TESTING LENIENCY
PROGRAMS EXPERIMENTALLY:
THE IMPACT OF "NATURAL" FRAMING

Jana Krajčová
Andreas Ortmann

CERGE-EI

Testing Leniency Programs Experimentally: The Impact of "Natural" Framing

Jana Krajčová
Andreas Ortmann

Testing Leniency Programs Experimentally: The Impact of "Natural" Framing[*]

Jana Krajčová[†], Andreas Ortmann[‡]

CERGE–EI[§]

October 2008

Abstract

We study the effects of loaded instructions in a bribery experiment. We find a strong gender effect: men and women react differently to real-world framing. The treatment effect becomes significant once we allow for gender-specific coefficients. Our paper contributes to the (small) literature on experimental tests of (anti-)corruption measures and adds evidence to the (mixed) results on gender effects and the on-going discussion on the need for sociodemographic controls.

Abstrakt

Skúmame rozdiely v rozhodovaní účastníkov experimentu zaoberajúcom sa korupciou a protikorupčnými optareniami, keď títo sú alebo nie sú oboznámení s kontextom. Naše výsledky poukazujú na významné rozdiely medzi správaním mužov a žien. Analyzovaný vplyv kontextu je významný v modeli, ktorý umožňuje rôzne koeficienty pre mužov a pre ženy. Náš článok tak prispieva k (skromnej) literatúre dokladujúcej experimentálne testovanie protikorupčných opatrení a rovnako aj k (rôznorodým) záverom o rozdieloch v správaní sa mužov a žien, či pokačujúcim diskusiám o význame sociodemografických charakteristík pri analýze experimentálnych dát.

[†]E-mail: jana.krajcova@cerge-ei.cz

[‡]E-mail: andreas.ortmann@cerge-ei.cz, aortmann@yahoo.com

[§]Center for Economic Research and Graduate Education–Economics Institute, a joint workplace of Charles University in Prague and the Academy of Sciences of the Czech Republic. Address: CERGE–EI, P.O. Box 882, Politických vězňů 7, Prague 1, 111 21, Czech Republic.

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

One promising anti-corruption measure is the leniency policy. 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 been analyzed in the literature mostly as an anti-cartel mechanism.

Spagnolo (2004), for example, theoretically examines the effects of leniency policies of various degrees – from moderate (which reduce or cancel the penalty for a criminal who reports) to full (which, in addition, pay a reward). He shows that reward-paying leniency programs provide a (socially) costless[1] and very efficient measure for cartel deterrence. Drawing on earlier versions of Spagnolo (2004), Apesteguia, Dufwenberg and Selten (2004) conducted an experiment that confirms the promising cartel-deterring properties of leniency policies.

Bigoni, Fridolfsson, Le Coq and Spagnolo (2007, 2008) conducted related experiments. They find that without leniency past convictions reduce the number of cartels but increase collusive prices. Their results also suggest that past expe-

---

[1]This is the case if the rewards are fully financed from fines imposed on other convicted members of the gang.

rience might have more important consequences for the perception of risk than an exogenous probability of detection, and that strategic risk plays a key role in the effectiveness of a leniency policy. In general, the deterrence is higher with leniency in place and rewards lead to almost complete deterrence.

The work of Bigoni et al (2007, 2008) contributes to a better understanding of the cartel-deterring properties of leniency policies and highlights the importance of proper policy design.

Leniency policies to deter cartels are not directly applicable as anti-corruption measures, since cartel deterrence is a simultaneous game while strategies, payoffs, and the move structure of anti-corruption measures are asymmetric.[2] A proper theoretical and experimental analysis is therefore called for.

To our knowledge the first theoretical work analyzing the various effects of leniency policies on corruption is Buccirossi & Spagnolo (2006). The authors show that poorly designed moderate policies may have a serious counter-productive effect: they might allow a briber to punish at relatively low cost a partner who does not respect an illegal agreement. In other words, some leniency policies might actually provide an enforcement mechanism for occasional illegal transactions.[3] Thus they can, contrary to the intention, increase corruption.

Buccirossi & Spagnolo's result together with the theoretical and experimental 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 (Dušek, Ortmann and Lízal 2005). They become especially useful when counterfactual institutional arrangements such as leniency programs need to be explored:

---

[2]For a more detailed discussion see Richmanová (2006).

[3]Occasional illegal transactions are essentially one-shot transactions.

they provide, for example, relatively cheap ways to examine the effects of such arrangements in controlled environments (see Dušek et al 2005, Apesteguia et al 2004, Buccirossi & Spagnolo 2006, Bigoni et al 2007, 2008, Richmanová & Ortmann 2008 and also Roth 2002).

In Richmanová & Ortmann (2008) we proposed a generalization of the Buccirossi & Spagnolo (2006) model in which we introduce the probabilistic discovery of evidence by auditing inspectors.[4] Our generalization makes the model more realistic and more readily applicable for experimental testing without changing the qualitative results of Buccirossi & Spagnolo.

We use the generalized Buccirossi & Spagnolo model for the experimental testing of leniency policies as an anti-corruption measure. The present paper, and the closely related work reported in Krajčová (2008), provide a new testbed for anti-corruption programs and address important methodological issues with which (anti-)corruption experiments are afflicted. Specifically, we are interested in to what extent home-grown priors that are related to corruption might translate into moral scruples and, for example, might induce subjects to make different decisions when loaded instructions are used that make it unambiguously clear what the context of the experiment is.[5]

---

[4]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 a probability of one. In Richmanová & Ortmann (2008), we argue that instead some evidence is created unintentionally and this can be discovered by the audit with some probability that is less than one.

[5]Altogether, we designed three treatments: a benchmark, which is common to both studies, Krajčová (2008) and the present paper, and in which all instructions are presented in neutral language; a context treatment, in which we use the same parameterization as in the benchmark but instructions are presented in real-world framing (the present paper); and, finally, a high-incentive treatment, which implements a new parameterization within neutral framing (Krajčová 2008). Krajčová (2008) examines the effect of a change in parameterization. It has been documented in the literature that a change in parameterization that does not affect the theoretical prediction might indeed have consequences for the 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 a considerable loss. Therefore, it is possible that subjects in the lab will not behave in accordance with the theoretical predictions, especially when the predictions

Abbink & Hennig-Schmidt (2006) is the study most closely related to the present paper. In a bribery setting, these authors used a between-subject design with one treatment providing instructions framed in neutral terms and the other treatment "loading" the instructions with real-world framing. The authors used a finitely repeated reciprocity game as the bribery setting. With a very low probability (0.3%) bribery would be detected leading to both parties involved in the transaction being excluded from the experiment and forfeiting their accumulated earnings. In addition, a completed transaction imposed a negative externality on the public, in Abbink & Hennig-Schmidt (2006) represented by the other subjects in the session.

The authors find a treatment effect which is, however, not statistically significant;[6] they argue that this result may be caused by the nature of the game: it is very simple, and thus, even with neutral wording, subjects may have deciphered that the experiment was about corruption and corruptibility. The generalized Buccirossi & Spagnolo game is more complex (e.g., it involves the realization of two random outcomes) and therefore is likely to be less susceptible to inferences about the true nature of the strategic interactions.

Our results confirm that loading the instructions with real-world framing might affect subjects' behavior in a significant manner. Moreover, we find a strong gender effect - male and female participants react to a bribery context differently. This treatment effect becomes significant once we allow for gender-specific effects.

The remainder of the paper is organized as follows. In the next section, we

are made under the assumption of risk neutrality. The author indeed finds a significant effect of the parametric change.

[6]This result is questionable. When looking at the evolutions of bribe offers and of permission frequencies, especially in the 20 central rounds, a clear difference between the two treatments is visible: for each round the average bribe offer is higher with neutral than with loaded instructions. 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 treatment effect is confirmed by both wilcoxon rank-sum test and effect sizes for the 20 central rounds. It would likely become significant, even without excluding the possibly problematic first and last five periods, with a larger number of observations.

discuss the generalized Buccirossi & Spagnolo model in detail, and we also describe and compare our two experimental treatments. In Section 3 we explain how we implemented our two treatments and in Section 4 we discuss our results. Section 5 concludes.

2 Experimental Design

The experiment implements the bribery game in Richmanová & Ortmann (2008). An entrepreneur has an investment possibility of net present value v, if a bureaucrat is willing to perform an illegal action, *Action a*. For doing so, the bureaucrat may require compensation in the form of a bribe, b.

