Experimental Testing of Leniency Programs: The Impact of Instructions Framing

Jana Richmanová∗, Andreas Ortmann† and L’ubomír Lížal‡
CERGE–EI§

April 2008

Abstract

This paper, together with Richmanova (2008), intends to shed some light on important methodological questions for experimental testing of leniency programs as anti-corruption measures. Following up on earlier work of Abbink & Hennig-Schmidt (2006) we study the effect of loaded instructions in a bribery experiment. We find a strong gender effect – men and women react differently to bribery context. Treatment effect becomes significant once we allow for gender-specific coefficients.

Keywords: corruption, anti-corruption mechanisms, optimal contract, monitoring

JEL classification:
1 Introduction

The severe consequences of corruption have been widely documented in the empirical literature. For example, Mauro (1995) and Tanzi (1998) have shown a negative effect of corruption on economic growth; Hwang (2002) has demonstrated that corruption, through tax evasion, reduces government revenues; and Gupta, Davoodi and Alonso-Terme (2002) conclude that corruption increases income inequality and poverty. The results of a cross-national study by Treisman (2000) suggest that historical circumstances play an important role in determining current levels of corruption. The author suggests that policy decisions work too slowly to outweigh the influence of recent decades.

The design and implementation of effective anti-corruption measures therefore remains an important concern.

One of the promising anti-corruption measures are leniency policies. Leniency policies award fine reductions of varying intensities to wrongdoers who "spontaneously" report an illegal agreement and thereby help to convict their accomplice(s). They serve as an enforcement mechanism as much as a means of deterrence in that, if appropriately designed and implemented, they have the potential to undermine the trust between wrongdoers. Leniency policies have, however, been analyzed in the literature mostly as an anti-cartel mechanism.

Spagnolo (2004) examines theoretically effects of leniency policies of various degrees – from moderate (which only reduce, or at most cancel, the penalty for a criminal who reports) to full (which, in addition, pay a reward). He shows that reward-paying leniency programs provide costless\(^1\) and very efficient measure for cartel deterrence.

Drawing on early versions of Spagnolo (2004), Apesteguia, Dufwenberg and Selten (2004) conducted an experiment that confirms these promising cartel-deterrence properties of leniency policies. The authors compare four treatments: Ideal (no cartel formation is possible), Standard (traditional antitrust with no fine reductions is in force), Leniency (a reduced fine is imposed on each reporting cartel member), and Bonus (a reward financed from fines paid by non-reporting members is awarded to each reporting cartel members). The lowest cartel formation is found under Leniency. Surprisingly, Bonus leads to the highest cartel formation and the highest reporting rates.

Bigoni, Fridolfsson, Le Coq and Spagnolo (2007) conducted another experiment. The authors use a repeated Bertrand game with uncertain duration to compare cartel formation and prices under five different schemes: Communication (communication is legal), Antitrust (with full fines to all members of detected cartel), Leniency (with zero fine to the member who self-reports), LenRing (with zero fine to the member who self-reports unless he was the initiator of communication), Re-

\(^1\)This is the case if the rewards are fully financed from fines imposed on other convicted members of a gang
ward (reward to the member who self-reports financed from fine paid by the other cartel member). They find that without leniency past convictions reduce number of cartels but increase collusive prices. The effect of leniency, compared to Antitrust, is similar. The authors also find that past convictions due reporting have much stronger deterrence effect than those due external investigation. Rewards lead to almost complete deterrence.

In another paper, Bigoni, Fridolfsson, Le Coq and Spagnolo (2008) investigate experimentally the role of risk attitudes for the design of antitrust policies. They design various treatments in order to be able to distinguish between strategic risk (or risk of being cheated upon and/or reported to the authority) and exogenous risk of detection. The results suggest that past experience might have more important consequences for the perception of risk than exogenous probability of detection, and that the strategic risk plays a key role for effectiveness of leniency policy. In general, deterrence is higher when leniency is in place.

Both Bigoni et al. (2007, 2008) papers are recent work in progress and which we were not aware of when running our experiments.

Leniency policies to deter cartels are not directly applicable as anti-corruption measures, as the former situation is (modelled as) a simultaneous game while strategies, payoffs, and move structure of the latter are asymmetric. A proper theoretical and experimental analysis is therefore called for.

To our knowledge the first theoretical work analyzing various effects of leniency policies on corruption is Buccirossi & Spagnolo (2006). The authors show that poorly designed moderate leniency policies may have a serious counter-productive effect: they might allow a briber to punish at relatively low cost a partner that does not respect an illegal agreement. In other words, leniency policies might actually provide an enforcement mechanism for occasional illegal transactions. Thus they can, contrary to the intention, increase corruption.

In Richmanova & Ortmann (2008) we have proposed a generalization of the Buccirossi & Spagnolo 2006 model. We introduce probabilistic discovery of evidence by auditing inspectors. In the original model, Buccirossi & Spagnolo assume that a briber and a bribee agree to produce hard evidence which serves as a hostage. Without hard evidence being produced, the occasional illegal transaction is not enforceable. An audit, if it takes place, discovers the evidence with probability one. In Richmanova & Ortmann (2008), we argue that instead some evidence is created unintentionally and this can be discovered by an audit with some probability which is less than one. Such generalization makes the model more realistic and more readily applicable for experimental testing without changing the qualitative result of Buccirossi & Spagnolo.

Buccirossi & Spagnolo’s result together with the theoretical and experimental

\footnote{For more detailed discussion see Richmanova (2006).}

\footnote{Occasional illegal transactions are one-shot transactions which are not implemented through repetition.}
evidence from the literature on cartel deterrence suggests that the potential of leniency policies to undermine trust between wrongdoers hinges upon proper design and implementation.

Experimental methods have been widely used albeit rarely to study corruption (Dusek, Ortmann & Lizal 2005). They become especially useful when counterfactual institutional arrangements such as leniency programs need to be explored: they provide relatively cheap way to examine their effects in controlled environment (see Dusek et al. 2005, Aspesteguia et al. 2004, Buccirossi & Spagnolo 2006, Bigoni et al. 2007, 2008, Richmanova & Ortmann 2008 and also Roth 2002).

We use the generalized Buccirossi & Spagnolo model for experimental testing of leniency policies as an anticorruption measure. This study, together with Richmanova (2008) [in prep], intends to shed some light on important questions regarding the proper experimental design and they serve as a point of departure for future study of leniency policies.

Richmanova (2008) examines the effect of change in parameterization. It has been documented in the literature that a change in parameterization which does not affect the theoretical prediction might indeed have consequences for behavior of subjects in the lab (e.g. Goeree & Holt 2001). In the generalized Buccirossi & Spagnolo game, the action bringing the highest possible payoff is also associated with a risk of considerable loss. Therefore, it is likely that subjects in the lab will not behave in complete accordance with the theoretical prediction, especially when the prediction is made under assumption of risk neutrality. The question we ask is, whether by making the risky choice more tempting and at the same time reducing the penalty we can bring the experimental results closer to the prediction.

In this paper, following up on earlier work of Abbink & Hennig-Schmidt (2006), we study the effect of loaded instructions in a bribery experiment.

Abbink & Hennig-Schmidt (2006) use a repeated reciprocity game in which a firm applies for a permit, and a public official is responsible for the issue of the permit. The firm decides whether to transfer some amount of money to the public official and what amount to transfer. Transfer can be accepted or rejected by the public official. If he rejects, the money is not transferred and the firm has to pay a small initiation cost. If he accepts, the amount he is about to receive is tripled. Afterwards a lottery is played: with a very low probability (.3%) a sudden death occurs: corruption is detected and both involved subjects are excluded from further experiment forfeiting all their cumulated earnings. In the last stage, the public official decides whether to issue the permit for the firm. If he issues the permit, a negative externality is imposed on the public which is represented by the rest of the subjects in the session.

No feedback about others’ decisions is provided to subjects. The experiment is conducted in two treatments: one with neutral wording and no context provided.

---

4This asymmetry represents a difference in utilities from bribe for the public official and for the firm.
to subjects, and the other with full bribery context in order to examine the effect of loading the instructions.

Abbink & Hennig-Schmidt find some treatment effect, it is not statistically significant, though. The authors conclude that this surprising result may be caused by the nature of the game: it is very simple, and as it was designed to capture all basic features of bribery, even with neutral wording, subjects may have deciphered what the experiment was about.

The generalized Buccirossi & Spagnolo game is more complex (e.g. involves realization of two random outcomes) and therefore is likely to be less susceptible to inferences about context. Hence, the experimental testing of the impact of instructions framing may provide important insights for further experimental testing of the impact of leniency policies.

