Four Essays on Corruption and Cooperation

Theory and Evidence

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2010

vorgelegt von

Jan Theodor Schikora

Referent:	Prof. Dr. Martin G. Kocher
Korreferent:	Prof. Dr. Johann Graf Lambsdorff
Promotionsabschlussberatung:	9. Februar 2011

Datum der mündlichen Prüfung: 27.01.2011

Namen der Berichterstatter: Martin Kocher, Johann Graf Lambsdorff, Joachim Winter

Acknowledgements

First, I would like to thank my supervisor Martin G. Kocher. Throughout my doctoral studies he supported all my projects and provided me with valuable comments. I am also very thankful to Johann Graf Lambsdorff who gave me priceless insights into the research area in which he truly is an expert. I also thank Joachim Winter who agreed to serve as the third supervisor to my dissertation and supported me with helpful comments.

I would also like to thank my co-author Johannes Maier with whom I learnt a lot about cooperation and efficiency going far beyond the subject of our research project. I have received support and valuable comments from many other friends, colleagues, and participants at conferences and seminars, especially from Klaus M. Schmidt and at the Brown Bag Seminar of MELESSA, for which I am very grateful.

In particular, I am grateful to the participants of the Workshop 'The Economics of Corruption - A University Training in Good Governance and Reform' at the University of Passau 2008, the 'MERSS' summer school at the University of Mannheim in 2007 and the 'BLESS' experimental summer school at the University Pompeo Fabra in 2008. I would also like to thank Toke Aid for his support during my stay at the University of Cambridge as a visiting PhD student. For financial support I would like to thank the German Research Foundation (DFG) and the Dekanat of the University of Munich. As my direct superiors I thank Florian Englmaier and Joachim Winter for their trust and confidence. I also thank the entire staff involved in conducting my experiments at MELESSA.

Lastly I thank Stephanie Penzel and my family, especially my mother, for their support and patience.

Jan Schikora.

Contents

List of Tables

List of Figures

Preface

This dissertation contains four separate studies in the fields of corruption and cooperation. The first chapter explicates the mechanism that links the fractionalization of society with its level of corruption within a theoretical model of relational contracting. The other three chapters describe experimental studies. The second and third chapters evaluate two popular anti-corruption policies, the 'Four-Eyes-Principle' and 'Whistle-Blowing', with respect to their effectiveness in decreasing the level of corruption and increasing social welfare. The last chapter considers the effect of endowment uncertainty on cooperative behaviour in a linear public goods game and explains it by specific conditional cooperation preferences.

Cooperation between decision-making agents is recognized as one of the single most important mechanisms in economic research and represents one of the cornerstones of economic development. Countless economic activities have been analysed with game theoretic models of cooperation. Experimental methods may not only demonstrate the deficiencies of standard economic theory in terms of explanation and predictive power, they may also help to improve existing models.

The public goods game (Isaac and Walker 1988) represents one of the most popular vehicles to experimentally analyse cooperative behaviour. It models the dilemma of the opposition between selfish preferences and social efficiency. Numerous experiments have shown that participants behave in a highly cooperative way in situations for which the standard economic theory of rational payoff maximization predicts strictly selfish behaviour. In my view, the most convincing approach to explaining the phenomenon of cooperation is the existence of conditional cooperative behaviour. This links the public goods game to games that are specifically designed to analyze trust and reciprocity, e.g. the gift exchange game (Fehr et al. 1993), the investment game (Berg et al. 1995) and the ultimatum game (Güth et al. 1982). Applications of these games even extend to criminal activities such as corruption. Administrative corruption, defined as 'the misuse of public power for public gain' (Klitgaard 1988) is recognized as the 'single most important obstacle to economic development' (The World Bank 2008). The key to effective anti-corruption policy with respect to institutional design is the understanding of the determinants, mechanisms and weaknesses of corruption. This demands the analysis of two different levels of cooperation common to most corrupt transactions.

Since any corrupt relationship is by definition illegal, the corrupt partners cannot rely on legal third parties, i.e. the courts, to enforce their contracts. The transaction between a client and a corrupt official depends on trust and reciprocity which may be fostered for example by repeated interaction. This kind of cooperation is similar to the mechanism modelled in the gift exchange game. In contrast to the original gift exchange game, where cooperation is efficient with respect to social welfare (measured e.g. in the sum of payoffs to all affected individuals), corrupt reciprocity is socially undesirable due to the reasonable assumption of its strong negative externality to the public (Shleifer and Vishny 1993, Rose-Ackermann 1999). In the case of instable corruption, it would be socially optimal for all agents to stay away from reciprocity and cooperation, and adhere to their selfish rationality. The situation involving all members of a society with respect to their choices for or against corrupt reciprocity can hence be seen as a reverse public goods dilemma in a broader sense.

The first chapter focuses on the role of the fractionalization of a society in determining the level of corruption. In a series of empirical cross-country studies social fractionalization, often (crudely) measured by the Ethno-Linguistic Fractionalization Index (ELF, see Appendix 1A) has been identified as an important determinant of the level of corruption measured by the Corruption Perception Index or similar indices (Alesina and Ferrara 2002, Mauro 1995, Bardhan 1997). In a cross regional analysis, providing a more controlled environment, Dincer (2008) finds an inversely U-shaped relationship between the two variables. However, none of these studies provides a model based theoretical explanation for the empirical evidence.

As a basis of our analysis we use an infinitely repeated version of a standard, multi-stage game of administrative corruption which captures the enforcement problem of the illegal transaction. In order to describe the main mechanism underlying the relationship between the social structure and corruption we define the structure of society in terms of its fractionalization as a vector of separated sub-networks whose members share information. Assuming that clients use simple punishment strategies as devices to maintain corrupt cooperation within relational contracting, we find that the maximum level of corruption is to be expected in societies that consist of a large but not maximal number of small (but not minimal) groups. This is due to the inversely U-shaped relationship between the relative size of a sub-network (measured in the number of group members relative to the size of society) and its members' ability to stabilize corrupt transactions. The (relative) size of a sub-group has two countervailing effects on the corruption level. On the one hand, the (average) probability of a successful corrupt transaction (expected frequency) increases in the number of group members. This is because the incentives for opportunism decrease due to growing stakes for the official. On the other hand, an increase in the relative sub-group size increases the (personal) costs for the clients through the internalization of a larger part of the negative externality. Thus, necessary compensation of a growing number of peers decreases the profitability of corruption.

This chapter provides a model-based explanation of the inversely U-shaped relationship between social fractionalization and corruption found in Dincer (2008). The results of our model are also in line with empirical observations of cross-country comparisons (Gunasekara 2008, La Porta et al. 1999 and Alesina et al. 2003). Our model can be extended to account for considerations of the influence of different types of social capital on corrupt behaviour.

Using the standard model of self-interested payoff maximization to analyse the mechanisms that underlie the determinants of corruption may only be reasonable in situations in which limit values and benchmark examinations are considered, and therefore simplifying assumptions such as infinite repetitions are justified. In finitely repeated interaction (or one shot games) of corruption, neither the standard self-interested model nor models of strong reciprocity relying on social preferences such as altruism (Andreoni and Miller 2002), inequity aversion (Fehr and Schmidt 1999) or intentions (Dufwenberg and Kirchsteiger 2004) can provide a consistent explanation or predictions for corrupt behaviour based on cooperation. This and the scarcity of reliable field data complicate any analysis of corruption that aims at deriving policy implications. In order to evaluate the usefulness of proposed anti-corruption measures, or those that are in force, controlled experiments are indispensible to complement empirical studies. The findings of several studies, e.g. Armantier and Boly (2008), demonstrate the external validity of laboratory corruption experiments (Dusek et al. 2004).

Following Abbink et al. (2002) with respect to experimental methodology, the second chapter describes an experimental approach to assess one of the most potent counter-corruption policy tools, the 'Four-Eyes-Principle' (business has to be conducted by at least two individuals, hence four eyes). Although proposed in various reports and lists of recommendations for anti-corruption measures by national and international organizations, it lacks any theoretical or empirical justification (Pörting and Vahlenkamp 1998, Rieger 2005, Wiehen 2005). In our laboratory experiment we replace a single decision-maker with a small group of two officials who decide jointly in the role of the official in order to model the introduction of the 'Four-Eyes-Principle'. We show that the introduction of the 'Four-Eyes-Principle' can lead to an increase rather than a decrease of the level of corruption. This result comes as a surprise when considering predictions from the standard self-interested model alone and ignoring effects stemming from the dynamics of group decision-making. Controlling for effects that are purely driven by differences in marginal incentives (i.e. effects stemming from the splitting of the benefits and costs of corruption between two officials) we find that group decision-making is dominated by the motive of individual (long term) profit maximizing, which has been identified as a main explanation for group decision-making (Kocher and Sutter 2007). Combining data of final choices (outcomes) with evidence from inside the decision-making process (i.e. the dynamics of individual choices), we show that groups follow strategies that foster reciprocity in a more sustainable way than their individual counterparts, which leads to a higher number of successful corrupt transactions. To explain the behavioural characteristics in more detail, we analyse the content of electronic chat messages exchanged during the joint decision-making process. In an average situation of disagreement between two officials, it is the official with the more corrupt agenda who dominates the decisions despite the honest official's veto power. Since corrupt reciprocity maximizes individual payoffs for the immediate transaction partners, arguments in favour of a strategy that fosters corrupt reciprocity are most persuasive. We interpret this result as support for the Persuasive Argument Theory attributing groups a higher ability to adhere to behaviour that generates maximum individual payoffs (Pruitt 1971). Our results show that the profit maximizing motive may dominate in the group decision-making process even though there is an obvious trade-off with social efficiency. Since they ignore the negative external effects of corruption, groups produce the least desirable outcomes in terms of welfare by their reciprocal behaviour. Against existing policy recommendations, the results of our experiment cast doubt on the usefulness of the introduction of the 'Four-Eyes-Principle'. This chapter does not only evaluate an anti-corruption measure, it also provides insights into the motivations, within groups and individuals, that underlie strategic decisions in social dilemmas in general and in the dilemma of corruption in particular.

The third chapter experimentally assesses the effectiveness of another tool of anti-corruption policy. The institutional enabling of whistle-blowing is seen as a powerful measure to contain corrupt activity (Drew 2003). Whistle-blowing is generally defined as 'the act of disclosing information in the public interest'. Despite its widespread use and perceived success, experimental evidence seems to cast doubt on its usefulness (Abbink 2006, Lambsdorff and Frank 2010, forthcoming). This may be because the analysis of whistle-blowing within the standard set-up of a corrupt transaction accounts for only one aspect of its total effect on social welfare. The standard game of corruption, often used to analyse the effect of a determinant on the frequency of corruption, only considers the direct but not the indirect consequences of corruption on social efficiency. While the direct consequences may be captured by the expected negative externality to the public resulting from a successful corrupt transaction, the indirect consequences include efficiency losses caused by honest clients leaving productive markets because of their fear of being exploited by corrupt officials.

Where in the standard game of corruption the official has a passive role with respect to the initiation of a corrupt transaction, we expand the standard model by allowing both sides of the transaction to activate corruption. The symmetry of corrupt engagement enables us to assess the potential effects of the introduction of the opportunity to blow the whistle on the two main negative consequences of corruption. First, whistle-blowing may reduce the negative effect of corruption hindering 'honest' clients to engage in productive activity by providing a tool against demanding corrupt officials. Second, whistle-blowing may affect the stability of the corrupt transaction and influence the number of successful deals and hence the amount of realized negative externalities.

We find that the total effect of symmetrically punished whistle-blowing (i.e. the punishment is independent of who has blown the whistle) is ambiguous. Confirming the findings of Lambsdorff and Frank (2010) and Abbink (2006), whistle-blowing increases the stability of a corrupt transaction. However we find that it also reduces the effect of corruption deterring productive activity, offering a safeguard for an 'honest' client against the extortion by a corrupt official. We demonstrate that asymmetric leniency for the official can offset the negative effect of whistle-blowing. Our results can be explained using simple arguments as to subjects' belief structures and payoff maximization. Moreover, we find that leniency is especially effective for male officials. The consideration of asymmetric punishment of illegal activities in general and leniency for whistle-blowing in particular should be considered in legislature. Our extended model of corruption provides the basis for experimental research targeting both direct and indirect effects of corruption.

While the focus of the first three chapters is on the negative consequences of cooperative behaviour in corruption, the fourth chapter, which is joint work with Johannes Maier, considers socially desirable cooperation. In an experimental public goods game using the Voluntary Contribution Mechanism (Isaac and Walker 1988) we study the effect of uncertainty as to others' endowments on contribution behaviour. In most applications of public goods provision it is more realistic to assume heterogeneity instead of homogeneity of initial endowments between cooperation partners. The own endowment can be private information, which means that endowment levels of fellow group members can be unknown. In situations of charitable giving, for example, endowments (individual wealth levels) are likely to be heterogeneous and information about them remains private, while information about actual contributions (donations) are often made publicly available. To quantify and explain the effect of endowment uncertainty on cooperative behaviour we use an adapted version of the experimental two-stage approach used by Fischbacher and Gächter (2010).

In the first stage of our experiment, subjects had to state their contribution preferences conditional on their group partners' contributions *and* endowments. Here we used an incentivized strategy method (Selten 1967). In the second stage, we quantify the effect of uncertainty in a repeated linear public goods game (ten periods) played in partner design with groups of three participants. While subjects knew their own endowment and the endowments of their two group partners in the certainty treatment, they only knew their own endowment, low or high, in the uncertainty treatment. However, they knew that the others were either high or low endowed. When we pool all observations of each treatment, we only find a small negative but insignificant effect of uncertainty on average contribution levels. When we separate observations according to participants' endowments, we find that subjects with high initial endowments contribute slightly more, while participants with low endowments contribute substantially less under uncertainty. These two opposing effects of uncertainty lead to lower contribution levels in poor and higher contribution levels in rich groups. The inequality in income levels between low and high endowed subjects therefore increases through uncertainty.

We explain our treatment effects by two mechanisms, the effect of deviating beliefs and the net effect of (strategic) over-contribution. We attribute both effects to conditional contribution preferences. One of our main results is that subjects are relative conditional contributors. In the context of heterogeneous endowments this means that they contribute more, the lower their group partners' endowments holding their absolute contributions constant. This and the findings of systematically deviating beliefs explain the former of the two mechanisms. Under uncertainty, low endowed subjects believe that they are in a richer group than they actually are and therefore contribute less in the repeated public goods game than they would have done, had they known the correct endowments of their group partners. High endowed subjects, on the contrary, believe that their group members are poorer on average and hence contribute more. The preference for relative conditional cooperation also explains the treatment differences in (strategic) over-contribution that remains when we substract the effect of deviating beliefs. The intuition for (strategic) over-contribution is that subjects contribute higher levels in repeated games than their stated preferences should allow in order to trigger positive reciprocity and thereby sustain cooperation (see e.g. Fischbacher and Gächter 2010). In contrast to groups in the certainty treatment and groups consisting of high endowed individuals in the uncertainty treatment, we do not find (strategic) over-contribution in groups of low endowed individuals under uncertainty. We attribute this to their fear of sending the wrong signals (i.e. being high endowed) by making large contributions. This fear may be due to participants' anticipation of others' relative conditional cooperation preferences. Overall, the combination of the two mechanisms explains a large fraction of our treatment effects. This paper not only explicates contribution behaviour under uncertainty, it also expands the knowledge of conditional cooperation preferences in general. Its results motivate future research on the theoretical foundations of conditional cooperation.

All four chapters contain their own introductions and appendices so they can be read independently.

Chapter 1

How do Groups stabilize Corruption?

1.1 Introduction

Confirming prevailing opinion, resigned World Bank President Paul Wolfowitz once called corruption 'the single most important obstacle to development' (The World Bank 2008). Most empirical as well as theoretical studies share the view that the negative effects dominate the positive effects of corruption (Huntington 1968 and Leff 1964, Aidt 2003, Bardhan 1997). The negative consequences of widespread corruption range from effects stemming from rent-seeking (Lambsdorff 2002), over the deterrence of (foreign) direct investments (Mauro 1995, Egger and Winner 2006), endogenous production of red tape (Guriev 2004) to inefficient (re)allocation of resources (Bohn 2004, Bertrand et al. 2007). These negative effects can be generalized as direct or indirect negative externalities to the public, created by the realization of a corrupt transaction.

Given any institutional set-up that is not able to prevent the occurrence of corruption, which we define as 'the misuse of entrusted power for private gain' (Klitgaard 1988), completely, the level of corruption depends critically on the stability of the illegal transaction and the degree of internalization of its negative externalities. Although it is of critical importance to the determination of the level of administrative corruption found in a country or region, the level of internalization of the negative external effects has received only limited attention in the literature. Rose-Ackerman (1999) notes that the internalization of the negative externality is likely to be connected with social fractionalization. She attributes this to information transmission and mutual cognizance of illegal activity between sub-group members. Moreover, there is anecdotal evidence that the composition of a society can significantly affect the ability of criminals to stabilize their corrupt transactions using their sub-group (network) membership (Schramm and Taube 2001). The empirical corruption literature provides some evidence on a causal relationship between the fractionalization of a society and its level of 'perceived' corruption[1]. Mauro (1995), Gunasekara (2008) and Alesina et al. (2003) find a positive correlation between the fractionalization of a society (country) with respect to its ethnic structure measured by the Ethno-Linguistic Fractionalization Index (ELF), see Appendix 1A, and the level of corruption measured in CPI (or comparable subjective indices). The relationship is found to be non-linear when shifting the analysis to a more controlled environment, considering cross-regional data. Using data from 48 US-American states, providing high control of (secondary) determinants (e.g. income, education etc.), Dincer (2008) finds an inversely U-shaped (hump-shaped) relationship between an index of fractionalization similar to the ELF and corruption measured by the relative numbers of convictions. States that are highly but not totally fractionalized seem to exhibit the highest levels of corruption. States with little or extremely high fractionalization show significantly lower levels. Although there are several approaches to explain the underlying mechanisms, e.g. through in-group favouritism or joint bribe maximization (Shleifer and Vishny 1993), there is no theoretical model-based explication of the empirical evidence.

In this paper we explain the hump-shaped relationship between social fractionalization and corruption in the framework of the New Institutional Economics of corruption. We define social structure as the number and size of (unconnected) informational sub-networks. These networks can be interpreted as the result of overlapping radii of trust (Realo et al. 2008, Fukuyama 1995).[2] We apply a simple repeated multi-stage game which centres around the commitment problem of the partners of a corrupt transaction being solved by the application of unforgiving punishment strategies in repeated interactions (Pechlivanos 1997). The focus is on the interaction between the effects of the existence and shape of sub-networks on the internalization of the negative external effect and the

[1]Most empirical studies work with indices, e.g. the Corruption Perception Index (CPI), published by Transparency International (TI). The CPI is based on the subjective evaluation by international as well as local experts on the level of corruption in a certain country or region (see Transparency International 2005 for a detailed description of the index).

[2]In a crude way (losing some of the information) the structure of society can be mapped into the one-dimensional index of Ethno-Linguistic Fractionalization (*ELF*, see Appendix 1A and Alesina et al. 2003).

stabilization of illegal corrupt transactions.

Analysing simple benchmark punishment strategies, we find that the variation in the size of a group (which defines the number of agents committing to a certain punishment rule) has two countervailing effects on the ability of individual sub-group members to stabilize their corrupt transactions. On the one hand, a larger group provides its members with additional power to stabilize a corrupt deal by raising the stakes (and thereby the threat-point) for a potentially defecting public official. On the other hand, the amount of internalization of the negative external effect, which is a by-product of cognizance, and thereby the cost of corruption is increasing with the size of a sub-network. We show that the balance between these two countervailing effects causes small but not minimal groups to maximize their members' ability to stabilize corruption. Through the implications of this result on the structure of an entire society, our model provides a theoretical rationalization of the empirical observations.

The remainder of the paper is structured as follows. In Section 1.2 we introduce the basic model of corruption and explain the main mechanism of bilateral punishment strategies. Section 1.3 introduces the existence of sub-networks and describes the main consequences of group-enforcement through multilateral punishment strategies in the infinitely repeated game. Section 1.4 summarizes and concludes.

1.2 Model

In its most general form a corrupt transaction is best described as a Principal-Agent-Client relationship which can be broken down into two distinct Principal-Agent (P-A) problems (Lambsdorff 2007). In the first, a benevolent government, representing the principal, delegates a (perfectly defined) task to his imperfectly controllable agents represented by potentially corrupt public officials.[3] This P-A problem is in essence not specific to corruption. Hence the New Institutional Economics (NIE) of corruption focuses on the P-A problem between the public official and a private entity (e.g. a firm) that may be willing to pay for preferential treatment. A central aspect of the analysis is

[3] With this specifications we exclude political or grand corruption.

the enforcement problem of a corrupt transaction.

1.2.1 Framework

Consider a society that consists of a small number (> 1) of potentially corrupt public officials (O) and a large number (N) of individuals (clients). All clients are potential bribers (B).[4] Assume that clients need some kind of permit (e.g. a licence) given out by the officials in order to engage in any kind of economic (e.g. productive) activity. Legally proceeded transactions never yield positive payoffs (e.g. the government-set price for the licence is equal to the net expected payoff for any given transaction). In terms of efficiency as well as social welfare, a licence should only be given out in case client B satisfies a set of conditions. Conditions are exogenous (defined by the benevolent authority) and can be thought of as safety requirements, quality standards, etc. Officials are required (by the benevolent authorities) to test whether conditions are met by the client and subsequently give out the licence or reject the request. Obtaining the licence without having to satisfy the set of conditions gives return E to the client. E may include opportunity costs of time or monetary costs saved e.g. by using sub-standard quality etc.[5] Hence, the client may have an incentive to distort the behaviour of the official who is neither controlled nor monitored perfectly (by the authorities)[6] by paying a bribe b.

A successfully completed corrupt transaction causes the negative externality D. D is directly proportional to the return of corruption ($D = iE$) where i depicts the factor of inefficiency common to all corrupt transactions. The damage D is assumed to spread equally across all (N) members of society. The assumption of a flat distribution of a fixed level of damage is a simplification that can be rationalized by considering personal damage as the certainty equivalent of the expected risk of damage caused by the realization of the corrupt transaction.

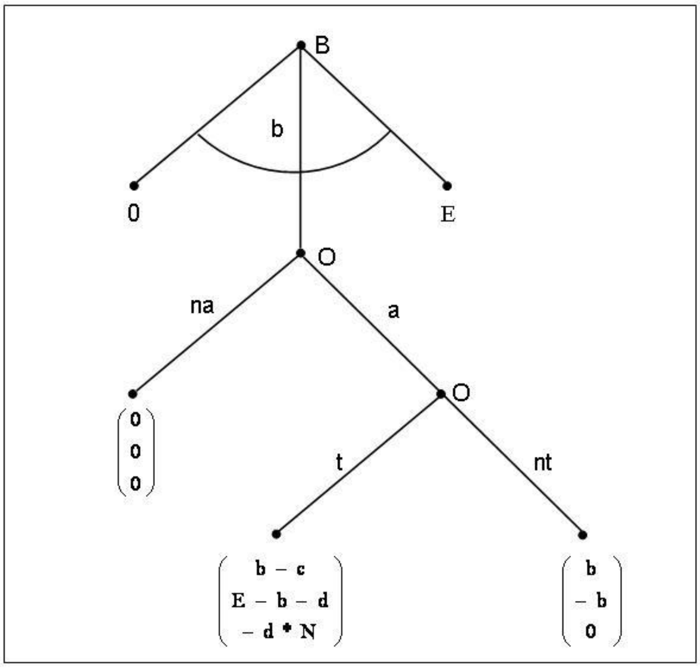
[4]This assumption follows the hypothesis of money maximizing individuals conducting illegal activity whenever this yields a positive payoff (Becker 1968).

[5]We assume that B's only motivation to satisfy the obligations set by the authorities is to receive the licence which enables production. There is no intrinsic motivation to satisfy the legal conditions. The reason for B's sub-optimal choice of production technology (in equilibrium) lies in the non-internalized external effect.

[6]We simplify our analysis considerably by assuming that the official does not face the risk of being fired.

The per-head-damage can be written as $d = \frac{D}{N} = \frac{iE}{N}$.
Setting $i > 1$ ensures that the total damage (in form of the externality) caused by a successful corrupt deal is always larger than the sum of benefits for the corrupt partners. From a social point of view (welfare perspective), O should always force all clients to satisfy government-set (and first-best-results inducing) requirements. This constitutes corruption as socially undesirable.
Delivering the corrupt service in exchange for a bribe (cooperating in the corrupt transaction) causes costs c to the official. These costs may include the moral costs of being responsible for causing damage to the public (fellow citizens) and real (technical) costs of hiding illegal activities from the authorities. Note that in this set-up the amount of these costs cannot be modified by the official. Nor do they depend on the profitability of the corrupt transaction (E). Costs c could also be interpreted as the certainty equivalent of the lottery between no punishment in the good state and (e.g. monetary) punishment in the bad state of the world (i.e. detection by the authorities).[7] By incurring c, the official is able to guarantee return E to the client.[8] Consider Figure 1.1 for the timing and payoff structure of the 3-stage game.

Figure 1.1: Extensive Form



[7]The assumption of constant costs of corruption serves as a simplification and may be changed in a more comprehensive model in which O may (endogenously) choose the amount of c to determine the probability of detection.

[8]This implies that the probability of being detected is assumed to be 0 if costs c are paid.

The first line in the payoff vectors depicts the payoff for the official, the second line that for the briber and the third line shows the (total) monetary external effect on the public. In Stage 1, client B offers bribe $b \epsilon [0, E]$[9] to the official O. The bribe can either be accepted or rejected (by O) in Stage 2. A bribe of 0 may be interpreted as a decision of honest acquisition of the licence. If O rejects the bribe (no acceptance, action 'na'), the game (round) ends and both agents receive their reservation utility of 0.[10] If O accepts the bribe (action 'a') she enters Stage 3 and decides whether to provide the corrupt service ('t') or to defect and pocket the bribe ('nt'). If O fulfils the costly task (of delivering the corrupt service to B), B's return E and the negative externality $D = d * N$ is realized.

It can easily be shown (see Appendix 1B for a short proof) that in a one shot game the only Sub-game perfect Nash Equilibrium (SNE) is characterized by B playing $b = 0$ (being honest) in Stage 1, rationally expecting O to choose 'a' in Stage 2 and 'nt' in Stage 3 for any amount of b. This is because O cannot credibly commit to cooperate. There is no corruption and no inefficiency without third party contract enforcement. Since most forms of corruption are illegal, the contractual exchange of bribes for illegal services cannot be protected or enforced by legal third parties (i.e. the courts) and thus relies on other forms of contract enforcement (Lambsdorff et al. 2005).