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*,[7] or perform *Action a*.[8]

If the bureaucrat chooses *Denounce*, an audit is carried out. The audit may (with probability β, $\beta \in (0,1)$), or may not (with probability $1 - \beta$), discover some evidence of bribery. If the bribery attempt is detected, the leniency policy guarantees that the bureaucrat will have to pay only a reduced fine whereas the entrepreneur will have to pay the full fine. In addition, bribe b is confiscated.[9] If the bribery is not detected, the bureaucrat will enjoy bribe b.

---

[7]*Nothing* denotes a passive action choice. For the bureaucrat, it means that he neither denounces nor respects (by providing the favor) the illegal agreement. For the entrepreneur, it means that she does not denounce in response to the bureaucrat's action.

[8]*Action a* means that the bureaucrat respects the illegal agreement and thus provides an (illegal) favor to the entrepreneur. That is, strictly speaking, not a corrupt action because it does not impose a negative externality on the public. According to Abbink, Irlenbusch & Renner (2002) it is not such a problem since people do not care much about the costs they impose on others.

[9]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.

If the bureaucrat chooses *Nothing* or *Action a*, the entrepreneur has another move. In both cases, she may choose between *Denounce* and do *Nothing*.

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

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

Figure 1 below summarizes the extensive form of the game and the expected payoffs.

The contribution of the generalized model lies in the introduction of probability β. In Buccirossi & Spagnolo (2006) it is assumed that, before the illegal transaction takes place, the bureaucrat and the entrepreneur agree on the 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 that commits each other to the desired outcome. It is furthermore assumed that, if an audit takes place, corruption is discovered and both culprits are convicted with a probability of one. Richmanová & 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 $\beta \in (0, 1)$. The basic structure of both the original and the modified game is the same except that in the original version the probability β is set to 1. The generalization makes the model
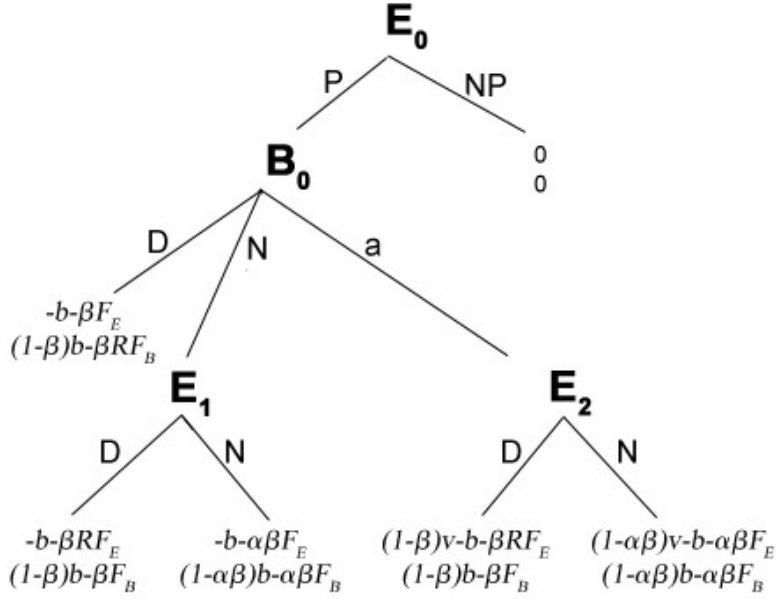
7

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*, b for bribe, v for the value of the project to the entrepreneur, α for the exogenous probability of an audit, β for the probability of conviction, F_E and F_B for full fines and RF_E and RF_B for reduced fines to the entrepreneur and to the bureaucrat, respectively.

more suitable for experimental testing, as no additional stage is needed in which subjects would have to agree on producing a hostage. In addition, the generalized model arguably resembles real-world situations more closely.[10]

Buccirossi & Spagnolo (2006) show that in the absence of a leniency program, occasional illegal transactions are not implementable.[11] The result carries over into the generalized model. After the introduction of a modest leniency program,[12] occasional illegal transactions are enforceable if the following three conditions are satisfied simultaneously. First, the *no-reporting condition* for the bureaucrat: the

---

[10]We realize that in such a game beliefs about the probability of detection might play an important role. However, we believe that the introduction of beliefs would make the game more complex than necessary for experimental testing. Instead, we view probability β as an empirical success rate, or effectiveness, of a detection technology that is known to subjects.

[11]Facing the full fine even after reporting, the entrepreneur cannot credibly threaten to report the bureaucrat in the case when 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 in the first place.

[12]Similarly to Spagnolo (2004), "modest" means that a leniency program does not reward for reporting, at best it cancels the fine.

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: the reduced fine and the 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 respects the illegal agreement.

These three conditions, given the value of the project together with the full and reduced fines, define a bribe range for which the occasional illegal transaction is implementable. Even though these conditions are modified in the 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 the per-round endowment for the entrepreneur. The per-round endowment exogenously defines the value of the bribe if the entrepreneur decides to pay it.[13] For each treatment we use two possible values of the per-round endowment: a low endowment that theoretically leads to a no-corruption equilibrium, and a high endowment that theoretically leads to a corruption equilibrium.

Following Abbink & Hennig-Schmidt (2006), the focus of the present paper is the question whether loaded instructions in a bribery experiment affect the behavior of subjects in a lab. For that purpose, we designed two treatments: a <u>B</u>enchmark (*B*) and a <u>C</u>ontext (*C*) treatment.[14]

---

[13]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.

[14]We also conducted two exploratory sessions of a partial context treatment (C- treatment), where we only provided context on the 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 of the game in this case and therefore it is hard to recognize all the possible effects in this treatment. Some results from this treatment are discussed in the appendix.

Table 1 below summarizes the parameterization chosen for both treatments.

Treatment	α	β	v	RF_E	RF_B	F_E	F_B	E_L	E_H	show-up
B and **C**	0.1	0.2	100	0	0	300	300	20	40	300

Table 1: Experimental parameterization. α and β denote the probability of an audit and of discovering evidence of bribery, respectively; v denotes the 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 are low and high per-round endowment, respectively; and show-up stands for the show-up fee.

The probabilities α and β were chosen such that they approximately correspond to the real world exogenous probabilities of an audit and to real-world conviction rates; at the same time they are intuitively comprehensible for subjects. The value of the project v was chosen together with the full fines F_E and F_B such that the subject faces a considerable gain from the investment but also severe punishment in the case of detection. We set reduced fines RF_E and RF_B equal to zero to analyze the case of full leniency programs that, according to Apesteguia et al (2004), have shown promising anti-cartel properties. Endowment determines the value of the bribe to be (not) paid. The "low endowment" of 20 leads (theoretically) to a no-corruption equilibrium, whereas the "high endowment" of 40 leads to a corruption equilibrium. Finally, the show-up fee was set such that we eliminate the possibility of earning a negative total from the experiment.

The parameterization does not differ between B and C treatments as we are interested purely in the effects of "neutral" and "loaded" instructions.

Extended game forms and expected payoffs resulting from our parameterization are illustrated in Figure 2 for both low- and for high-endowment periods. The branches identifying the equilibrium choices of risk-neutral agents are in bold.

The instructions for the B treatment were presented in a completely context-free fashion. Subjects were called Participant X and Participant Y, actions were denoted by neutral letters and the realization of "detection" or "no detection" as
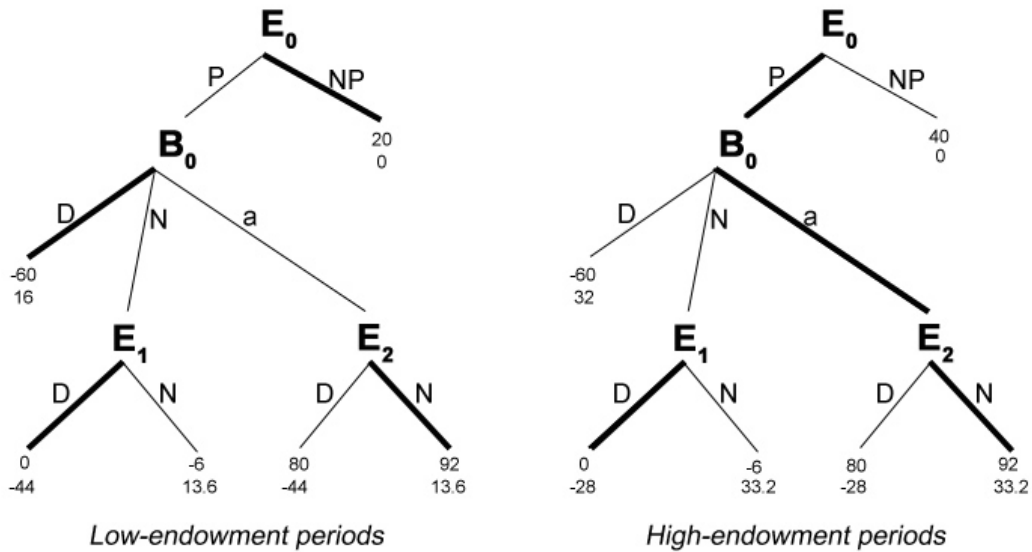
Figure 2: Expected payoffs from the corruption game.