Altogether, we design three treatments: a benchmark, which is common for both studies, Richmanova (2008) and the present paper, and in which all instructions are presented in completely neutral language; a context treatment, in which we use the same parameterization as in the benchmark but instructions are presented in full bribery context; and, finally, a high-incentives treatment, which implements new parameterization within neutral framing (Richmanova 2008).

In all three treatments we observe deviations from the theoretical prediction. Richmanova (2008) finds significant effect of the parametric change, however, not always towards the theoretical prediction. In the present paper, we find a strong gender effect - male and female participants show different reactions to bribery context. The effect of context significantly affects behavior of subjects once we allow for gender specific effects.

The remainder of the paper is organized as follows. In the next section we discuss the generalized Buccirossi & Spagnolo model in detail, and we also describe and compare two experimental treatments. Section 3 talks about experimental implementation and in Section 4 we discuss the results. Section 5 concludes.

2 Experimental Design

The experiment implements the bribery game in Richmanova & Ortmann (2008) in which a bureaucrat and an entrepreneur are matched. The entrepreneur has an investment possibility of net present value $v$, the success of which requires the bureaucrat to perform an illegal Action $a$. For doing so, the bureaucrat may require a compensation in form of a bribe $b$.  

5This result is, however, disputable. When looking at the evolutions of bribe offers and of permission frequencies, especially in twenty central rounds, a clear difference between two treatments is visible. The first rounds might likely be affected by a learning effect of the first kind (e.g., subjects becoming familiar with the lab setting rather than reacting to incentives, see e.g. Hertwig & Ortmann 2001) and, the last five rounds by a possible termination effect – which is acknowledged by the authors. This difference might become significant (even without excluding possibly problematic first and last five periods) with larger number of observations.
The timing of the game is as follows. First, the entrepreneur decides whether to Pay or Not Pay a bribe. If she does not pay a bribe, the game ends. If she does, the bureaucrat chooses one of three possible actions: Denounce, do Nothing, or perform Action a.

If the bureaucrat chooses Denounce, an audit is carried out. The audit may (with probability $\beta$, $\beta \in (0,1)$), or may not (with probability $1 - \beta$), discover some evidence of bribery. In the former case, bribery is detected and the leniency policy guarantees that the bureaucrat will have to pay only a reduced fine whereas the entrepreneur will have to pay a full fine. After detection, they in addition forfeit their gains from the illegal transaction – which in this particular case means that bribe $b$ is confiscated.\(^6\) In the latter case, bribery is not detected and the bureaucrat enjoys his illegal gain - bribe $b$.

If the bureaucrat chooses Nothing or Action a, then the entrepreneur moves next. In both cases he chooses between Denounce and do Nothing.

If the entrepreneur chooses Denounce and the ensuing audit discovers evidence (which, again, happens with probability $\beta$), then she will have to pay a reduced fine whereas the bureaucrat will have to pay a full fine and, in addition, their illegal gains will be confiscated. If no evidence is discovered both agents will keep their illegal gains.

If the entrepreneur chooses Nothing, then an audit may still occur with some nonzero probability $\alpha$. If the audit detects bribery (which happens with probability $\beta$), both parties are subject to sanction, which consists of the confiscation of illegal gains plus full fine. Note that illegal gains include bribe $b$ in any case and value $v$ only in case when the bureaucrat has chosen to perform Action a.

Figure 1 summarizes the extensive form of the game and respective expected payoffs. Notation is as follows: $b$ denotes bribe, $v$ value of the project to the entrepreneur; $\alpha$ denotes exogenous probability of audit if neither party reports (after reporting, the probability of audit is 1); $\beta$ denotes the probability that audit discovers some evidence sufficient for conviction; $F_E$ and $F_B$ denote full fines and $RF_E$ and $RF_B$ reduced fines to the entrepreneur and to the bureaucrat, respectively.

The contribution of the generalized model lies in introduction of probability $\beta$. In Buccirossi & Spagnolo (2006) it is assumed that, before the illegal transaction takes place, the bureaucrat and the entrepreneur agree on production of hard evidence. Without hard evidence being voluntarily produced by both of them the illegal transaction is not enforceable. In essence it is assumed that both involved are holding a hostage which commits them to the desired outcome. It is furthermore assumed that, if an audit takes place, corruption is discovered and both culprits are convicted with probability one. Richmanova & Ortmann (2008) assume instead that some hard evidence is created unintentionally along the way and that this evidence may be discovered by an audit with probability $\beta \in (0,1)$.

\(^6\)Note that in this case the illegal transaction has been detected without Action a being performed and therefore there is no gain to the entrepreneur to be confiscated.
Figure 1: Extensive form of the corruption game in the generalized model. $P$ stands for Pay, $NP$ for Not Pay, $D$ for Denounce, $N$ for doing Nothing, $a$ for performing Action $a$.

The basic structure of both, the original and the modified game, is the same, except that in the original version the probability $\beta$ is set to 1. The generalization makes the model more suitable for experimental testing, as now no additional stage in which subjects would have to agree on producing a hostage is needed. In addition, in Richmanova & Ortmann (2008) we argue that the generalized model resembles real-world situations more closely. \footnote{We realize that in such a game beliefs about probability of detection might play an important role. However, we believe that introduction of beliefs would make the game more complex than necessary for experimental testing. Instead, we view probability $\beta$ as an empirical success rate, or effectiveness, of a detection technology that is known to subjects.}

Buccirossi & Spagnolo (2006) show, that in the absence of a leniency program, occasional illegal transactions are not implementable. \footnote{Facing the full fine even after reporting, the entrepreneur cannot credibly threaten to report the bureaucrat in case he would not deliver. Therefore, the bureaucrat would keep the bribe and not perform Action $a$, knowing that it is not profitable for the entrepreneur to punish him. Consequently, the entrepreneur would not enter the illegal agreement at the first place.} The result carries over into the generalized model. After the introduction of a modest leniency program, \footnote{Similarly to Spagnolo (2004), modest means that leniency program never rewards for reporting, at best it cancels the fine.} occasional illegal transactions are enforceable if the following three conditions are satisfied simultaneously. First, the no-reporting condition for the bureaucrat: the reduced fine must be such that the bureaucrat prefers performing Action $a$ to Denouncing once the bribe has been paid. Second, the credible-threat condition for the entrepreneur: reduced fine and full fine must be set such that the entrepreneur can credibly threaten to report if the bureaucrat does not deliver. Third, the
credible-promise condition: the entrepreneur must be able to credibly promise not to report if the bureaucrat obeys to the illegal agreement.

These three conditions, given the value of the project together with full and reduced fines, define a bribe range for which the occasional illegal transaction is implementable. Even though these conditions are modified after the introduction of probability $\beta$ in generalized model, the qualitative result remains unaffected.

We used the generalized version of the game for experimental testing of the theoretical prediction under two different scenarios: when the occasional illegal transaction is implementable in equilibrium, and when it is not. Implementability is a function of per-round endowment for the entrepreneur. Per-round endowment exogenously defines the value of the bribe\textsuperscript{10} if the entrepreneur decides to pay it. For each treatment we use two possible values of the per-round endowment: low endowment which theoretically leads to no-corruption equilibrium, and high endowment which theoretically leads to corruption equilibrium.

Following up Abbink & Hennig-Schmidt (2006), we study whether loaded instructions in a bribery experiment affect the behavior of subjects in a lab. For that purpose, we designed 2 treatments: a benchmark (B) and a full-context (C) treatment.\textsuperscript{11}

Table 1 below summarizes the parameterization chosen for both treatments.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$v$</th>
<th>$RF_E$</th>
<th>$RF_B$</th>
<th>$F_E$</th>
<th>$F_B$</th>
<th>$E_L$</th>
<th>$E_H$</th>
<th>show-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>B, C</td>
<td>0.1</td>
<td>0.2</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>300</td>
<td>300</td>
<td>20</td>
<td>40</td>
<td>300</td>
</tr>
</tbody>
</table>

Table 1: Experimental parameterization. $\alpha$ and $\beta$ denote probability of audit and of discovering evidence of bribery, respectively; $v$ denotes value of the project to the entrepreneur, $RF_E$ and $RF_B$ denote reduced fines and $F_E$ and $F_B$ full fines to the entrepreneur and to the bureaucrat, respectively; $E_L$ and $E_H$ low and high per-round endowment; and show-up stands for the show-up fee.

The probabilities $\alpha$ and $\beta$ were chosen such that they approximately correspond to the field exogenous probability of audit and to the field conviction rate; and at the same time they are intuitively comprehensible for subjects. Value of the project $v$ was chosen together with full fines $F_E$ and $F_B$ such that subject face a considerable gain from the investment but also severe punishment in case of detection. We set reduced fines $RF_E$ and $RF_B$ equal to zero to analyze the case of full leniency.