1.2.2 Stabilizing corruption through repeated interaction

Among several theories on possible ways of stabilizing the corrupt transaction provided by the economic literature on corruption (see Appendix 1D for a short summary) repetition of the transaction seems to be the main driver of stability (Schramm and Taube 2001, Pechlivanos 1997). Consider the infinitely repeated game ($G_\infty(\delta)$), where δ is the discount factor for future periods, common to all agents. All agents maximize their discounted present value of the stream of expected monetary payoffs over their infinite life time by choosing actions from their action space, $max_{a_i^t} \; E \sum_{t=1}^{T} \delta^{t-1} g_i(a_i^t(h_{t-1}))$. Each agent has complete and costless access to the entire history of information (h_t) about her own and her (direct) transaction partner's (partners') behaviour. a_i^t is the

[9] We do not assume a budget constraint but restrict the amount of transfer to the gross product (return) of the corrupt transaction.

[10] We assume transaction costs of 0 for simplicity.

action space and $g_i(.)$ the payoff depending on strategies which may be dependent on the past behaviour of (a sub-set of) players, where $i = \{O, B\}$. The period-wise repetition of the SNE of the one-shot game ('b=0'; 'a'/'na'; 'nt') remains an equilibrium in the infinite repetition of the game, but is not the unique equilibrium. Being interested in maximal levels of corruption, we consider (only) the most collusive equilibrium (in pure strategies) that is backed by the strongest of incentive compatible punishment strategies and treat the results as benchmarks.[11]

By applying the Folk Theorem, a corrupt deal can be stabilized by a Grim Trigger Strategy (Axelrod 1984) played by B.

B has to pay a bribe high enough to ensure incentive compatibility for O and low enough to yield a positive payoff for herself. The Grim Trigger strategy of B can hence be formulated as:

'Pay the equilibrium bribe b^* as long as O *has never defected (i.e. played 'nt') before and play '$b = 0$' forever, otherwise.'*

We call it a 'Bilateral Punishment Strategy' (BPS).

The Incentive Compatibility Constraint of O under BPS (ICC_{BPS}) embodies a condition under which the official prefers to deliver the corrupt service rather than defect on B. The expected life time pay-off from cooperation appears on the left and the immediate payoff plus the continuation value from cheating on the right hand side of equation (1.1).[12]

$$ICC_{BPS}: \qquad b - c + \sum_{i=1}^{T} \delta^{i-1}(b_i - c) \geq b + 0;\ T \rightarrow \infty : \frac{1}{1-\delta}(b-c) \geq b \tag{1.1}$$

Solving for the minimal incentive compatible bribe $b*$ yields $b^*_{BPS}(\delta, c) = \frac{1}{\delta}\, c$.

To satisfy ICC_{BPS}, the bribe b has to be sufficiently larger than c. In order to make it worthwhile for the client to enter the corrupt relationship (instead of engaging in honest activity), the Participation Constraint (PC_{BPS}) of B has to be satisfied. Corruption takes place only if the surplus of the briber E is large enough to cover the (equilibrium)

[11] Weaker punishment strategies such as e.g. 'Tit for Tat' (Axelrod 1984) in which only the behaviour of the last period is relevant for the actions played by the relevant agent are therefore not considered in order to keep the analysis as simple as possible.

[12] Note that the negative externality does not occur in any of O's considerations since we use the assumtions of the model of self interested payoff maximization.

bribe $b*$ and *B's* personal expected loss from the negative externality (d).

$$PC_{BPS}: \qquad E - \frac{iE}{N} - b^*_{BPS} \geq 0 \tag{1.2}$$

Solving for the minimal surplus E^* yields: $E^*_{BPS} = c\,\frac{N}{\delta(N-i)}$. E^*_{BPS} can be interpreted as the minimal corrupt return of a transaction for which the briber will rationally choose to bribe the official.

1.2.3 Level of corruption

We define the level of corruption as its (relative) frequency.[13] The minimum return E^* can be taken as a direct measure of the maximum level of corruption for the following reasons. E is likely to differ across types of economic activities. Some types are more suitable for creating the opportunity to extract rents through corruption than others. This is likely to be true across (Bardhan 1997) but also within economic sectors (e.g. special purpose construction offers more scope to extract rents through corruption than the maintenance of infrastructure, Klitgaard 1988).

We assume that E takes values between $\underline{E}$ and $\overline{E}$ according to the distribution function $f(E)$. Furthermore we assume that E is stable across individuals of society (all individuals can, in expectation, extract the same rent from a given type of economic activity through the use of corruption). The expected total number of realized corrupt deals (per capita as well as for the entire society of identical clients when multiplied by N) is captured by the sum of all deals that yield (in expectation) a return $E \geq E^*$. Ordering all economic transactions according to their potential for corrupt rent extraction, the frequency '*Frq*' of per capita (as well as overall corruption within a society) can be defined as $Frq(E^*) = 1 - F(E^*)$, with the cumulative distribution function $F(E^*) = \int_{\underline{E}}^{E^*} f(z)dz$. Independent of the specific form of $f(E)$ it is clear that the level of corruption (per capita) is strictly decreasing in E^* ($\frac{\partial Frq(E^*)}{\partial E*} < 0$).[14]

The time structure of our game implicitly assumes that B holds all bargaining power

[13] This is in line with the notion of CPI being strongly correlated with the relative frequency of transactions involving corruption.

[14] Any cumulative distribution function satisfies: $\frac{\partial F(E)}{\partial E} > 0$.

since she makes a 'take it or leave it' offer. Irrespective of how bargaining power is distributed between client and official, the level of corruption that can be stabilized by some punishment strategy will be defined by the relative mass of potential transactions that yield a corrupt return E that is high enough to pay the equilibrium bribe b^* and compensate for expected personal losses. Hence E^* always defines the maximum level of corruption, independent of the distribution of bargaining power since the transaction with a corrupt return E^* is the 'last' (marginal) type of transaction for which corruption is profitable and will be undertaken. See Appendix 1E for a detailed explanation.

1.3 Analysis

1.3.1 Heterogeneity of social structure

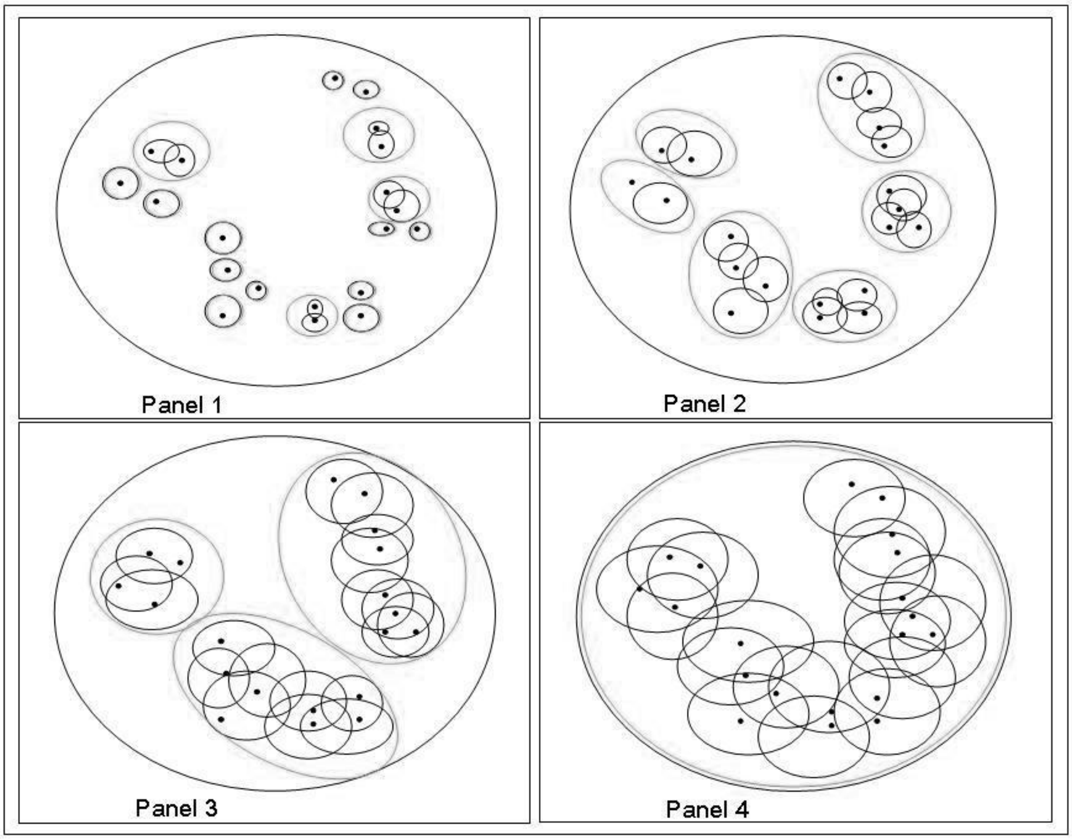
The New Economic Sociology describes corruption as the result of the clash between particularized and universal norms where the institutional economics of corruption distinguishes between trust to insiders and trust (or responsibility) to outsiders (Lambsdorff et al. 2005). In our view trust between individuals ultimately determines the level of corruption through its implications on the structure of society with respect to its fractionalization into sub-groups enabling the exchange of information between its members. Exchange of information enables joint cooperation, but also determines the degree of the internalization of external effects. This idea can be visualized by depicting the intensity of trust as the sizes of the radii of trust around individuals (Fukuyama 1995, 1999, Putnam 2000, Realo et al. 2008, Alesina and La Ferrara 2000). Considering a given pattern of individual proximity[15], the sizes of the individual radii of trust define the shape of society (i.e. the number and size of its sub-groups) in the following way. Two individuals are linked as soon as their radii of trust overlap. A sub-network is defined by all individuals linked either directly or indirectly.

Figure 1.2 contains four panels as examples of social structures defined by different sizes of radii of trust. In Panel 1, very small radii of trust leave society highly fraction-

[15]By given we mean the pre-determined geographical position of each individual on a plain. Fixed proximity is justified by exogenous variables such as heritage.

alized. There are only small sub-groups (depicted by the larger circles), which may be interpreted as nuclear entities (e.g. families). As the radii of trust increase, the size of networks grows, while their numbers decrease (Panel 2 and 3). There are fewer but larger networks sharing trust and information. At a certain size, the radii of trust overlap in such a way that a single network is created which includes all members of society (Panel 4). Social fractionalization graphically depicted by the set union of the plains of radii of trust can be defined by a vector containing the number and size (in the number of individuals) of all separable sub-networks.[16]

Figure 1.2: Examples of group structures



1.3.2 Multilateral Punishment Strategies (MPS)

The existence of sub-networks defined by their members' trust has an important effect on the level of corruption that can be stabilized in our model since mutually trusting

[16]The assumption of exclusiveness of network membership based on trust can be justified by the existence of natural barriers such as language, dialects or cultural differences driven by ethnic or religious differences as they tend to foreclose link formation to outsiders and increase the (prohibitively high) costs of setting up bridging links between members of other sub-networks (Alesina et al. 2003).

individuals are likely to share information. We use a mechanism of contract enforcement close to the one described in Greif (1993), which was motivated by historic (14^{th} century) sources of an ethnic group of Jewish traders (Maghribi). These solved their enforcement problems in their cross Mediterranean trade (modelled as a legal P-A relationship) through a reputation mechanism based on joint punishment strategies enabled through an information network.

The lack of sufficient repetition of bilateral contracting led to the demand for an institution securing mutually profitable trade. The threat of spreading the information of a potential breach of contract between a principal and a specific agent to the entire network increased the stake for any agent decreasing her incentives for opportunistic defection. Defecting on a single network member would have entailed the loss of all potential future gains from trade with any member of the network. The common knowledge (within the network) of the deployment of Multilateral Punishment Strategies (MPS) can explain the larger set of transactions that can be stabilized within the informational network as compared to the use of bilateral contracting alone.

Where Greif (1993) considers a fixed structure of society (the ethnic group of Maghribi Traders), we are interested in the differences in the ability to stabilize corrupt transactions between members of societies of different structures with respect to the existence and size of sub-networks. The necessity to use informal institutions such as Multilateral Punishment Strategies (MPS) to enforce socially efficient trade in medieval times stems from the lack of legal international institutions. These problems are in general solvable by appropriate and functional international legal institutions including commercial courts and law enforcement agencies. The same is not true for corrupt transactions which are illegal by definition.

While Greif (1993) notes that the limit to the (relative) number of deals that can be stabilized under MPS stems from the quality of the available information exchange technology (i.e. the reliance on sending long distance letters) the main limit to stabilization in the set-up of corruption is determined by the (degree of) internalization of the negative external effect. To include the structure of society into our model, consider the simple 3-stage corruption game with the time and pay-off structure described above (Figure 1.1). We assume that the society consists of N bribers to be fractionalized into k groups of size n_i ($i = \{1, 2, .., k\}$), where $\sum_{i=1}^{k} n_i = N$.

To establish the main systematic relation between the properties of a society and the level of corruption, we first consider a specific sub-group of society and denote its size (number of individual members) by n (dropping the subscript i). The crucial feature in our model of social structure is that information flows perfectly inside a sub-network[17]. This means that it cannot be purposefully ignored or limited by any individual member and does not spill over to the members of other sub-groups. As in the Maghribi Traders' problem, sub-groups in our set-up can act like transaction networks. By sharing information about the behaviour of an agent (i.e. the official) and committing to a certain (punishment) strategy, sub-group members are able to increase the expected future pay-off which is at stake for the official in case of misbehaviour (defection).

Again we assume that information storage is perfect in a sense that all potential bribers of a certain sub-group can (free of cost) track the complete history of the behaviour of a specific official O towards any of their fellow sub-group members. However, they do not have any (relevant) information about the official's behaviour towards (any unlinked) outsiders (members of other sub-groups). We assume that the structure of society is constant. Sub-group members cannot be admitted or expelled from a sub-group (links cannot be established or severed). Apart from the history of the behaviour of an official, information transmission in our model involves the details about the illegal deal including the size of the bribe b, the size of the return E and thereby the size of the per-head damage d caused by successful realization. This gives each sub-group member the potential to serve as a key witness in a criminal law suit. To prevent sub-group members from informing the authorities in order to press compensation for their personal damage on a legal basis they have to be compensated for (at least) their expected personal loss caused by the successful corrupt transaction.[18]

It is clear that sub-group members will always accept a 'take it or leave it' offer by the active briber covering their expected total damage. Distributional issues of the extracted rent $(E - b^* - d)$ within a sub-group have no effect on the marginal corrupt transaction that can be stabilized (characterized by E^*). Hence the frequency and level of corruption (per head) within a sub-group of size n is ultimately determined by E^* as long as all

[17]This assumption is similar to the conditions imposed in Greif (1993).

[18]We assume that the expected monetary punishment (lawsuit) is strong enough to ensure that B always (for all relevant values of E) prefers to compensate all sub-group members fully. Hence, retaliations of unsatisfied group members do not occur in equilibrium.

sub-group members are being compensated (obtain d each). This means that a (certain type or class of) corrupt transaction will only be undertaken if it yields a return E high enough to cover both expenditures, bribe b^* and $n * d$.[19]

In the one shot version of an adapted game between a particular member of a sub-group of a certain size and an official, the uniqueness of the SNE of Section 1.2.1 remains.[20] To incorporate the effects of information transmission between sub-group members within the repeated game, we consider ICC and PC under an adapted version of Grim Trigger, the benchmark punishment strategy.

Our Multi-lateral Punishment Strategy (MPS) builds on B's threat of cancelling potential future economic transactions not only when defected on herself but also when observing (having observed) misbehaviour towards any of her fellow sub-group members.[21] Again, all agents in the economy live for infinite periods and maximize their expected present value of life time payoff by discounting their continuation payoff with the common factor δ.[22]

In the simplest case, the MPS for B can be formulated as follows.
'Pay the equilibrium bribe b^* as long as O *has never defected on* any *of* B*'s fellow sub-group members. Never engage in any corrupt transaction with* O *again and choose another* O *who is known to have never defected*[23] *on* B *or any of* B*'s fellow sub-group members otherwise.'*[24]
Appendix 1B provides more details on the incentive compatibility of the MPS.

[19]Modelling the peer-compensation as a fixed payment is the easiest vehicle of transporting the idea of joint payoff maximization by all members of a sub-group. In reality direct large-scale compensation payments on a project basis are unlikely. In expectation all group members engage in the same class of corrupt transactions in their infinite life time, making direct compensation redundant. The direct compensation makes sure that only those transactions are realized that cover all fellow sub-group members' losses (the sub-group's share of the total negative externality).

[20]Consider the same backward induction argument as applied in the one shot game of the situation of bilateral contracting in Section 1.2.1 (and Appendix 1B).

[21]This is consistent with the mechanism in Greif (1993).

[22]We ignore the potential problem of renegotiation between B and O by assuming that collusion between officials is impossible. This assumption is especially reasonable if staff relocation among officials is common and officials do not belong to the same sub-group. The existence of network ties between an official and a client provides a different form of contract enforcement device (Kingston 2007) that is not considered in this paper.

[23]In equilibrium there is no defection, hence there must always be at least one O who has never defected before as long as we assume that there are more officials than needed.

[24]As under BPS, there are more forgiving strategies, e.g. 'Tit for Tat', which are not considered in order to keep the analysis as simple as possible (benchmark argument).

1.3.3 Corruption and group size

Compared to the case with only two agents (ICC_{BPS}), the Incentive Compatibility Constraint under the Multilateral Punishment Strategy ICC_{MPS} changes with respect to the expected future loss of the benefit from potential corrupt transactions in case of defection of the official. Incentive compatibility has to include all O's potential transactions with any of B's fellow sub-group members.[25] Since we assume clients to be homogeneous, we consider a representative agent (B) in a sub-group of size n in order to derive the ICC_{MPS}.

$$ICC_{MPS}: \quad b - c + \sum_{i=1}^{T} \delta^i n(b_i - c) \geq b + 0;\ T \ \to \infty:\ b - c + \frac{\delta}{1-\delta} n(b-c) \geq b \quad (1.3)$$

Solving for the minimal incentive compatible bribe yields $b^*(\delta, c, n)_{MPS} = (\frac{1-\delta}{n\delta} + 1)\ c$. The larger the sub-group, the lower the equilibrium bribe has to be in order to hinder O from defection: $\frac{\partial b^*}{\partial n} = -\frac{1-\delta}{n^2} c < 0$. For the full effect of n on the level of corruption, consider B's new Participation Constraint PC_{MPS}. B will only participate if her profits are still positive after paying b^*_{MPS} and compensating all her sub-group members for the damage realized by the corrupt transaction.

$$PC_{MPS}: \quad E - n\frac{iE}{N} - b^*_{MPS} \geq 0 \quad (1.4)$$

This yields the minimum return: $E^*_{MPS}(\delta, N, c, i, n) = cN\frac{1+\delta(n-1)}{n\delta(N-in)}$. For a given size of the sub-group n and under the technical assumption of $N > in$, the partial effect of the size of society N is negative: $\frac{\partial E^*}{\partial N} = -\frac{(1+\delta(n-1))}{\delta(N-in)^2}\ i\ c < 0$ if $N > i\ n$. The larger the society, the wider the total damage is spread and the less of it has to be internalized through the compensation of all sub-group members. There are two countervailing effects of sub-group size n on E^*:

$$E^*_n(n) \ \equiv \frac{\partial E^*(n)}{\partial n} = \frac{cN(i(2+\delta(n-2))n - N(1-\delta))}{\delta n^2(N-in)^2} \ \lesseqqgtr 0. \quad (1.5)$$

On the one hand we can identify a positive effect, the *Coalition Effect* (CE). The larger the sub-group, the more future potential earnings are at stake for O when deciding

[25] The assumption that all n transactions potentially take place in any of the periods implies sufficient (time) capacities available to the official.

between cooperation (t) and defection (nt). This decreases the equilibrium bribe b^*_{MPS} needed for incentive compatibility.
On the other hand, there is a negative effect, the *Internalization Effect* (IE), which captures the direct internalization of the sum of negative external effects relevant for B's fellow sub-group members. The number of sub-group members (group size) increases the sum of compensation payments needed for the successful engagement in corruption.

Corruption maximizing group-size

We can show that the relationship between E^* and n is U-shaped for all relevant coefficient values. This means that an intermediate group size balances the trade-off between the two countervailing effects.

First, we show that an increase in the size of the sub-group (whose members use *MPS*) increases the level of corruption for low values of n. Consider the marginal effect of n on $E^*(n)$ at $n = 1$: $E^*_n(n = 1) = c\,\delta N\,\frac{((2-\delta)i-(1-\delta)N)}{((1-\delta)i-\delta t)^2}$.

$$\text{If } N > i\frac{2-\delta}{(1-\delta)}:\ E^*_n(n=1) < 0. \tag{1.6}$$

For group enforcement through *MPS* to be effective (E^* is smaller in a group when using MPS than in the case of BPS), the size of society N needs to be large enough compared to the factor of inefficiency of corruption (i) and the discount factor δ.[26]

Second, we show that, due to the inefficiency of corruption, group enforcement (through MPS) does not provide more stability than BPS when sub-group size n is large relative to the size of society N. Note that in our set-up of perfect information transmission inside a sub-group, sub-group members cannot choose to use a bilateral punishment strategy since information cannot be withheld. This may be rationalized by the inability of group members to hide criminal activities from their peers.[27] If sub-group size n reaches values smaller but close to N ($N = \frac{N}{i}$) which is, by definition ($i > 1$), strictly smaller than N, we can show that E^* approaches infinity under MPS and hence

[26]With $\delta = 0.9$ (a depreciation rate of 10%) society must be only ten times larger than the factor of inefficiency.

[27]The U-shaped curve of $E^*(n)$ indicates that MPS is individually optimal only up to a certain sub-group size n_o above which BPS would be optimal if feasible (if information could be blocked).

no corruption occurs.[28]

$$\lim_{n \to \frac{N}{i}} E^* = \lim_{n \to \frac{N}{i}} cN \frac{1+\delta(iN-1)}{n\delta(N-in)} = \lim_{z \to 0} \frac{c\,(1+\delta(\frac{N}{i}-1))}{z} \to \infty. \tag{1.7}$$

Third, we show that there is a unique group size n^* that minimizes E^* and hence maximizes the level of corruption. $E^*_n(n) \equiv \frac{\partial E^*(n)}{\partial n} = \frac{cN(i(2+\delta(n-2))n-N(1-\delta))}{\delta n^2(N-in)^2} = 0$. Solving for n^* yields:

$$n^*(i,\delta,N) = \frac{i(1-\delta)+\sqrt{i^2+\delta(1-\delta)(N-i)i}}{\delta i}. \tag{1.8}$$

Since the second derivative is positive ($\frac{\partial^2 E^*(n)}{(\partial n)^2} > 0$) for the relevant range of n[29], subgroup size $n^*(i,\delta,N)$ characterizes the global (and unique) minimum in E^*.

Characteristics of corruption maximizing group size

The size of society N affects the corruption maximizing sub-group size in the following way. The larger N, the broader the spreading of the inefficiency causing a lower per capita damage d. While having no effect on the *CE*, a bigger society decreases the *IE*, as less of the total negative effect needs to be internalized under a given sub-group size. Hence the balancing sub-group size n^* increases with N.

$$n^*_N \equiv \frac{\partial n^*}{\partial N} = \frac{1-\delta}{2\sqrt{i^2+\delta(1-\delta)(N-i)i}} > 0 \tag{1.9}$$

The opposite is true for the factor of the inefficiency of corruption. While the degree of inefficiency is not accounted for in the *CE* it increases the (total) amount of the internalized damage (*IE*) and thus shifts n^* in the opposite direction. The larger the inefficiency relative to the return E, the smaller the size of the sub-group that maximizes corrupt activity.

$$n^*_i \equiv \frac{\partial n^*}{\partial i} = -\frac{(1-\delta)N}{2i\sqrt{i^2+\delta(1-\delta)(N-i)i}} < 0 \tag{1.10}$$

Since the function $E^*(\delta,N,c,i,n)$ is convex and continuous in n (for at least

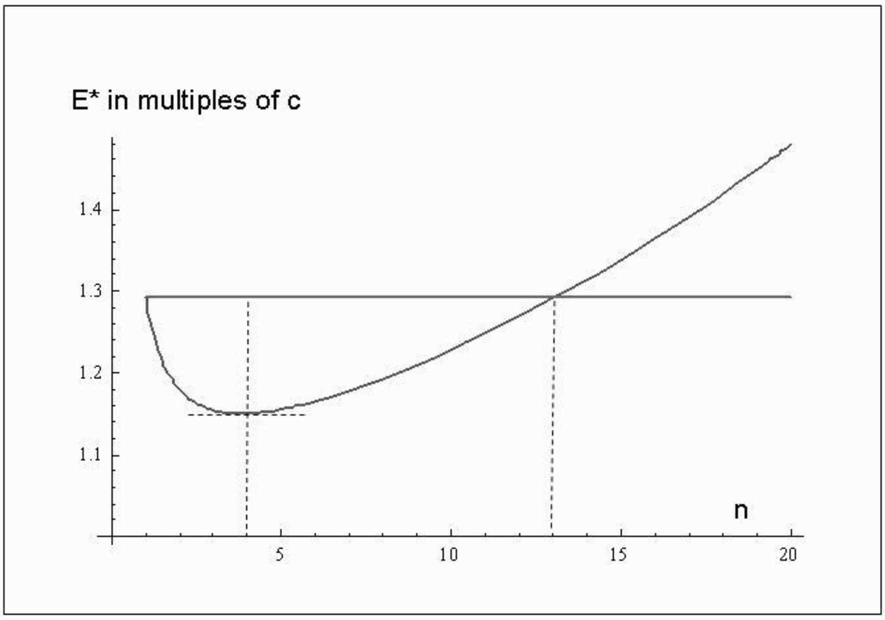
[28] If BPS is feasible, the maximum amount of corruption is determined by E^*_{BPS} for all groups of size $n > n_o$

[29] $E^*_{nn}(n) \equiv \frac{\partial^2 E^*(n)}{(\partial n)^2} = \frac{cN((1-\delta)(2(N-in)^2-in(n-2in))+i^2\delta n^3)}{\delta n^3(N-in)^3} > 0$ for $n \leq \frac{2N}{3i}$. Specifically $E^*_{nn}(n^*) > 0$

$1 < n < \frac{2N}{3i})$[30], decreases in n for small sub-group sizes ($n < n^*$), has a unique minimum of corrupt activity at n^* and increases in n for $n > n^*$, we conclude that $E^*(n)$ must be a U-shaped function in n.

It is trivial to check numerically that, for any reasonable set of parameters (e.g. $\delta \geq 0.5$, discounting is not excessive and $i \geq 1.2$, the inefficiency caused by corruption is significant), the corruption maximizing group size n is small relative to N. In this set of parameters, the corruption maximizing sub-group size is at least *ten* times smaller than the size of the society ($n^* \leq \frac{N}{10}$). As an example, Figure 1.3 depicts the function E^*

Figure 1.3: Exemplary form of E^*



over n where E^* is displayed in multiples of the direct costs of corruption c with the parameter values $delta = 0.9$, $N = 100$ and $i = 1.5$. In this example the minimum return $E^*(n)$, characterizing a maximum number of corrupt transactions, is produced by sub-groups of size $n^* = 4$. Using MPS, groups of size n between 2 and 12 produce corruption levels that are higher than those attainable under BPS. If group size is larger than 13, we find the same per capita corruption levels as in a totally fractionalized society ($n = 1$), if BPS is feasible and lower levels if not. The decline of $E^*(n)$ in n for $n < n^*$ can be interpreted as the result of the stabilizing effect of increasing the radii of trust, linking more individuals to the sub-group. The section to the right of the minimum (n^*) shows that a further increase in the radii of trust deters individuals from engaging in

[30]Recall that $\frac{\partial^2 E^*(n)}{(\partial n)^2} > 0$ for this range.

corrupt transactions because of increasing costs of internalization. The U-shaped curve of $E^*(n)$ translates immediately into a hump-shaped curve in the relation between per capita frequency of corruption and sub-group size n.