"outcome A" or "outcome B", respectively.

In the <u>C treatment</u>, the roles that subjects were assigned were called *"Entrepreneur"* and *"Bureaucrat"*; actions were called *"<u>P</u>ay bribe"*, *"<u>N</u>ot <u>P</u>ay bribe"*, *"<u>D</u>enounce"*, *"do <u>N</u>othing"* and *"Provide the favor <u>a</u>"*; and the realizations of random outcomes were called *"corruption has been detected"* and *"corruption has not been detected"*. Figure 3 below provides a comparison of the wording for the two treatments, with the neutral wording always in the upper row.

3   Implementation

The experiment was conducted in November and December 2006 at CERGE-EI in Prague, using a mobile experimental laboratory.[15]

Participants were recruited from the Faculty of Social Sciences of Charles University in Prague, from various faculties of the Czech Technical University in Prague and of the University of Economics in Prague. Students were approached via posters

---

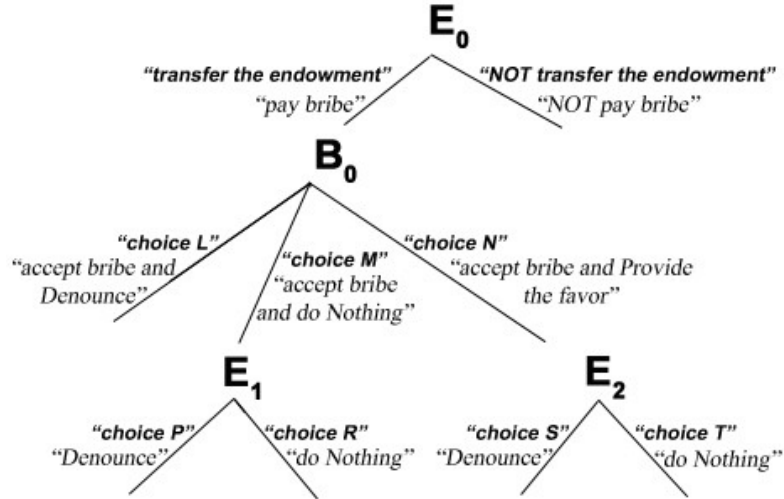[15]http://home.cerge-ei.cz/ortmann/BA-PEL.htm

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.

distributed on campus and via e-mail.[16]

We conducted four sessions of each treatment. Twelve participants, six in the role of Participant X – the entrepreneur – and six in the role of Participant Y – the bureaucrat – interacted in each session. In each session, all subjects participated in six rounds during which they kept the role that was assigned to them at the beginning of the first round.[17] 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).

---

[16]By e-mail, we also directly invited students from these schools who participated earlier in unrelated experiments conducted at CERGE-EI.

[17]After each Participant X interacted exactly once with each Participant Y, the roles were switched for another six rounds. Subjects were not informed about the switch of roles in advance in order to avoid a 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. The decisions in the last six rounds are likely affected by subjects' experience from the first six rounds and therefore we do not report them in the main text. A comparison of the before-switch and after-switch data is provided in the appendix. For both treatments, we observe more transferring in the after-switch data. In the B treatment, we also observe more denouncing in both the second and the third stage. In the C treatment, the differences for the second- and the third-stage data are very small.

Treatment	Subject Source[18]	M/F ratio[19]	mean (age)	mean (RA score)	mean (final pay)	Irreg[20]
B	FSS	8/4	20.9	29.7	320	1
B	FSS	10/2	21.75	28.8	330	0
B	CTU	11/1	22.9	34.7	330	0
B	FSS	9/3	22.3	26.4	323.3	0
C	CTU	9/3	21.9	33.7	340	0
C	UE	7/5	22.9	28	318.3	1
C	CTU	10/2	23	31.4	318.3	0
C	UE	7/5	21.7	28.1	315	0

Table 2: Summary of the demographic characteristics of subjects for all eight sessions.

Table 2 summarizes the demographic characteristics of subjects participating in the experiment. The majority of our subjects were male, reflecting the composition of the subject pools that we drew on. Mean age ranged between 20.9 and 23, over all sessions the minimum is 18 and the maximum is 29. We also measured subjects' risk aversion using a questionnaire based on Holt & Laury (2002). Mean RA score in the sample ranged between 26.4 and 34.7, over all sessions the minimum is 13 and maximum 51.[21] Average final payoffs for the B treatment ranged from 320 to 330, with the minimum being 300 and the maximum 400; for the C treatment it ranged between 315 and 340, with the minimum being 300 and the maximum 400.[22]

---

[18]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. We control for imbalance of the subject pool by including the econ and gender dummies in the econometric analysis.

[19]Male/Female ratio in the session.

[20]Irreg stands for a dummy variable for session irregularities. It identifies any unusual activities by subjects or any irregularities on the experimenter's side. In the first B treatment session an experimenter effect is possible; 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 a CTU student participated in a UE session. We do not believe that these matter but wanted to control nonetheless. After running the preliminary regressions we concluded that they indeed did not matter.

[21]The higher the score the more risk averse the subject is. The maximum possible RA score is 60 which, using the 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 $-.13$. An RA score of 23 corresponds to risk-neutrality.

[22]At that point 400 CZK corresponded to about 16 USD, in purchasing power up to twice as much. Subjects were informed during recruitment that their final payoff from the experiment might be zero, but could not be negative. The non-negativity of the final payoff was ensured by the show-up fee.

Each session began with general instructions. Afterwards, subjects were asked to fill in Risk-aversion and Demographics questionnaires, for 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. The computerized part of the experiment started only after every participant answered all testing questions correctly.[23] Each session concluded with a questionnaire asking for the subject's feedback on the experiment.[24]

All instructions were read aloud by the experimenter. As a part of the instructions, subjects received a pictorial representation of the game with a 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).[25]

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 their current per-round endowment. Afterwards, each pair interacted sequentially.[26] Between the second and the third stage, Participants X were asked about their choices at each node of the third stage if they were to reach it. After they made their conditional choices, they learned the 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 in rounds when the third stage was not reached.

At the end of each round subjects were given feedback about their action(s),

---

[23]This was common knowledge.

[24]For filling this last questionnaire, subjects were paid an additional 50-200 CZK (corresponded to about 2-9 USD) - the amount varied between sessions. This mechanism was used to adjust average earnings per session to the levels promised during recruitment.

[25]Originals (in Czech) of all materials that subjects received during the experiment are available at http://home.cerge-ei.cz/richmanova/WorkInProgress.html.

[26]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.

the action(s) of the player they were paired with, the realization of the random outcome (detection vs. no detection, or outcome A vs. outcome B) 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 the decision in every round is made as if in a one-shot game. This payment procedure was common knowledge ex-ante.

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

4   Results

In Figure 4, the results from low- and high-endowment periods 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 in bold face.

For the aggregate first-stage data we observe surprisingly small differences between the two treatments. In addition, in both treatments, the frequencies of choosing *Pay* are higher in the low-endowment periods than in the high-endowment periods, which contradicts the theoretical prediction. Intuitively, subjects seem to be willing to transfer their (low) endowment to have a chance of receiving a high payoff, but their willingness to risk their endowment is limited. Instead of risking the high endowment, they seem to prefer the sure outcome.