\textsuperscript{10}This way we reduce the cognitive demand on subjects, the only decision they have to make is whether they want to transfer their per-round endowment or not.

\textsuperscript{11}We also conducted two exploratory sessions of a partial context treatment (C- treatment), where we only provided context on types of roles. In this treatment Participant X was called "Entrepreneur" and Participant Y "Bureaucrat”. All actions and realizations of random outcomes were denoted by neutral letters, as in the B treatment. We do not report these data in the main text as it is not possible to control for subjects’ interpretation the game in this case and therefore it is hard to recognize all possible effects in this treatment. Some results from this treatment are discussed in the appendix.
programs which, according to Apesteguia et al. (2004), have shown promising anti-cartel properties. Endowment determines the value of bribe to be (or not) paid. The ”low endowment” of 20 leads (theoretically) to no-corruption, whereas the ”high endowment” of 40 leads to corruption equilibrium. Finally, the show-up fee was set such that we eliminate the possibility of negative total earning from the experiment.

The parameterization does not differ among B and C treatments as we are interested purely in the effect of the instructions’ framing.

Extended forms of the game together with expected payoffs resulting from chosen parameterization are illustrated in Figure 2 for low- and for high-endowment periods, respectively. The branches identifying the equilibrium choices of risk-neutral agents are bolded.

![Figure 2: Expected payoffs from the corruption game.](image)

The instructions for the B treatment were presented in completely context-free fashion. Subjects were called Participant X and Participant Y, actions were denoted by neutral letters and realization of ”detection” or ”no detection” as ”outcome A” or ”outcome B”, respectively.

The instructions for the C treatment were presented in full context. Subjects were called an ”Entrepreneur” and a ”Bureaucrat”; actions were called ”Denounce”, ”do Nothing” and ”Provide the favor”; and realizations of random outcomes were called ”corruption has been detected” and ”corruption has not been detected”, respectively. Figure 3 below provides the comparison of wording for two treatments, neutral wording in the upper row and loaded below.
Figure 3: Neutral vs. loaded instructions wording. For each branch, the upper line provides the neutral labels of the B treatment (bolded); below are the loaded labels of the C treatment.

3 Implementation

The experiment was conducted in November and December 2006 at CERGE-EI in Prague, using a mobile experimental laboratory.\(^{12}\)

Participants were recruited from the Faculty of Social Sciences of the Charles University in Prague, from different faculties of the Czech Technical University in Prague and of the University of Economics in Prague. Students were approached via posters distributed in campus and via e-mails. By e-mail, we also directly invited students from these schools who participated earlier in unrelated experiments conducted at CERGE-EI.

We conducted four sessions of each treatment. Twelve participants, six in the role of Participant X – or the entrepreneur – and six in the role of Participant Y – or the bureaucrat – interacted in each session. In each session, all subjects participated for six rounds during which they kept the role which was assigned to them at the beginning of the first round.\(^{13}\) Participants were anonymously re-matched so that no subject was matched twice with the same co-player. This was common knowledge. The incentive compatibility of this matching scheme is discussed in Kamecke (1997).

Table 2 summarizes some of the demographic characteristics of subjects partic-

\(^{12}\)http://home.cerge-ei.cz/ortmann/BA-PEL.htm

\(^{13}\)After each Participant X interacted exactly once with each of the Participants Y, the roles were switched for another six rounds. Subjects were not informed about the switch of roles in advance in order to avoid possible impact on their behavior in the first six rounds. Before the beginning of the seventh round the announcement about the switch of roles appeared on their screens. Decisions in the last six rounds are likely to be affected by subjects' experience from the first six rounds and therefore we do not report them in the main text. Comparison of the pre-switch and after-switch data is provided in the appendix.
ipating in the experiment. The majority of our subjects are male, reflecting the composition of the subject pools that we drew on. Mean age ranges between 20.9 and 23, over all sessions the minimum is 18 and maximum 29. We also measured subjects’ risk aversion using a questionnaire based on Holt & Laury (2002). Mean RA score in the sample ranges between 26.4 and 34.7, over all sessions the minimum is 13 and maximum 51.\(^{14}\) Average final payoffs for the B treatment ranges from 320 to 330, with the minimum over four sessions of 300 and maximum of 400; for the C treatment it ranges between 315 and 340, with the minimum of 300 and maximum of 400.\(^{15}\)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Subject Source (^{16})</th>
<th>M/F ratio (^{17})</th>
<th>mean (age)</th>
<th>mean (RA score)</th>
<th>mean (final pay)</th>
<th>Irreg (^{18})</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>FSS</td>
<td>8/4</td>
<td>20.9</td>
<td>29.7</td>
<td>320</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>FSS</td>
<td>10/2</td>
<td>21.75</td>
<td>28.8</td>
<td>330</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>CTU</td>
<td>11/1</td>
<td>22.9</td>
<td>34.7</td>
<td>330</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>FSS</td>
<td>9/3</td>
<td>22.3</td>
<td>26.4</td>
<td>323.3</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>CTU</td>
<td>9/3</td>
<td>21.9</td>
<td>33.7</td>
<td>340</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>UE</td>
<td>7/5</td>
<td>22.9</td>
<td>28</td>
<td>318.3</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>CTU</td>
<td>10/2</td>
<td>23</td>
<td>31.4</td>
<td>318.3</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>UE</td>
<td>7/5</td>
<td>21.7</td>
<td>28.1</td>
<td>315</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2: Summary demographic characteristics of subjects.

Each session begun with general instructions. Afterwards, students were asked to fill-in Risk-aversion and Demographics questionnaires by which they earned their show-up fee. Then the instructions to the computerized part of the experiment were distributed. Understanding of the instructions was tested by a brief questionnaire (Keren, Willemsen 2008). The computerized part of the experiment started only after every participant answered all testing questions correctly.\(^{19}\) Each session concluded with a questionnaire asking for subject’s feedback on the experiment.\(^{20}\)

\(^{14}\)The higher the score the more risk averse the subject is. The maximum possible RA score is 60 which, using standard CRRA utility function \(x^{1-r}\), approximately corresponds to a relative risk aversion coefficient of .17. The minimum possible RA score is 0, which approximately corresponds to a relative risk aversion coefficient of -.13. A RA score of 23 corresponds to risk-neutrality.

\(^{15}\)Subjects were informed during the recruitment that there is a chance that their final payoff from the experiment will be zero, but never negative.

\(^{16}\)For each session, subjects were recruited from one source. FSS stands for the Faculty of Social Sciences in Prague, CTU for the Czech Technical University in Prague, UE for the University of Economics in Prague.

\(^{17}\)Male/Female ratio in the session.

\(^{18}\)Irreg stands for a dummy variable for session irregularities – identifies any unusual activities by subjects or any irregularities on the experimenter’s side. In the first B treatment session it is possible experimenter effect; in the second C treatment session, one of the subjects reports “building engineering” as a field of study in a demographics questionnaire, which may mean that CTU student participated in UE session. We do not believe that they matter but wanted to control nonetheless.

\(^{19}\)This was common knowledge.

\(^{20}\)For filling this last questionnaire, subjects were paid an additional 50-200 CZK (corresponds to about 2-9 USD) - the amount varied between sessions. This mechanism was used to adjust average earnings per session to levels promised during recruitment.
All instructions were read aloud by the experimenter. As a part of the instructions subjects received a pictorial representation of the game with minimum use of game-theoretic terminology. Probabilistic outcomes were presented in both, probabilistic terms and frequency representation (see e.g. Gigerenzer & Hoffrage 1995, or Hertwig & Ortmann 2004).\textsuperscript{21}

The experiment was computerized using Z-tree software (Fischbacher 2007). At the beginning of each round, each participant was notified of her/his role. Participants X also learned current per-round endowment. Afterwards, each pair interacted sequentially.\textsuperscript{22} Between the second and third stage, Participants X were asked what would be their choices in each node of the third stage if they would reach it. After they made their conditional choices, they learned actual decision of their co-player and they were asked to confirm, or to change, their previous choice. This mechanism allowed us to collect some additional data also in rounds when the third stage was in fact not reached.

At the end of each round subjects were given feedback about their action(s), action(s) of the player they were paired with, realization of the random outcome (detection vs. no detection; or outcome A vs. outcome B, respectively) and their resulting payoff. At the end, one round was randomly chosen to determine the final payoff from the computerized part of the experiment. This mechanism was chosen in order to ensure that decision in every round is made as if in a one-shot game.