1.3.4 Structure of society and level of corruption

We have shown that, irrespective of the behaviour of the other sub-groups in a society, sub-groups that are small relative to society are able to (and will) produce the highest levels of corruption per capita.[31] This implies the following relationship between the structure of (the entire) society and the level of corruption. Let n_g denote the number of members in sub-group g ($g = \{1, 2, ..., k\}$) of k different groups in a society of size N where $\sum_{g=1}^{k} n_g = N$. The relative frequency of corrupt transactions is calculated as $Frq_{soc} = \frac{\sum_{g=1}^{k} n_g(1-F(E^*(n_g)))}{N}$.[32] We can show that a society S^* which is fractionalized into ($k^* = \frac{N}{n^*}$) sub-groups of size n^* ($n_g = n^*$) exhibits the maximum level of corruption. For $n^* << N$, the problem of a positive balance after dividing is negligible.[33] Departing from the situation of $n_g = n^*$ we consider the effects of deviations in terms of changes in the size and number of sub-groups on corruption, holding the size of society constant. In the following we discuss the properties of S^* with respect to its connection to the notion of ELF as well as the implication of deviations of its structure on the level of corruption.

First, we hold the number of sub-groups constant ($k = k^*$). Regard society S_{newk} which deviates from S^* in the characteristics (i.e. the size) of at least one sub-group. For any sub-group larger than n^* ($n_m > n^*$) at least one (other) sub-group with size $n_l < n^*$ is needed to hold the size of society constant (at N^*). Consider all other sub-groups maintaining size n^* ($n_{g \neq l,m} = n^*$). It is easy to see that S_{newk} will exhibit lower levels of corruption, since $E^*(n_l) > E^*(n^*)$ (the *IE* dominates the *CE*) as well as $E^*(n_m) > E^*(n^*)$ (the *CE* dominates the *IE*) as shown in Section 1.3.3 (n^* defines the global minimum in E^* and hence the global maximum in the frequency of corruption levels per capita). By

[31] Recall that the level of corruption within a society is defined by its expected frequency, and the return E is distributed according to the function $F(E)$. Hence the frequency of corruption is defined by $1 - F(E^*)$.

[32] Sub-groups are assumed to be homogeneous, therefore all individuals within a certain sub-group produce the same (relative) amount of successful corrupt transactions (per head) in expected terms.

[33] According to the definition of the Ethno-Linguistic Fractionalization Index (see Appendix 1A), S^* would be characterized by $ELF^* = 1 - \sum_{g=1}^{k^*}(\frac{n_g^*}{N})^2 = 1 - k^*(\frac{n^*}{N})^2$.

the convexity of the ELF[34], any society of size N and with a number of sub-groups k will exhibit a fractionalization index of $ELF < ELF^*$ (being less fractionalized) if any $n \neq n^*$. The further away a society shifts from the characterization of S^* in terms of a lower ELF, the lower its level of corruption. Moreover, the total level of corruption is strictly decreasing in the sum (since n^* constitutes a unique minimum in E^*) as well as in the variance (since $E^*(n)$ is convex) of the sum of (individual) differences between n_g and n^*. Since $Var(n-n^*) = Var(n)$, the ELF and the total level of corruption must go into the same direction considering societies of equal size and equal numbers of sub-groups.

Second, we consider societies of the same size but different numbers of sub-groups (of the same size). Here we have to consider two cases. If in society S_{newn}, the number of sub-groups is smaller than k^* ($k_{newn} < k^*$), the size of each sub-group must be larger than n^* ($n_{newn} > n^*$). This means that each member of any sub-group exhibits a lower (per capita) level of corruption since ($E^*(n_{newn}) > E^*(n^*)$). S_{new-} would show an $ELF_{S_{newn}} < ELF_{S^*}$[35] indicating a less fractionalized society.
If the number of sub-groups in S_{newn} is larger than in S^* ($k_{newn} > k^*$), $n_{newn} < n^*$, the (average) critical value of the rent of coruption must be larger than in S^* ($E^*(n_{gnewn}) > E^*(n^*)$) because of the shape of $E^*(n)$. It is clear that S_{newn} is more fractionalized than S^* ($ELF_{newn} > ELF^*$) using the reverse argument from above. The larger the sum of (individual) differences between n and n^* (in either way), the lower the frequency of corruption. Moreover, due to the convexity of $E^*(n)$ ($E^*_{nn}(n) > 0$), the difference in the level of corruption decreases at an increasing rate.

Third, we compare societies allowing for different numbers as well as different sizes of sub-groups. A society being characterized by $ELF > ELF^*$ (higher fractionalization) must consist of a larger number of sub-groups ($k > k^*$) than S^* (for all combinations of sub-group sizes). Holding the size of society constant ($N = N^*$) there must be at least two sub-groups (l,m) with $n_{l,m} < n^*$. Hence the overall level of corruption must be lower than in S^* since $E^*(n_{l,m}) > E^*(n^*)$. Compared to S^*, a society (of the same size) exhibiting an $ELF < ELF^*$ may either consist of a smaller number of sub-groups ($k < k^*$ and hence $n > n^*$) or accommodate sub-groups of unequal size (or both). In all (three) cases there are at least two sub-groups (l,m) which exhibit lower (per capita)

[34]The more variance in the sub-group size, the smaller ELF holding k constant.

[35]$ELF_{S_{newn}} = 1 - \sum_{g_{new-}=1}^{k_{newn}} (\frac{n_{g_{newn}}}{N})^2 = 1 - k_{newn}(\frac{n_{newn}}{N})^2 < 1 - k^*(\frac{n^*}{N})^2 = ELF_{S^*}$, since $N_{newn} = k_{newn}n_{newn} = K^*n^* = N^*$ and $n_{newn} > n^*$.

levels of corruption since $n_{l,m} \neq n^*$ and hence $E^*(n_{l,m}) > E^*(n^*)$.[36]

These results provide a simple theoretical explanation for the inversely U-shaped relationship between the social fractionalization of a society and the level of corruption found by Dincer (2008) using data from 48 (US-)American states. Our model interprets this observation as an effect stemming from the distribution of citizens (mass) between the sub-groups of a society. Regions with a structure close to that of S^* exhibit the highest levels of corruption, while societies that are either more fractionalized ($ELF > ELF^*$) or less fractionalized ($ELF < ELF^*$) show lower levels. Generally, the larger the distance to S^* in terms of the sum of differences between n and n^*, the lower the level of corruption.

1.3.5 Stability and convergence of group size

The internalization of part of the negative external effect through the compensation of all sub-group members leads to efficient behaviour within sub-groups but causes large inefficiencies for the society as a whole. Hence, even if we consider the size of the radii of trust as endogenous, i.e. subjects can sever or establish links to citizens in their immediate proximity (by increasing or decreasing their radii of trust, see Figure 1.2) we still face a prisoner's dilemma type situation (Fudenberg and Tirole 1991). Regard a situation of (k^*) sub-groups of corruption maximizing size n^*. It would be socially efficient to increase the radii of trust (e.g. by promoting trust to outsiders) in order to increase the internalization and thus reduce the relative number of corrupt transactions and thereby the amount of inefficiency resulting from the negative external effects.

However, for the members of a particular sub-group it cannot be individually optimal to increase their radii of trust as long as their sub-group size is greater than or equal to n^*. A unilateral step of increasing the radii of trust by the members of a particular sub-group would yield a lower level of corruption in society (and thereby decrease the degree of realization of the negative external effect) but at the same time decrease the expected individual corrupt rents for all members of this sub-group.
Given the behaviour of the members of all other groups, the positive effect of the decrease

[36] If the sub-group size n is not constant, the relationship between the level of corruption and the ELF is not monotonic since the ELF does not allow for a unique matching.

in damages caused by the realization of fewer corrupt transactions per head is lower than the loss of individual rent creation through corruption. Therefore there cannot exist a member of any sub-group of size $n^* \leq n \leq (N-1)$ (or $n^* \leq n \leq (n_o - 1)$ in case of the feasibility of BPS) who has an incentive for the establishment of additional links to outsiders. This means that a society fractionalized into groups of corruption maximizing size will not transform into any other structure by endogenous link formation.

1.4 Conclusion

Our analysis of the effect of the social structure on the level of corruption focuses on the solution of the enforcement problem in the Principal-Agent relationship between a bribing client and a corrupt official through the use of multilateral punishment strategies. The social structure is defined by a society's fractionalization into sub-groups. These are visualized by multilateral links between individuals stemming from the overlapping of radii of trust of different sizes. The critical feature of a sub-network is the perfect (and unimpeded) flow of information between all of its members.

Considering the deployment of (bilateral and) multilateral punishment strategies, we show that the relative size of a sub-group determines the relative mass of economic transactions for which corruption is profitable (for a representative member) and hence defines the (per capita) level of corruption. For a broad range of parameters, the (unique) maximum level of corruption is found for sub-groups that are small relative to the size of society since these balance the trade-off between two countervailing effects in the enforcement problem.

Translating this result into the relationship between the social structure of an entire society and its (per head) level of corrupt activity we conclude that a society fractionalized into a large number of relatively small sub-groups presents the maximum level of corruption. The further away from these characteristics a society moves (in terms of lower or higher fractionalization, e.g. measured by the Ethno-Linguistic Fractionalization index), the lower the total level of corruption. This explains the inversely U-shaped relationship between social fractionalization and corruption found by Dincer (2008).

As explicated in this paper, the corruption maximizing sub-group size is small relative to the size of society. This result is also consistent with the observations in cross country studies of a strong positive correlation between ethnic fractionalization and the level of corruption (Alesina et al. 2003, Lipset and Lenz 2000, Tonoyan 2005).[37]

Our model shows that, even in situations where there is no special (e.g. parochial) relationship between a briber and an official[38], the structure of a society affects the frequency of corruption. The main driver of the results is the mechanism of multilateral punishment strategies, which is successful only because the corrupt agents use the value of the official's reputation to increase the power of the stabilizing effect of repeating the corrupt transaction. This links the model to the experimental literature on indirect reciprocity. In studies by Seinen und Schram (2006) and Bonein and Serra (2007) the level of cooperation is found to be determined not only by direct interaction between transaction partners but also by ego's information on alter's relevant cooperative behaviour towards third parties in a repeated game. The main argument in the explanation of (enhanced) strategic cooperation is similar to that used in our model. Information about the transaction partner's past cooperative behaviour to outsiders serves as a device to capitalize on the threat of reputational damage in the case of defection (non-cooperative behaviour), see also Novak and Sigmund (1998). However, it is information transmission that ultimately limits the level of corruption by forcing individuals to take into account larger parts of the negative externality that they are causing by engaging in corruption.

In order to refine the model and enhance its explanatory power, the model might be enriched by additional dimensions of characteristics of sub-groups. Where we assume perfect transmission of information inside a group, the rate of transmission should also depend on characteristics such as absolute sub-group size. This would capture the idea of the qualitative distinction between bonding and bridging social capital (Fukuyama 1995, Harris 2007). We leave this extension to future research.

---

[37] Several studies (Mauro 1995, Bardhan 1997) indicate that northern European countries, which are characterized by very low levels of fractionalization, exhibit the lowest levels of corruption where Sub-Sahara-African as well as South-East-Asian countries (such as the Chad, Sierra Leone, Cambodia and the Philippines) in contrast feature a highly fractionalized society and the highest corruption levels (measured in the Corruption Perception Index).

[38] Our model ignores nepotism (i.e. an official belongs to the same sub-group as the bribing client), which is, in terms of the enforceability of a corrupt transaction, equivalent to the vertical integration of the official into a group of bribers (see Appendix 1D and Klitgaard 1988).

1.5 Appendix 1

Appendix 1A: Definition of the *ELF*

The Ethno Linguistic Fractionalization Index measures the probability that two individuals, randomly drawn from the population, belong to different ethnic groups (Bossert et al. 2006).

$ELF = 1 - \sum_{g=1}^{k}(\frac{n_g}{N})^2$

Let k be the number of different ethnic groups in society, n_g the size of group g (measured by the number of members) and N the size of society. The larger the *ELF*, the more fractionalized the society. The *ELF* is crude in the sense that societies consisting of different structures can lead to the same *ELF*. However, it is the most practical measure of fractionalization and widely used in empirical studies.

Appendix 1B: Equilibrium properties

In order to check for the equilibrium property of E^* and b^* as well as O and B playing according to the *MPS*, consider the following arguments. Since the formal proofs would be technically equivalent to the ones in Greif (1993) we will outline the intuition only.

***B's* behaviour**

Under *MPS*, only an 'honest' official (an official who has never reportedly cheated on B or any of B's fellow sub-group members) will be hired by B and all her sub-group members in future periods, whereas an official who has cheated (at least once) on B or any of B's fellow sub-group members will in equilibrium never be hired (by them) again. Implicitly, O's expected future payoff through bribery determines the magnitude of the equilibrium bribe b^*. The larger the potential future payoffs, the smaller is b^*. Hence an 'honest' official has a brighter prospect of future gains from corruption, so that a briber will always prefer an honest official to an official who is known to have cheated at least once, since the cost, i.e. the incentive compatible bribe is lower. In our model, the difference in expected future payoff between an official who has cheated and an official who has not, is extreme, since (on the equilibrium path) a cheater will never be bribed again, which yields 0-profit for all future periods. A model that would allow for imperfect information transmission (which would not destroy the entire future perspective of a cheating official) and some probability of leaking information to outsiders would capture the argument of the re-hiring of the (more) 'honest' official(s) in a more realistic way.

O's behaviour

The official will always choose 'a' as long as '$b > 0$'. Moreover, she will choose 'nt' for all $b < b^*$ and 't' for all $b \geq b^*$ (*ICC*). The existence of several officials causes competition for bribes (no monopoly for the official), hence there must be unemployment[39] and positive rents for those officials who serve bribing clients.

However unemployed officials who have a clean record of corrupt cooperation behaviour with respect to at least one sub-group, may try to under-bid the equilibrium bribe b^*. This would lead to a violation of the *ICC*, since any $b < b^*$ leads to defection of the official in the first period of the relationship. Hence such a bribe will never be (rationally) chosen (or accepted when proposed) by any B.

Appendix 1C: Uniqueness of the Sub game perfect Nash Equilibrium

We show the uniqueness of the SNE by using an argument of backward induction. Denote by I_i the information set in stage i ($i \epsilon \{1, 2, 3\}$). Let $p(I_i)$ be the probability of reaching stage 'i' and $q(`s'|I_i)$ the conditional probability of the relevant agent choosing action 's' once having reached stage i. An information set contains all relevant information about all relevant players' behaviour up to stage i.

First, we show that there cannot be an equilibrium in which O chooses 't' in Stage 3. Consider a strategy-set $EQU1 = [s_1, s_2, s_3]$ in which $p(I_3) > 0$ and $q(`t'|I_3) > 0$. Compare the payoff, resulting from the realization of this strategy-set ($PO(EQU1)$) to that of an alternative set ($EQU1_{new}$) which consists of equal strategies up to stage 3 but for

[39] An unemployed official does not engage in corruption.

which $q(\text{'}t\text{'}|I_3) = 0$ yielding payoff $PO(EQU1_{new})$. Since $b > b - c$, $PO(EQU1_{new}) > PO(EQU1)$ and $EQU1$ cannot constitute an equilibrium.

Second, we show that O will never choose 'na' if $b > 0$. Consider the strategy-set $EQU2 = [s_1, s_2, s_3]$ in which $p(I_2) > 0$, $q(\text{'}t\text{'}|I_3) = 0$, $q(\text{'}b > 0\text{'}|I_1) = 1$, $q(\text{'}nt\text{'}|I_3) = 1$ and $q(\text{'}na\text{'}|I_2) > 0$. Compare $PO(EQU2)$ to the payoff of the strategy set $EQU2_{new}$ which differs from the former only in $q(\text{'}na\text{'}|I_2) = 0$. Since $b > 0$, $PO(EQU2_{new}) = b > 0 = PO(EQU2)$ so that $EQU2$ is not an equilibrium. It is clear that O is indifferent between 'a' and 'na' if $b = 0$.

Third, we show that B setting a positive bribe cannot belong to the equilibrium path. Compare the payoff levels of the strategy set $EQU3$ exhibiting $q(\text{'}t\text{'}|I_3) = 0$, $q(\text{'}na\text{'}|I_1) = 0$ and $q(\text{'}b > 0\text{'}|I_1) > 0$ with a similar set that differs only in $q(\text{'}b > 0\text{'}|I_1) = 0$: $EQU3_{new}$. Since $PO(EQU3) = -b < 0 = PO(EQU3_{new})$, $EQU3$ cannot describe an equilibrium. Hence the unique Sub-game perfect Nash Equilibrium is characterized by ['b=0'; 'a'/'na', 'nt'].

Appendix 1D: Alternative explanations of corrupt stability

Knapp (1986) emphasizes that third-party enforcement of a corrupt transaction can be circumvented by middlemen who perform the 'dirty work' of the transaction between a client and an official. These arrangements are usually labelled as commissioned services. Despite being a widely used practice, the deployment of middlemen does not solve the stability problem in the corrupt model, since it only shifts the vulnerability to opportunism to a third party. A corrupt transaction then consists of two unenforceable transactions, one between the client and the middleman and the other between the middleman and the official. Hence, the theoretical problem of contract enforcement is rather duplicated than eliminated.

Some individuals or groups try to solve the problem by vertical integration, i.e. integrating a public official into their own social network. This may be done in two ways. A group of clients may form informal relationships with officials through reciprocal gift exchange (Klitgaard 1988), or they may encourage closely related agents such as family members to enter administrative offices which enable them to stabilize corrupt deals through parochial exchange in the future.

Another option is the use of assets as collateral. The application of collateral can be seen as a simple form of a second (legal) game linked to the actual corrupt transaction. Kingston (2007) considers the stabilization of a one-shot bribery game by using the expected future revenue of another game (which is stabilized by infinite repetition) as collateral for the corrupt transaction. The corrupt parties make their strategy choices of one game contingent on the (expected future) outcome of the other game.

Schramm and Taube (2001) point out that certain social structures in general and specific networks in particular (which may or may not feature criminal elements as their main objective) can help to stabilize corruption. Unlike our model, they consider official and briber to belong to the same (e.g. ethnically or regionally pre-defined) sub-group and therefore share distribution channels (see Ogilvie 2004, Greif 2006, Kandori 1992). The mechanism of stabilizing corruption is similar to the vertical integration argument. Along this argumentation favouritism of group members over outsiders, altruism towards sub-group members or positive externalities within sub-networks arising from deals between sub-group members are likely to be the main drivers.

In China, repeated transactions between contractors within Guanxi Networks seem to be central to the enforcement mechanism. In general, these networks stem from common heritage, family ties and other predominantly exogenous factors (Schramm and Taube 2001, Lambsdorff 2007). Guanxi Networks are not primarily built to facilitate illegal activities but present alternative means of stabilizing economic transactions (legal and illegal) in an environment that does not provide sufficient legal protection of commercial activity. Group membership may decrease transaction costs (which can be especially high in illegal activities) but also provides the framework for credible threats, sanctions and punishment strategies against defective and opportunistic behaviour. Indirect reciprocity as found in experiments (Seinen and Schram 2006) may play a role, as agents may condition their own reciprocal behaviour on information obtained in situations that did not involve themselves but third parties, e.g. fellow group members.

Appendix 1E: Bargaining power

In order to show the insignificance of the distribution of bargaining power for the resulting level of corruption, consider the following two benchmark cases.

If B holds all bargaining power (which is implicitly assumed in our set-up since B can make a take-it or leave-it offer), all types of deals that yield a return $E \geq E^*$ are undertaken with $b = b^*$. B gets the net surplus of $E - E^*$. Any $b > b^*$ would be incentive compatible for O but would shift rent away from B, which cannot constitute an equilibrium. Any $b < b^*$ in turn would not satisfy ICC and thus would yield a negative payoff ($-b < 0$) for B, which cannot be optimal for her either. B maximizes her payoff while satisfying ICC by paying b^* for all deals that yield $E \geq E^*$.

If O holds all bargaining power, she will (implicitly) demand $b = E\,\frac{N-i}{N}$ as long as $E \geq E^*$. If $E < E^*$, either ICC or PC is violated. The marginal ('last') feasible deal (in terms of the ordering of expected rent from corruption) is again the one that yields E^* which gives a 0-payoff to B (weakly satisfying PC) and ensures incentive compatibility for O.

There is no reason to believe that for any distribution of bargaining power which can be understood as a linear combination of the two benchmark cases, the marginal ('last') stabilizable deal yields a return different from E^*.
Hence E^* qualifies as a measure for the total level of corruption within a society independent of the distribution of bargaining power between B and O.

Chapter 2

Bringing the Four-Eyes-Principle to the Lab

2.1 Introduction

With almost daily media attention of high profile scandals, corruption, generally defined as 'the misuse of public office for private gain' (OECD 2010), has been recognized as a major problem. In general, a corrupt transaction is illegal and exerts a large negative external effect on outsiders, which is usually assumed to be larger than the sum of benefits to the agents who are directly involved. This defines corruption as socially inefficient (Klitgaard 1988, Rose-Ackerman 1999).

In addition to traditional views of deterring an agent from engaging in criminal activity by varying the amount of penalties and the probability of detection (Becker 1968), the New Institutional Economics (NIE) of corruption concentrates on finding an institutional design that optimally exploits the instability of the corrupt transaction between a client and an official (Schulze and Frank 2003).[1] The instability of a single corrupt transaction stems from the enforcement problem between a bribing agent and a potentially corrupt official. Its illegal nature precludes the assistance of legal third parties, i.e. the courts (Lambsdorff 2007). The occurrence of corruption therefore relies heavily on trust and

[1]There is a considerable amount of theoretical research on the Principal-Agent relationship between a (benevolent and non-corrupt) government and its public officials (Groenendijk 2004). Ignoring the lack of (legal) enforcement of a corrupt transaction between O and B boils the problem down to the analysis of an ordinary Principal-Agent model with a specific application.

reciprocity and is difficult to explain in standard theoretical models. Nonetheless, national as well as international organizations such as Transparency International, the OECD and several national fiscal authorities publish lists of (institutional) policy recommendations containing measures to curb corruption. Along with 'staff rotation' (analyzed in Abbink 2004), the introduction of the Four-Eyes-Principle (4EP), 'a requirement that business has to be effectively conducted by at least two individuals (four eyes)', is one of the most prominent examples (Pörting and Vahlenkamp 1998, Rieger 2005, Wiehen 2005, Hussein 2005). As a result of general problematic tractability (let alone predictability) of corrupt behaviour, a theoretical analysis of the effectiveness of the 4EP does not exist. Nor is there any kind of traceable empirical evidence to support its usefulness.

Not only in the corrupt context, but also on a general level, the distinction between individuals and small groups as decision-makers has been widely ignored in the theoretical literature. Differences between the behaviour of individuals deciding alone or in a group have only recently been addressed in the field of experimental economics, where results seem ambiguous. Some studies find that the behaviour of groups is closer to standard equilibrium predictions derived from the self interested model of payoff maximization (e.g. Bornstein and Yaniv 1998, Blinder and Morgan 2005), other studies (e.g. Kocher and Sutter 2005, 2007, Cason and Mui 1997) provide experimental evidence to the contrary. Kocher and Sutter (2007) conclude that the direction of the group decision-making effect critically depends on the nature of the task determining which of two countervailing motives, the profit maximizing motive or the competitive motive, dominates. The basic set-up of our laboratory experiment is close to those used in Abbink (2004) and Lambsdorff and Frank (2010). Our experiment is designed to assess the effects of the introduction of the 4EP on observed levels of corruption. Within this framework we model the 4EP as replacing a single official (deciding individually) with a group[2] of two officials deciding jointly according to a decision-making process that secures veto power for non-corrupt officials.

Using four different treatments we can separate two countervailing effects of the introduction of the 4EP. One is due to the difference in marginal incentives resulting from the division of the transfer between the jointly deciding officials. The other effect is determined by the group decision-making process alone (keeping marginal incentives constant).

[2]Although the entity consists of only two participants we call it a group rather than a team.

Rejecting predictions taken from (self interest based) arguments within the standard model of corruption, we find that the introduction of the 4EP increases the frequency of successful corrupt transactions unambiguously. We substantiate this hypothesis in three stages. First, we consider only outcomes (actual corruption levels). Second, we investigate behaviour in the decision-making process, i.e. we compare initial and final choices. Third, we analyse the content of electronic text messages exchanged during the decision-making process between jointly deciding officials. Our results strongly suggest the dominance of the profit maximizing motive (Kocher and Sutter 2007, Blinder and Morgan 2005). Groups reveal more functional behaviour with respect to conditional responding. By their higher (joint) cognitive capacity, groups of officials seem to be more capable of maximizing their payoffs by following strategies that are shown to lead to a higher frequency of corrupt transactions based on mutual reciprocity. Our explanations of the observed effects are in line with the argumentation of the persuasive argument theory (Pruitt 1971). Since groups perform better in solving the enforcement problem between briber and official, the introduction of the 4EP moves behaviour further away from the theoretical prediction of selfish behaviour, which, in the corrupt context falls in line with the social optimum. Therefore our results cast doubt on the usefulness of the introduction of the 4EP and its justification as a recommended measure against corruption.

The remainder of the paper is structured as follows. Section 2.2 describes the experimental set-up giving details on the specifications of all four treatments. Section 2.3 analyses the effects of the introduction of the 4EP in the framework of the NIE of corruption and forms hypotheses. Section 2.4 gives details on the procedure of the experiment. In Section 2.5 we describe the main findings, provide a detailed explanation of the empirical strategies and interpret the results. Section 2.6 summarizes and concludes.

2.2 Experimental set-up

2.2.1 Corruption and the NIE

In its most general form, a corrupt transaction can be described as a Principal-Agent-Client relationship, in which the principal, represented by the government or any kind of benevolent[3] authority, deals with its clients (private entities, e.g. firms) through its agents, the potentially corrupt (not perfectly monitored) public officials. Clients may have an incentive to transfer side-payments (i.e. bribes) to relevant officials in order to alter their behaviour with respect to their duties (i.e. fulfilling legal procedures clearly defined but not perfectly controlled by the authorities).

The main mechanism of the instable and therefore interesting part of the relationship is best explained in a simple 3-Stage game:
In Stage 1, a client (e.g. a potentially bribing firm) B decides on the level of bribe b=0,1,...,12 Experimental Monetary Units (EMU) to be given to a potentially corrupt official O. The limit of 12 EMU reflects B's budget constraint. At this point the amount of b is tripled. The factor 3 captures the idea of a difference in the marginal utilities of money between a (wealthy) client and a (poor) official (Abbink et al. 2002).