As to the second-stage data, it is only relative percentages that can be compared across treatments, as the numbers of subjects that entered this stage of the game varied. In the B treatment, subjects split their choices evenly 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
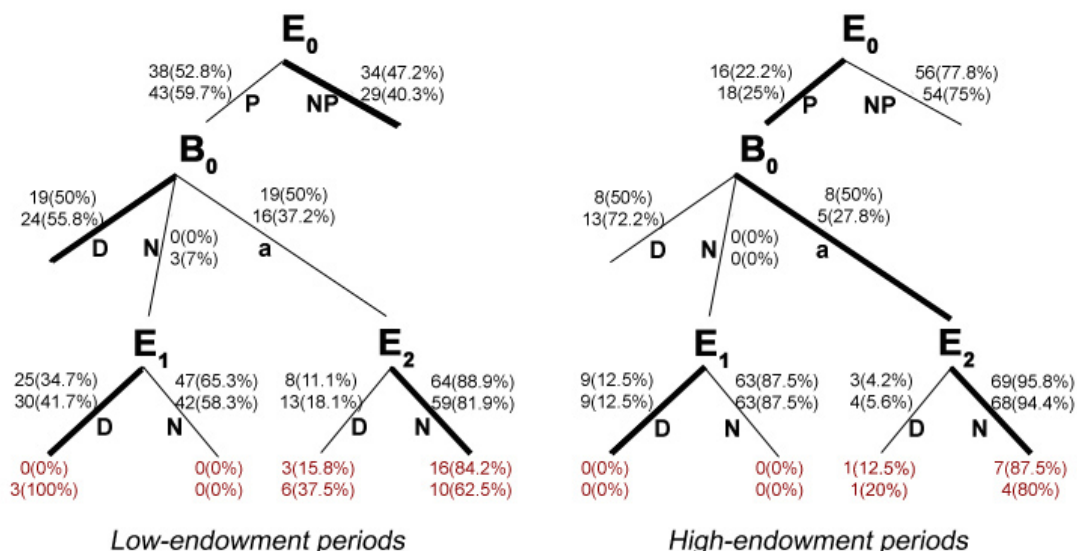
15

Figure 4: Experimental results. For each branch of the extensive form of the game, the upper row always displays the frequency of the action in the B treatment, while the lower row displays the frequency of the action in the C treatment (with the corresponding percentage in parentheses). For stages E_1 and E_2, above the branches, we present the conditional choices subjects were asked to report before they made their actual choice. The frequencies of real choices, which depend on the preceding decision of Participant Y, are presented at the bottom part of each figure.

and might explain why we do not observe a stronger inclination to either choice. Also note that in both treatments *Denounce* is the only action through which the bureaucrat can avoid a negative expected round-payoff with certainty.[27]

In the C treatment, choices are shifted in favor of *Denounce*. Arguably, in the high-endowment periods, this result contradicts the theoretical prediction, but it is in line with our conjecture – knowing the context of their action choice, reporting corruption might be more attractive for subjects.

In line with the theoretical prediction and also our intuition, *Nothing*[28] was almost never chosen.

As to the third-stage data, conditional choices provide mixed evidence. In both

---

[27]See Figure 2 and Table 1 for more details. Even though the subject could possibly earn a negative round payoff, each subject also received a show-up fee which ensured a non-negative final payoff.

[28]Payoffs for Participant Y resulting from *Nothing* and *Action a* are the same, but taking into account the likely decisions of Participant X in the following stage, he is more likely to collect a higher payoff after choosing *Action a*.

treatments, subjects seem to prefer playing *Nothing* in either case. For the E_1 node that contradicts the theoretical prediction, while it is in line with the theoretical prediction for the E_2 node. When we look at the sequential choices, the results seem in line with the theoretical prediction for both treatments, inferring from relatively few observations.[29] We observe essentially 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 few independent observations (especially so for the high-endowment periods)[30] to perform a reliable formal analysis. Therefore, we only perform statistical and regression analyses of the first-stage data.

4.1 Analysis of the first-stage data

We performed standard non-parametric tests with the null hypothesis of no differences in the distributions of choices under the two treatments. We also computed the effect sizes 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 $1, 3$, and 5 (low-endowment) and from periods $2, 4$, and 6 (high-endowment) are clustered

---

[29]When we asked the subjects to make their real choices in the B treatment, only one of them changed her/his decision in the E_2 node from *Denounce* to *Nothing* (after observing what Participant Y has chosen) in the low-endowment period. In the C treatment, three subjects changed her/his decision in the E_2 node – 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 the E_1 node from *Nothing* to *Denounce* after Participant Y played *Nothing*. All four cases occurred in low-endowment periods.

[30]Recall that Figure 4 presents the aggregated data from all relevant periods, therefore it contains repeated observations for individual subjects.

by subject to correct standard errors for likely within-subject correlation; 2) first-period data analysis – only first-period data (for the low-endowment case) and only second-period data (for the high-endowment case) are analyzed; 3) averaged data analysis – averaged data for periods $1, 3$, and 5 and for periods $2, 4$, and 6 are analyzed; and 4) dominant-choice data analysis – for each endowment value (low or high) each subject makes choices in three periods, and the dominant choice is the one that is played more often.

Clustered data have the advantage of using all the available information, 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. The analysis of averaged, first-period, and dominant-choice data can be found in Appendix 2, part A, as a robustness check of the main results. By and large, there are no major findings in these robustness tests.

In addition to the robustness checks based on different "data handling", we also run a few exploratory sessions of treatments in which the experimental conditions are only slightly modified compared to the benchmark and the context treatment. The results from the analysis on the extended data set is provided in Appendix 2, part B, as an additional robustness check of the main results. By and large, there are no major findings in these robustness checks. Pooling slightly different treatments leads to noisier results, which is not very surprising.

4.1.1 Statistical analysis

In Table 3 below we report the results of three standard non-parametric tests in order to identify the differences in the distributions of choices under the two treatments. Specifically, we test the null hypothesis of no differences between the

two treatments using the averages of the binary transfer variable[31] over periods
1, 3, and 5 and 2, 4, and 6.

periods	Ranksum[32]	Ksmirnov[33]	Fisher[34]
1,3,5	-.526 (.599)	.083 (.846)	(.947)
2,4,6	-.715 (.475)	.167 (.513)	(.218)

Table 3: Non-parametric tests.

According to Wilcoxon rank-sum, Kolmogorov-Smirnov, and Fisher's exact
tests, we cannot reject the hypothesis of no differences in the distributions of choices
under the two treatments at the 5% significance level.

To assess the magnitude of the effect for practical purposes, we in addition
compute two standardized measures of effect size: Cohen's d and odds ratio, again,
using the averages of the binary transfer variable over periods 1, 3, and 5 and
2, 4, and 6. We also want to look separately at male and female data to discover
possible gender effects. The results for the full sample and for the male and female
subsamples are reported in Table 4 below.

Periods	Sample	B			C			effect size	
		N	mean	std.dev.	N	mean	std.dev.	odds ratio	Cohen's d
1,3,5	full	24	.528	.4495	24	.597	.4282	1.131	.1571
	male	18	.519	.4461	17	.667	.4082	1.285	.346
	female	6	.556	.5018	7	.429	.4600	.772	-.264
2,4,6	full	24	.222	.3764	24	.25	.3147	1.126	.0807
	male	18	.296	.4105	17	.275	.3170	.929	-.057
	female	6	0	0	7	.190	.3253	NA[35]	.826

Table 4: Effect-size indices.

Cohen (1998) defines effect sizes of $d = 0.2$ as *small*, $d = 0.5$ as *medium*, and

---

[31]Transfer has a value of one if Participant X chooses *Pay* and a value of zero if s/he chooses *Not Pay* in the respective period.

[32]Ranksum stands for the two-sample Wilcoxon rank-sum (or Mann-Whitney) test. We report the normalized z statistic and corresponding p-value below.

[33]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.

[34]Fisher stands for Fisher's exact test. We report the resulting p-value.

[35]A division-by-zero problem occurs, due to no variation in this subsample.

$d = 0.8$ as *large*. For the full sample, the results suggest only a 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 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 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 the C and B treatments. Effect-size measures for the male and female subsamples suggest that this result 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 Gender differences

Before the estimation, we want to look more closely into gender-specific data. Figures 5 and 6 provide the summary data separately for men and women.

For low-endowment periods, in the first stage of the B treatment the difference in the 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 males and of females shift in opposite directions – two thirds of men, 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 the E_1 node we
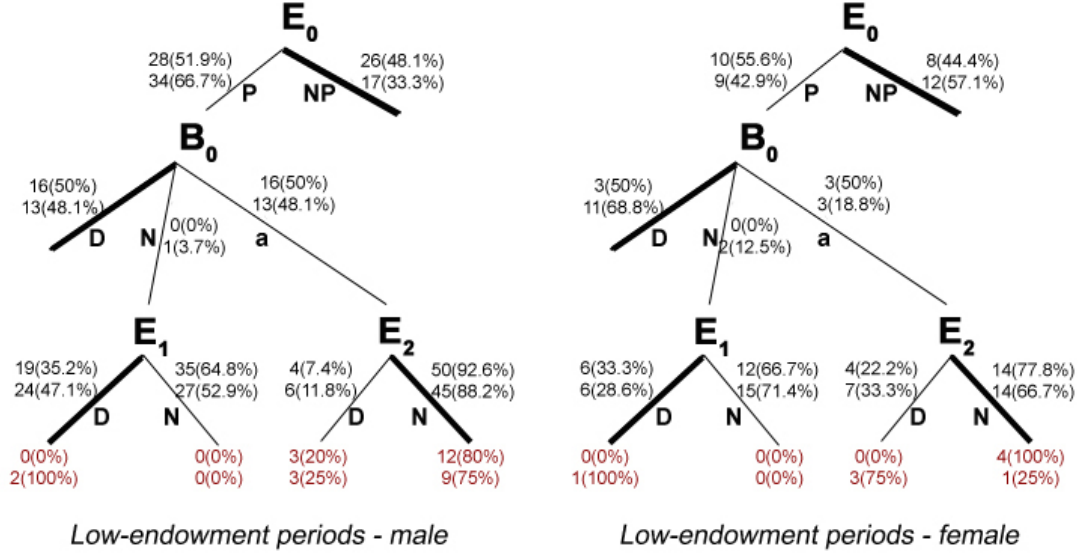
Figure 5: Experimental results for male and for female subjects in low-endowment periods. For each branch of the extensive form of the game, the upper row always displays the frequency of the action in the B treatment, while the lower row displays the frequency of the action in the C treatment (with the corresponding percentage in parentheses). For the nodes E_1 and E_2, above the branches, we present the conditional choices subjects were asked to report before they made their actual choice. Frequencies of real choices, which depend on the preceding decision of Participant Y, are presented at the bottom part of each figure.