Participants were paid anonymously in cash right after each session. We used the Czech crown as a currency unit throughout the whole experiment.

4 Results

In Figure 4, the results from low- and high-endowment periods, respectively, are presented. Each figure integrates the results from both treatments – the B treatment data in the upper rows and C treatment data below. The equilibrium choices for each case are bolded.

From the aggregate first stage data we observe surprisingly small differences between the two treatments. Moreover, the difference is just the opposite to what we expected – the frequencies of choosing \textit{Pay} are somewhat higher in the C treatment. In addition, in both treatments, the frequencies of choosing \textit{Pay} are higher in the low-endowment periods than in the high-endowment periods, which is in contradiction with the theoretical prediction. Intuitively, subjects seem to be willing to transfer their endowment in order to get a chance of receiving high payoff, but they are more willing to put at stake a low endowment than a high one. Instead of risking the high endowment they seem to prefer choosing sure outcome.

\textsuperscript{21}Originals (in Czech) of all materials that subjects received during the experiment are available at http://home.cerge-ei.cz/richmanova/WorkInProgress.html.

\textsuperscript{22}Choices were made by clicking the respective buttons on the screen. Subjects were notified that once they make their choice it would not be possible to take it back.
As to the second stage data, it is only relative percentages that can be compared across treatments, as different number of subjects in fact entered this stage of the game. In the B treatment, it is an equal split between playing Denounce or Action a for both, low- and high-endowment periods. The difference in expected payoffs resulting from Denounce and Action a is however very small and that may be the reason why we do not observe stronger inclination to either choice. Also note that in both treatments Denounce is the only action that ensures a non-negative payoff.

In the C treatment, choices are shifted in favor of Denounce, which in the high-endowment periods is in contradiction with theoretical prediction but is in line with our prediction – knowing the context behind their action choice, reporting corruption might be becoming more attractive for subjects. This is one of the interesting results, suggesting that context indeed does make a difference in a bribery experiment.

In line with the theoretical prediction and also intuition, Nothing23 was almost never chosen.

As to the third stage data, conditional choices provide mixed evidence. In both treatments, subjects seem to prefer playing Nothing in either case. This is for E₁ stage in contradiction, and for E₂ stage in line with the theoretical prediction.

---

23Payoffs for Participant Y resulting from Nothing and Action a are the same, but taking into account likely decisions of Participant X in consecutive stage, he is more likely to collect higher payoff after choosing Action a.
When we look at the sequential choices (in red), the results are in line with the theoretical prediction for both treatments – but only as much as we can tell from this few observations. \(^{24}\) We observe basically no framing effect for high-endowment periods. For low-endowment periods, we observe a small shift in favor of *Denounce*, which is in line with our expectations.

Note that for the second and the third stage data we have too little independent observations (especially so for the high-endowment periods)\(^ {25}\) to perform reliable formal analysis. Therefore, we only perform statistical and regression analysis of the first stage data.

### 4.1 Analysis of the first stage data

In the following two subsections we report the results from the formal analysis of the first-stage data. We performed standard non-parametric tests with the null hypothesis of no differences in distributions of transferring rate under two treatments. We also computed the effect size indices to measure the magnitude of the treatment effect. Finally, we report the results from the estimation of a linear probability model in which we control for some demographic characteristics of subjects.

Due to the panel nature of the data, we considered four different approaches to formal regression analysis: 1) clustered data analysis – data from periods 1, 3, 5 (low-endowment) and from periods 2, 4, 6 (high-endowment) are clustered by subject to control for likely within-subject correlation of choices; 2) first-period data analysis – only first-period data (for low-endowment case) and only second-period data (for high-endowment case) are analyzed; 3) averaged data analysis – averaged data for periods 1, 3, 5 and for periods 2, 4, 6 are analyzed; and 4) dominant-choice data analysis – for each value of endowment (low or high) each subject makes choices in three periods, dominant choice is the one which is played more often.

Clumped data have the advantage of using all the available information while controlling for likely correlation between subject’s choices, while the other three approaches use only a part of the information we have. Therefore, in the main text we discuss the results for clustered data. Analysis of averaged, first-period, and dominant-choice data can be found in the Appendix 2, part A, as a robustness check of the main results.

In addition to the robustness checks based on different “data handling”, we also

---

\(^{24}\)When we asked the subjects to make their real choices, in the B treatment, only one of them changed her/his decision in \(E_2\) stage from *Denounce* to *Nothing* (after observing what Participant Y has chosen) in low-endowment period. In the C treatment, three subjects changed her/his decision in \(E_2\) stage – from that two switched from *Nothing* to *Denounce* after Participant Y played *Action a* and one from *Denounce* to *Nothing* after Participant Y played *Action a* – and one subject changed her/his decision in \(E_1\) stage from *Nothing* to *Denounce* after Participant Y played *Nothing*. All four cases occurred in low-endowment periods.

\(^{25}\)Recall that Figure 4 presents aggregated data from all relevant periods, therefore contains repeated observations for individual subjects.
run a few exploratory sessions of treatments in which the experimental conditions are only slightly modified compared to the benchmark and the context treatment, respectively. The results from the analysis on the extended data set is provided in the Appendix 2, part B, as an additional robustness check of the main results.

4.1.1 Statistical analysis

In the Table 3 below we report the results of three standard non-parametric tests in order to identify the differences in distributions of choices under two treatments. Specifically, we test a null hypothesis of no differences between the two treatments using averages of binary transfer variable\(^{26}\) over periods 1,3,5 and 2,4,6, respectively. According to all three, Wilcoxon rank-sum, Kolmogorov-Smirnov, and Fisher’s exact test, we cannot reject the hypothesis of no differences in distributions of choices under the two treatments at the 5% significance level.

<table>
<thead>
<tr>
<th>periods</th>
<th>Ranksum(^{27})</th>
<th>Ksmirnov(^{28})</th>
<th>Fisher(^{29})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,3,5</td>
<td>-0.596 (.599)</td>
<td>.083 (.846)</td>
<td>(.947)</td>
</tr>
<tr>
<td>2,4,6</td>
<td>-0.715 (.475)</td>
<td>.167 (.513)</td>
<td>(.218)</td>
</tr>
</tbody>
</table>

Table 3: Non-parametric tests.

To assess the magnitude of the effect for practical purposes, we, in addition, compute two standardized measures of effects size: Cohen’s d and odds ratio, again, using averages of binary transfer variable over periods 1,3,5 and 2,4,6, respectively. The results for the full sample and for the male and female subsamples are reported in Table 4 below.

<table>
<thead>
<tr>
<th>Periods</th>
<th>Sample</th>
<th>B</th>
<th>C</th>
<th>effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>mean</td>
<td>std.dev.</td>
<td>N</td>
</tr>
<tr>
<td>1,3,5</td>
<td>full</td>
<td>24</td>
<td>.528</td>
<td>.4495</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>18</td>
<td>.519</td>
<td>.4464</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>6</td>
<td>.566</td>
<td>.5018</td>
</tr>
<tr>
<td>2,4,6</td>
<td>full</td>
<td>24</td>
<td>.222</td>
<td>.3764</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>18</td>
<td>.296</td>
<td>.4105</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4: Effect-size indices.

\(^{26}\)Transfer has value of one if Participant X chose Pay and value of zero if s/he chose Not Pay in respective period.

\(^{27}\)Ranksum stands for the two-sample Wilcoxon rank-sum (or Mann-Whitney) test. We report the normalized z statistic and corresponding p-value below.

\(^{28}\)Ksmirnov stands for the Kolmogorov-Smirnov test. We report the statistic and below the corresponding p-value from testing the hypothesis that average transfer is lower in the B treatment.

\(^{29}\)Fisher stands for the Fisher’s exact test. We report the resulting p-value.

\(^{30}\)Division-by-zero problem occurs, due to no variation in this subsample.
Cohen (1998) defines effect sizes of $d = 0.2$ as small, of $d = 0.5$ as medium, of $d = 0.8$ as large. For the full sample, the results suggest only small effect. However, when we look at the male and female subsamples separately, the effect size appears larger than in the full sample. It is also noticeable that the effects for the male and for the female subsamples have the opposite directions, which naturally results in a very small total effect. We observe very similar results when looking at the odds ratio – the effect is smaller in the full sample than it appears in the two subsamples. These results suggest a non-negligible gender effect.

Altogether, both, statistical tests and effect-size measures, suggest that there are only minor differences between the first-stage choices in C and B treatments. Effect-size measures for the male and female subsamples suggest that it might be caused by counteracting gender effects. Therefore, further analysis which would control for gender and for other subjects’ characteristics is called for.

4.1.2 Econometric analysis

In this section we report the results from econometric analysis controlling for some of the subjects’ characteristics and for the treatment effect.