In Stage 2, O can either 'accept' the transfer b and keep the tripled amount ($3*b$) to herself or 'reject' it. In the latter case the game ends, O keeps only b to herself (as some 'benefit' from pro social behaviour) while the rest (the remaining amount of $2*b$) is used for the 'public benefit'.[4] If O accepts the bribe, she gets the tripled transfer for sure. She automatically enters Stage 3 where she decides between two options. The first option includes initiating an increase of B's payoff by 16 EMU (delivering the corrupt service) at the fixed costs of 4 EMU (to herself). In this case, the 'public' suffers substantially (by −24 EMU). Note that independent of the size of the transfer, the negative externality is always larger than the sum of payoffs for the agents B and O so that a successful corrupt transaction is always inefficient by construction.[5] In the second option O arranges nothing (implicitly defects on B), saving costs for herself and the public from the negative externality. The costs of delivery may be interpreted as the certainty equivalent of the lottery: punishment in the case of detection and corrupt success, or as the practical costs of engaging in a criminal activity, e.g. hiding illegal activities from colleagues and superiors.[6]

[3]This excludes political or 'grand' corruption (see Klitgaard 1988).

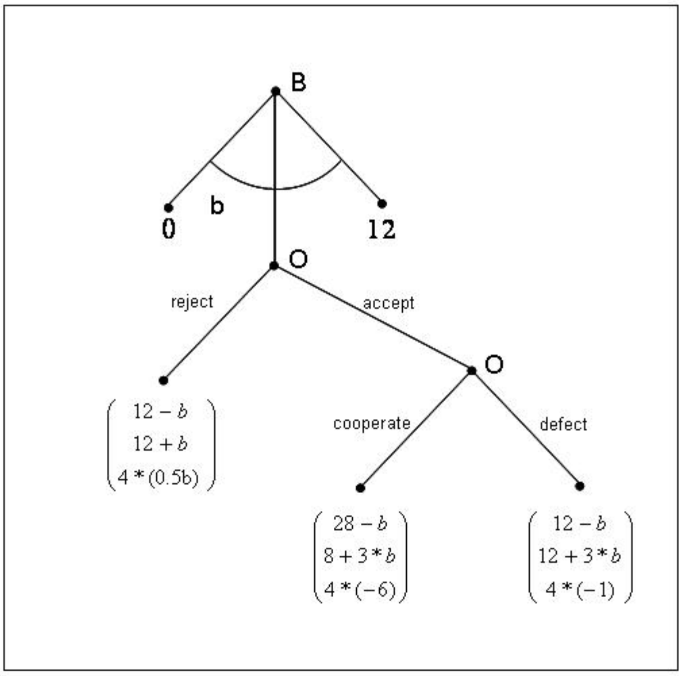
[4]This implies the assumption of equal marginal utility between O and the public.

[5]See Bardhan (1997) for a review of (empirical) evidence supporting the assumption of the negative externality and inefficiency of corruption.

[6]Abbink (2004) uses a fourth stage which accounts for probability of detection. Using a lottery instead of a fixed amount to model the cost of corruption adds another problem of individual differences (risk aversion). This would require to disentangle potential treatment effects from differences in risk aversion.

If O decides against delivering the corrupt service in Stage 3, B does not get the bonus, O does not bear the costs of 4 EMU, and only a minor negative externality to the public (4 EMU) is realized. Within the standard self interested model it is never optimal, neither in a one shot nor in a finitely repeated game, for O to deliver the corrupt service and hence for B to transfer a positive bribe, see Appendix 2A for a short proof. Figure 2.1 represents the basic set-up in its extensive form.

Figure 2.1: Extensive form representation



2.2.2 The 4EP

In order to investigate the effects of the introduction of the 4EP on the level of corruption in this set-up, we consider two participants instead of one making the decisions of the official in Stage 2 and Stage 3. We define the level of corruption as the frequency of successful corrupt transactions relative to the total number of possible transactions.[7]

We use four different versions of the game as treatments of the experiment. In all four treatments, subjects play their version of the corruption game for ten successive periods. After each period they learn about their group partners'/partner's (payoff relevant)

[7]Note that in our set-up this number is directly proportional to the sum of negative externalities, but not directly proportional to the average (total) payoff of the participants.

choices.[8] All treatments are run in a partner design so that all subjects remain in their respective unit (of B and O subjects) for all ten periods of the experiment. Separating potentially countervailing effects and still being able to compare outcomes across treatments, we have to consider different group sizes within the treatments. Find Figures 2.7 and Figure 2.8 in Appendix 2B for representations of the full extensive forms of the games played in the respective treatments.

In our experiment, the 'public' is modelled in two different ways. In one set of sessions (mode 1) we model the externality on the public as payments (reductions of payments) to four randomly chosen participants of the (same session of the) experiment. For technical reasons a particular subject can never be hit more than once per period. In the second set of sessions (mode 2) we model the externalities as increases or decreases in the amount of a donation to the public aid organization 'Doctors without Borders'. We chose 'Doctors without Borders' to obtain results as comparable as possible to the findings of Lambsdorff and Frank (2007, 2010, forthcoming), who use this organization in their experiment. The use of donations to a charity in experiments goes back to Eckel and Grossman (1996). As expected, we find from answers to our post-experimental questionnaire that virtually all subjects approve of this organization and take this as evidence for our working hypothesis that the reduction of a real donation represents a valid model for the reduction of public welfare. The total amount of added or deducted payments is equal across the two modes. Using two different modes of modelling the externality including real outsiders allows us to address the problem of a 'super-game' considering the possibility of participants forming expectations on the behaviour of participants outside their own group.

IDT1[9]

In Treatment 1, the 'Individual Decision-making Treatment 1' (IDT1), we consider units of two subjects, one in the role of the official O and one in the role of the potential briber B. The 3-Stage corruption game (see Figure 2.1) is played for ten consecutive periods. At the end of each period, all participants get to know their own payoffs. Additionally,

[8]By choosing a repeated instead of a one-shot set-up, we focus on the strategic component of the reciprocal transaction of corruption which we expect to be affected by the number of participants within a decision-making entity (Lambert-Mogiliansky et al. 2006).

[9]See Figure 2.7 in Appendix 2B for the extensive form representation.

type B subjects get information about the (Stage 2 and Stage 3) decisions of their transaction partners (of type O). While through this information all subjects know about the negative or positive externalities they have helped to cause to the public (four randomly chosen participants in mode 1 or 'Doctors without Borders' in mode 2) they do not learn about the magnitude of the spill-overs that may have been caused to them by the decisions of the subjects outside their unit in mode 1. Nevertheless, there could be violations of the independence assumption due to considerations of a 'super-game'. Participants may condition their choices on their beliefs about the other participants' corrupt behaviour which may affect them through the negative externality. To check for the existence of effects stemming from this 'super-game' and provide a robustness check, we use mode 2 in which the public is modelled as a 'real' third party, the recipients of the donation towards the public aid organisation. At least for this mode independence of observations across units is warranted.

TDT1

In the second treatment, the 'Team Decision-making Treatment 1' (TDT1), we form units of three subjects, one B and two O types. The B type decides in Stage 1 about her bribe b which is tripled and then transferred to both officials of her (3-player-)unit. Note that although the amount goes to two players, it is subtracted only once from B's account. The parameters of the game are set in such a way that the incentives for the officials are equal to the ones in IDT1, given the amount of bribe. This way we can separate the true 'Group Decision-making Effect' (GDE) from effects stemming from the partition of the bribe between the two officials.

In Stage 2 and Stage 3, the two officials of a unit make their decisions jointly. In both stages they decide independently first. If they do not come to an unambiguous decision (e.g. one official decides for 'reject', the other for 'accept' in Stage 2), they learn about each other's choice and decide again. If there is still no agreement, they get the opportunity to communicate with each other via a real time electronic 'chat' in which they can, for one minute, exchange electronic messages (see the translated instructions in Appendix 2C).

If there is still no mutual consent, the corruption-unfriendly choice is taken ('reject' in Stage 2 and 'defect' in Stage 3). This rule reflects the veto-power of officials who do not want to engage in a corrupt transaction in Stage 2 and those who do not want to reciprocate in Stage 3 (in both cases avoiding damage to the public). The idea is that an official cannot force her colleagues to engage in a corrupt transaction but can try to convince them. As a consequence of the decision-making rule, a corrupt transaction can only be successful if both officials (finally) choose 'accept' in Stage 2 and 'cooperate' in Stage 3.

IDT2

The 'Individual Decision-making Treatment 2' (IDT2) differs from IDT1 only in the number of possible transactions between a particular B-O pair. In this treatment we consider units of four, two type B and two type O participants. Every type B participant sends only one transfer to one of the two officials in her unit per period. This means that playing the game for ten periods makes five possible transactions per pair, producing two transactions per period and four-player-unit. The reason for running this treatment is to control for possible effects in the behaviour of subjects stemming from playing in a larger group and interacting less frequently with a particular transaction-partner. The decisions of the participants within a unit of four yield only one independent observation.

In Stage 1, one of the (potential) bribers (B_1) decides about her transfer b_1 to one of the officials (O_1), and the other briber (B_2) decides about her transfer b_2 to the remaining official (O_2). All transfers are tripled and shown to the respective officials. In Stage 2 and Stage 3 each of the officials decides independently. The respective pairs change every period so that in the following period B_1 decides about the size of her transfer to O_2 while B_2 interacts with O_1.

TDT2

In the 'Team Decision-making Treatment 2' (TDT2) we form again four-player-units. Each of the two type B players sends one transfer each to the group of two officials who decide jointly in Stage 2 and Stage 3, according to the same decision-making process explained for TDT1. Contrary to the case of TDT1, the transfer is split equally so that each of the officials receives only half of the tripled amount of the transfer chosen by the respective briber ($3 * 0.5 * b$). This means that each of the type B participants makes one decision, while each of the two officials has to decide in two separate situations per period. As a consequence, each type O subject receives two payoffs per period, while B receives only one.

Here the parameters are set in such a way that one transaction in TDT2 corresponds to two transactions in IDT2 in terms of payoffs for type O subjects. This means that O's individual incentives in a certain situation (transaction) are not equal to those in IDT1, TDT1 or IDT2, since a certain transfer b leads to double the amount of revenue reaching a single official in IDT2 as compared to the revenue reaching each of the subjects in the role of the official in TDT2. Not only gains (tripled transfers) but also costs (4 EMU) are shared equally[10] between the two officials.

2.2.3 Related literature

In some respects the set-up of the corruption game is similar to that of the gift exchange game (Fehr et al. 1993), which analyzes reciprocity in labour markets. Workers are found to increase their effort when being paid a premium in addition to the market clearing wage. The investment (trust) game (Berg et al. 1995) also shares important elements with the corruption game with respect to design. In this game, the transfer sent by an investor is tripled. Part of the revenue may be sent back by a responder. As in the corruption game, the gains through trust and reciprocity are linear. Several studies find positive levels of reciprocity in one-shot versions and even higher levels in repeated versions of these games (Dufwenberg and Gneezy 2000, Gächter and Falk 2002).

The main and most important difference between these games and the corruption game is the fact that cooperation in the form of positive reciprocity is beneficial for all members of society (i.e. the society as a whole), whereas in the case of corruption, cooperation is only beneficial for the client and the official but not for the public in general (and those who are hit by the negative externality in particular). The negative externality is (usually) assumed to be high enough to result in a net social loss. This difference may have strong effects on the level of cooperation and positive reciprocity, as

[10]The assumption of an equal split of the transfer reflects equal bargaining power and similarity of the officials. There is no reason to believe that endogenous distribution would yield a different sharing rule.

well as on its behavioural explanations and motivations.

Contrary to theoretical predictions applying the standard self interested model of payoff maximization, Abbink et al. (2000, 2002), Abbink (2004, 2006), Abbink and Hennig-Schmidt (2006), Jacquemet (2005) and Lambsdorff and Frank (2007, 2010, forthcoming) find a large amount of cooperation in a series of laboratory experiments using one-shot as well as (finitely) repeated versions of set-ups comparable to ours.[11]

These findings can neither be explained (or predicted) by the standard self interested model (see Appendix 2A), nor by models of reciprocity based on social preferences. Altruism (Andreoni and Miller 2002) fails as an explanation since a successful corrupt transaction does not only decrease the payoff of outsiders (resources of potential reference groups), but also decreases the sum of payoffs, leading to an efficiency loss. It is also unconvincing to use an argument based on inequity aversion (Fehr and Schmidt 1999, Charness and Rabin 2002), since the outcome of a successful reciprocal transaction reduces the welfare of those who are potentially (by the externality modelled in mode 1) or certainly (by the externality modelled in mode 2) worse off than the corrupt partners. Any equilibrium would critically depend on the assumptions on the identity of the reference group which is likely to be heterogeneous across and possibly not even consistent within subjects (across periods). The clear strategic background of the situation makes an explanation of reciprocity based on intentions (Falk and Fischbacher 2006 and Dufwenberg and Kirchsteiger 2004) difficult, since positive transfers are unlikely to be considered as displays of kindness. More convincing explanations may be expected in the notion of 'weak' reciprocity based on repetition and reputation formation (Fehr and Fischbacher 2003, Kreps et al. 1982) or of impulses of reciprocity (Strassmair 2009).

However, even though behaviour in the basic set-up may be (partly) explainable by a theoretical model ex post, all theories lack the ability to develop point predictions with respect to the effect of the introduction of the 4EP, since this involves the additional problem of group decision-making. We know from the experimental literature on group decision-making that decisions made in a small group or team can differ substantially from individual decisions even if individual marginal incentives are equal (Chalos and Pickard 1985, Levine and Moreland 1998). Experimental evidence suggests that the

[11] The common feature is the trade-off between reciprocity based maximization of individual long-term payoffs and a combination of impulses of myopic payoff-maximization and preferences for social efficiency.

direction of the effect of group decision-making is ambiguous and depends highly on the nature of the particular situation. The majority of studies find that decisions made in small unitary groups (which is the case in our study) act more in line with the predictions of the self interested model of payoff maximization (Blinder and Morgan 2005, Bone et al. 1999, Bornstein and Yaniv 1998, Kugler et al. 2007), while there is also contrary evidence (e.g. Cason and Mui 1997). Kocher and Sutter (2007) show (using a gift exchange game) that decisions made by a small group may be either more or less in line with selfish preferences. The total effect of group decision-making seems to depend on what kind of motivation dominates the decision-making process, the competitive or the profit maximizing motive.

2.3 Hypotheses

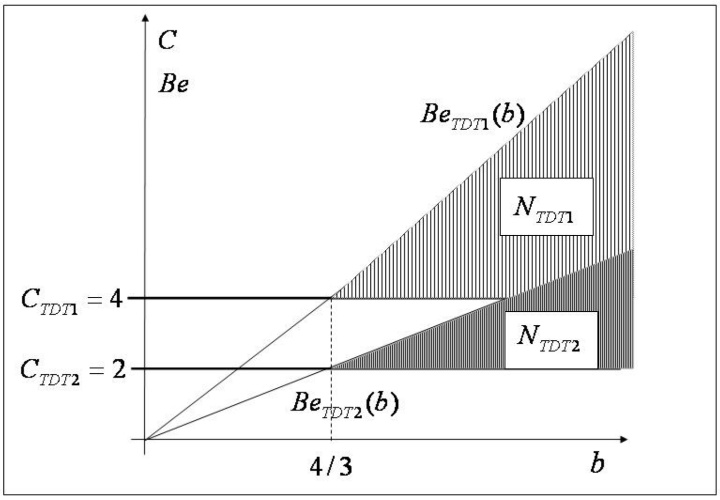
In our analysis, we distinguish between two main effects of the introduction of the 4EP with respect to the officials' behaviour. First, the introduction of the 4EP causes a bribe to be divided between two officials instead benefiting just one. Keeping B's behaviour (i.e. the amount of transfer) constant, the splitting of the bribe causes each official to receive half of what a single official would have got. We call the officials' immediate reaction to the lower benefit from a bribe in TDT1 and TDT2 the 'Bribe Splitting Effect' (BSE). Second, we consider the pure effect of group decision-making when we hold marginal effects constant and call it the 'Group Decision-making Effect' (GDE).

2.3.1 Bribe Splitting Effect

From a series of experiments using comparable set-ups (e.g. Abbink 2004), we know that the probability of success increases with the level of transfer (bribe). The phenomenon relates closely to the findings of reciprocity in the gift exchange game (Fehr et al. 1993). For simplicity we assume that the bribe is shared equally between the two officials, ignoring potential distributional issues.[12] The existence and magnitude of an effect stemming from the splitting of a bribe depends on what officials condition their behaviour on. Officials may consider the monetary benefits of the bribe only or they may include the 'intentions' of the briber. Leaving intentions out, the correlation between the size of the bribe and the probability of positive reciprocity may be explained by the trade-off between marginal benefits and marginal costs of engaging in corruption. Moreover, considering the repeated set-up, the minimum amount of transfer (in an early period) needs to be large enough to give O an incentive (in terms of expected benefits in the future) to make her incur personal costs with the objective to trigger B's reciprocity in future periods.

Figure 2.2 illustrates the monetary benefits (Be) and costs (C) on the y-axis as a function of the level of transfer on the x-axis for treatments TDT1 and TDT2. In TDT2, the benefits as well as the costs are split between the officials and hence are both half as large as in TDT1. The net monetary benefits (differences between benefits and costs) are labelled as N_{TDT1} and N_{TDT2}.

Figure 2.2: Monetary Costs and Benefits of corruption in TDT1 and TDT2



For $b < \frac{4}{3}$, corrupt behaviour cannot be rationalized since costs are larger than benefits.[13] For $b > \frac{4}{3}$ the monetary benefit is always larger in TDT1 than in TDT2. The difference is increasing in b. If O conditions her corruptibility on the net benefit of the transaction (at all), the probability of corrupt success must be weakly larger in TDT1

[12] In our anonymous setting there is no reason to believe that there should be any other kind of distribution rule to be agreed by both officials.

[13] For $b < \frac{4}{3}$ not even the technical costs c are covered.

than in TDT2 for all $b > \frac{4}{3}$.[14] The difference in probabilities should be even larger considering not only the monetary but also the moral costs of inflicting (monetary) harm to other members of society (i.e. other participants of the experiment in mode 1 or recipients of the donation in mode 2). Unlike the technical costs (c), these are likely to apply to both officials at the full scale, since the approval of both is needed to finalize a corrupt transaction and hence they should both be held morally accountable (see Appendix 2A under 'Responsibility and veto Power').

However, if subjects condition their behaviour on intentions and equilibrium outcomes alone (i.e. they consider the 'kindness' of B's decision only in the sense that it leads to a certain outcome, given that the transaction is successful) there should not be any difference between the conditional behaviour of type O subjects in TDT1 and TDT2. By the construction of the experiment[15], strategies leading to equalized outcomes between B and O (ignoring the negative externalities to the public) are equal across treatments (require the same actions for both types), see Section 2.5.1. Hence outcome-based models of inequity aversion (e.g. Fehr and Schmidt 1999) would yield the same predictions across treatments assuming the irrelevance of the negative externality.

Hypothesis 1: "Holding bribe levels constant, there will be no difference in corruption levels between TDT1 and TDT2, if officials condition their reciprocal behaviour exclusively on intentions or consider equalization of payoffs only."

Hypothesis 1 will be rejected if the actual amount of bribe in a particular situation has an effect on the probability of success of a corrupt transaction (being different in TDT1 as compared to TDT2). In this case we call the effect the Bribe Splitting Effect (BSE).

2.3.2 Group Decision-making Effect

In order to measure the effect of group decision-making (GDE) separated from BSE, we have to compare the behaviour of subjects deciding alone and subjects deciding

[14]This allows for individual heterogeneity and does not even exclude participants who would never engage in a corrupt transaction ($p_j(b;\ c) = 0$ for any value of b and c)

[15]Including the difference in the number of transactions played per period by the different types.

within a group in situations in which all relevant decision-makers face the same marginal monetary incentives. The comparison between IDT1 and TDT1 satisfies this condition. So we compare the expected number of successful corrupt transactions, $E(N_{success})_{IDT1}$ (one official decides individually) with the expected number of successful corrupt deals, $E(N_{success})_{IDT1}$ (two officials decide jointly) facing the same bribe b.

In IDT1, the probability of success of a corrupt transaction is $p_i(b)$ for the deal in which official O_i is relevant. Assume that the probability of corrupt success (reciprocity) is positively dependent on the relevant bribe b (see e.g. Abbink 2004).
In TDT1, officials O_i and O_j decide jointly in Stage 2 and Stage 3. The group decision-making process provides veto power for non-corrupt and non-reciprocal behaviour ('reject' in Stage 2 and 'defect' in Stage 3).
In this case the probability of success is $p_i(b)p_j(b)$ (both officials have to decide in favour of corruption), if decisions are completely independent. Since $p_i(b_1) \leq 1$ and $p_j(b_2) \leq 1$, $E(N_{success})_{IDT1} = p_i(b) \geq p_i(b)p_j(b) = E(N_{success})_{TDT1}$. As long as the individual behaviour of officials is independent of the decision-making process (including the observation of or the belief on the behaviour of the other official), the expected number of successful corrupt transactions should be weakly greater under IDT1 than under TDT1.

Hypothesis 2: "If decisions are completely independent of the decision-making process we will observe lower (relative) numbers of successful corrupt transactions in TDT1 than in IDT1."

Experimental evidence shows that individual decisions are far from independent when made inside a group or team. The decisions made by small groups tend to be more in line with the predictions of the self interested model of payoff maximization when considering bargaining situations in which competition plays a relevant role, while in games representing social dilemmas, e.g. the public goods or the gift exchange game, groups may even move further away from the predictions of standard game theory (Cason and Mui 1997, Levine and Moreland 1998). Kocher and Sutter (2007) track this back to two opposing motivations driving the effect of group decision making, the competitive and the profit maximizing motive. The total effect depends on which of these dominates. The set-up of a corrupt transaction cannot easily be categorized into one of the two situations. Where reciprocity-based cooperation increases the level of social welfare in most applications (as it usually helps to overcome a social dilemma), corrupt reciprocity decreases social welfare (and efficiency) by design. For the official it is always individually (in the short term) as well as socially optimal not to reciprocate in the corruption game.

Competitive Motive

The pure fact of group membership may cause a shift in individual preferences towards a decision that reflects higher awareness of competition with other groups or individuals (minimal group paradigm, Tajfel and Turner 1986). In our situation this may push groups to behave more in line with the predictions of standard game theory. Members of a group may follow strategies that increase the difference of pay-off levels between in- and out-group members. Especially under mode 1 (fellow subjects are hit by negative external effect), this may help cooperation within and hinder cooperation across groups[16].

An important question, which cannot easily be answered in our setting, is who exactly type O subjects consider as their (refernce) group members. This could be the fellow official (who a subject forms a group with), but it could also be the unit of all participants involved in the transaction, including the type B subject(s).

Being primarily interested in behavioural differences between individual decision-making treatments (ID treatments) and group decision-making treatments (GD treatments), the in-group-effect within the transaction unit (including respective B and O participants) should be considered as irrelevant, since it should be equally present in all treatments. What distinguishes individual from group decision-making treatments is the additional in-group-effect between jointly deciding officials in TDT1 and TDT2. The creation of a sub-group by letting officials decide jointly may result in more competitive behaviour towards their type-B transaction (unit) partners, which may result in myopic profit maximization at the cost of corrupt reciprocity. We denote this motivation as the Competitive Motive (CM).

Profit Maximizing Motive

The *Profit Maximizing Motive* (PMM) causes groups to make decisions that yield larger payoffs (in the long run) while, if necessary, shifting behaviour even further away from equilibria predicted by standard theory (Kocher and Sutter 2007). Despite being in-

[16] This may involve the understanding of participants being involved in a reverse public goods dilemma.

efficient on a social level (by the reverse public goods dilemma), a successful corrupt transaction yields the largest individual payoffs for the transaction partners (unit), given the behaviour of other groups. Groups may be more capable of suppressing short-sighted impulses of behaviour which may maximize myopic payoffs but ultimately decreases total individual payoffs of all transaction partners. This behaviour includes free-riding or defecting in social dilemmas (e.g. the public goods game using the voluntary contribution mechanism) and failing to foresee the breakdown of future cooperation (reciprocal relationships). The Persuasive Argument Theory (PAT, see Pruitt 1971, Bishop and Myers 1974, Burnstein et al. 1973) predicts that groups are more successful in finding strategies that maximize their members' long term payoffs. Explorative and knowledge capacities in groups (containing more than one individual) are expected to be greater than those of a single individual. Chalos and Pickard (1985) proclaim that groups are better in processing information load. In games where payoff maximizing strategies are as complicated as in the repeated corruption game, we expect groups to develop and follow more successful strategies than individuals with respect to maximizing their members' monetary payoffs when we assume that groups and individuals exhibit equal preferences with respect to the trade-off between individual and social welfare maximizing.

Hypothesis 3: "If the group decision-making process is dominated by the CM, outcomes in TDT1 will be closer to the game theoretical predictions than those in IDT1. If the PMM dominates group decision-making, groups will produce higher levels of corruption by following strategies that are more successful in maximizing their members' individual payoffs."[17]

2.3.3 *B*'s behaviour

The introduction of the 4EP may not only affect the behaviour of the officials but also that of the bribers. The direction of the effect depends entirely on the beliefs about the (effects on the) behaviour of the official(s).[18]

[17]Note that the comparison between IDT1 and TDT1 shuts off any potential effect arising from the splitting of the bribe between the two officials deciding jointly in a group since marginal incentives for all subjects are the same in both treatments.

[18]Relying on the assumptions of standard game theory (within the self interested model), we do not expect bribers across treatments to adhere to different beliefs about the behaviour of groups and

Even if we assume very restrictive belief structures it is difficult to formulate consistent hypotheses. If subjects of type B expect groups of officials to be less likely to reciprocate (given a certain amount of bribe) than individual officials, B's reaction can (still) go into both directions. On the one hand, bribers who want to initiate a corrupt transaction may be discouraged by their anticipation of a higher probability of failures and therefore choose '0'-transfers more often in the GD treatments. On the other hand, there might be an increase in the bribe level in the GD treatments, coming from bribers who anticipate that groups of officials are more demanding to be 'convinced' to act in a reciprocal way than individuals are. Therefore the total effect is ambiguous and depends on which of the effects dominates the decisions.

The argument for a belief structure that assumes higher reciprocity (for any bribe level) within groups of officials leads to similarly ambiguous predictions. In contrast to the inconclusive predictions on B's expected reaction on the anticipation of the GDE, we can form hypotheses on the direction of the effect stemming from B's anticipation of the BSE. This can be quantified by comparing average bribe levels between TDT1 and TDT2. If bribers anticipate the BSE (correctly) they may send larger transfers in order to compensate the splitting of the bribe.[19]

Hypothesis 4: "The bribe level (and distribution) will be different in TDT1 and TDT2 if type B subjects anticipate officials to behave according to the BSE and react accordingly."

2.3.4 Gender effects

An especially important finding in the empirical literature on corruption is the relationship between gender and corruption. Dollar et al. (2001) and Swamy et al. (2001) find that female participation in market transactions leads to lower levels of corruption. Sung (2003) however advocates that the findings might be at least partly due to inconsistencies

individuals, since in all treatments the Sub game-perfect Nash Equilibrium (see Appendix 2A) is unique and predicts neither positive transfer levels nor positive reciprocation.