observe the opposite effect of context on men than on women. In the E_2 node, 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 refraining from making transfers but the framing effect seems to increase the transfer 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 men and women 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.
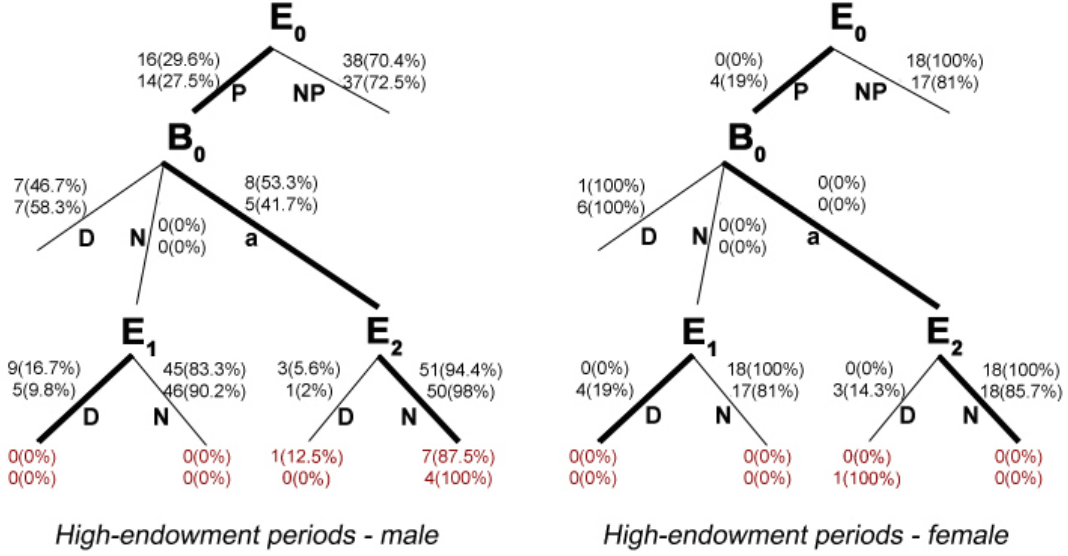
Figure 6: Experimental results for male and for female subjects in high-endowment periods. For each branch of the extensive form of the game, the upper row always displays the frequency of the action in the B treatment, while the lower row displays the frequency of the action in the C treatment (with the corresponding percentage in parentheses). For the nodes E_1 and E_2, above the branches, we present the conditional choices subjects were asked to report before they made their actual choice. Frequencies of real choices, which depend on the preceding decision of Participant Y, are presented at the bottom part of each figure.

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

4.1.3 Econometric analysis

In this section we report the results from econometric analysis controlling for some of the subjects' characteristics and for the treatment effect. We are also trying to confirm gender-specific effects.

During the experiment we distributed questionnaires in order to collect basic demographic data. Specifically, we have information about subjects' age, gender, university and field of study.[36] We also measure each subject's risk aversion.

---

[36]In addition, we collected data on: size of subject's household, number of cars in the household, and whether the subject himself has his own car and what is its approximate value, all of which serve as proxies for income. We also asked the subjects whether they considered themselves as

The dependent variable was defined as a 0/1 dummy variable *translog* identifying *Pay* being chosen (value of 1) or not (value of 0) in a particular period. 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, the violation of which leads to inconsistent estimates if non-linear models are employed. Another advantage of the linear probability model lies in the straightforward interpretation of the estimated coefficients. We ran clustered robust estimation to correct standard errors for likely within-subject correlation and for heteroskedasticity.

In the appendix, we provide a discussion of the robustness checks we conducted in addition to the clustered regressions analysis. As the theoretical prediction differs for low- and for high-endowment periods,[37] these two groups were analyzed separately.

We start with a basic minimal model:[38]

$$P(translog = 1|x) = \beta_0 + \beta_1 \cdot age + \beta_2 \cdot male + \beta_3 \cdot econ + \beta_4 \cdot Ctreat,$$

technical types compared to their peers. We recorded the occurrence of any inconsistencies in the after-instructions questionnaire, which served as a simple test of understanding of the basic structure of the game, and in the risk-aversion questionnaire. At the end of the session we asked our subjects whether they understood the experiment. Finally, we recorded some general information about each session – the time of day 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. The demographic and the risk-aversion questionnaires are based on Rydval (2007).

[37]Recall that in periods $1, 3$, and 5 the endowment was low, while in periods $2, 4$, and 6 the endowment was high.

[38]The second approach we used was $P(translog = 1|x) = \beta_0 + \beta_1 \cdot ra\_score$, where ra_score is a risk aversion score computed based on data from the risk-aversion questionnaire. Preliminary analysis suggests that age, male and econ predict ra_score well (all three are jointly significant at the 5% level, age and male with a negative sign on the coefficient, age with a 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 ra_score only, and the other including male, age and income - as candidates for minimal models for our analysis. However, in $P(translog = 1|x) = \beta_0 + \beta_1 \cdot ra\_score$ ra_score never appeared significant and only rarely we observed the joint significance of the estimated models. Therefore, we omit a discussion of these results.

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

The results from the estimation are summarized in Table 5, denoted as Model 1. This model is, however, not significant. In the next step, we extend the basic minimal model by interaction terms with *male* to allow for gender-specific effects. This leads to Model 2:

$$P(translog = 1|x) = \beta_0 + \beta_1 \cdot age + \beta_2 \cdot male + \beta_3 \cdot econ + \beta_4 \cdot Ctreat+$$

$$+\beta_5 \cdot male * age + \beta_6 \cdot male * econ + \beta_7 \cdot male * Ctreat.$$

The results from the estimation of Model 2 are also summarized in Table 5.

	periods 1,3,5		periods 2,4,6	
Model	1	2	1	2
age	-.0287	.1280	.0220	.0913
	(.302)	(.007)	(.381)	(.000)
male	.0686	3.3442	.1706	2.5462
	(.646)	(.010)	(.055)	(.014)
econ	-.1601	-.6307	-.0731	.2210
	(.212)	(.000)	(.503)	(.001)
Ctreat	.0559	-.7156	.0230	-.0375
	(.657)	(.004)	(.809)	(.644)
age∗ male	-	-.1852	-	-.0941
		(.002)		(.032)
econ∗ male	-	.5354	-	-.3395
		(.002)		(.019)
Ctreat∗ male	-	.7983	-	.0036
		(.006)		(.983)
const	1.2342	-1.4593	-.3400	-2.1070
	(.068)	(.139)	(.553)	(.000)
mean $\widehat{p}$(y=1)	.5625	.5625	.2361	.2361
# of obs.	144	144	144	144
joint p-value	(.488)	(.000)	(.078)	(.000)

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 $\widehat{p}$(y=1)* denotes the mean predicted probability of a transfer being made.

Model 2 is strongly significant and this confirms a 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 – *age*, *male*, and *econ* – and their interaction terms are significant at the 5% level. Interestingly, the treatment dummy together with its interaction term is only significant for the low-endowment periods. This suggests that only for the low-endowment periods the presentation of the game matters.

The mean predicted probability of transfer in the low-endowment periods is .56; in the high-endowment periods it is only .24, which is considerably lower. This result contradicts the theoretical prediction.[39]

For the low-endowment periods, *age* has a positive sign on the coefficient for female, but a negative sign for male. *Econ* has a negative sign on the coefficient for both male and female. 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.

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

For the high-endowment periods, both *age* and *econ* have a positive sign on the coefficient for female, but negative for male. Similarly to 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.