During the experiment we distributed several questionnaires in order to collect basic demographic data. Specifically, we have information about subjects’ age, gender, university and field of study. We also measure each subject’s risk aversion. All these characteristics provide us with data that could potentially help to explain subjects’ decisions.

Dependent variable was defined as a 0/1 dummy variable $translog$ identifying $Pay$ being chosen (value of 1) or not (value of 0) in particular period. In order to make use of all observations (over all relevant periods) and at the same time to control for likely dependence of subject’s observations over periods we estimate a clustered linear probability model. We prefer a linear probability model to other non-linear alternatives, as it does not rely on very specific distributional assumptions, violation of which leads to inconsistent estimates if non-linear models are employed. Another advantage of the linear probability model lies in straightforward interpretation of estimated coefficients.

In the appendix, we provide a discussion of the robustness checks we conducted in addition to the clustered regressions analysis. As the theoretical prediction.

---

Footnote 31: In addition, we collected the data as: size of subject’s household, number of cars in the household, whether the subject himself has his own car and what is its approximate value, all of which serve as a proxy for income. We also asked the subjects whether they considered themselves being technical type compared to their peers. We recorded occurrence of any inconsistencies in the after-instructions questionnaire, which served a simple test of understanding the basic structure of the game, and in the risk-aversion questionnaire. At the end of the session we asked our subjects whether they did understand the experiment. Finally, we recorded some general information about each session – time of day when it started and any session irregularities if they occurred. After running some preliminary regressions we, however, conclude that none of these variables is significant for explaining subjects’ decisions.
differs for low-\textsuperscript{32} and for high-\textsuperscript{33} endowment periods, these two groups were analyzed separately.

We start with a basic minimal model:\textsuperscript{34}

\[
P(\text{trans log} = 1 | x) = \beta_0 + \beta_1 \cdot \text{age} + \beta_2 \cdot \text{male} + \beta_3 \cdot \text{econ} + \beta_4 \cdot \text{Btreat}
\]

where \textit{age} corresponds to subject’s reported age, \textit{male} is a dummy variable defined based on subject’s reported gender, and \textit{econ} is a dummy variable identifying subject having (value of 1) or not (value of 0) economic background which is defined based on subject’s reported field of study. As we are mainly interested in the treatment effect, we also include B-treatment dummy \textit{Btreat} in the model.

The results from the estimation are summarized in Table 3, denoted as Model 1. This model is, however, not significant. Therefore, based on the results from effect-size computations, we extend the basic minimal model by interaction terms with \textit{male} to allow for gender-specific effects. This leads to Model 2:

\[
P(\text{trans log} = 1 | x) = \beta_0 + \beta_1 \cdot \text{age} + \beta_2 \cdot \text{male} + \beta_3 \cdot \text{econ} + \beta_4 \cdot \text{Btreat} + \\
+ \beta_5 \cdot \text{male} \times \text{age} + \beta_6 \cdot \text{male} \times \text{econ} + \beta_7 \cdot \text{male} \times \text{Btreat}
\]

The results from the estimation of Model 2 are also summarized in Table 3.

Model 2 is strongly significant. This confirms strong gender effect. Therefore, in the discussion that follows, we will concentrate on the results from Model 2.

For both, low- and high-endowment-period data, the joint p-value of the model is .000. All demographic characteristics – \textit{age}, \textit{male}, and \textit{econ} – and their interaction terms are significant on 5% level. Interestingly, treatment dummy together with its interaction term is only significant for the low-endowment periods, not for high. This suggests that only for the low-endowment periods the presentation of the game matters. Recall, however, that from descriptive data, the transferring rates for high-endowment periods are considerably lower than for low, and so for both treatments.

\textsuperscript{32}Recall that in periods 1,3, and 5 the endowment was low.

\textsuperscript{33}Recall that in periods 2,4, and 6 the endowment was high.

\textsuperscript{34}Second approach we used was \(P(\text{trans log} = 1 | x) = \beta_0 + \beta_1 \cdot \text{ra_score}\), where \textit{ra_score} is a risk aversion score computed based on data from the risk-aversion questionnaire. As after the preliminary analysis we concluded that age, male and econ predict \textit{ra_score} relatively well (all three jointly significant on 5% level, age and male with negative sign on coefficient, age with positive; our proxy for income appeared insignificant, which is reasonable given our population sample), it was natural to consider these two sets of independent variables - one including \textit{ra_score} only, and the other including male, age and income - as a candidates for minimal models for our analysis. However, in \(P(\text{trans log} = 1 | x) = \beta_0 + \beta_1 \cdot \text{ra_score}\) \textit{ra_score} never appeared significant and only rarely we observed joint significance of estimated models. Therefore, we omit the discussion of these results.
### Table 5: Results from estimation of the linear probability model(s). The first row of each cell reports estimated coefficients. The second row reports the corresponding p-value. Mean $\hat{p}(y=1)$ denotes mean predicted probability of transfer being made.

Mean predicted probability of transfer in the low-endowment periods is 0.56, in the high-endowment periods it is only 0.24 which is considerably lower. This result is in contradiction with the theoretical prediction.\(^{35}\)

First, we will discuss in more detail the regression results for the low-endowment-periods.

Age has a positive sign on the coefficient for female, but negative for male. This means that an additional year of age increases the probability of transfer for female, but reduces the probability for male subsample.

Econ has a negative sign on coefficient for both male and female. This means that having economic background shifts the probability of transfer closer towards the theoretical prediction for low-endowment periods. The negative effect of having economic background is greater for female (-0.63) than for male (-0.1) subjects.

Treatment dummy Btreat has a positive sign for female but negative for male subjects. This suggests negative impact of corruption context on transferring decision of women but positive impact for men, which is an intriguing result.

In addition, also the intercept is negative for women and positive for men. This means that with the same characteristics, women are less likely to make the transfer than men.

For illustration, consider an example of a twenty-two year old woman who is not studying economics and participates in the C treatment. Then the predicted probability that she will make the transfer is 0.641. For a man with the same

\(^{35}\)Recall that in the equilibrium Participant X always transfers high endowment and never transfers low.
characteristics, the predicted probability is .714, which is greater by .073. If we look at a woman of the same age who is studying economics than the probability of transfer drops dramatically to .01,\textsuperscript{36} whereas for men the drop is much smaller – to .618. The difference in the probability for men and for women with economic background is much greater – .608. In the B treatment, women are more likely to transfer than men (and than women in the C treatment). The effect of economic background is similar as in the C treatment – it reduces the probability of transfer and more so for women than for men.

Second, we will look in more detail at the regression results for the high-endowment-periods.

Similarly as for periods 1, 3, 5, \textit{age} has a positive sign on the coefficient for female, but negative for male. Thus an additional year of age increases the probability of transfer for female, but reduces the probability for male subjects also in high-endowment periods.

The effect of having economic background is, however, different for high-endowment periods. \textit{Econ} has a negative sign on coefficient only for men. For women it is positive. This means that having economic background shifts the probability of transfer closer towards the theoretical prediction only for female subjects.

Treatment dummy \textit{Btreat} has a positive sign for both female and male subjects, which is yet another difference from low-endowment periods. This suggests negative impact of corruption context on transferring decision – in high-endowment periods subjects are less likely to transfer when they are fully aware of context. Note, however, that even though the sign reflects the expected impact of context, the coefficient is not significant.

Similarly as for low-endowment periods, the intercept is negative for women and positive for men. Thus, also when the endowment is high, having the same characteristics, women are less likely to make the transfer than men.

For illustration, consider again a twenty-two year old woman who is not studying economics and participates in the C treatment. Then the predicted probability that she will make the transfer is $-1.43$.\textsuperscript{37} For a man, the predicted probability is .339, which is greater by .482. If we look at a woman of the same age who is studying economics than the probability of transfer increases to .078, whereas for men it drops to .22. The difference in the probability for men and for women with economic background is now smaller – .142. Here the effects of context and of economic background do not impact as consistently as in low-endowment periods. In addition, the treatment effect is not significant.

Finally, in Figures 5 and 6, for illustration of the above discussed results, we

\textsuperscript{36}We, however, need to keep in mind that number of women in the sample is considerably smaller than number of men.

\textsuperscript{37}This particular combination of values of the explanatory variables however does not occur in our sample. Negative predicted probability is a problem which sometimes occurs when using linear probability models.
provide the summary data separately for men and for women.

