovsky, Matthias Sutter, David V. Budescu and Patrick Bernau:** Teams make you smarter: Learning and knowledge transfer in auctions and markets by teams and individuals

2010-16 **Martin Gächter, Peter Schwazer and Engelbert Theurl:** Stronger sex but earlier death: A multi-level socioeconomic analysis of gender differences in mortality in Austria

2010-15 **Simon Czermak, Francesco Feri, Daniela Rützler and Matthias Sutter:** Strategic sophistication of adolescents - Evidence from experimental normal-form games

2010-14 **Matthias Sutter and Daniela Rützler:** Gender differences in competition emerge early in live

2010-13 **Matthias Sutter, Francesco Feri, Martin G. Kocher, Peter Martinsson, Katarina Nordblom and Daniela Rützler:** Social preferences in childhood and adolescence - A large-scale experiment

2010-12 **Loukas Balafoutas and Matthias Sutter:** Gender, competition and the efficiency of policy interventions

2010-11 **Alexander Strasak, Nikolaus Umlauf, Ruth Pfeifer and Stefan Lang:** Comparing penalized splines and fractional polynomials for flexible modeling of the effects of continuous predictor variables

2010-10 **Wolfgang A. Brunauer, Sebastian Keiler and Stefan Lang:** Trading strategies and trading profits in experimental asset markets with cumulative information

2010-09 **Thomas Stöckl and Michael Kirchler:** Trading strategies and trading profits in experimental asset markets with cumulative information

2010-08 **Martin G. Kocher, Marc V. Lenz and Matthias Sutter:** Psychological pressure in competitive environments: Evidence from a randomized natural experiment: Comment *An extended version with the title "Psychological pressure in competitive environments: New evidence from randomized natural experiments" is forthcoming in Management Science*

2010-07 **Michael Hanke and Michael Kirchler:** Football Championships and Jersey sponsors' stock prices: An empirical investigation

2010-06 **Adrian Beck, Rudolf Kerschbamer, Jianying Qiu and Matthias Sutter:** Guilt from promise-breaking and trust in markets for expert services - Theory and experiment

2010-05 **Martin Gächter, David A. Savage and Benno Torgler:** Retaining the thin blue line: What shapes workers' intentions not to quit the current work environment

2010-04 **Martin Gächter, David A. Savage and Benno Torgler:** The relationship between stress, strain and social capital

2010-03 **Paul A. Raschky, Reimund Schwarze, Manijeh Schwindt and Ferdinand Zahn:** Uncertainty of governmental relief and the crowding out of insurance

2010-02 **Matthias Sutter, Simon Czermak and Francesco Feri:** Strategic sophistication of individuals and teams in experimental normal-form games

2010-01 **Stefan Lang and Nikolaus Umlauf:** Applications of multilevel structured additive regression models to insurance data

University of Innsbruck

Working Papers in Economics and Statistics

2011-28

Ronald Peeters, Marc Vorsatz, Markus Walzl

Truth, trust, and sanctions: On institutional selection in sender-receiver games

Abstract
We conduct a laboratory experiment to investigate the impact of institutions and institutional choice on truth-telling and trust in sender-receiver games. We find that in an institution with sanctioning opportunities, receivers sanction predominantly after having trusted lies. Individuals who sanction are responsible for truth-telling beyond standard equilibrium predictions and are more likely to choose the sanctioning institution. Sanctioning and non-sanctioning institutions coexist if their choice is endogenous and the former shows a higher level of truth-telling but lower material payoffs. It is shown that our experimental findings are consistent with the equilibrium analysis of a logit agent quantal response equilibrium with two distinct groups of individuals: one consisting of subjects who perceive non-monetary lying costs as senders and non-monetary costs when being lied to as receivers and one consisting of payoff maximizers.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)