The treatment dummy *Ctreat* has a negative sign for both female and male sub-

---

[39]Recall that in the equilibrium Participant X always transfers high endowment and never transfers low.

jects, which is yet another difference from low-endowment periods. This suggests a negative impact of a corruption context on the transferring decision – in high-endowment periods subjects are less likely to transfer when they are fully aware of the context. Note, however, that even though the sign reflects the expected impact of context, the coefficient is not significant.

5 Discussion

Some of the results confirm our expectations whereas some do not.

In the aggregate data, we find only a 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) gender effects that cancel each other out and are responsible for the reduced overall effect of non-neutral framing.

For the aggregate second-stage data, the treatment effect shows in an increased denouncing rate, which is in line with our expectations. For male and female subsamples, as much as we can tell given the low number of observations, denouncing rates are lower or the same[40] in the B treatment. Also for the aggregate third-stage data the treatment effect goes in the predicted direction.

Different attitudes of men and of women towards corruption have previously been reported by, for example, Alatas et al (2006). These authors find significant differences in the behavior of men and women in a corruption experiment. Their results, however, appear to be culture-specific.[41]

The observed negative impact of non-neutral framing on the transferring deci-

---

[40]This refers to all possible cases, when we are looking separately at male and female subsamples for high- and for low-endowment periods.

[41]The authors run the experiment in Melbourne (Australia), Delhi (India), Jakarta (Indonesia), and Singapore. Only the Australian data confirm a significant gender effect.

sions of women, together with the positive impact on the denouncing decisions 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 (e.g., Swamy, Knack, Lee and Azfar, 2001, or Dollar, Fisman and Gatti, 2001).

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

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

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

This phenomenon, in addition, appears robust. In Krajčová (2008) the same result is found for the high-incentive treatment.[42]

There are several possible explanations. One of them is the "preference for inclusion" conjectured by Cooper & Van Huyck (2003). They find that subjects are significantly more likely to make "non-rational" choices that allow their co-player to make a choice – and thereby to affect the final payoffs – when given extensive form

---

[42]Recall that the benchmark treatment is the same for both the present paper and Krajčová (2008).

27

versions of a game. In our game, "inclusion" introduces the risk of a significant loss. Our results suggest that subjects with a "preference for inclusion" were willing to transfer and to continue playing the game as long as the stakes were not too high (low endowment).

We also 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. When we computed the theoretical prediction for a (modestly) risk averse subject, we found that under some (reasonable) assumptions, our chosen parameterization can lead to a no-corruption equilibrium also for the high-endowment periods.[43] That is, for risk-averse subjects, it might be in fact optimal not to transfer a high endowment.

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

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

---

[43]We assume the standard CRRA utility function $u(x) = x^{(1-r)}$. The average risk-aversion coefficient in our sample is about 0.03, and the maximal is about 0.1. As the bribery game involves nodes with negative payoffs, some assumptions need to be made about the utility function in the negative domain. The prospect theory suggests that in the negative domain, the steepness of the utility function might be about twice as much as in the positive domain. For illustration, we computed the theoretical prediction for a risk-neutral subject assuming two different levels of (dis)utility from paying a 300 CZK penalty after detection: $u(-300) = -u(450)$; and $u(-300) = -u(600)$. For low endowment, the theoretical prediction is the same as for a risk-neutral subject. For high endowment it changes. For an extremely risk-averse participant ($r = 0.1$), the disutility of 450 still predicts a corruption equilibrium, however, the disutility of 600 predicts a no-corruption equilibrium. For an average risk-aversion coefficient ($r = 0.03$), the disutility of 450 is sufficient to change the theoretical prediction.

and with a better gender-balanced sample, our results might become more conclusive.[44] Altogether, we conclude that subjects seem to engage in all sorts of social considerations in a bribery game, including moral scruples, which should not be dismissed by experimenters looking for relevant policy implications.

---

[44]Ortmann and Tichy (1999) also report some evidence of differences in the (cooperative) behavior of men and women. Also the gender composition of the subject pool in the experimental session matters. When controlling for past experience, gender differences, however, disappear.

References

Abbink, K., Hennig-Schmidt, H., (2006). Neutral versus Loaded Instructions in a Bribery Experiment, *Experimental Economics 9(2)*, 103-121.

Abbink, K., Irlenbusch, B., Renner, E., (2002). An Experimental Bribery Game, *Journal of Law, Economics, and Organization 18(2)*, 428-454.

Alatas, V., Cameron, L., Chaudhuri, A., Erkal, N., Gangadharan, L., (2006). Gender and Corruption: Insights from an Experimental Analysis. *The University of Melbourne Research Paper No. 974.*

Apesteguia, J., Dufwenberg, M., Selten, R., (2007). Blowing the Whistle. *Economic Theory 31*, 143–166.

Bigoni, M., Fridolfsson, S.-O., Le Coq, C., Spagnolo, G., (2007). Fines, Leniency, Rewards and Organized Crime: Evidence from Antitrust Experiments. [manuscript dated November 15, 2007, not for distribution]

Bigoni, M., Fridolfsson, S.-O., Le Coq, C., Spagnolo, G., (2008). Risk Aversion, Prospect Theory, and Strategic Risk in Law Enforcement: Evidence From an Antitrust Experiment. [manuscript dated February 1, 2008, not for distribution]

Buccirossi, P., Spagnolo, G., (2006). Leniency Policies and Illegal Transactions, *Journal of Public Economics 90*, 1281-1297.

Cohen, J., (1988). Statistical Power for the Behavioral Sciences, 2nd edition. *Lawrence Erlbaum Associates Inc, Hillsdale.*

Cooper, D., J., Van Huyck, J.,B., (2003). Evidence on the Equivalence of the Strategic

and Extensive Form Representation of Games, *Journal of Economic Theory 110*, 290-308.

Dollar, D., Fisman, R., Gatti, R., (2001). Are Women Really the "Fairer" Sex? Corruption and Women in Government. *Journal of Economic Behavior and Organization 46(4)*, 423-29.

Dušek, L., Ortmann, A., Lízal, L., (2005). Understanding Corruption and Corruptibility through Experiments. *Prague Economic Papers 14*, 147-162.

Fischbacher, U., (2007). Z-tree: Zurich Toolbox for Ready-made Economic Experiments - Experimenter's Manual. *Experimental Economics 10(2)*, 171-178(8).

Gigerenzer, G., Hoffrage, U., (1995). How to Improve Bayesian Reasoning without Instruction: Frequency Formats. *Psychological Review 102*, 684–704.

Goeree, J., K., Holt, C., A., (2001). Ten Little Treasures of Game Theory, and Ten Intuitive Contradictions. *American Economic Review 91*, 1402-1422.

Gupta, S., Davoodi, H., Alonso-Terme, R., (2002). Does Corruption Affect Income Inequality and Poverty? *Economics of Governance 3(1)*, 23-45.

Hertwig, R., Ortmann, A., (2001). Experimental Practices in Economics: A Methodological Challenge for Psychologists? *Behavioral and Brain Sciences 24(3)*, 383-451.

Hertwig, R., Ortmann, A., (2004). The Cognitive Illusions Controversy: A Methodological Debate in Disguise That Matters to Economists. *in Zwick, R., Rapoport, A. (eds.), Experimental Business Research, Kluwer Academic Publishers, Boston, MA*, 361- 378.

Holt, C., A., Laury, S., K., (2002). Risk Aversion and Incentive Effects, *The American*

Economic Review 92(5), 1644-1655.

Hwang, J., (2002). A Note on the Relationship Between Corruption and Government Revenue. *Journal of Economic Development 27(2)*, 161-178.

Kamecke, U., (1997). Rotations: Matching Schemes that Efficiently Preserve the Best Reply Structure of a One Shot Game, *International Journal of Game Theory 26(3)*, 409-417.

Krajčová, J., (2008). Testing Leniency Programs Experimentally: The Impact of Change in Parameterization, *CERGE-EI Working Paper Series*, forthcoming.

Mauro, P., (1995). Corruption and Growth. *Quarterly Journal of Economics 110*, 681-712.

Ortmann, A., Lízal, L., (2003). Designing and Testing Incentive-compatible and Effective Anti-corruption Measures, grant proposal accepted by the Grant Agency of the Czech Republic. Grant No. 402/04/0167.

Ortmann, A., Tichy, L., K., (1999). Gender Differences in the Laboratory: Evidence from Prisoner's Dilemma Games, *Journal of Economic Behavior & Organization 39*, 327-339.

Richmanová, J., (2006). In Search of Microeconomic Models of Anti-Corruption Measures – A Review, *CERGE-EI Discussion Paper No. 2006-157*.