Figure 5: Experimental results for male and for female in low endowment periods. Above each branch of the extensive form of the game, the upper row displays the frequency of respective action being chosen in the B treatment; and lower row displays the frequency of respective action being chosen in the C treatment (both with the corresponding relative percentage in parentheses). For stages $E_1$ and $E_2$, above the branches, we present the conditional choices subjects were asked about before making real sequential choice. Frequencies of real choices, which depend on preceding decision of Participant Y, are presented at the bottom part of each figure (in red).

For low-endowment periods, in the first stage of the B treatment the difference in behavior of men and of women does not appear substantial – slightly more than half of each makes the transfer. However, in the C treatment, the transferring decisions of male and of female shift in opposite directions – two thirds of male, whereas less than a half of women decide to make the transfer. This suggests that the corruption framing affects men and women differently.

Similarly in the second stage, we can clearly see from the descriptive data that facing a full context, women become much more likely to report. Men’s decisions seem to remain unaffected.

The results from the last stage are not so clearly distributed. In $E_1$ stage we observe the opposite effects of context on men than on women. In $E_2$ stage, the direction of the effect does not vary with gender. In general, both, men and women, prefer doing Nothing to Denouncing.

In the first stage of the high-endowment periods, the results are somewhat different. We still observe considerably more women to refrain from making transfer but the framing effect seems to increase the transferring rate. Recall, however, that the results from the regression analysis suggest that these four observations might be just random realization. We observe almost no framing effect in the male subsample. In general, both, male and female, prefer not making the transfer.
In the second stage, the female subsample in the role of Participant Y is very small. In both treatments, all the women choose *Denounce*. For the male subsample, we observe some (possible) treatment effect, which shifts the choices more in favor of playing *Denounce* in the C treatment.

In the third stage percentage of men choosing *Denounce* slightly decreases with framing, for female it goes slightly up. In both subsamples, the prevailing choice is doing *Nothing* though.

Opposite directions of the framing effect for men and for women we observe in some nodes explain why allowing gender-specific coefficients improved the econometric model.

5 Conclusion

Overall, some of the results confirm our expectations whereas some do not.

In aggregate data, we find only small and statistically insignificant treatment effect, which is in line with Abbink & Hennig-Schmidt (2006) but not with our expectations. Once we look at the male and female subsamples separately, we discover (significant) counteracting gender effects, which are responsible for reduced overall effect of wording.
For the aggregate second-stage data, the treatment effect shows in increased denouncing rate, which is in line with our prediction. For male and female subsamples, as much as we can tell, given low number of observations, denouncing rates are lower or same\textsuperscript{38} in the B compared to the C treatment. Also for the aggregate third-stage data the treatment effect goes in predicted direction.

Gender effect is not a novel result. For example Alatas, Cameron, Chaudhuri, Erkal and Gangadharan (2006) find significant differences in the behavior of men and women in a corruption experiment. Their result, however, appears to be culture-specific.\textsuperscript{39}

The observed negative impact of corruption context on transferring decision of women together with positive impact on denouncing decision of women are in line with earlier findings of women being less likely to engage in, as well as less tolerant of (thus more likely to act against), corruption than men. Such results have been documented, for example, by Swamy, Knack, Lee and Azfar (2001), or Dollar, Fisman and Gatti (2001).

For low-endowment periods, we find positive impact of bribery context on transferring rates of mens. This suggest the opposite treatment effect to what we expected, but only for the male subsample. Women react to context by reduced transferring. The (significant) result for male subsample is very surprising and difficult to understand.

For high-endowment periods, treatment effect appears insignificant. We find (slightly) reduced transferring rate for male and (slightly) increased transferring rate for female subsample. The result for female subsample is counterintuitive, however, the results of the t-test suggest that it might be only due to random realization.

Another interesting result is that for both low- and high-endowment periods more than 50% of subjects does not play equilibrium. Recall that theoretically, the optimal strategies are to transfer when the endowment is high and to not transfer when the endowment is low. For both treatments we observe just the opposite – relatively high transferring rates for low- and relatively low transferring rates for high-endowment periods.

This phenomenon, in addition, appears robust. In Richmanova (2008) the same result is found for the high-incentives treatment.\textsuperscript{40}

There are several possible explanations. One of them might root in a so called "preference for inclusion" reported, for example, by Cooper & Van Huyck (2003). They find that subjects presented with an extensive form game, are significantly more likely to make choices that allow their co-player to make a choice – and thereby

\textsuperscript{38}This refers to all possible cases, when we are looking separately at male and female subsamples for high- and for low-endowment periods.

\textsuperscript{39}The authors run the experiment in Melbourne (Australia), Delhi (India), Jakarta (Indonesia), and Singapore. Only Australian data confirm significant gender effect.

\textsuperscript{40}Recall that the benchmark treatment is the same for both, the present paper and Richmanova (2008).
to affect final payoffs – rather than choosing a terminal node. In an extensive form 
game this "]\(\)\(\)inclusion\] is more salient. In our game, "inclusion" introduces 
a risk of significant loss. Altogether, it might have resulted in subjects with the "]\(\)\(\)preference for inclusion\] willing to transfer and continue playing the game, but only 
willing to risk the low endowment. High endowment they might have preferred to 
keep with certainty.

Furthermore, nature of the game implies that the endowment has explicit payoff 
consequences in the second and third stage of the game only for Participant Y, not 
for Participant X. Therefore, strategic importance of the endowment level might 
have been less obvious to Participants X than we thought.

Finally, it is important to note that the theoretical prediction is computed under 
the assumption of risk neutrality, which, as also suggested by the data from the 
risk-aversion questionnaire, is not likely to hold in our sample.

Altogether, our results suggest that context indeed plays an important role 
for subject’s behavior in a bribery game. More importantly, the effect on male 
participants might be very different than the effect on female participants. Some of 
our results are not significant, but this might be caused by relatively small sample 
and gender-unbalanced subject pool. With more subjects, possibly observed over 
more periods, and with better gender-balanced sample, our results might become 
more conclusive.

\[\text{\footnotesize Ortmann and Tichy (1999) also report some evidence of differences in (cooperative) behavior of men and women. Also gender composition of subject pool in experimental sessions matters. When controlling for past experience, gender differences, however, disappear.}\]


Richmanova (2008), Experimental Testing of Leniency Programs: The Impact of Change in Parameterization, in prep.


Spagnolo, G., (2004). Divide et Impera: Optimal Leniency Programs, *C.E.P.R. Disc...


APPENDIX 1

Comparing the data from periods before and after the switch of roles.

In Figure 7, we present the data from pre- and after-the-switch-of-roles periods for low and for high endowment of the **B** treatment, respectively.

In both cases, we observe somewhat higher transferring rate in the second six periods. Similarly as in the first part of the experiment, the transferring rate is higher in periods when the endowment is low than when it is high. In **B_0** stage, more subjects were choosing safe option (with no possibility of loss) after the switch of roles. This means, for low-endowment periods a shift towards, but for high-endowment periods shift further away from the theoretical prediction. In **E_2** stage, results from pre- and after-switch data are very similar and for both, low and high endowment, they are in line with the theoretical prediction. In **E_1** stage, we observe a shift towards equilibrium after the switch of roles.

In Figure 8 below, we present the data from pre- and after-the-switch-of-roles periods from low- and high-endowment periods of the **C** treatment, respectively.

In the C treatment, transferring rate drops after the switch of roles, more so in periods when endowment is high. This is just an opposite effect as in the B treatment. The transferring rate is higher when endowment is low in both cases, before and after the switch of roles, which contradicts the theoretical prediction. In **B_0** stage, a higher fraction of subjects was choosing safe option (with no possibility of loss) after the switch of roles. This is a similar result as in the B treatment – for low-endowment periods it means a shift towards, but for high-endowment periods shift further away from the theoretical prediction. In **E_1** and **E_2** stage, the results
from pre- and after-switch data are similar for low-endowment periods (more so in $E_1$ than in $E_2$) stage. In high endowment periods we observe no difference at all.

Figure 8: Pre- vs. after-the-switch-of-roles data in the C treatment. Pre-switch data are in the upper rows and after-switch data are below.
APPENDIX 2

Robustness checks

We performed two types of robustness checks of our estimation results. The first regards the way we treated individual observations over rounds when running regressions – this is discussed in subsection Handling of the data. The second regards the experimental design – we also run several sessions of alternative treatments that were shown not to affect the behavior of subjects significantly – this is discussed in subsection Pooling the sessions.

A. Handling of the data

Throughout the analysis we have defined three alternative dependent variables each of which captures slightly different information about the first-stage data.

translog – as described in the main text, it is a 0/1 dummy variable identifying transfer being made (value of 1) or not (value of 0) in particular period.

atranslog – is average value of translog for one individual over periods 1, 3, 5 (low-endowment periods) or 2, 4, 6 (high-endowment periods).

ltranslog – defines a dominant choice of a subject in periods 1, 3, 5 or 2, 4, 6. For a subject who has chosen Pay two or three times out of total three periods of interest, the dominant choice is 1; for a subject who has chosen Not Pay two or three times out of total three periods of interest, the dominant choice is 0.