[19]Note that such a reaction can not be explained by models of inequity aversion, since a strategy that aims at equalizing payoffs would not proclaim different levels of transfers across these treatments.

related to omitted variable biases.[20]

Avoiding these problems, Rivas (2007) as well as Lambsdorff and Frank (forthcoming) use controlled laboratory experiments to find that women tend to reciprocate less often with the corrupt partner in the role of the official even if risk aversion, a potential source of behavioural gender difference, can be ruled out as a driving force (Schubert et al. 2009).[21]

The findings lead to the conclusion that the presence of women may destabilize trust and reciprocity-backed stability of the corrupt transaction. It provides a strong argument for policies aiming at increasing women participation in corruption-sensitive sectors within public procurement. Lambsdorff and Frank (forthcoming) show that the gender effect is especially strong when B has a direct opportunity of negative reciprocity, i.e. costly punishment of defective behaviour (by the tool of whistle-blowing). This may be explained by the fact that female participants are less inclined to reciprocate negatively and therefore do not anticipate negative reciprocity in the behaviour of their transactions partners.

In our set-up B's only opportunity of negative reciprocity is by reducing bribes in future periods. Relying on the results of corruption experiments using comparable set-ups we do not expect a strong gender effect in B's behaviour as long as O's gender is unknown. For the individual decision making (ID) treatments, we expect a gender effect only if the female lack of anticipation of negative reciprocity extends to cross-period reciprocity (i.e. anticipating lower transfers in periods following a defected transfer situation). In order to control for gender effects within officials in the GD treatments, we need to distinguish between pure female groups (both officials are women), mixed groups and pure male groups. Even if we assume a gender effect within the officials' decision-making, the interaction of a male and a female official, when deciding jointly, is

[20]The real but unobserved driver may be the quality and level of development of institutions, affecting both, women labour market participation and corruption. Assuming a positive correlation between institutional quality and women (labour market) participation, we would expect an over-estimation of the predicted effect. Furthermore, women might self-select into sectors where corruption is less rampant. This may be motivated by a vector of non-measurable variables presenting even more complex problems of misspecification.

[21]Note that experiments on corruption that model the individual cost of reciprocation by adding a lottery including a large loss of payoff with small probability as a fourth stage of the corruption game (instead of modelling it as a certainty equivalent) cannot distinguish corruption-specific gender effects from gender effects that are caused through differences in individual risk aversion without controlling for individual risk attitudes separately.

unclear.

Hypothesis 5: "If the gender effect is present in our repeated set-up, female officials in the ID treatments and pure female groups of officials in GD treatments may produce lower levels of successful corrupt transactions than male officials in ID treatments and pure male (and mixed groups) in the GD treatments."

2.3.5 Total effect

In order to evaluate the usefulness of the implementation of the 4EP, we consider the total effect, i.e. the combination of the BSE and the GDE. The total effect can be directly measured by the comparison of IDT1 (or IDT2) and TDT2.

The introduction of the 4EP can only help to reduce the level of corrupt activity, if the conditions of either of the two following situations hold. First, the 4EP will certainly reduce corruption, if the BSE and the GDE are both positive. Second, it will reduce corruption, if a positive BSE over-compensates a negative GDE (which is dominated by the Profit Maximizing Motive).

Only if the GDE is negative and over-compensates the BSE, the introduction of the 4EP is counter-productive even without considering the costs of the installation of such an institution.

Hypothesis 6: "If the GDE is negative (PMM is stronger than CM) and dominates the BSE, the average rate of corrupt success will be greater in TDT2 than in either IDT1 or IDT2. In this case the 4EP is counter-productive."

2.4 Procedure

All 8 sessions (two sessions for each treatment) were programmed and conducted at the experimental laboratory MELESSA at the University of Munich. It used the program Z-Tree (Fischbacher 2007) and the organizational software Orsee (Greiner 2004). Each session was conducted with 24 subjects (a total of 192 participants). Subjects were randomly assigned a type, (B or O) and randomly allocated into units of two in IDT1, into units of three (one type B and two type O participants) in TDT1 and into units of four (two type B and two type O participants each) in IDT2 and TDT2. In all treatments group members stayed together in their units for all 10 periods (partner design) where full anonymity was ensured.[22] Every period was paid where 1 EMU was worth 5 eurocents. Payoffs were summed up over all 10 periods and paid out in private at the end of the experiment. The whole experiment took less than 90 minutes. The instructions were kept completely neutral, avoiding any language indicating the subject of research in order to concentrate on the specific features of the model and minimize the differences between the instructions of the treatments. Abbink and Hennig-Schmidt (2006) show that framing has no significant effect on behaviour in the corruption game. All this was common knowledge to all participants. Understanding of the (rather complicated) set-up was insured by partly reading out the instructions, answering questions in private and checking of several control questions. At the end of the experiment subjects filled in a questionnaire including demographic information. Payoffs lay between 4 Euros and 25 Euros excluding a show up fee of 4 Euros which is standard to experiments at MELESSA. Average earnings amounted to 14.37 Euros. A total of 249.10 Euros was paid out as a donation to the organization 'Doctors without Borders' as a result of the decisions made by the participants in the treatments where we chose mode 2 as a model of the negative externality (one session in IDT2 and one session in TDT2).

2.5 Results and Interpretation

As noted in Section 2.2, the negative external effect has been modelled in two different ways. According to predictions using the standard self interested model, there should be no effect of either of the two models and hence no difference between the set-ups, neither in the behaviour of the official (or the group of officials) nor in that of the briber. In our experiment, the difference might have played a role for the size and even direction of the GDE. In mode 1, the negative external effect hits four randomly chosen participants of the same session who themselves make decisions that may result in negative payments to

[22] Note that the interaction of officials was conducted via a chat which did not allow for any form of identification, see the Appendix 2C for the full instructions of TDT2.

other participants. This interaction between groups may produce an unwanted additional in-group effect and destroy the idea of the external effect as an unreciprocated reduction in the payoff of unrelated third parties. In particular, subjects may justify their own corrupt behaviour by their belief of others' corrupt interaction. Corrupt behaviour might even be considered as a payoff equalizing equilibrium ('super-game'). In order to rule out effects stemming from these considerations we applied mode 2 in one session of IDT2 and one session of TDT2. In mode 2, the 'super-game' problem is eliminated by modelling the negative externality as a reduction of a (fixed) amount of donation to the public aid organization 'Doctors without Borders'.[23]

Checking for differences we compare all relevant variables, i.e. the average total transfer-level, the transfer levels after success and failure of a corrupt deal (measure of the client's reciprocity), the average relative number of successful deals, the percentage of rejected bribes and the percentage of zero-value transfers between observations of the two modes and find no significant difference, taking averages over all 10 periods for each (relevant) subject and applying (pair-wise) two-sided Mann Whitney U-tests[24] on group levels ($p \geq 0.363$; $N \geq 16$). We take this as sufficient evidence for the assumption that the design of the externality (mode) does not have any significant impact on the outcomes of the relevant decisions. We therefore pool the data from these sessions in the respective treatments for the entire analysis. The absence of a 'super-game' effect may be explained by subjects having difficulties in forming beliefs of higher orders (see Anderson and Holt 1997, Hung and Plott 2001).

The main objective of the experiment is to evaluate the introduction of the 4EP by comparing the main performance variables in IDT2 and in TDT2. This may have been problematic since these treatments do not only differ in the decision-making process and splitting of the bribe, but also in the numbers of transactions per period for the officials. Where in TDT2 an official interacts with each of the two bribers in her unit of four once in a period, an official in IDT2 only interacts once in every two periods with any of the bribers in her unit of four. This makes a large difference with respect to the 'horizon' of bilateral (IDT2) or multilateral (TDT2) repetition. Considering repetition as

[23]We chose this organization to be able to compare our results to those obtained in Lambsdorff and Frank (2007) who use donations to this public aid organization to model the negative externality of their corruption experiment.

[24]Unless stated otherwise (exact, highest or lowest) $p-$values and numbers of observations (N) apply to the two-sided Mann Whitney U-test.

a major determinant of reciprocal behaviour (see Abbink 2004) there may be differences in corrupt behaviour not related to any of the effects discussed in Section 2.3 (GDE and BSE). A comparison between the outcomes of IDT2 and TDT2 may still be interpreted as capturing the total effect of the introduction of the 4EP if we take into account that it may lead to an increase in the frequency of multilateral transactions. This consideration is not unreasonable assuming that the available number of officials is held fixed across the situations modelled in the treatments.

To check whether there is an effect of the amount of repetition we compare the outcomes of IDT1 and IDT2. Although we do not find any significant differences in the main variables (U-tests; $p > 0.236$; $N \geq 24$), we do not pool the data from these treatments, but apply separate tests.

2.5.1 Descriptive results

Table 2.1: **Performance variables**

	Corrupt success		*Payoff*		*Neg. Externality*		*Transfer level*	
Treatment	mean	std.dev.	mean	std.dev.	mean	std.dev.	mean	std.dev.
IDT1	0.25	0.09	13.13	3.92	2.90	1.01	3.18	1.52
IDT2	0.25	0.06	12.72	3.78	3.49	2.12	3.16	1.25
TDT1	0.47	0.07	7.32	4.26	5.13	2.91	3.58	1.15
TDT2	0.42	0.05	9.40	4.05	5.38	1.97	4.15	1.24

All means are calculated as averages across periods and (relevant) participants of the respective treatment.

Table 2.1 shows the values of the four main performance variables in all treatments. *Corrupt success* depicts the average share of successful transactions per unit ($\frac{N_{success}}{N_{total}}$).[25] *Payoff* represents the average payoff level (in Euros) after the reduction of the negative externality[26] (final payoff) per subject. *Neg. Externality* describes the level of the

[25] Note that units contain two, three or four individuals, which makes a comparison between absolute levels of corrupt success inconclusive.

[26] To be able to accurately compare payoff levels between all treatments, we subtract the relevant share of the reduction of the donation from the actual payoff in the externality mode 2.

(relevant) negative externality in Euros. *Transfer level* measures the average amount of bribe (in EMU) transferred by type B participants per period and unit.

Corruption levels

Comparing corrupt success rates (*Corrupt success*) between treatments, we can identify a relatively small but significant Bribe Splitting Effect. The difference in corrupt success rates between TDT1 and TDT2 (0.05) is significant ($p = 0.082$; $N = 28$). We reject **Hypothesis 1**. Corrupt success does not seem to depend exclusively on intentions or final outcomes. Moreover we find a substantial Group Decision-making Effect (**Hypothesis 2**). The negative difference in the corrupt success-levels between IDT1 and TDT1 amounts to 0.22 ($p = 0.034$; $N = 40$) and strongly suggests the dominance of the Profit Maximizing Motive (**Hypothesis 3**). The large and significant difference (0.17) between IDT1/IDT2 and TDT2 (IDT1 vs. TDT2: $p = 0.002$; N=36, IDT2 vs. TDT2: $p = 0.042$; N=24) indicates a negative total effect of the introduction of the 4EP even if we control for the difference in the number of repetitions (between IDT1 and TDT2) confirming **Hypothesis 6**[27].

Figure 2.3: Success-probabilities over Periods

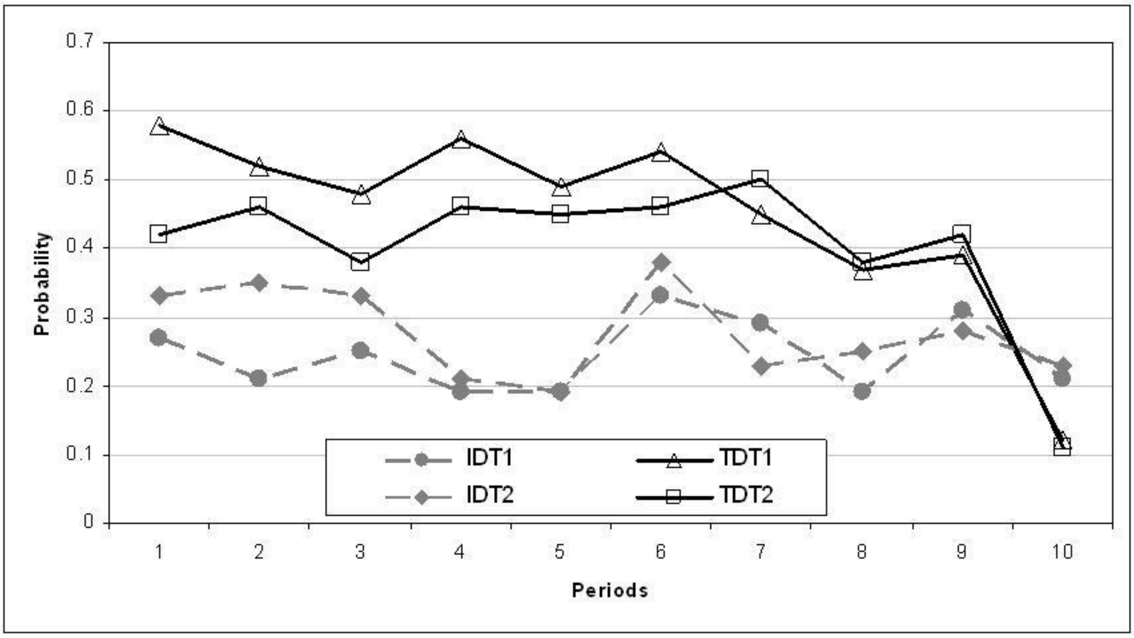


Figure 2.3 demonstrates the differences in the dynamic development of success-

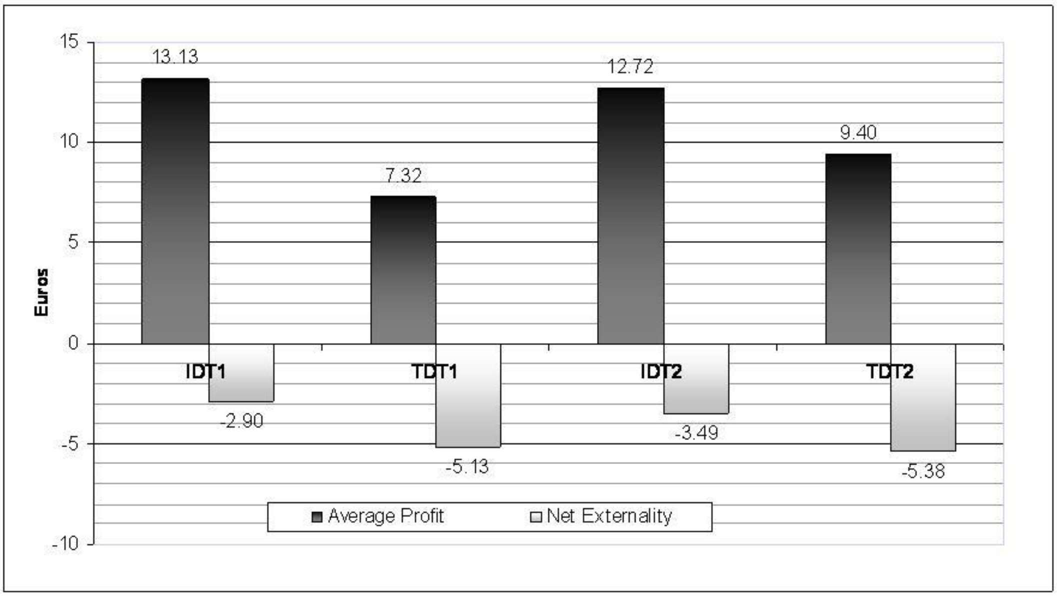
[27]The Profit Maximizing effect dominates the Competetive Motive in group decision-making. The resulting effect is stronger than the Bribe Splitting Effect.

probabilities over the 10 periods between the treatments. Subjects in the GD-treatments start with very high levels of successful corruption, which decline gradually, showing a large 'end-game effect' (end of repetition in the 10^{th} period), while corruption levels stay relatively stable (at a lower level) in both ID treatments. We interpret this observation as first evidence for the hypothesis of groups following systematically different strategies compared to individuals.

Payoff and externalities

The average net payoff[28] is more meaningful in terms of a welfare comparison between the treatments than the level of corruption. Figure 2.4 shows the average total payoff level (*Payoff*) per subject in comparison to the average negative external effects (caused by the average subject, *Externality*) for all four treatments.

Figure 2.4: Payoff levels and levels of externalities



The bilateral differences can be explained through two effects related to the transfer level and corruptibility. First, the level of transfer increases the payoff independent of the outcome since bribes are tripled (accounting for the assumed difference in marginal utility between O and B) and hence enhance efficiency and payoffs. Second, the level of successful corrupt transactions (which is positively correlated with the transfer level, see Section 2.5.2) reduces payoff levels through the impact of the negative externality. The

[28]As net payoff we define the sum of payoffs after subtraction of the externality caused by others' decisions.

relatively large (but marginally insignificant) difference in payoff levels between TDT1 and TDT2 of 2.08 Euros ($p = 0.157$; $N = 28$) is due to lower levels of transfers and higher levels of successful corruption. The relatively low difference in payoff levels between both ID treatments and TDT2 can be explained by the negative difference in bribe levels and the positive difference in the level of corrupt success. While we can report a significant difference in payoff levels only between IDT1 and TDT2 ($\Delta = 3.73$; $p = 0.071$; $N = 36$) and not between IDT2 and TDT2 ($\Delta = 3.32$; $p = 0.153$; $N = 24$), all differences between ID treatments and TDT1 are significant on the 1%-level ($p \leq 0.002$ $N \geq 28$). Taking cumulated payoffs as a measure for welfare suggests a negative total effect of the introduction of the 4EP. To consolidate this result we can report large differences between the treatments in the distribution of payoffs across subjects. Table 2.2 depicts the values of the Gini-coefficient[29] for all treatments.

Table 2.2: **Gini Coefficients**

	IDT1	IDT2	TDT1	TDT2
Gini	0.14	0.15*	0.18	0.16*
Observations	48	48	48	48
Bribers/Officials	24/24	24/24	16/32	24/24

* The values of the Gini-coefficients of sessions run in mode 2 are adjusted for the negative externality.

In both GD treatments some units of subjects accumulate large payoffs by sending high transfers and profiting from positive reciprocity at the expense of other participants (mode 1) whose units do not engage in corruption (as often) but are hit by the negative external effect just the same. This leads to more inequality in payoffs indicated by higher values in the Gini-coefficients. The difference in the Gini-coefficients between TDT1 and TDT2 may be due to the different numbers of officials and bribers.

Payoff maximizing strategies of B

Assuming that subjects aim at maximizing personal payoffs, type B subjects' strategies

[29]The Gini-coefficient measures the distribution of wealth within a certain population on a scale between 0 and 1 where a value of 0 corresponds to complete equality and a value of 1 signifies complete inequality.

vary systematically in their effectiveness between the treatments. We use a simple linear panel regression (random effects) to derive the profitability of bribing. To account for dynamic aspects and potentially decreasing (or increasing) profitability of transfers we use the following specification for 88 (all) participants of type B. We cluster errors on the unit level to take possible dependence of behaviour between type B participants into account.

$$M1: \quad PP_{it} = \beta_0 + \beta_1 b_{it} + \beta_2 b_{it}^2 + \gamma D_i + \delta D_i * b_{it} + \zeta D_i * b_{it}^2 + \epsilon_{it} \tag{2.1}$$

Index i stands for respective subject, index t for the respective period. The dependent variable PP signifies the individual payoff (in EMU) excluding subtractions from the externality caused by subjects outside a subject's unit. Independent variable b indicates the transfer paid by briber i in period t. The quadratic variable b^2 captures a potential non-linear effect of the transfer on period payoffs (decreasing marginal payoffs). Vector D stands for the treatment Dummy variables D_{IDT2}, D_{TDT1} and D_{TDT2}. Accordingly, vectors $D*b$ and $D*b^2$ contain interaction effects of b with treatment dummies for TDT2, TDT1 and TDT2.
We find (see Table 2.8 in Appendix 2B for the full list of estimated coefficients for M1[30]) that the individual payoff per period is strictly increasing in the level of transfer sent to the (group) of official(s) in all treatments.

There is no significant evidence for a decreasing profitability of bribing. Neither β_2 nor any element of ζ is different from 0 (t-tests; $p \geq 0.262$) in either of the treatments. The main finding is that bribing is far more profitable in the GD than in the ID treatments. On average an additional unit of transfer increases the payoff per period by 1.896 EMU in IDT1 and 2.029 in IDT2 ($\Delta = 0.1321$; t-test: $p = 0.284$). By contrast, the marginal effect is 3.94 EMU per unit in TDT1 and 2.63 EMU in TDT2. The differences in the marginal effects between ID and GD treatments are all highly significant (t-tests for differences between IDT1 and TDT1 and TDT2: $p < 0.001$ and F-tests for differences between IDT2 and TDT1/TDT2; $p < 0.001$).

Payoff maximizing strategies of *O*

[30]Using pooled OLS (with and without period dummies) did not yield qualitatively different results. To account for the censored independent variable we also ran the model with Tobit yielding similar results.

The differences in the individual profitability from engaging in corrupt activities between the treatments are not as large for the officials. We estimate a simple OLS regression, measuring the marginal effect of N, the number of successful corrupt transactions (number of choices in which the (group of) official(s) has cooperated) on 'Payoff', the total payoff of each (group of) official(s).[31]

$$M2: \quad \text{Payoff}_i = \beta_0 + \beta_1 N_i + \gamma D_i + \delta D_i * N_i + \epsilon_i \tag{2.2}$$

Vectors D and $D * N$ have analoguous interpretations as D and $D * b$ in Model M1. The results of the regression are reported in Table 2.9 of Appendix 2B. Using OLS as well as Tobit (as a robustness check), we find a strong positive effect of the number of successful transactions on the total payoff in all treatments. The profitability of being corrupt is significantly higher in both GD treatments than in the ID treatments (t-tests for IDT1 vs. TDT1/TDT2 and F-tests for IDT2 vs. TDT1/TDT2; $p \leq 0.004$). While officials in the ID treatments earn on average only 15.892 (IDT1) and 15,075 (IDT2) EMU more for an additional successful corrupt transaction, the rate is at 19,035 (TDT1) and 20.923 (TDT2) EMU considerably higher in the GD treatments (Differences between ID and TD treatments are all significant; t-test for IDT1 vs. TDT1/TDT2, F-test for IDT2 vs. TDT1/TDT2: $p \leq 0.001$).

Altruism

To accept a bribe instead of rejecting it (in Stage 2 of the game), means to keep the benefit of the tripled transfer for oneself instead of sharing it with the public. Moreover, the corruption game is designed in such a way that for any given level of transfer, the socially optimal decision (maximizing the sum of payoffs) is to reject the bribe in Stage 2. The number of non-zero rejections of bribes in Stage 2 is extremely low in all four treatments.[32] In only 6.75% of possible cases (i.e. if $b > 0$) (groups of) type O subjects reject a non-zero bribe in IDT1 compared to 6.22% in TDT1, 5.40% in IDT2 and 5.48% in TDT2. Displays of altruism towards the public at a high personal cost ($2b$) are

[31] Since officials within a unit in the GD treatments decide jointly they receive the same payoff and are therefore treated as a single observation. Officials who are in the same unit but decide independently (IDT2) are treated as individual observations but we cluster their standard errors in the regression.

[32] Accepting a 0-level bribe should be (weakly) dominated by the option 'reject', since it inflicts damage to the public with no personal gain. Therefore an interpretation as altruistic behaviour cannot be justified.

rare and confirm the findings of Büchner et al. (2008) with respect to altruism in the context of negative externalities. Moreover, we do not find a significant difference across treatments (pair-wise U-tests between all treatments: $p \leq 0.472$; $N \geq 24$).

Transfer levels

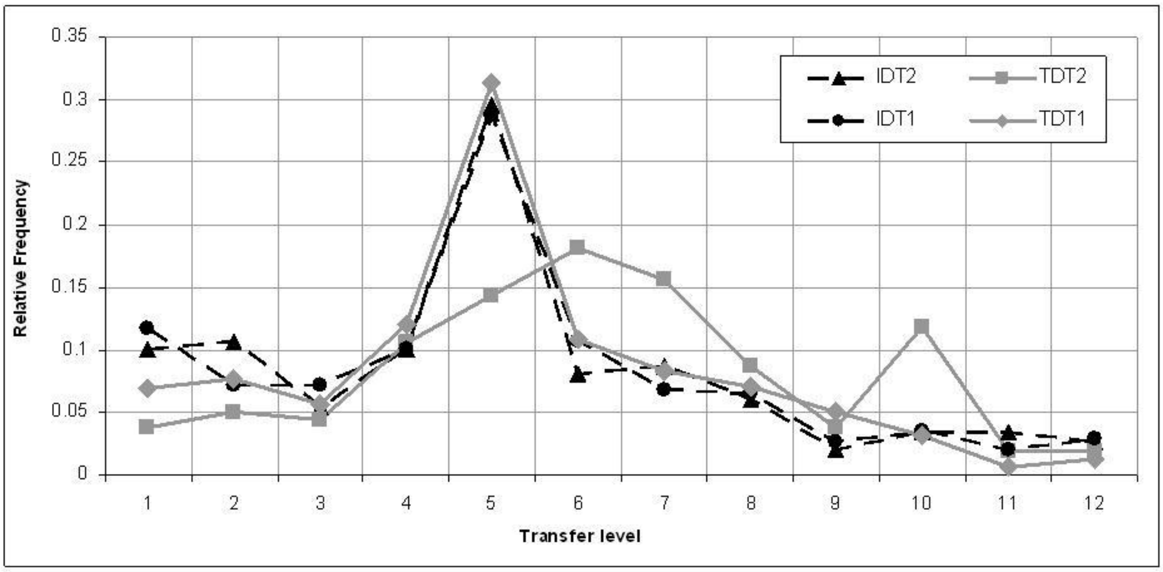
Average bribe levels (including 0-transfers) are substantially (and significantly, U-tests in all treatments: $p < 0.001$, $N \geq 12$) larger than 0 for all treatments and almost identical within the ID treatments (3.18 and 3.16 EMU). Transfers are at 3.58 only insignificantly larger in TDT1 than in the individual decision-making treatments (TDT1 vs. IDT1/IDT2: $p = 0.351/0.464$, $N = 36/28$). At 4.15 EMU, the average transfer level in TDT2 is significantly (U-tests: TDT2 vs. TDT1/IDT1/IDT2; $p \leq 0.041$; $N \geq 24$) larger than those in any of the other treatments. The large difference in transfer levels between TDT1 and TDT2 suggests that bribers anticipate different behaviour from officials and react accordingly. Taking into account that success levels in corruption are significantly lower in TDT2 than in TDT1 despite the positive difference in transfer levels, we conclude, assuming realistic beliefs, that bribers anticipate the BSE and 'react' by trying to 'convince' officials by transferring larger bribes (**Hypothesis 4**).