Richmanová, J., Ortmann, A., (2008). A Generalization of the Buccirossi & Spagnolo (2006) Model, *CERGE-EI Discussion Paper No. 2008-194*.

Roth, A., (2002). The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics, *Econometrica 70(4)*, 1341-1378.

Rydval, O., (2007). The Impact of Financial Incentives on Task Performance: The Role of Cognitive Abilities and Intrinsic Motivation, *PhD dissertation, CERGE-EI.*

Spagnolo, G., (2004). Divide et Impera: Optimal Leniency Programs, *C.E.P.R. Discussion Paper No. 4840.*

Swamy, A., Knack, S., Lee, Y., Azfar, O., (2001). Gender and Corruption. *Journal of Development Economics 64(1)*, 25-55.

Tanzi, V., (1998). Corruption Around the World: Causes, Consequences, Scope and Cures. *IMF Working Paper 98/63.*

Williamson, O., (1983). Credible Commitments: Using Hostages to Support Exchange. *American Economic Review 73*, 519-540.

APPENDIX 1

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

In Figure 7, we present the data from the before- and after-the-switch-of-roles periods for low and high endowments of the B treatment.
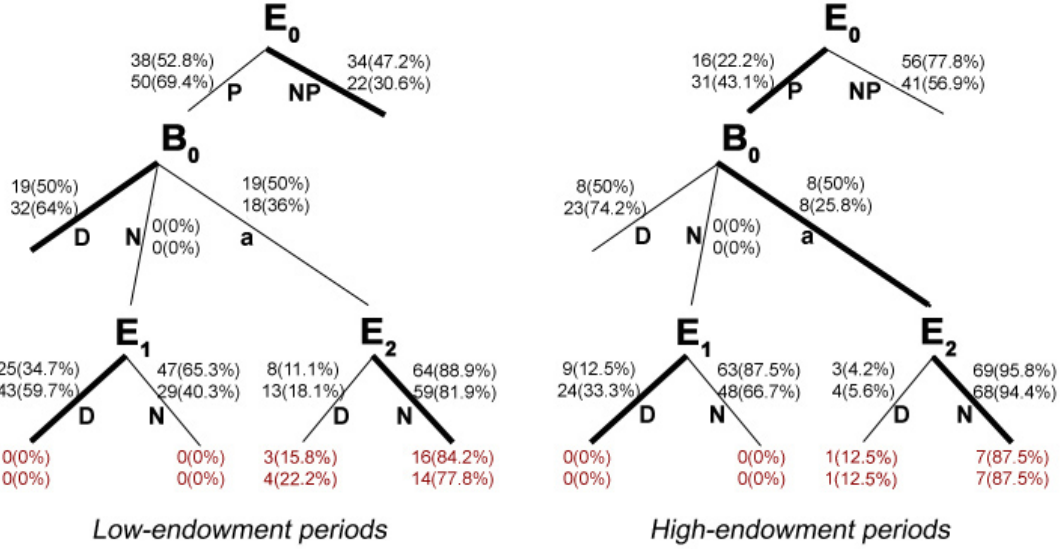

Figure 7: Before- vs. after-the-switch-of-roles data in the B treatment. Before-switch data are in the upper rows and after-switch data are below.

In both cases, we observe a somewhat higher transfer rate in the second six periods. Similarly as in the first part of the experiment, the transfer rate is higher in periods when the endowment is low than when it is high. In the B_0 node, more subjects chose the 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 a shift further away from, the theoretical prediction. In the E_2 node, results from before- and after-switch data are very similar and for both low and high endowment, and they are in line with the theoretical prediction. In the E_1 node, we observe a shift towards the equilibrium after the switch of roles.

In Figure 8 below, we present the data from before- and after-the-switch-of-roles

periods from the low- and high-endowment periods of the C treatment.

In the C treatment, the transfer rate drops after the switch of roles, more so in periods when the endowment is high. This is just the opposite effect as in the B treatment. The transfer rate is higher when the endowment is low in both cases, before and after the switch of roles, which contradicts the theoretical prediction. In the B_0 node, a higher fraction of subjects chose the 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 a shift further away from, the theoretical prediction. In the E_1 and E_2 nodes, the results from before- and after-switch data are similar for low-endowment periods (more so in the E_1 than in the E_2 node). In high endowment periods we observe no difference at all.
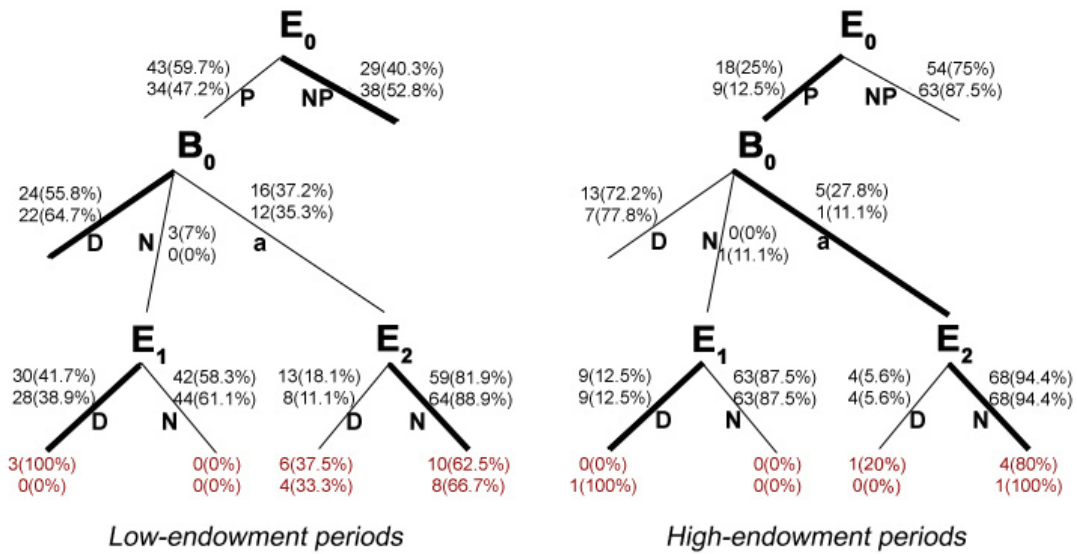

Figure 8: Before- vs. after-the-switch-of-roles data in the C treatment. Before-switch data are in the upper rows and after-switch data are below.

35

APPENDIX 2

Robustness checks

We performed two types of robustness check of our estimation results. The first regards the way we treated individual observations over rounds when running regressions – this is discussed in the subsection *Handling of the data*. The second regards the experimental design – we run several sessions of alternative treatments in which we introduce only minor changes that do not appear to significantly affect behavior of subjects – this is discussed in the 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 is a 0/1 dummy variable identifying transfer being made (value of 1) or not (value of 0) in a particular period.

Atranslog is the average value of *translog* for one individual over periods $1, 3$, and 5 (low-endowment periods) or $2, 4$, and 6 (high-endowment periods).

Ltranslog defines the dominant choice of a subject in periods $1, 3$, and 5 or $2, 4$, and 6. For a subject who has chosen *Pay* two or three times out of a 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 a clustered (robust)[45] linear probability model estimation with the binary variable *translog* as a dependent variable.

---

[45]Standard errors are corrected for heteroskedasticity and for within-subject correlation.

Regressions on Averaged data – in this case, we run an ordinary least squares estimation of *atranslog*. We analyze only averaged data, where higher values of *atranslog* correspond to more transfers being made and thus to a stronger preference for this choice.[46]

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 the part of the data that is not affected by the experience from previous rounds.[47]

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.

First we look at effect size measures, whether they give robust results for all four approaches to the data. The results are summarized in Table 6 below.

		B		C		effect size	
	Data	mean	std.dev.	mean	std.dev.	odds ratio	Cohen's d
1,3,5	1st period	.583	.5036	.625	.4945	1.072	.0841
	average	.528	.4495	.597	.4282	1.131	.1571
	dominant	.5	.5108	.583	.5036	1.166	.1635
	all periods	.528	.5027	.597	.4939	1.131	.1385
2,4,6	2nd period	.292	.4643	.25	.4423	0.856	-.0926
	average	.222	.3764	.25	.3147	1.126	.0807
	dominant	.25	.4423	.25	.4423	1	0
	all periods	.222	.4187	.25	.4361	1.126	.0655

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 the effect is zero (but we need to keep in mind that only part of the available information is used). Except for 2nd period data, also the direction of effect is the same in all cases.

---

[46]We also run poisson regressions on a count variable (counting the number of transfers made by an individual in the relevant three periods). The qualitative results are the same as with OLS and *atranslog*.