Then, using one of the three types of dependent variable, we conducted four different types of regression analysis.

Clustered regressions – as discussed in the main text, we run clustered (robust) linear probability model estimation with binary variable translog as a dependent variable. Using this approach we control for subject’s unobservable characteristics likely causing dependence of choices among periods and, at the same time, we are using full information from the data.

Regressions on Averaged data – in this case, we run ordinary least squares estimation of atranslog. In this case we analyze only averaged data, where higher values of atranslog correspond to more transfers being made and thus to stronger preference for this choice.

Regressions on the 1st- or 2nd-period data – we estimate LPM only on the 1st- and 2nd-period translog (for low- and high-endowment periods, respectively). In this approach we are omitting part of the information, however we only use part of the data that is not affected by the experience from previous rounds.42

Regressions on Dominant Choice – we estimate LPM using ltranslog as a dependent variable. Thus in this case, we are only looking at the dominant choice of each subject.

42We realize that for 2nd-period data it may not be completely true in case subjects fail to realize that it is a different game they are playing in the high-endowment periods.
First we look at effect size measures, whether they give consistent results for all
four approaches to the data. The results are summarized in the Table 6 below.

<table>
<thead>
<tr>
<th>Data</th>
<th>B</th>
<th></th>
<th>C</th>
<th></th>
<th>effect size</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>std.dev.</td>
<td>mean</td>
<td>std.dev.</td>
<td>odds ratio</td>
<td>Cohen’s d</td>
</tr>
<tr>
<td>1,3,5</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;-period</td>
<td>.583</td>
<td>.5036</td>
<td>.625</td>
<td>.4945</td>
<td>1.072</td>
</tr>
<tr>
<td></td>
<td>average</td>
<td>.528</td>
<td>.4496</td>
<td>.597</td>
<td>.4282</td>
<td>1.131</td>
</tr>
<tr>
<td></td>
<td>dominant</td>
<td>.5</td>
<td>.5108</td>
<td>.583</td>
<td>.5036</td>
<td>1.166</td>
</tr>
<tr>
<td></td>
<td>all periods</td>
<td>.528</td>
<td>.5027</td>
<td>.597</td>
<td>.4939</td>
<td>1.131</td>
</tr>
<tr>
<td>2,4,6</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;-period</td>
<td>.292</td>
<td>.4643</td>
<td>.25</td>
<td>.4423</td>
<td>0.856</td>
</tr>
<tr>
<td></td>
<td>average</td>
<td>.222</td>
<td>.3764</td>
<td>.25</td>
<td>.3417</td>
<td>1.126</td>
</tr>
<tr>
<td></td>
<td>dominant</td>
<td>.25</td>
<td>.4423</td>
<td>.25</td>
<td>.4423</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>all periods</td>
<td>.222</td>
<td>.4187</td>
<td>.25</td>
<td>.4361</td>
<td>1.126</td>
</tr>
</tbody>
</table>

Table 6: Effect-size indices.

In all cases, the effects are small (Recall that Cohen (1998) defines effect sizes
of $d = 0.2$ as small), for high-endowment dominant choice data it is zero (but we
need to keep in mind that only part of available information is used). Except of
2<sup>nd</sup>-period data, also the direction of effect is the same in all cases. This suggests
that initially, the transferring rate was lower for high-endowment periods in the
context treatment but in later periods it increased. When we look at the male
and female subsamples, the results are also consistent for all four approaches –
suggesting counteracting gender effect (we omit reporting all numbers here as they
are very similar to the results for averaged data reported in Table 4 in the main
text).

Tables 7 and 8 below summarize the main results from the estimation for low-
and high-endowment periods, respectively. For all four approaches, the models
which do not allow for gender-specific effects are not significant. Therefore in the
discussion that follows we will concentrate only on models containing interaction
terms.

For the low-endowment periods, the results from the averaged, 1<sup>st</sup>-period and
dominant-choice data analysis confirm the results from clustered regression. We
find directions of all the effects being the same, the explanatory variables are signi-
ficant in most cases and there are no dramatic differences in coefficients sizes.
Only econ and econ*male are not significant in the 1<sup>st</sup>-period data case. They
both become significant once we include the information from future rounds – for
clustered, averaged and dominant-choice data.

For the high-endowment periods, only the results form averaged and dominant-
choice data analysis confirm the results from clustered regressions – the treatment
dummy is not significant, neither is its interaction term, directions of all the effects
are the same, and sizes of the coefficients are comparable. For the 2<sup>nd</sup>-period data
we fail to get significant results. This suggests that the behavior in the second
period is different, harder to be explained by demographic characteristics. To be
able to say whether in later rounds the behavior really stabilizes, we would need to
observe more high-endowment periods.
<table>
<thead>
<tr>
<th></th>
<th>clustered</th>
<th>averaged</th>
<th>1st-period</th>
<th>dominant</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>-.0287</td>
<td>.1280</td>
<td>-.0287</td>
<td>.1280</td>
</tr>
<tr>
<td></td>
<td>(.302)</td>
<td>(.007)</td>
<td>(.317)</td>
<td>(.822)</td>
</tr>
<tr>
<td>male</td>
<td>.0686</td>
<td>4.1425</td>
<td>.0686</td>
<td>4.1425</td>
</tr>
<tr>
<td></td>
<td>(.646)</td>
<td>(.002)</td>
<td>(.656)</td>
<td>(.004)</td>
</tr>
<tr>
<td>econ</td>
<td>-.1601</td>
<td>-.6307</td>
<td>-.1601</td>
<td>-.6307</td>
</tr>
<tr>
<td></td>
<td>(.212)</td>
<td>(.000)</td>
<td>(.226)</td>
<td>(.000)</td>
</tr>
<tr>
<td>Btreat</td>
<td>-.0559</td>
<td>.7156</td>
<td>-.0559</td>
<td>.7156</td>
</tr>
<tr>
<td></td>
<td>(.657)</td>
<td>(.004)</td>
<td>(.666)</td>
<td>(.006)</td>
</tr>
<tr>
<td>age*male</td>
<td>-</td>
<td>-.1852</td>
<td>-</td>
<td>-.1852</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.002)</td>
<td></td>
<td>(.003)</td>
</tr>
<tr>
<td>econ*male</td>
<td>-</td>
<td>.5354</td>
<td>-</td>
<td>.5354</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.002)</td>
<td></td>
<td>(.004)</td>
</tr>
<tr>
<td>Btreat*male</td>
<td>-</td>
<td>-.7983</td>
<td>-</td>
<td>-.7983</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.006)</td>
<td></td>
<td>(.010)</td>
</tr>
<tr>
<td>const</td>
<td>1.2901</td>
<td>-2.1749</td>
<td>1.2901</td>
<td>-2.1749</td>
</tr>
<tr>
<td></td>
<td>(.048)</td>
<td>(.047)</td>
<td>(.055)</td>
<td>(.061)</td>
</tr>
<tr>
<td>mean p(y=1)</td>
<td>.5625</td>
<td>.5625</td>
<td>.5625</td>
<td>.5625</td>
</tr>
<tr>
<td># of obs.</td>
<td>144</td>
<td>144</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>joint p-value</td>
<td>.488</td>
<td>.000</td>
<td>.5193</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 7: Results from clustered regressions vs. regressions on averaged, 1st-period, and dominant-choice data from low-endowment periods.