The distribution of the size of transfers reveals even more information about B's behaviour. Figure 2.5 shows the relative frequency of transfer levels for all treatments. Transfers are almost identically distributed in IDT1, TDT1 and IDT2. There are only few low ($b < 4$ EMU) and high ($b > 8$ EMU) transfers. We observe a very strong mode at $b = 5$. This particular observation may, e.g. be explained by subjects behaving according to preferences based on inequity aversion (Fehr and Schmidt 1999). The strategy [$b = 5$ EMU; 'accept'; 'cooperate'] leads to equal payoffs for B and O within a unit in all four treatments.[33]

The distribution in TDT2 depicts a significantly different pattern. We compare the distribution of bribes in TDT2 to those in all three other treatments with a Kolmogorov-Smirnov test ($p < 0.001$; observations of strictly positive bribes: $N \geq 382$; all observations: $N = 480$). In TDT2 probability mass is shifted towards the higher

[33]For the results to be explained by social preferences we either need to assume that B and O's reference group excludes the public (other participants in mode 1 or recipients of donations from 'Doctors without Borders' in mode 2), or assume a certain structure of beliefs on the (corrupt) behaviour of the other units (only valid for mode 1).

Figure 2.5: Relative distribution of transfer levels



end of transfer levels. The second mode at $b = 10$ EMU can be explained by a reaction to the bribe splitting. Within a certain situation between a particular briber and two officials, the strategy [‘b = 10 EMU’; ‘accept’; ‘cooperate’] leads to equalized payoffs. However, since the total payoff for a certain period consists of two payments for each of the officials, while the briber only receives one, this strategy does not equalize outcomes with respect to the total period payoff. It is common knowledge (and was made explicitly clear with the help of several control questions in the instructions) that it is the strategy [‘b = 5 EMU’; ‘accept’; ‘cooperate’] that yields equal outcomes (in expectations) for all transaction partners, just as in the other treatments. While the monetary benefits as well as the monetary costs are split in TDT2 (compared to TDT1) we may interpret the higher levels of transfer levels in TDT2 as a premium compensation for the ‘moral’ costs of causing damage to the public, which applies to both participants, since they both have full (moral) responsibility for the corrupt outcome.[34]

2.5.2 Conditional reciprocity

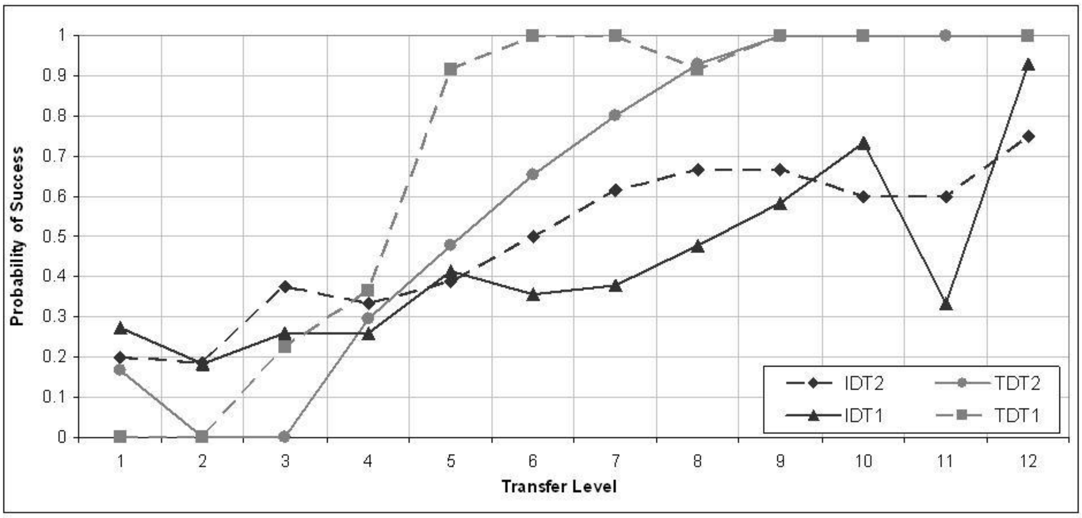
Conditional reciprocity of *O*

Throughout the experimental literature on trust and reciprocity in general (Fehr et al.

[34]See Appendix 2A for an explanation of full responsibility.

1993) and on the corruption game in particular (see e.g. Abbink et al. 2002, Abbink 2004, Lambsdorff and Frank 2010), the scale of reciprocation has been found to depend critically on the first mover's behaviour (in our case the level of b). In order to explain differences in the level of corruption between our treatments we need to control for the level of transfer. Figure 2.6 shows the probability of a successful corrupt deal (in %) for any positive bribe level (1-12 EMU) for all treatments.

Figure 2.6: Success rates conditional on Transfer



In all treatments success rates increase with the level of b (Spearman rank correlation coefficients $\rho \geq 0.85$; $p < 0.001$ for all treatments). This is not surprising, since the cost of corruption is fixed, and future gains from successful reciprocity increase with the size of the bribe, assuming that bribe levels are positively correlated across periods. While the relationship between success and transfer levels seems almost linear in IDT1 and IDT2, we observe a different pattern for the GD treatments. In these treatments the conditional probability is substantially higher for large transfers ($b > 5$ EMU) and slightly lower for small ones ($b < 5$ EMU). We find significant positive differences in probabilities between GD and ID treatments for larger transfer levels $5 < b \leq 11$, considering transfer levels separately (TDT1 vs. IDT1/IDT2: $p \leq 0.021$, TDT2 vs. IDT1/IDT2: $p \leq 0.094$; $N \geq 12$). The negative differences for $b < 5$ are not significant for pair-wise comparisons except for $b = 2$ (IDT1/IDT2 vs. TDT1/TDT2: $p \leq 0.08$; $N = 22$). Differences at $b = \{0; 1; 3; 4; 5; 12\}$ are not significant at any relevant level between ID and TD treatments ($p \geq 0.127$; $N \geq 5$).

Parametrically, the most straight-forward way to quantify the differences in the prob-

ability of a successful corrupt transaction conditional on the relevant transfer between the treatments is to use a linear panel regression (random effects) controlling for clustered standard errors on the unit level. Since we are primarily interested in the causal relationship between the level of transfer (b) and the success levels (SC), we do not distinguish between a corrupt deal that failed in Stage 2 or in Stage 3. Treating the decisions 'reject' and 'defect' equally with respect to the outcome of a corrupt deal (success or failure), we do not have to take the selection process of reaching Stage 3 into account.[35]
We use the following specifications for the linear probability model:

$$M3: \quad \text{Prob}(SC_{it} = 1|\psi X) = \beta_0 + \beta_1 b_{it} + \gamma D_i + \delta D_i * b_{it} + \theta Z_i + \epsilon_{it} \tag{2.3}$$

ψ stands for the vector of coefficients. X represents independent variables. Again, vectors D and $D * b$ stand for treatment dummies and interaction terms of treatment dummies with the transfer b, just as in Model M1. Vector Z contains individual demographic characteristics (e.g. age, gender[36], an interaction term between gender and the level of transfer, etc.) obtained from the questionnaire. Since we do not find any significant effects with any of these characteristics we do not report them in the regression output (Table 2.3).[37] M3 in Table 2.3 reports the results (coefficients and standard errors) of the linear probability model.

In all treatments, we find that an additional unit in transfer (b) increases the probability of the corrupt success significantly (1%-level). The effect is significantly stronger in both GD than in the ID treatments (t-tests for IDT1 vs. TDT1/TDT2: $p \leq 0.003$, F-tests for IDT2 vs. TDT1/TDT2: $p < 0.001$). There is no indication for the confirmation of the hypothesis that women are less reciprocal in a corrupt transaction (than men) when we consider observations from all four treatments. The behaviour of female officials (in ID treatments) and 'all-female' groups of officials (in GD treatments) does not appear to be different from that of their male ('all-male' and mixed group-) counterparts with respect to corrupt reciprocity (all coefficients including a gender dummy remain highly

[35] Treating the outcomes of 'reject' and 'defect' differently would require a Heckman-selection process explaining the selection of cases in which Stage 3 is reached (Heckman 1979).

[36] For officials in the GD treatments we use a dummy for 'all-female' groups and do not distinguish between 'all-male' and mixed groups.

[37] As a robustness check we ran the panel regression with a series of specifications, including a regression excluding Z and a set of pooled OLS regressions including dummy variables for periods. None of these specifications yield results qualitatively different from those reported in the left part of Table 2.3.

insignificant; t-tests: $p \geq 0.397$). Female participants do not seem to behave more in line with the predictions of the standard self-interested model (**Hypothesis 5**). An explanation may be found in the absence of the possibility of direct punishment of defective behaviour by the bribers (e.g. negative costly retaliation), which is believed to be a main determinant of the gender effect found in Lambsdorff and Frank (forthcoming).[38]

Table 2.3: **Output of (M3) and (M4)**

	Dependent variable: SC			
	(M3) Lin. Prob		(M4) Probit	
	Coefficient	Stand. error	Coefficient	Stand. error
Constant	0.0582***	0.0215	1.5832***	0.1052
D_{IDT2}	0.0981***	0.0351	0.5442***	0.1536
D_{TDT1}	−0.0343**	0.016	−0.4637**	0.2091
D_{TDT2}	−0.0722***	0.0232	−1.3272***	0.3162
b	0.0593***	0.0132	0.2123***	0.0171
$D_{IDT2}{*}b$	−0.0208	0.0175	−0.0981	0.0872
$D_{TDT1}{*}b$	0.0521***	0.0162	0.2438***	0.0402
$D_{TDT2}{*}b$	0.0462***	0.0133	0.3402***	0.0540
	Pseudo $R^2 = 0.36$		–	

*** denotes significance at the 1%-level.

Number of subjects: 96, Number of clusters: 64, Number of periods: 10

The non-linear relationship of success-probabilities and transfer levels observed in Figure 2.6 can be quantified by a simple maximum likelihood model. To account for differences in the marginal effect of an additional unit in transfer on the success probability across transfer levels we run the following Probit model in its panel version (random effects).[39] We use the same set of independent variables and repeat all robustness checks

[38] See Chapter 3 for a comprehensive discussion of the gender effect.

[39] See Pereira et al. (2006) and Gneezy and List (2006) for examples of the use of a panel version of maximum likelihood models in comparable settings, i.e. repeated gift exchange games.

(pooled version etc.) applied to the linear probability model (M3).

$$M4: \quad \text{Prob}(SC_{it} = 1|\psi X) = \phi(\beta_0 + \beta_1 b_{it} + \gamma D_i + \delta D_i * b_{it} + \theta Z_i) \quad (2.4)$$

Again, ψ stands for the vector of coefficients and X for independent variables. As expected, qualitative results (direction and significance of the evaluated marginal effects at the mean of of transfers $\bar{b} = 3.46$) do not change compared to the results from the linear probability model, see (M4) in Table 2.3. Table 3.10 in Appendix 2B reports marginal effects of the relevant[40] variables as well as predicted conditional probabilities of success of model M4. The Probit model shows that marginal effects are lower in the TD than in the ID treatments for low transfers, $b < 3$, while they are higher for $b \geq 4$. Consequently, the predicted success levels (probabilities) conditional on the transfer level are lower in the TD treatments than in the ID treatments for $b \leq 4$ while they are larger for $b \geq 6$ (see Table 2.10 in Appendix 2B).

The pattern shown in Figure 2.6 and quantified in M4, i.e. a stronger curvature of the probabilistic cumulative distribution function for GD than for ID treatments, may be explained by differences in the strategies between groups and individuals. On the one hand, groups of officials seem to 'defect' (or 'reject') more often in the case of low transfer levels. On the other hand, they seem to be more likely to reward high transfers than their individual counterparts by corrupt reciprocity. We interpret this as strategic signals of unwillingness to return the corrupt favour in less profitable transactions (aiming at inducing a higher transfer in the following periods) and signals of willingness to reciprocate for high transfers (aiming at receiving further high transfer in future periods in exchange for cooperation). This strategy seems to aim at the extraction of a maximum amount of cumulative bribes. In all treatments, a large fraction of non-zero transfers over all ten periods (between 36% in IDT1 and 52% in TDT2) fall into the interval for which the probability of success is significantly larger in the GD than in the ID treatments. Hence the strategies followed by groups seem to be more successful in the sense of higher reciprocal stability between briber and official than the strategies applied by individuals. We interpret this as a piece of strong evidence for the dominance of the Profit Maximizing Motive in group decision-making (**Hypothesis 3**).

[40]Again we do not report any coefficients that are not significant, e.g. a dummy variable for gender.

Conditional reciprocity of B

The stabilization of the corrupt transaction does not only depend on the reciprocal behaviour of the official but also on that of the briber. In order to test for conditional reciprocity on the side of the briber, we compare the average size of transfer between periods that follow a successful corrupt transaction and those that follow a failed[41] transaction. We apply a method introduced by Abbink (2004) that requires calculating the difference in average bribe transfers after a successful and after an unsuccessful deal (ignoring the history of events up to this point in time) for every type B participant. The resulting variable is called R ($R = \frac{\sum b_s}{N_s} - \frac{\sum b_f}{N_f}$, where s stands for success and f for failure). Replicating the qualitative results reported by Abbink (2004), briber's reciprocity is strictly positive for an average of 83% (79.2% in IDT1, 82.3% in IDT2, 87.5% in TDT1 and 84.3% in TDT2) of (the total of 88) type$B$ participants in all treatments. This indicates a strong positive correlation between former corrupt success and the magnitude of the bribe in the next period.[42] Additional to the simple R-values we also calculate a adjusted R-values which consider only observations of type B subjects who exhibit at least one case of corrupt success and at least one case of corrupt failure.[43] We can reject the hypothesis that R- (adjusted R-) values are different from 0 for all treatments (Wilcoxon matched pairs signed rank test, $p < 0.001$; $N \geq 14$).

We find substantially higher R-values for both GD treatments than for the ID treatments, which further strengthens the hypothesis concerning the success of strategies followed by groups of officials with respect to payoff maximization through corrupt reciprocity (maximizing the benefit from transfers). This also explains why corruption is more profitable for officials in the GD treatments (see Section 2.5.1). Type B subjects may anticipate type O's strategies with respect to reciprocity and adapt by developing more accentuated reciprocity in their own behaviour. We do not find any significant effect of gender (splitting observations according to B's gender in all treatments and applying pair-wise U-tests, female vs. male: $p \geq 0.476$; $N \geq 14$) with respect to bribers' reciprocity.

[41] Again, we do not distinguish between cases of rejected bribes and cases of defections.

[42] The variable R is a very crude measure of reciprocity in the sense that it does not control for potential determinants of the decision, e.g. the number of successful transactions before $t-1$ or the magnitude of transfers in former situations. Nevertheless it provides some (comparable) measure of B's reciprocity.

[43] For groups without observations of both choices the interpretation of R is useless. Therefore we exclude those groups in the statistical tests.

Table 2.4: ***R*-value and adjusted *R*-value**

	IDT1	IDT2	TDT1	TDT2
R-Value	1.846***	2.036***	2.857***	2.720***
	(0.23)	(0.36)	(0.45)	(0.39)
adjusted R-Value	2.445***	2.300***	3.064***	3.142***
	(0.22)	(0.32)	(0.37)	(0.28)

Standard deviations in parentheses

*** denotes significance of a Wilcoxon matched pairs rank sum test at the 1%-level

2.5.3 Switching behaviour

So far we have analyzed revealed behaviour by considering outcomes alone. Our data on the group decision-making process allows for a more detailed analysis of the reasons and motivations underlying observed treatment differences. Since there are no significant differences in the behaviour of officials between the treatments in Stage 2 (see Section 2.5.1) we concentrate on O's Stage 3 behaviour.

The data on initial choices of individual officials in the GD treatments demonstrate that (at least part of) the higher levels of reciprocity within groups are due to mechanisms within the decision-making process and not based on differences in individual preferences. First, we identify situations in which officials within a group initially revealed opposing opinions on a decision, i.e. one official in the group chose to 'defect' and the other to 'cooperate' in the first step of the decision-making process. Second, we compare the (relative) numbers of successful corrupt transactions and failures following initial disagreement. For simplicity we pool cases of final disagreement and final consent against reciprocation (since both cases lead to a failed deal because of the veto power rule). Tables 2.5 and 2.6 show average percentages of corrupt success and failure conditional on initial consent (or the lack of it).

Table 2.5: **Success and Initial Consent, TDT1**

	Successful corruption	Failed Corruption	Total
No initial Consent	23.13%	8.75%	31.88%
Initial Consent	23.75%	44.37%	68.12%
Total	46.88%	53.12%	100%

Averages are derived from 160 transactions (16 independent groups of officials in 10 periods) in TDT1

Table 2.6: **Success and Initial Consent, TDT2**

	Successful corruption	Failed Corruption	Total
No initial Consent	20.42%	18.33%	38.75%
Initial Consent	21.66%	39.59%	61.25%
Total	42.08%	57.92%	100%

Averages are derived from 240 transactions (12 independent groups of officials in 10 periods) in TDT2

Assuming independence of decisions (i.e. no influence of the process on final decisions) we would expect 100% of transactions without initial consent to fail because of the veto power of the non-reciprocating official. On the contrary, we find that the final decision was made in favour of (corruption-stabilizing) reciprocity in 72.6% ($\frac{23.13}{31.88}$, TDT1 in Table 2.5) and 52.70% ($\frac{20.42}{38.75}$, TDT2 in Table 2.6) of cases in which the two officials initially disagreed. Assuming that initial decisions reflect the true underlying preferences, this means that the decision-making process alone is responsible for a large share of the treatment effects with respect to corrupt success levels. We conclude that (in both treatments) those officials who are in favour of engaging in, or maintaining, a successful corrupt relationship dominate the outcome of the decision-making process although their decision-adversaries hold veto power. We take this finding as evidence for the Persuasive Argument Theory (Pruitt 1971) which suggests that those participants (in the role of O) who provide the most valuable ideas for maximizing long term individual payoffs during the experiment (which in our case is the maintenance of the corrupt relationship through reciprocity, see Section 2.5.2) dominate the decisions within a group.

2.5.4 Content analysis

In addition to the arguments derived from the comparison of outcomes between the treatments (see Section 2.5.1 and 2.5.2) and the analysis of choices in the different phases of the group decision-making process (see Section 2.5.3), we are able to get some insight into the mechanism of group decision-making by considering the content of the messages[44] exchanged during the decision-making processes of Stage 2 and Stage 3. 22 (out of 28, 16 in TDT1 and 12 in TDT2) groups exchanged electronic messages.[45] First, we separate messages and identify 132 distinct statements.[46] We allocate each statement (sent in either of the two stages) into four main categories: 'Neutral' (statements that do not contain any traceable argument, e.g. 'Hello, nice game'); 'Social' (statements including arguments against the cooperation in the corrupt transaction mentioning the negative externality, e.g. 'We have to consider the effect on the others, we should not cooperate'); 'Strategic' (arguments in favour of the stabilization of the reciprocal relationship with the objective of payoff maximization, e.g. 'Let us cooperate, otherwise we won't get any profit in the next period(s)') and 'Strategic Neg.' (arguments against cooperation in a certain period to implicitly demand larger transfers in future periods, e.g. 'Do not re-transfer, then he [the briber] will know to give more next time').[47] We add a 5^{th} category 'Social/Strategic' to account for (mostly twisted) statements that included both, other-regarding (social) and strategic (payoff maximizing) arguments. Table 2.7 reports the relative frequencies of statements of the respective categories subdivided by the final outcome of the respective transaction in terms of success and failure. Of all statements, only 12.2%[48] contain other-regarding arguments (Social and Social/Strategic). Their low frequency is noteworthy, and so is their lack of effectiveness (only 37.7%[49]) of transactions finally fail).

[44] We analyse all electronic chat messages exchanged by officials in the GD treatments.

[45] 6 groups either did not encounter a situation of initial disagreement or ignored the possibility of writing messages.

[46] A 'conversation' between two officials may yield more than one statement since it may be split into single entries.

[47] All examples are translated (word by word) into English from the original statements in German.

[48] $8.3 + 3.9\%$, Table 2.7

[49] $\frac{3.1+1.5}{12.2}\%$, Table 2.7

Table 2.7: **Success and Content**

	Neutral	Social	Strategic	Strategic Neg.	Social/Strategic	Total
Success	8.3%	0.8%	27.3%	1.5%	6.8%	44.7%
Failure	31.8%	3.1%	12.1%	6.8%	1.5%	55.3%
Total	40.1%	3.9%	39.4%	8.3%	8.3%	100%

Percentages are derived from 132 statements in TDT1 and TDT2

An explanation may be that in more than 75% of all situations a social argument was followed (in the same chat conversation) by a statement arguing in favour of strategic reciprocity. 82% of these situations ended with a successful corrupt transaction. The majority (56.0%[50]) of statements contained arguments in favour of some kind of strategic reciprocity. Additional to 63 statements of positive reciprocity there were 11 separate statements arguing in favour of strategic defection aimed at extracting larger bribes in future periods. In 19 (out of all 28 or 22 relevant) groups of officials we found at least one statement in favour of strategic reciprocity (positive or negative).

The dominance of arguments in favour of payoff maximization is demonstrated not only by the relative frequency but also by the effectiveness as to corrupt success (71.5%[51] of statements including an argument for strategic (positive) reciprocity ended in a successful corrupt transaction). This provides another piece of evidence for the hypothesis that the Profit Maximizing Motive is the driving force in the decisions made in groups. Arguments that seem persuasive in the pursuit of payoff maximizing are adopted and corresponding suggestions (i.e. maintenance of strategies aiming at payoff maximizing through corrupt reciprocity) realized, while arguments in favour of social efficiency (and fairness) are neglected, since they would lead to individually costly strategies. Again the argumentation is in line with the Persuasive Argument Theory (Pruitt 1971).

We leave it to further research to separate the effect of the decision-making process from effects stemming exclusively from the nature of the exchange of arguments via electronic chat messages. For our purpose of evaluating the effectiveness of the 4EP the effort of distinguishing between those two would lead to an even more artificial setting and therefore would not help to derive conclusions.

[50] $39.4 + 8.3 + 8.3\%$, Table 2.7

[51] $\frac{27.3+6.8}{39.4+8.3}\%$, Table 2.7

2.6 Conclusion

The results of our experiment are interesting in two respects. First, they serve as an assessment of the usefulness of the Four Eyes Principle. Second, they provide an insight into the mechanism of group decision-making. With our experiment, using the framework of a simple 3-Stage game which is standard in the experimental corruption literature, we show that the introduction of the Four Eyes Principle, which is generally promoted as one of the most effective tools to curb corruption (Pörting and Vahlenkamp 1998, Rieger 2005, Wiehen 2005), can be counter-productive. We find that it increases the relative number of successful corrupt transactions as well as the amount of bribes being transferred resulting in reduced welfare (measured by the sum of participants' payoffs) and equality (measured in the distribution of payoffs across participants).

Moreover, we find two opposing effects of the introduction of the Four Eyes Principle. One, the Bribe Splitting Effect, is caused by the splitting of the transfer between two officials, which reduces the level of corruption by changing the trade-off between its costs and its benefits. The other, the Group Decision-making Effect, increases the level of corruption. We are able to separate these two effects by the use of four different treatments and show that the Group Decision-making Effect is negative and over-compensates the Bribe Splitting Effect. This leaves a negative total effect from the introduction of the Four Eyes Principle with respect to the level of corruption and resulting social efficiency.

To explain the direction and magnitude of the Group Decision-making Effect (leading to higher conditional rates of reciprocity) we proceeded in three steps. First, the differences in revealed strategies between groups and individuals (i.e. final outcomes) can be identified by non-parametric tests and quantified by (parametric) regression analyses of conditional levels of corruption. Groups of officials reciprocate more often for high transfers and less often for low transfers than individual officials. These functional strategies lead to a higher number of successful corrupt transactions in the group decision-making treatments. Second, the analysis of behaviour within the group decision-making process provides further evidence. Contrary to predictions, in most cases initial disagreement between jointly deciding officials leads to a successful corrupt transaction despite the veto-power of non-corrupt officials. Third, we analyze the content of electronic chat-messages, exchanged during the decision-making process. Arguments in favour of strategic reciprocity (i.e. initiating or maintaining only corrupt transactions that yield a large payoff through high transfers) dominate the decision making-process not only quantitatively but also in terms of effectiveness (outcomes). The results of this 3-step analysis suggest the dominance of the 'Profit Maximizing Motive' in the group decision-making process (Kocher and Sutter 2007). This is in line with the Persuasive Argument Theory (Pruitt 1971).

Moreover, we show that the Profit Maximizing Motive drives group decisions further away from the theoretical predictions in a situation where there is, in contrast to the gift exchange game (Fehr et al. 1993), a trade-off between unit (group)-level payoff maximization and social efficiency. While groups of officials seem to be better in maximizing their unit(group)-level payoffs through maintaining corrupt relationships based on trust and reciprocity (explainable by higher cognitive capacity, Chalos and Pickard 1985), they do not manage to take the negative externalities (produced by successful corruption) into account, which ultimately aggravates the social dilemma of corruption.

So, policies that prescribe group decision-making should be restricted to situations for which the pursuit of maintaining reciprocal and payoff maximizing strategies are in line with the policies' objectives. Our results cast serious doubt on the usefulness of the Four Eyes Principle for situations where this condition does not apply. Looking into the black box of the mechanism that underlies group decision-making through the analysis of processes in combination with content analyses within controlled laboratory experiments may help to interpret behavioural patterns and to discover determinants of situations where the strategic use of group decision-making might reduce social inefficiencies. Future research should be directed at the theoretical foundation of the mechanisms found within group decision-making.

2.7 Appendix 2

Appendix 2A: Proofs

Equilibrium in the 3-Stage Game

Proof by (backward Induction).

Denote by $I_{i,n}$ the information set in stage i ($i \in \{1,2,3\}$) of period n ($n \in \{1,2,...,10\}$). Let $p(I_{i,n})$ be the probability of reaching the respective stage and $q(`s'|I_{i,n})$ the conditional probability of the relevant agent choosing action 's' once reached Stage (i,n). An information set contains all relevant information about ego's and alter's behaviour up to the respective stage. Furthermore let $PO(I_{i,n})$ be the (sum of) payoff(s) gained up to the arrival of stage (i,n).

First we show that there cannot be an equilibrium in which O chooses 'cooperate' in Stage 3 of the last (10^{th}) period. Consider a Strategy-Set $EQU1 = [s_{1,1}, s_{2,1}, s_{3,1}, s_{1,2}..., s_{3,10}]$ in which the third stage of period 10 is reached with some probability ($p(I_{3,10}) > 0$) and O cooperates with some probability ($q(`cooperate'|I_{3,10}) > 0$). Compare the payoff, resulting from the realization of Strategy-Set $EQU1$ ($PO(EQU1)$) to the one of an alternative Strategy-Set $EQU1_{new}$ which consists of the same strategies up to $I_{3,10}$ but for which $q(`cooperate'|I_{3,10}) = 0$ yielding payoff $PO(EQU1_{new})$.