[47]We realize that for 2nd period data this may not be completely true if subjects fail to realize that it is a different game they are playing in the high-endowment periods.

This suggests that initially, the transfer 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 robust for all four approaches – suggesting a 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. For all four approaches, the models that 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.

	Periods 1,3,5							
	clustered		averaged		1st period		dominant	
age	-.0287	.1280	-.0287	.1280	-.0069	.1255	-.0470	.1141
	(.302)	(.007)	(.317)	(.011)	(.822)	(.034)	(.160)	(.058)
male	.0686	3.3442	.0686	3.3442	.2748	2.8556	.0758	3.4441
	(.646)	(.010)	(.656)	(.015)	(.117)	(.067)	(.664)	(.042)
econ	-.1601	-.6307	-.1601	-.6307	-.1269	-.3627	-.1529	-.8570
	(.212)	(.000)	(.226)	(.000)	(.416)	(.313)	(.342)	(.000)
Ctreat	.0559	-.7156	.0559	-.7156	.0381	-.7183	.0733	-.6802
	(.657)	(.004)	(.666)	(.006)	(.794)	(.007)	(.621)	(.010)
age*male	-	-.1852	-	-.1852	-	-.1446	-	-.1983
		(.002)		(.003)		(.035)		(.009)
econ*male	-	.5354	-	.5354	-	.2467	-	.8200
		(.002)		(.004)		(.534)		(.000)
Ctreat*male	-	.7983	-	.7983	-	.8433	-	.7390
		(.006)		(.010)		(.009)		(.025)
const	1.2342	-1.4593	1.2342	-1.4593	.6265	-1.7304	1.6046	-1.0004
	(.068)	(.139)	(.077)	(.162)	(.382)	(.194)	(.047)	(.444)
mean $\widehat{p}$(y=1)	.5625	.5625	.5625	.5625	.6042	.6042	.5417	.5417
# of obs.	144	144	48	48	48	48	48	48
joint p-value	.488	.000	.519	.000	.370	.001	.370	.000

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

For the low-endowment periods, the results from the *averaged*, 1st *period* and *dominant-choice* data analysis confirm the results from the *clustered* regressions. We find the directions of all the effects the same, the explanatory variables are significant in most cases and there are no dramatic differences in coefficient sizes. Only *econ* and *econ*male* are not significant in the 1st *period* data case. They

both become significant once we include the information from later rounds – for *clustered*, *averaged* and *dominant-choice* data.

	Periods 2,4,6							
	clustered		averaged		2nd period		dominant	
age	.0220	.0913	.0220	.0913	.0052	.0456	.0486	.1369
	(.381)	(.000)	(.396)	(.000)	(.873)	(.397)	(.108)	(.000)
male	.1706	2.5462	.1706	2.5462	.2312	2.0233	.1133	3.2569
	(.055)	(.014)	(.063)	(.021)	(.051)	(.241)	(.327)	(.010)
econ	-.0731	.2210	-.0731	.2210	-.1495	.1772	-.1039	.3316
	(.503)	(.001)	(.516)	(.002)	(.334)	(.274)	(.483)	(.002)
Ctreat	.0230	-.0375	.0230	-.0375	-.0514	.0480	-.0143	-.0563
	(.809)	(.644)	(.815)	(.663)	(.699)	(.627)	(.908)	(.663)
age*male	-	-.0941	-	-.0941	-	-.0625	-	-.1215
		(.032)		(.043)		(.397)		(.023)
econ*male	-	-.3395	-	-.3395	-	-.3710	-	-.5021
		(.019)		(.026)		(.133)		(.016)
Ctreat*male	-	.0036	-	.0036	-	-.1750	-	-.0395
		(.983)		(.984)		(.424)		(.863)
const	-.3400	-2.1070	-.3400	-2.1070	.1145	-1.1202	-.8378	-3.1605
	(.553)	(.000)	(.565)	(.000)	(.875)	(.370)	(.214)	(.000)
mean $\widehat{p}$(y=1)	.2361	.2361	.2361	.2361	.2708	.2708	.25	.25
# of obs.	144	144	48	48	48	48	48	48
joint p-value	.078	.000	.095	.000	.175	.027	.183	.000

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

For the high-endowment periods, only the results from *averaged* and *dominant-choice* data analysis confirm the results from *clustered* regressions – the treatment dummy is not significant, neither is its interaction term, the directions of all the effects are the same, and the sizes of the coefficients are comparable. For the 2nd *period* data the estimated model is not significant. This suggests that the behavior in the second period is different, and more difficult to explain 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.

B. Pooling the sessions

In addition to the benchmark treatment B we conducted two plus two 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, with six subjects in the role of Participant X and six subjects in the role of Participant Y. The computer program always played a (subgame perfect) optimal strategy. Subjects 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 the instructions, a so-called Backwards Induction Tutorial, intended to explain the basic principles of using backwards induction.

In addition to the full-context C treatment, we conducted two sessions with partial context – the 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 – the 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 the distributions of choices – Fisher's Exact test and Wilcoxon rank-sum test. We find no evidence of significant differences in the distributions of choices between A, AI and B treatments, nor between C- and C treatments.

Afterwards, we performed two types of pooled analysis: 1) pooling the data from A and B treatments vs. pooling the data from C- and C treatments; and 2) pooling the data from A, AI and B treatments vs. pooling the 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- and 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, the significance of *econ* (and its interaction term) disappears.

	Periods 1,3,5					
	B vs. C		B,A vs. C,C-		B,A,AI vs. C,C-	
age	-.0287	.1280	-.0191	.0854	-.0093	.0784
	(.302)	(.007)	(.380)	(.073)	(.641)	(.040)
male	.0686	3.3442	.0162	2.4523	.0076	2.2189
	(.646)	(.010)	(.890)	(.039)	(.940)	(.030)
econ	-.1601	-.6307	-.1754	-.3061	-.1343	-.0944
	(.212)	(.000)	(.089)	(.025)	(.164)	(.532)
CCtreat	.0559	-.7156	.0609	-.2708	.0736	-.2738
	(.657)	(.004)	(.550)	(.218)	(.449)	(.178)
age*male	-	-.1852	-	-.1235	-	-.1080
		(.002)		.(023)		(.018)
econ*male	-	.5354	-	.1801	-	.0005
		(.002)		(.318)		(.998)
CCtreat*male	-	.7983	-	.3586	-	.3877
		(.006)		(.154)		(.101)
const	1.2342	-1.4593	1.0848	-.9740	.8308	-.9774
	(.068)	(.139)	(.035)	(.339)	(.071)	(.243)
mean p(y=1)	.5625	.5625	.5787	.5787	.5714	.5714
# of obs.	144	144	216	216	252	252
joint p-value	.488	.000	.439	.066	.675	.2194

Table 9: Results from estimation on basic vs. extended data sets for low-endowment periods. CCtreat is a dummy identifying context-type treatment(s) – C, or C and C- treatments.

	Periods 2,4,6					
	B vs. C		B,A vs. C,C-		B,A,AI vs. C,C-	
age	.0220	.0913	.0310	.1134	.0253	.0758
	(.381)	(.000)	(.100)	(.000)	(.160)	(.007)
male	.1706	2.5462	.0620	2.3301	.0144	1.4000
	(.055)	(.014)	(.461)	(.001)	(.867)	(.081)
econ	-.0731	.2210	-.1751	-.1029	-.1424	.0227
	(.503)	(.001)	(.070)	(.616)	(.113)	(.876)
CCtreat	.0230	-.0375	-.0780	-.2268	-.1172	-.2955
	(.809)	(.644)	(.331)	(.072)	(.133)	(.026)
age*male	-	-.0941	-	-.1019	-	-.0601
		(.032)		(.001)		(.099)
econ*male	-	-.3395	-	-.0768	-	-.2101
		(.019)		(.740)		(.248)
CCtreat*male	-	.0036	-	.1282	-	.1958
		(.983)		(.436)		(.243)
const	-.3900	-2.1070	-.3103	-2.1335	-.1354	-1.2879
	(.553)	(.000)	(.469)	(.000)	(.737)	(.029)
mean p(y=1)	.2361	.2361	.2593	.2593	.2817	.2817
# of obs.	144	144	216	216	252	252
joint p-value	.078	.000	.045	.000	.075	.040

Table 10: Results from estimation on basic vs. extended data sets for high-endowment periods. CCtreat is a dummy identifying context-type treatment(s) – C, or C and C- treatments.

As regards the treatment dummy, on the one hand, the significance for low-endowment periods disappears, but on the other hand, the treatment dummy becomes significant for high-endowment-period data.