<table>
<thead>
<tr>
<th></th>
<th>clustered</th>
<th>averaged</th>
<th>2nd-period</th>
<th>dominant</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>.0220</td>
<td>.0913</td>
<td>.0220</td>
<td>.0913</td>
</tr>
<tr>
<td></td>
<td>(.381)</td>
<td>(.000)</td>
<td>(.396)</td>
<td>(.000)</td>
</tr>
<tr>
<td>male</td>
<td>.1766</td>
<td>2.5498</td>
<td>.1766</td>
<td>2.5498</td>
</tr>
<tr>
<td></td>
<td>(.055)</td>
<td>(.008)</td>
<td>(.063)</td>
<td>(.012)</td>
</tr>
<tr>
<td>econ</td>
<td>-.0731</td>
<td>.2210</td>
<td>-.0731</td>
<td>.2210</td>
</tr>
<tr>
<td></td>
<td>(.503)</td>
<td>(.001)</td>
<td>(.516)</td>
<td>(.002)</td>
</tr>
<tr>
<td>Btreat</td>
<td>-.0230</td>
<td>.0375</td>
<td>-.0230</td>
<td>.0375</td>
</tr>
<tr>
<td></td>
<td>(.809)</td>
<td>(.644)</td>
<td>(.815)</td>
<td>(.663)</td>
</tr>
<tr>
<td>age*male</td>
<td>-</td>
<td>-.0941</td>
<td>-</td>
<td>-.0941</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.032)</td>
<td></td>
<td>(.043)</td>
</tr>
<tr>
<td>econ*male</td>
<td>-</td>
<td>-.3395</td>
<td>-</td>
<td>-.3395</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.019)</td>
<td></td>
<td>(.026)</td>
</tr>
<tr>
<td>Btreat*male</td>
<td>-</td>
<td>-.0036</td>
<td>-</td>
<td>-.0036</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.983)</td>
<td></td>
<td>(.984)</td>
</tr>
<tr>
<td>const</td>
<td>-.3170</td>
<td>-2.1445</td>
<td>-.3170</td>
<td>-2.1445</td>
</tr>
<tr>
<td></td>
<td>(.556)</td>
<td>(.000)</td>
<td>(.568)</td>
<td>(.000)</td>
</tr>
<tr>
<td>mean p(y=1)</td>
<td>.2361</td>
<td>.2361</td>
<td>.2361</td>
<td>.2361</td>
</tr>
<tr>
<td># of obs.</td>
<td>144</td>
<td>144</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>joint p-value</td>
<td>.078</td>
<td>.000</td>
<td>.095</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 8: Results from clustered regressions vs. regressions on averaged, 1st-period, and dominant-choice data from high-endowment periods.
B. Pooling the sessions

In addition to the benchmark treatment B we conducted 2 plus 2 sessions of "automatic" treatments A and AI. Under both treatments, A and AI, we used the same game and same parameterization as in the B treatment. The only difference was that in automatic treatments, each subject played against a computer program, six subjects in the role of Participant X and six subjects in the role of Participant Y. The computer program was always playing (subgame perfect) optimal strategy. Subject were acquainted with these facts in the instructions.

The only difference between A and AI treatments was that in AI subjects received, as a separate part of instructions, a so called Backwards Induction Tutorial, intended to explain basic principles of using backwards induction.

In addition to the full-context C treatment, we conducted two sessions with partial context – C- treatment. In the C- treatment, the subjects receive only limited information about the context – Participant X is called "Entrepreneur" and Participant Y is called "Bureaucrat". Actions are, however, denoted by neutral letters – same as in the B treatment.

Before pooling the data from different treatments we performed basic statistical tests in order to discover significant differences in distributions of choices – Fisher’s Exact test and Wilcoxon rank-sum test. We find no evidence of significant differences in distributions of the 1st-period choices between A, AI and B treatments, as well as between C- and C treatments.

Afterwards, we performed 2 types of pooled analysis: 1) pooling the data from A and B treatments vs. pooled data from C- and C treatments; and 2) pooling the data from A, AI and B treatments vs. pooled data from C- and C treatments. Note that in 1) both pools contain the same number of subjects, which is not the case after we extend the benchmark-type pool by data from AI.

See Tables 9 and 10 for the regression results for low- end high-endowment periods, respectively.

Clearly, pooling slightly different treatments leads to noisier results, which is not very surprising.

For both, low- and high-endowment periods, we are loosing on significance of econ (and its interaction term).

As regards the treatment dummy, we are loosing significance for low-endowment periods on one hand, but on the other hand it becomes significant for high-endowment-period data.
Table 9: Results from estimation on basic vs. extended data sets for low-endowment periods. BAtreat in each model is a dummy identifying a benchmark-type treatment – B; or B and A; or B, A, and AI treatments, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Periods 1.3.5</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B vs. C</td>
<td>B,A vs. C,C-</td>
<td>B,A,Al vs. C,C-</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>-0.0287 (.302)</td>
<td>-0.0191 (.380)</td>
<td>-0.0093 (.073)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1280 (.007)</td>
<td>0.0854 (.073)</td>
<td>0.0404 (.040)</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>0.0686 (.646)</td>
<td>0.0162 (.890)</td>
<td>0.0076 (.940)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.1425 (.022)</td>
<td>2.8109 (.029)</td>
<td>2.0067 (.020)</td>
<td></td>
</tr>
<tr>
<td>econ</td>
<td>-0.1601 (.212)</td>
<td>-0.1754 (.089)</td>
<td>-0.1343 (.164)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.6307 (.000)</td>
<td>-0.3061 (.025)</td>
<td>-0.0944 (.532)</td>
<td></td>
</tr>
<tr>
<td>BAtreat</td>
<td>-0.0559 (.657)</td>
<td>-0.0609 (.550)</td>
<td>-0.0736 (.449)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.7156 (.004)</td>
<td>0.2708 (.218)</td>
<td>0.2738 (.170)</td>
<td></td>
</tr>
<tr>
<td>age*male</td>
<td>-0.0345 (.022)</td>
<td>-0.1235 (.023)</td>
<td>-0.0005 (.018)</td>
<td></td>
</tr>
<tr>
<td>econ*male</td>
<td>0.5354 (.002)</td>
<td>0.1801 (.318)</td>
<td>0.0005 (.998)</td>
<td></td>
</tr>
<tr>
<td>BAtreat*male</td>
<td>-0.7933 (.006)</td>
<td>-0.3586 (.154)</td>
<td>-0.3877 (.101)</td>
<td></td>
</tr>
<tr>
<td>const</td>
<td>1.2901 (.048)</td>
<td>1.1457 (.025)</td>
<td>0.9044 (.054)</td>
<td></td>
</tr>
<tr>
<td>mean p(y=1)</td>
<td>.5625</td>
<td>.5787</td>
<td>.5714</td>
<td></td>
</tr>
<tr>
<td># of obs.</td>
<td>144</td>
<td>216</td>
<td>252</td>
<td></td>
</tr>
<tr>
<td>joint p-value</td>
<td>.488</td>
<td>.439</td>
<td>.675</td>
<td></td>
</tr>
</tbody>
</table>

Table 10: Results from estimation on basic vs. extended data sets for high-endowment periods. BAtreat in each model is a dummy identifying a benchmark-type treatment – B; or B and A; or B, A, and AI treatments, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Periods 2.4.6</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B vs. C</td>
<td>B,A vs. C,C-</td>
<td>B,A,Al vs. C,C-</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>0.0220 (.381)</td>
<td>0.0310 (.100)</td>
<td>0.0253 (.160)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0913 (.000)</td>
<td>0.1134 (.000)</td>
<td>0.0758 (.007)</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>1.106 (.055)</td>
<td>0.0620 (.461)</td>
<td>0.1044 (.867)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.5488 (.008)</td>
<td>2.4589 (.000)</td>
<td>1.0512 (.051)</td>
<td></td>
</tr>
<tr>
<td>econ</td>
<td>0.0731 (.503)</td>
<td>-0.1029 (.616)</td>
<td>-0.1424 (.113)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2210 (.001)</td>
<td>-0.2268 (.133)</td>
<td>0.227 (.876)</td>
<td></td>
</tr>
<tr>
<td>BAtreat</td>
<td>-0.0030 (.809)</td>
<td>0.0780 (.331)</td>
<td>0.1172 (.133)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0375 (.644)</td>
<td>0.2268 (.072)</td>
<td>0.2955 (.026)</td>
<td></td>
</tr>
<tr>
<td>age*male</td>
<td>-0.0941 (.032)</td>
<td>-1.019 (.001)</td>
<td>-0.0601 (.099)</td>
<td></td>
</tr>
<tr>
<td>econ*male</td>
<td>-0.3395 (.019)</td>
<td>-0.0768 (.740)</td>
<td>-0.2101 (.248)</td>
<td></td>
</tr>
<tr>
<td>BAtreat*male</td>
<td>-0.0036 (.983)</td>
<td>-1.282 (.436)</td>
<td>-1.958 (.243)</td>
<td></td>
</tr>
<tr>
<td>const</td>
<td>-0.3170 (.556)</td>
<td>-2.3604 (.000)</td>
<td>-1.5033 (.526)</td>
<td></td>
</tr>
<tr>
<td>mean p(y=1)</td>
<td>.2361</td>
<td>.2593</td>
<td>.2817</td>
<td></td>
</tr>
<tr>
<td># of obs.</td>
<td>144</td>
<td>216</td>
<td>252</td>
<td></td>
</tr>
<tr>
<td>joint p-value</td>
<td>.078</td>
<td>.045</td>
<td>.075</td>
<td></td>
</tr>
</tbody>
</table>

33