Since the payoff for period 10 is larger for $EQU1_{new}$, since $8 + 3 * b < 12 + 3 * b$, $EQU1$ cannot constitute a Sub-game perfect Nash equilibrium. Second we show that, in the last (10^{th}) period, B will never choose any Strategy-Set that includes the action '$b > 0$' in Stage 1. Consider again a Strategy-Set $EQU2 = [s_{1,1}, s_{2,1}, s_{3,1}, s_{1,2}, ..., s_{3,10}]$ in which $p(I_{1,10}) > 0$, $q(`cooperate'|I_{3,10}) = 0$ and $q(`b > 0'|I_{1,10}) > 0$.[52] Again compare $PO(EQU2)$ to $PO(EQU2_{new})$, the payoff of a Strategy-Set that differs from the former only in $q(`b > 0'|I_{1,10}) = 0$. Since $12 - b \leq 12$, payoff $PO(EQU2)$ must be smaller than $PO(EQU2_{new})$ so that $EQU2$ cannot constitute an equilibrium. Hence only a Strategy-Set featuring $[s_{1,1}, ..., s_{9,1}, `b = 0', `accept'/`reject', `defect']$ can characterize an equilibrium.

[52]Given that $q(`cooperate'|I_{3,10}) = 0$ must be satisfied, O will never choose 'reject' in Stage 2 for '$b > 0$' and is indifferent between 'reject' and 'accept' if '$b = 0$', see Appendix 1C in Chapter 1.

Consider now a period-set $PS = \{k, ..., 10\}$ of (the last 10-k) consecutive periods for which the above stated last period's equilibrium Strategy-Set is played. Assume $q(\text{'}cooperate\text{'}|I_{3,k-1}) > 0$ for the period $k-1$. By the same line of arguments as for the last (10^{th}) period we can easily repeat the task up to the point of excluding all strategy sets that do not exhibit the strategy characteristics of the (Stage Game) equilibrium in the 10^{th} period [$\text{'}b = 0\text{'}, \text{'}accept\text{'}/\text{'}reject\text{'}, \text{'}defect\text{'}$]. Letting k decrease from 9 down to 1, it is obvious that the Stage Game Nash Equilibrium remains the only Sub Game perfect Nash Equilibrium in the (finitely) repeated game.

Responsibility and Veto Power

Consider two officials, O_i and O_j, who decide jointly in Stage 3 between 'cooperate' (c) and 'defect' (d). Consider O_i's preferences to be represented by the utility function $U_i(s_1, s_2)$ where s_1 is her own action and s_2 that of O_j. O_i's decision is pivotal only if $s_2 = c$. If not ($s_2 = d$), the outcome is (d) giving utility $U_i(d, d)$, independent from O_i's choice. This means that giving up responsibility works only in the cases where the socially optimal choice is taken anyway. Therefore, no official who is deciding jointly in a group can (for herself) deny responsibility for her group's corrupt behaviour since this would need the approval of both officials. For these reasons we believe that a lack of individual responsibility within a group cannot be applied as an argument in favour of the prediction of higher levels of corruptibility among officials within a group.

Appendix 2B: Figures and Tables

Extensive forms of games in all treatments

In all treatments except TDT2 both, O and B, decide once in every period. In TDT2 only B decides once per period while each O decides twice.

Figure 2.7: Extensive forms of TDT1 and IDT1

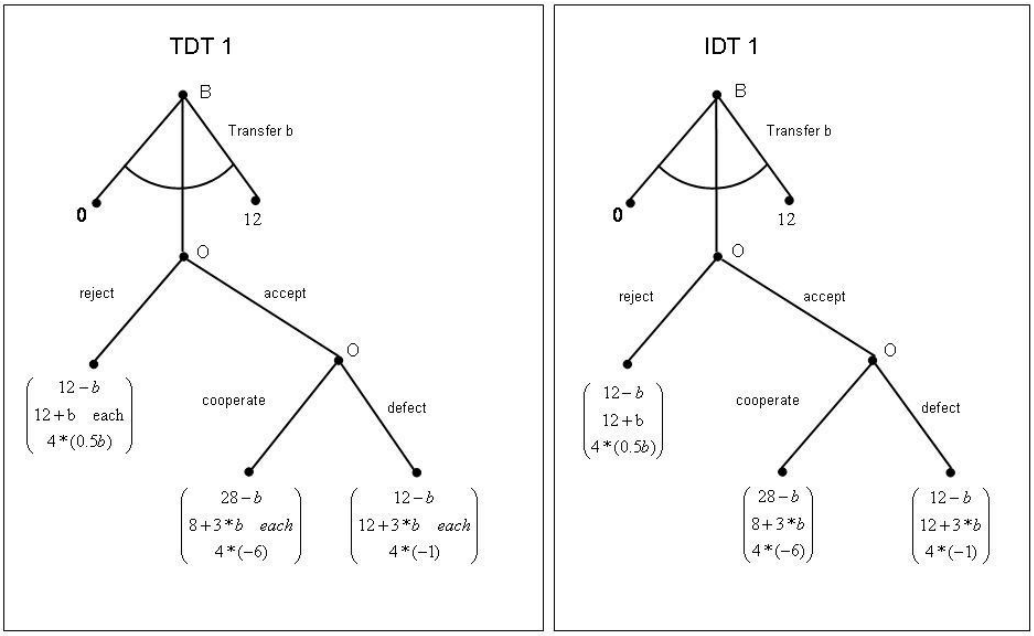


Figure 2.8: Extensive forms of TDT2 and IDT2

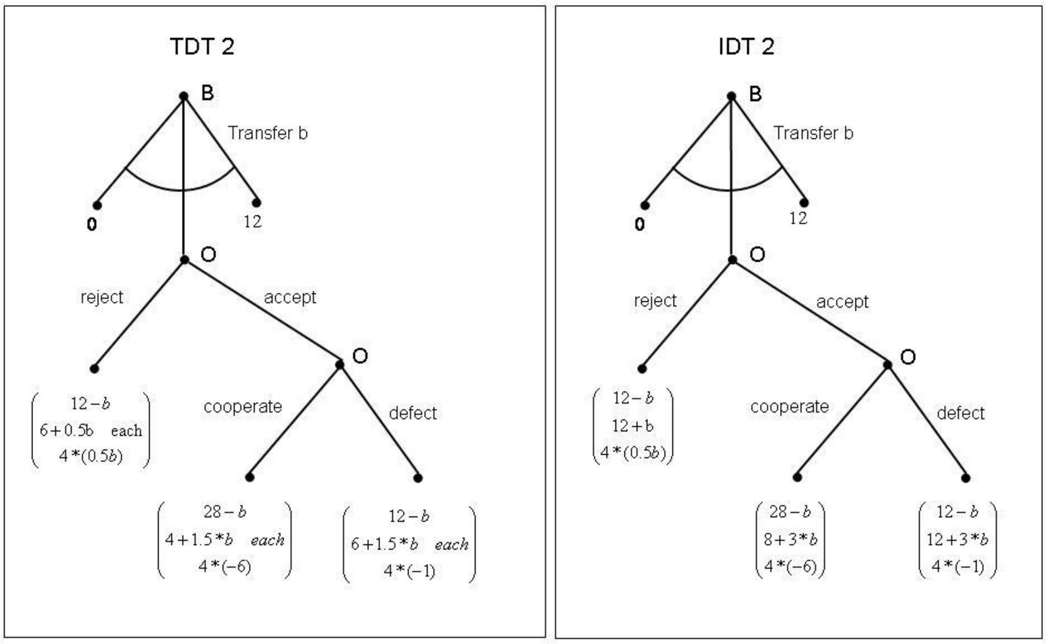


Table 2.8: **Model M1, Random effects estimation**

	Dependent variable: PP			
	Linear (OLS)		Tobit	
	Coefficient	Stand. error	Coefficient	Stand. error
Constant	22.2525***	1.02	22.1728***	1.92
b	1.8964***	0.28	1.8776***	0.35
b^2	−0.2727	-0.29	−0.2734	-0.32
D_{IDT2}	−0.1221	0.19	−0.1456	0.19
D_{TDT1}	−6.3437***	0.84	−6.3332***	1.14
D_{TDT2}	−3.0786***	0.46	−3.1232***	0.65
D_{IDT2}*b	0.1321	0.17	0.1432	0.19
D_{TDT1}*b	2.0542***	0.34	2.0318***	0.42
D_{TDT2}*b	0.7415***	0.19	0.7332***	0.21
D^2_{IDT2}*b	−0.0208	0.14	−0.0328	0.10
D^2_{TDT1}*b	−0.2528	0.21	−0.2421	0.28
D^2_{TDT2}*b	0.4451	0.31	0.4721	0.52
	overall $R^2 = 0.54$		-	

Number of periods: 10, Number of observation: 88, Number of clusters: 64 (heteroskedasticity) robust standard errors in OLS

Table 2.9: **Model M2**

	Dependent variable: Payoff			
	OLS		Tobit	
	Coefficient	Stand error	Coefficient	Stand error
Constant	225.2521***	54.92	224.5843***	44.47
N	15.8923***	2.48	15.5952***	3.11
D_{IDT2}	−7.1245	10.29	−7.0319	8.39
D_{TDT1}	−156.3421***	42.87	−157.2822***	44.27
D_{TDT2}	−133.17***	39.72	−134.0318***	41.90
D_{IDT2}*N	−0.8177	1.17	−0.7849	1.06
D_{TDT1}*N	3.1425***	0.73	3.1929***	0.75
D_{TDT2}*N	5.0308***	0.95	5.1121***	0.78
	adjusted $R^2 = 0.43$		-	

*** denotes significance at the 1%-level; Number of observations: 76; Number of clusters: 64 (heteroskedasticity) robust standard errors

Table 2.10: **Model M4**

	Dependent variable: SC					
	transfer level					
	0	1	2	3	4	5
b	0.026***	0.034***	0.045***	0.055***	0.065***	0.073***
D_{IDT2}*b	−0.002	-0.002	-0.003	-0.004	-0.004	-0.005
D_{TDT1}*b	0.018***	0.025***	0.032***	0.039***	0.046***	0.052***
D_{TDT2}*b	0.015***	0.020***	0.026***	0.032***	0.037***	0.043***
$\hat{SC}_{IDT1}$	0.093	0.119	0.154	0.198	0.254	0.319
$\hat{SC}_{IDT2}$	0.092	0.116	0.147	0.192	0.241	0.303
$\hat{SC}_{TDT1}$	0.013	0.044	0.091	0.172	0.239	0.352
$\hat{SC}_{TDT2}$	0.009	0.034	0.073	0.146	0.195	0.318
	transfer level					
	6	7	8	9	10	11
b	0.080***	0.083***	0.082***	0.078***	0.071***	0.062***
D_{IDT2}*b	−0.005	-0.006	-0.006	-0.005	-0.005	-0.005
D_{TDT1}*b	0.057***	0.059***	0.058***	0.056***	0.051***	0.044***
D_{TDT2}*b	0.047***	0.048***	0.048***	0.045***	0.041***	0.036***
$\hat{SC}_{IDT1}$	0.392	0.471	0.555	0.638	0.715	0.785
$\hat{SC}_{IDT2}$	0.370	0.445	0.523	0.598	0.672	0.737
$\hat{SC}_{TDT1}$	0.510	0.627	0.723	0.889	0.921	0.998
$\hat{SC}_{TDT2}$	0.471	0.542	0.698	0.802	0.897	0.967
	Number of subjects: 96					
	Number of units: 64					

*** denotes significance at the 1%-level

Marginal effects are calculated at the respective values of transfer and at the means of the remaining independent variables

$\hat{SC}$ denotes the estimates for the success probabilities conditional on the respective transfer levels.

Appendix 2C: Instructions from TDT2 (translated from German)

Thank you very much for your appearance. In the next 90 minutes you will take part in an experiment in the laboratory of MELESSA. If you read the following instructions carefully, you can (depending on your decisions) earn money, additional to the show-up fee of 4 Euros. Additional to the money you can earn for yourself, you will affect the amount of donation to the public aid organization 'Doctors without Borders'. The money you will earn during the experiment will be added to the show-up fee and paid out in cash at the end of the experiment. The money that is going to be donated will be transferred to the donations account of 'Doctors without Borders'.

During the experiment you are not allowed to communicate with the other participants. If you have questions, please approach one of the experimenters by raising your hand. In the case of violation of this rule we have to exclude you from any payments.

During the experiment we will refer to Experimental Monetary Units (EMU) instead of Euros. Your income will be calculated in EMU. In the end of the experiment the total amount will be exchanged in Euros.

The exchange rate is **1 EMU = 5 Eurocents**.

All 24 participants are assigned to **groups of four**. Neither the experimenters nor the other participants know which group you are in. Your decisions remain completely anonymous.

The Decision Situation

There are two types in this experiment: type A and type B. The types play different roles and make decisions that affect their own income, the income of the other participants of the experiment and the amount of donation transferred to the organization 'Doctors without Borders'. The type of a participant is allocated randomly.

A group of four consists of two type A and two type B participants who stay together for the entire experiment.

The experiment has 10 periods.

Procedure:

All of the 10 periods consist of at most 3 Stages.

Stage 1

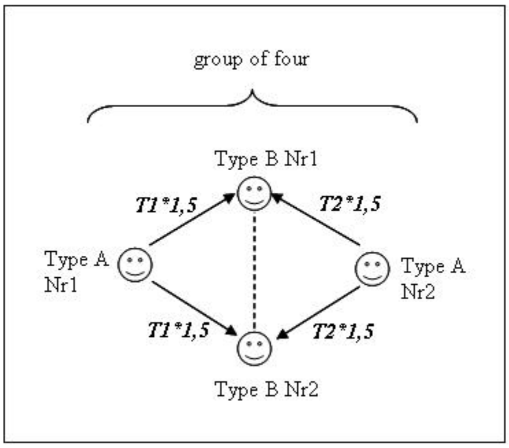
In the first Stage, **every participant of type A** (type A Nr 1 and type A Nr 2) decides on the size of their transfer (**T1** denotes type A Nr 1's transfer and **T2** denotes type A Nr 2's transfer) which has to lie between 0 and 12 EMU.

Next, the amount of the transfers is **tripled** and **then split** equally between the two type B participants (type B Nr 1 **and** type B Nr 2) of the group of four. If T1 is for example 6 EMU, type B Nr 1 receives 9 EMU ($0.5 * 6 * 3$ EMU) and type B Nr 2 receives 9 EMU.

Hence there are 2 situations per group in any period:

Situation 1: Type A Nr1 transfers T1 to the two type B participants (where T1 is **first tripled and then shared**)

Situation 2: Type A Nr2 transfers T2 to the two type B participants (where T2 is **first tripled and then shared**)



Stage 2

In Stage 2 the two type B participants decide jointly on how to react on the transfer of the respective type A participant. They have (in both situations) two alternatives.

1^{st} Alternative: Both decide (for a specific transfer, e.g. T1) jointly for 'keep': In this case Stage 3 is entered

2^{nd} Alternative: One or both decide in favour of 'distribute': In this case, the respective type A participant (e.g. type A Nr 1) does not get a bonus (and receives only 12 - T1 EMU). The type B participants both get 6 EMU plus half of the value of the transfer ($6 + 0.5 * T1$ EMU). Moreover, the amount of $2 * T1 + 24$ EMU is transferred as a donation to the organization 'Doctors without Borders'.

A joint decision between the two subjects is found as follows.

First, **each** of the two type B participants decides **individually** whether to 'keep' or to 'distribute' the particular transfer.

If the decision is not unanimous (one type B participant wants to 'keep' and the other wants to 'distribute' the transfer), the decision of the fellow participant appears on his or her own screen.

Next, the participants decide **once again separately**. If there is still no agreement, the two type B participants can exchange messages via an electronic 'chat' (see explanation below) for one minute. After this the participants decide for the **last time**.

Note that only if both type B participants decide in favour of 'keep' the third Stage is actually reached.

Since there are two type A participants in every group of four (type A Nr 1 and type A Nr 2), each of the type B participants has to decide (jointly with the other type B participants) in two situations: once for T1 and once for T2.

Stage 3

In Stage 3 (which is only reached if both type B subjects have chosen 'keep') the two type B participants decide **again jointly** whether to initiate a re-transfer or not.
Again, both type B subjects decide **separately first**.
If the decision is not unanimous (one type B participant wants to initiate the re-transfer and the other does not), the decision of the other participant is shown on the screen. Then the participants can **decide again separately**. If there is still no consent, the participants enter again a 'chat' in which they can exchange electronic messages for one minute. After this, there is a final decision.

1. **Case**: **Both** type B participants decide **in favour of** a re-transfer. Both carry the costs of 2 EMU each (independent of the amount of the respective transfer). They both get 6 EMU plus one and a half times the value of the transfer, less the costs of 2 ($6 + 1.5 * T1 - 2$ EMU). The respective type A participant (type A Nr 1) receives a Bonus of 16 EMU in addition to the 12 EMU of initial endowment ($16 + 12 - T1$). In this case there is no donation to the organization 'Doctors without Borders'.

2. **Case**: **One or both** type B participants decide **against** a re-transfer. In this case, there are no personal costs for the two type B participants (they get $6 + 1.5 * T1$ each), the respective type A participant does not receive a bonus (and gets $12 - T1$), and the donation to the organization is 20 EMU.

In the end, all participants are shown their personal income in the period. Please note, that the type A participants can thereby reconstruct whether or not the type B participants chose for or against the re-transfer.
These (maximal) 3 stages are repeated 10 times (10 periods). Since the members of groups stay together, participants *always* **interact with the same persons** in the same roles for the **entire experiment**. (Type A Nr 1 remains type A Nr 1. type A Nr 2 remains type A Nr 2 etc.)

Chat:

Type B subjects potentially have the possibility to communicate via real time **electronic messaging** (Chat) with their fellow type B subject to agree on a joint decision (e.g. 'keep' or 'distribute') in Stage 2 and Stage 3.
The content of the communication is generally free to choose but there are some restrictions. You are not allowed to make statements about personal characteristics such as your name, age, address, gender, subject of study or any information that might lead to your identification. Moreover, strong language is strictly forbidden. Anyone who violates these rules of communication will be automatically expelled from the experiment and will not get any payments **for the entire experiment**.
Each participant in the chat can send as many messages to the other participant as he wishes or is able to send within the time limit of one minute.
Every message appears automatically on the screens of both type B participants of a group of four but cannot be seen by any other participant of the experiment.

Payoff table

The following table shows the kind of consequences the decisions of the participants lead to - in terms of their own payoff, the payoff of the other participants and the organization 'Doctors without Borders' (Example for T1).

The following table can be read as follows. Generally we start from the top and go down cell by cell. If a participant chooses a certain alternative, only those cells that lie directly beneath it are relevant for the next period.

The payoff table is analogous for situations in which T2 (Transfer of Type A Nr2) is relevant.

1. Stage	Type A Nr1 chooses ***T1***		
2. Stage	At least one Type B participant (Type B Nr1 **or** Type B Nr2 **or both)** decide in favour of ***'distribute'*** for ***T1***	**Both** Type B participants (Type B Nr1 **and** Type B Nr2) decide in favour of ***'keep'*** for ***T1***	
3. Stage	Stage 3 is not reached	Type B Nr1 or Type B Nr2 or both decide against a **re-transfer** to Type A Nr1	**Both** (Type B Nr1 and Type B Nr2) decide in favour of a **re-transfer** to Type A Nr1
Payoff **Type A Nr1**	**12 –T1**	**12 –T1**	12 + 16–T1 = **28 –T1**
Payoff **Type B Nr1/Type B Nr2**	**6 + 0,5*T1**	**6 + 1,5*T1**	6 - 2 + 1,5*T1 = **4 + 1,5*T1**
Donation to 'Doctors without Borders'	**2*T1+ 24**	**20**	**0**

Note that each **type B participant** receives **two** payments because two situations are relevant for each of them, one with type A Nr 1 (T1 is relevant) and one with type A Nr 2 (T2 is relevant). These two are added up for any period.

For the **type A participants** only **one** situation per period is relevant so that there is only one payment per period.

Each participant gets information at the end of each period about his own personal payoff. Type A subject can infer whether type B subjects have chosen to initiate a re-transfer or not.

Note that the sum of payments (exchanged in Euros) to the organization is actually donated to 'Doctors without Borders'.

Timing

Stage 1: Type A Nr1 chooses T1 and Type A Nr2 chooses T2

Stage 2: Type B Nr1 and Type B Nr2 decide (T1 and T2) each time jointly about 'keep' or 'distribute'

Stage 3:

Situation (for T1): Only reached if in this situation, both Type B participants chose 'keep' in Stage 2. In this case, both Type B participants decide jointly whether to initiate a re-transfer or not.

Situation (for T2): is analogous, only for T2.

At the end of each of the ten periods, each participant gets information about his/her own payoff in the respective period. At the end of the last (10^{th}) period, participants get to know their final income and their payment in Euros.

The following **control questions** will help you to get a better understanding of the situation. All the necessary information can be found in the payment table.
Please answer all the control questions and raise your hand when you have finished. An experimenter will come to your place to check your solutions.

Question 1

Assume that **you are type A Nr1** and you chose a transfer of **4 EMU** (T1). The other participant of type A (type A Nr2) has chosen a transfer (T2) of **10 EMU**.
Situation 1 (T1): One of the participants of type B in your group of four (type B Nr1) decides to **'distribute'** your transfer (T1). The other participant of type B (type B Nr2) wants to **'keep'** your transfer (T1). (Therefore Stage 3 is not reached.)
Situation 2 (T2): Both type B participants chose to **'keep'** the transfer of type B Nr 2 (T2) in Stage 2 and decide **against** a re-transfer in Stage 3.
a) What is the payoff of type B Nr1 in the situation (1) with type A Nr1 (you)?
Your answer:___________________________
b) What is the payoff of type B Nr2 in the situation (1) with type A Nr1 (you)?
Your answer:___________________________
c) What is your (type A Nr1) **total** payoff in this period?
Your answer:___________________________
d) What is the total payoff of type B Nr1 for all situations relevant to him/her?
Your answer:___________________________
e) What is the total payoff of type B Nr2 for all situations relevant to him/her?
Your answer:___________________________
f) What is the total payoff of type B Nr2 in this period?
Your answer:___________________________
g) What is the amount of donation to 'Doctors without Borders' caused by the situation relevant to you (type A Nr1)?
Your answer:___________________________
h) What is the amount of donation to 'Doctors without Borders' caused by the situation relevant to type A Nr2?
Your answer:___________________________
i) What is the total amount of donation to 'Doctors without Borders' in this period?
Your answer:___________________________
j) What is the total amount of payoff generated by the decisions of your group of four?
Your answer:___________________________

Question 2

Assume that you (type A Nr1) and type A Nr2 have both chosen a transfer of **0** (T1 is 0 EMU and T2 is 0 EMU). Neither participant of type B (neither type B Nr1 nor type B Nr2) wants to 'keep' any of the two transfers in Stage 2.

a) What is your (type A Nr1) total payoff in this period?

Your answer:__________________________

b) What is the total payoff of type A Nr2 in this period?

Your answer:__________________________

c) What is the total payoff of type B Nr1 in this period?

Your answer:__________________________

d) What is the total payoff of type B Nr2 for all situations relevant to him/her?

Your answer:__________________________

e) What is the total amount of donation to 'Doctors without Borders' in this period?

Your answer:__________________________

f) What is the total amount of payoff generated by the decisions of your group of four?

Your answer:__________________________

Question 3

Assume that you (type A Nr1) have chosen a transfer of **5 EMU** (T1 is 5 EMU) and type A Nr2 has also chosen a transfer of **5 EMU** (T2 is 5 EMU). **Both** participants of type B (type B Nr1 and type B Nr2) decide to **'keep'** the transfer and initiate a re-transfer in Stage 3.

a) What is your (type A Nr1) total payoff in this period?

Your answer:__________________________

b) What is the total payoff of type A Nr2 in this period?

Your answer:__________________________

c) What is the total payoff of type B Nr1 in this period?

Your answer:__________________________

d) What is the total payoff of type B Nr2 for all situations relevant to him/her?

Your answer:__________________________

e) What is the total amount of donation to 'Doctors without Borders' in this period?

Your answer:__________________________

f) What is the total amount of payoff generated by the decisions of your group of four?

Your answer:__________________________

Chapter 3

Bringing Good and Bad Whistle-Blowers to the Lab

3.1 Introduction

For several decades, corruption, defined as 'the misuse of public office for private gain' (Klitgaard 1988), has been considered as a major obstacle to growth and development (Bardhan 1997, Mauro 1995). A large number of activities falling under this definition can be modelled as a Principal-Agent-Client relationship between the government, its imperfectly controlled agents (public officials) and their private clients (individuals or firms). This relationship can be further divided into two disjoint principal agent problems, one between the government and its official and the other between the official and the client. While the first is of less importance, a major objective of the New Institutional Economics (NIE) of corruption (Lambsdorff 2007) is to understand the illegal and therefore (legally) unenforceable transaction between a private entity (e.g. a firm paying a monetary bribe) and a potentially corrupt official who may reciprocate a payment with the delivery of a corrupt service. The goal is to draw conclusions for the design of institutions which will optimally destabilize and hence minimize corruption. Apart from obvious institutional measures such as applying harsh punishment and increasing detection probabilities of corrupt behaviour (Becker 1968, Klitgaard 1988, von Rose-Ackerman 2006), an effective way to destabilize corruption is to enable whistle-blowing. In our analysis, we define whistle-blowing as 'the act of disclosing information in the public interest' [1]. Whistle-blowing can hence be seen as the act of ending a corrupt transaction and all of its consequences by incurring non-trivial personal costs (Drew 2003).[2]

In their experimental studies, using the framework of a standard corruption game (see Chapter 2), Lambsdorff and Frank (2010) and Abbink (2006) find that the possibility of whistle-blowing leads to an increase rather than a decrease of the number of successful corrupt transactions.[3] These results are at odds with the fact that whistle-blowing policies are in widespread use and perceived as successful measures in the abatement of the negative consequences of corruption (Hall and Davies 1999, Spagnolo 2006, Buccirossi and Spagnolo 2006, Hussein 2005). We explain this by the fact that the standard game of corruption (as used in Abbink 2006, Lambsdorff and Frank 2010) accounts for only one of the two main negative consequences of corruption potentially affected by whistle-blowing. The first consequence of a successful corrupt transaction (which is modelled explicitly in the standard game of corruption) concerns the direct negative externality on the public which is directly proportional to the level of corruption (Bardhan 1997, Rose-Ackermann 1999).[4] The second, indirect, consequence (which is disregarded in the standard corruption game) concerns the 'crowding out' of legal productive activities (Klitgaard 1988). Rampant corruption is likely to drive 'honest' clients out of productive markets, forcing them to engage in alternative activities (of lower productivity).[5] This may be due to individual (firm-level) differences in the moral costs, ability (criminal energy) or legal framework (foreign versus domestic firms) affecting corrupt transactions (see Foreign Corrupt Practices Act in Egger and Winner 2006, Bardhan 1997).

The direct effect of corruption is independent of who initiates the transaction, the client or the official, while the indirect effect requires an actively corrupt official. The compactness

[1]Another definition is 'the disclosure to the public or to authorities, usually by an employee, of wrongdoing in a company or government department' (Drew 2003)

[2]By assuming that the briber consists of a single decision maker, we shut off considerations on behavioural interactions within groups of agents (e.g. group pressure). Hence we do not consider that whistle-blowing may exploit the differences in individual corruptibility within a group of decision makers (e.g. inside a firm or within a group of officials). In this paper we concentrate on the incentives and behavioural reactions of decision-makers represented by single individuals only.

[3]Neither Lambsdorff and Frank (2010) nor Abbink (2006) focus on the mechanism of whistle-blowing or on the assessment of its effectiveness.

[4]The most common example is the expected damage to outsiders caused by the deployment of sub-standard quality in construction projects realized with the help of administrative corruption.

[5]This line of argument is closely related to the rent seeking literature of corruption (Lambsdorff 2002).

of the standard game of corruption does not allow for corrupt initiation by the official. This approach is useful to obtain insights into individual behaviour within a certain corrupt situation (concentrating on the corruptibility of a public official with respect to her reaction to a temptation), but it may not be able to capture the indirect consequence of corruption. In particular, it may fail to account for an honest client who may prefer to stay away from a productive market for fear of encountering an official who demands a bribe (and engage in an alternative activity of lower productivity but without the risk of encountering corruption). The introduction of whistle-blowing may affect the behaviour of both decision-makers and hence determine the magnitude of both negative consequences of corruption.

In order to capture both negative consequences of corruption and hence be able to assess the full impact of whistle-blowing experimentally, we expand the standard game of corruption in two ways. First, we add an option to the action space of the client in form of the legal procedure of a first best productive activity. Second, we give both agents, the client and the official, an opportunity to actively initiate a corrupt transaction. In the control treatment of our experiment we show that our model is able to capture both negative effects of corruption. This permits the possibility of whistle-blowing for both decision-makers and enables us to consider the additional aspect of asymmetry of leniency policies towards whistle-blowing. We compare three treatments. In the control treatment, neither of the decision-makers has an opportunity to blow the whistle. In the second, the symmetric whistle-blowing treatment, both decision makers have the opportunity to blow the whistle. Blowing the whistle has symmetric consequences in the sense that it leads to the loss of all privileges obtained by engaging in corruption for both players, the client and the official, independent of who was the one that blew the whistle. In the third treatment, we consider leniency for whistle-blowing officials.

Analyzing treatment differences in the outcomes of our extended corruption game which was played for ten successive periods, we show that symmetrically punished whistle-blowing has two countervailing effects on the three main performance variables, efficiency (measured in the number of first best choices), level of corruption (frequency) and social welfare (measured in total payoff[6]). On the one hand, symmetric whistle-blowing

[6]We consider total payoff as a measure for welfare which combines the consequences of efficiency and the level of corruption.

increases the level of corruption by providing the stabilizing tool of negative retaliation after defective behaviour confirming results of Lambsdorff and Frank (2010, forthcoming) and Abbink (2006). On the other hand, whistle-blowing, in both specifications, increases the number of legal transactions by providing the client with an effective safeguard against the attempts of the official to force participation in a corrupt transaction. These two effects almost keep their balance in terms of resulting payoff levels. Since it is not straightforward to trade off productivity (efficiency) gains with higher levels of corruption, a conclusion as to the usefulness of symmetric whistle-blowing is not feasible.

Providing asymmetric leniency for whistle-blowing to the official, allowing the official to keep most of the benefit from an un-finished corrupt transaction, creates an incentive for 'insured' defection which shuts off the (corruption-) stabilizing effect of whistle-blowing while the effect of increased productivity remains. Hence, the introduction of whistle-blowing with asymmetric leniency increases welfare in terms of a higher sum and a more balanced structure of payoffs. Our results are in line with the theoretical findings of Lambsdorff and Nell (2007), who show that by allowing asymmetric punishment (interpretable as leniency for the official) within the corrupt deal, the stability of a corrupt transaction can be weakened as the incentives for cheating (on the side of the official) are increased.[7] Most patterns of behaviour can be explained in terms of payoff maximizing strategies and first order belief structures.[8]

The remainder of the paper is structured as follows. Section 3.2 illustrates the basic set-up and explains its main divergences from the existing experimental literature on corruption. It describes the set up of the control treatment as well as the two treatment specifications of the experiment in detail. The procedure of the experiment is outlined in Section 3.3. In Section 3.4 we analyze the model in theoretical terms and derive the main hypotheses. Section 3.5 interprets the results and tests the hypotheses. Section 3.6 summarizes and concludes.

[7]Lambsdorff and Nell (2007) propose a legal institution that distinguishes between (self-reported) corrupt behaviour of the official and of the briber with respect to punishment. In addition, they advocate different penalties to those who reciprocate a corrupt transaction and those who only pocket a bribe without providing a corrupt service, thus giving incentives for defection within the corrupt transaction.

[8]This involves the relaxation of the assumption of common knowledge of rationality.

3.2 Model

According to the NIE of corruption, the weakest link of a corrupt transaction is its lack of legal enforceability.[9]. A corrupt transaction is usually modelled as a two player, 3-stage game in which a client decides about the amount of bribe (Stage 1) to be sent to an official who decides whether or not to accept it (Stage 2). If accepted, Stage 3 is entered, in which the official decides whether or not to reciprocate the bribe by delivering a (pre-defined) corrupt service. Reciprocation produces a large negative externality to the public and incurs substantial personal costs (e.g. costs that are caused by keeping the transaction secret as well as 'moral' costs created by harming the public and risking detection).[10] In order to assess the effectiveness of different institutions of whistle-blowing considering both systematic consequences of corruption, the negative externality and the crowding out of productive activity, we need to extend the standard game of corruption in two ways. First, we need to provide a real alternative to corruption for the client and, second, we enable both, the client and the official, to activate a corrupt transaction. Therefore we add a third option in Stage 1 and a fourth stage to the 3-stage game.

3.2.1 Representation of the basic game

Figure 3.1 shows the extensive form of the game used as a basis for our analysis. It does not include the possibility of whistle-blowing and hence serves as the control treatment.

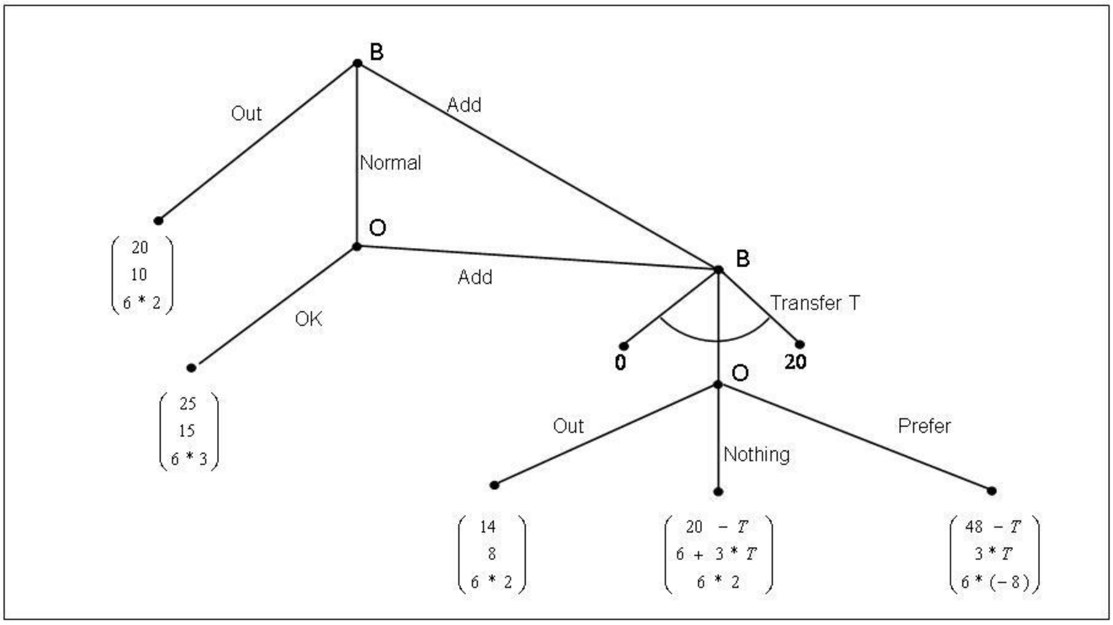
The first line in the outcome vectors depicts the payoff of the potentially corrupt client (B). The second line is for the payoff of the official (O). The third line stands for the monetary (sum of) external effect to the public. In our experiment this means that the payoff of six randomly chosen participants of the experiment is either increased or reduced according to the decisions made by the client and the official.[11]

[9]Since corruption is illegal in all relevant countries (i.e. countries that rely on some form of legal administration), legal third parties, i.e. the courts, cannot be used to enforce a corrupt transaction, see Chapters 1 and 2.

[10]In most models, especially those that are designed to be tested experimentally, the risk of detection is modelled as a certainty equivalent rather than a lottery which would add another dimension to the game. By choosing a certainty equivalent, we do not have to control for differences in risk preferences.

[11]The model of the negative externality, hitting potentially corrupt actors instead of a real third party, may be problematic as it may lead to unintended effects on beliefs and thereby on strategies of participants. In Chapter 2, however, we show that the nature of the public (third party), passive or

Figure 3.1: Extensive form representation of the basic set-up



Our set-up is best explained using the following example. A private entity (e.g. a firm) wants to engage in some kind of productive activity for which it needs to obtain a permit (licence) provided by the authorities through a public official.

In Stage 1, client B chooses between three alternatives. The choice 'Out' leads to the end of the game. We consider this as a failed transaction as it represents a second best outside option in the sense that both agents' payoffs as well as the positive externality to the public[12] are relatively low.

The choice 'Add' represents B's decision to use a transfer T to convince O to provide preferential treatment, i.e. to realize the corrupt transaction in Stage 4.

The choice 'Normal' represents the willingness of B to engage in a legal procedure to obtain the licence. Only if client B chooses 'Normal', Stage 2 is entered and O decides between providing the licence (option 'Ok'), which leads to a successful 'legal' transaction (giving moderate payoffs to the deciding agents and a relatively large positive externality to the public), and demanding a bribe (option 'Add'). If 'Add' is chosen by, either client B in Stage 1, or official O in Stage 2 (i.e. a corrupt transaction has been initiated), B decides in Stage 3 about the amount of transfer T, which is restricted to integers between 0 and 20 EMU (budget constraint), to be given to O.

---

active, does not significantly affect behaviour. Therefore we ignore this issue in our analysis.

[12]We assume that any kind of productive activity yields positive spill-overs to other agents of the economy.

Having received T, it is again O's turn to decide in Stage 4.

If O chooses 'Out', the corrupt transaction fails, the bribe is transferred back to B and both agents fall back to their second-best outcome with reduced payoffs (sunk transaction costs). In case O accepts the bribe (chooses 'Nothing' or 'Prefer'), the transfer is tripled, capturing the assumed difference in marginal utility between a rich client and poor official.[13]

By choosing the option 'Nothing' O defects on B's implicit corrupt demand while keeping the benefit of the transfer ($3 * T$) for herself. B, as well as the public, fall back to their outside option (reduced by the costs of the transfer T for B).

By choosing 'Prefer', O reciprocates the transfer, causing substantial costs to herself (6 EMU)[14] and a large negative externality (-48 EMU) to the public while the client receives a large bonus (28 EMU).

Contrary to the standard (3-stage) game of corruption where the client can only choose between corruption and an implicit (unattractive) outside option, B has the opportunity to enter a relatively attractive honest activity. Since this requires the approval of the official, the latter is provided with the opportunity to actively initiate a corrupt transaction by (implicitly) demanding a bribe from an honest client. This captures the possibility of extortion of 'honest' clients by corrupt officials and thereby captures the indirect negative effect of corruption.

Applying the principle of backward induction within the standard model of payoff maximization (assuming rationality and common knowledge of rationality), it is easy to find the unique Sub game perfect Nash Equilibrium (SNE) in pure strategies.

Maximizing her pay-off, O will never choose 'Prefer' in Stage 4, since the payoff obtainable by choosing 'Nothing' is always larger, independent of the history of choices made before reaching Stage 4.

The choice between 'Out' and 'Normal' depends on the size of the transfer T. O will choose 'Out' if $8 > 6 + 3 * T$. Since only integers can be chosen in our experiment, O will decide in favour of 'Out' as long as $T = 0$ and 'Normal' otherwise. B will anticipate

[13]The assumption of differences in the marginal utility is standard to the experimental literature of corruption (Abbink 2004) and links the literature to the trust and gift exchange game (Berg et al. 1995).

[14]Note that the costs of 6 EMU represent the total costs of reciprocation (delivering the corrupt service). We may interpret these costs as costs that arise in the effort to conceal information from colleagues and the authorities in order to keep a corrupt transaction secret or as a certainty equivalent for the expected monetary loss from possible detection by the authorities (which is not modelled in our set-up.)

this in Stage 3 and minimize her losses by choosing $T = 1$, yielding a payoff of 19 EMU for B and 9 EMU for O. O will therefore choose 'Ok' over 'Add', once she has reached Stage 2 (anticipating B's behaviour) and B will therefore choose 'Normal' in Stage 1, yielding the highest possible payoff level.

Proposition 1: The unique SNE is characterized by the strategy set: ['Normal'; 'Ok'; 'T=1'; 'Nothing'].
See Appendix 3A for the proof of a similar situation.

The standard self-interested model does not predict the occurrence of corruption, neither in the one shot version of the game nor in a finitely repeated version of it. This (unrealistic) result is consistent with the theoretical predictions of the standard 3-stage corruption model of single-sided initiation of corruption used in the set-ups of Abbink et al. (2002), Abbink (2006) and Lambsdorff and Frank (2010).

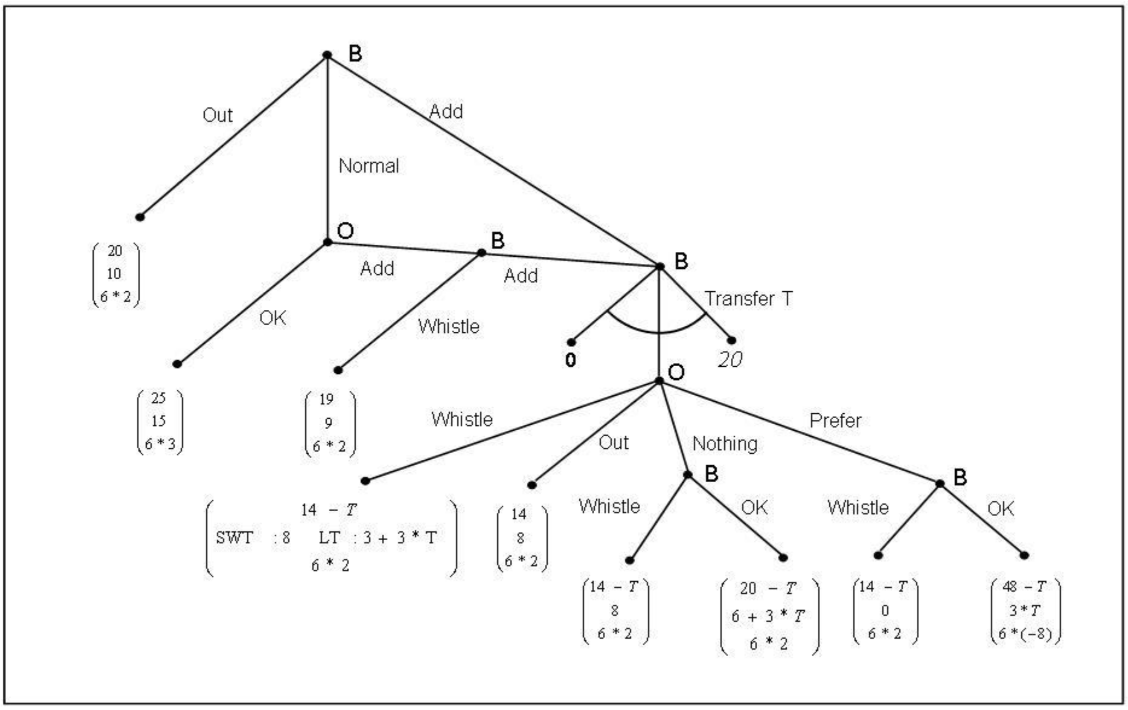
3.2.2 Treatment specifications

With this paper, we provide the first experimental study on corruption in which both agents, the official and the briber, have an opportunity to actively initiate a corrupt transaction (Dusek et al. 2004). Our extensions to the standard game of corruption alone enable us to explore the systematic effects of the introduction of whistle-blowing on the relevant performance variables considering asymmetry of its consequences with respect to punishment (leniency).

In the control treatment (CT), whistle-blowing is not possible. Participants are randomly assigned to their roles (B and O) and then randomly matched in pairs. One participant in the role of the official and one participant in the role of the client play the specification of the game, see Figure 3.1, for ten consecutive periods (repeated game in partner design).

In the symmetric whistle-blowing treatment (SWT), see Figure 3.2, we introduce the possibility of whistle-blowing to the basic game in the following way. At each relevant stage, i.e. a stage after which a player has acted in a way to initiate or to maintain a corrupt transaction, the other player can blow the whistle or proceed with the corrupt transaction. Whistle-blowing at any stage ends the game and leads to the loss of all benefits obtained (or obtainable) through the corrupt transaction for both agents. Whistle-blowing is symmetric in the sense that its consequences (personal losses of the benefit from the corrupt transaction) are independent of the player who has chosen to blow the whistle.

Figure 3.2: Extensive form representation of SWT and LT



The first opportunity to blow the whistle is given to B in case O (implicitly) demands a bribe by choosing 'Add' in Stage 2. This decision leads to the failure of the transaction, leaving both agents with reduced payoffs compared to a Stage 1 failure (transaction costs). The second opportunity to blow the whistle is for O. Once B has chosen T, O gets the chance to blow the whistle in addition to the alternatives she holds in the basic set-up (CT). This ends the corrupt transaction as well. In SWT this is not attractive. Both agents lose their (potential) benefits. Compared to the choice 'Out', 'Whistle' does not increase O's own material profit. However, it may be used by O as a punishment device against B. The last opportunity of whistle-blowing is for B. In the case that O has chosen to keep the bribe with (action 'Prefer') or without (action 'Nothing') delivering the corrupt service, B can decide whether to accept O's decision (action 'Ok') or blow the whistle (action 'Whistle'). Again blowing the whistle causes the loss of all corruption-related benefits for both players. O loses the benefit from the transfer (which amounts to $3T$) and B loses everything except the value of the (reduced[15]) outside option less the transfer T.

The leniency treatment (LT), see Figure 3.2, differs from SWT only in O's incentives for whistle-blowing in Stage 5. While the options and consequences of whistle-blowing remain the same for B, O can keep the full marginal benefits ($3 * T$) of the transaction in case she blows the whistle in Stage 5 when offered a bribe T and has to bear costs which are relatively small (3 EMU) and independent of the size of the bribe. This captures the idea of exemption from punishment for key witnesses by granting leniency.
Solving both treatment specifications of the 6-Stage-game by backward induction within the standard self-interested model of payoff maximization (assuming rationality and common knowledge of rationality) yields the same predictions as in the basic set-up in terms of the occurrence of successful corruption. All strategy sets that include any kind of whistle-blowing belong to strictly dominated strategies, since all forms of whistle-blowing are designed to reduce both decision makers' payoffs compared to the immediate alternative. This leaves all forms of whistle-blowing off the predicted and unique equilibrium path in any (finite) repetition of the game.[16]

Proposition 2: The unique SNE in SWT as well as in LT is characterized by the strategy set: ['Normal'; 'Ok'; 'Add'; 'T=1'; 'Nothing'; 'Ok'], assuming rationality, common knowledge of rationality and selfish preferences.
The proof for the specification of LT can be found in Appendix 3A.

3.2.3 Related literature

The main behavioural mechanism of stability in the standard corruption game is determined by trust and reciprocity. So, the methods used in experiments on corruption are closely related to those used in the gift exchange and the trust (investment) game

[15]We assume considerable transaction costs in this case.

[16]Note that the relevance of the equilibrium critically depends on the assumption of rationality and common knowledge of rationality which will be relaxed in the arguments of Section 3.4.

(Fehr et al. 1993, Berg et al. 1995). Rejecting predictions from standard economic models, countless studies find positive levels of reciprocity in one-shot versions of these games and even higher levels if participants interact repeatedly (Dufwenberg and Gneezy 2000, Gächter and Falk 2002).

The main and most important difference between these games and the corruption game lies in the fact that cooperation in the form of positive reciprocity is beneficial for all members of society, whereas in the case of corruption it is only beneficial for the client and the official. The negative externality to the public caused by a successful corrupt transaction is (usually) assumed to be high enough to result in a net social loss (i.e. a negative sum of payoffs) from a corrupt transaction.

While positive reciprocity in the gift exchange or the trust game may be explained by altruism (Andreoni and Miller 2002), inequity aversion (Fehr and Schmidt 1999, Charness and Rabin 2002) or intentions (Rabin 1993, Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006), the main arguments involved in the respective models are not valid for the corruption game. Altruism (Andreoni and Miller 2002) would not predict positive levels of corruption because of its large negative externality creating inefficiencies and its (potential) reduction of the welfare in general and that of the worst-off in particular. Inequity aversion (e.g. Fehr and Schmidt 1999, Charness and Rabin 2002) may only help to explain corrupt reciprocity under very restrictive assumptions. Corrupt reciprocity may only lead to more equitable outcomes, if all participants of the experiment are believed to act in a corrupt way.[17] The arguments put forth in models based on intentions (e.g. Dufwenberg and Kirchsteiger 2004 or Cox et al. 2007) are also of limited plausibility because of the obvious strategic environment. Bribes are not likely to be conceived as kind actions but calculated behaviour.[18] None of the respective models seems to be able to provide a consistent and convincing theoretical explanation of experimental findings. In particular, they cannot be used to derive relevant predictions on the behaviour of the corrupt transaction partners, neither in the standard 3-stage game nor in our (even more

[17]Consider corruption as the equilibrium of a 'super-game' in which the public is not passive but consists of potentially corrupt pairs of clients and officials. This assumption, however, is contrary to the findings of Chapter 2 where we show that corrupt reciprocity does not depend on the model of the public that suggests that inequity aversion is unlikely to provide a comprehensive explanation for the behaviour in the corrupt game.

[18]Strassmair (2009) shows that strategic intentions need not spoil the kindness of a gift. However her results and methodology cannot be used to make point predictions for behaviour in our set-up.

complicated) set-up.

Reciprocation may be better explained (and predicted) by simple arguments on reputation formation and (off-equilibrium) first order beliefs about the rationality and resulting behaviour of fellow participants (Fehr and Fischbacher 2003, Kreps et al. 1982). In a series of laboratory experiments using one-shot as well as (finitely) repeated versions of the standard game of corruption, Abbink et al. (2000, 2002), Abbink (2004, 2006), Abbink and Hennig-Schmidt (2006) and Lambsdorff and Frank (2007, 2010, forthcoming) find high levels of cooperation between corrupt partners. The unstable corrupt transaction seems to be sustained by some form of reciprocity. The common feature of these experiments is the trade-off between reciprocity-based maximization of individual long-term payoffs and a combination of myopic short term maximization of personal gains and a preference for social efficiency (Dusek et al. 2004).

In our extended set-up, the main mechanism of corrupt stability still relies on positive reciprocity. The additional options of whistle-blowing may be interpreted as (costly) punishment devices. Fehr and Gächter (1998, 2000a) as well as Sutter et al. (forthcoming) explicate that direct punishment can be effective in enhancing positive reciprocity in social dilemmas. In the context of corruption, Abbink (2006) as well as Lambsdorff and Frank (2010, forthcoming) find that punishment devices designed as whistle-blowing opportunities may stabilize corrupt reciprocity (being counter-productive). Since their set-ups are not designed to capture the full negative effect of whistle-blowing in terms of social efficiency, concentrating on the level of corruption alone, our experimental study provides the first assessment of whistle-blowing in the corrupt context with respect to its total effect on social welfare. In addition we test the hypothesis of the usefulness of asymmetry in leniency rules for whistle-blowing suggested by Lambsdorff and Nell (2007), who provide theoretical evidence for the effectiveness of asymmetric punishment in the context of corruption.

3.3 Procedure

We conducted five sessions (one session for CT and two sessions each for SWT and LT) programmed in Z-Tree (Fischbacher 2007) at the experimental laboratory (MELESSA) at the University of Munich. We used the organizational software Orsee (Greiner 2004). Our experiment included a total of 102 participants (who were randomly picked from the MELESSA subject pool). Types (B and O) were randomly assigned to subjects who were randomly matched in pairs (one B and one O per pair). Pairs stayed together for all ten periods (partner design). Full anonymity was ensured throughout the experiment. All periods were payoff relevant. The period payoffs were summed up and paid out in private at an exchange rate of 6 Eurocents per EMU, additional to a show up fee of 4 Euros. The (potential) negative externalities or positive spill-overs caused by the decisions of other group members (see Figures 3.1 and 3.2) were summed over all 10 periods and subtracted from the total sum of payoffs at the end of the experiment. By revealing the total (external) damage or addition only after all decisions were made, allows us to consider any (B-O-) pair as an independent observation. This implies the assumption of irrelevance of one pair's beliefs on the other pairs' behaviour with respect to the adjustment of its corrupt behaviour, which is reasonable according to the findings in Chapter 2.[19] The instructions were partly read aloud (See Appendix 3E for the full instructions in SWT). By providing a sufficient amount of time (approximately 60 minutes) to read and work through the instructions, including a series of examples and control questions, we made sure that the rather complicated set-up was fully understood by all participants before they entered the actual computerized experiment. The entire experiment took less than 90 minutes. A questionnaire containing socio-economic questions was filled out after the computerized experiment was finished and before payments were conducted in private.[20]

Payoffs lay between 6 Euros and 25 Euros (excluding a show-up fee of 4 Euros) at an average of 14.53 Euros. The framing of the instructions was kept neutral throughout,

[19]Inter-dependencies (with respect to relevant variables) may present a problem when considering equilibria within the 'super-game' regarding strategies of and beliefs on the behaviour of all participants within a session. In Chapter 2 (which uses a similar set-up with respect to the potential effects of the model negative externality), we do not find any significant difference in the relevant performance variables (and treatment effects) between sessions using a reduction in other participants' payoffs (an 'active' public) and sessions using a reduction in the amount of donation for the public aid organization 'Doctors without Borders' (a 'passive' public) as a model for the negative externality. This indicates the irrelevance of 'super-game' considerations. Therefore we treat pairs as independent observations.

[20]Payments were made by an assistant who did not appear as an experimenter.

avoiding any language indicating the subject of research. Abbink and Hennig-Schmidt (2006) find no significant effect of loaded language (framing) on the main variables of the corruption game, providing an argument for the robustness of the use of neutrally framed experiments in the context of corruption as long as the essential features of corruption are salient and explained sufficiently to the participants. A neutral setting allowed the easiest comparison between the treatments and across studies.

3.4 Analysis and hypotheses

The main purpose of this paper is to evaluate the effectiveness of whistle-blowing in general and different institutions of whistle-blowing in particular. Therefore we compare our three treatments with respect to the following variables.

First, we examine the average fraction of successful 'legal' transactions resulting from the decisions of a pair of O and B participants ('Normal' in Stage 1 and 'Ok' in Stage 2) over all 10 periods within a treatment. This provides a measure, both for the indirect consequence of corruption and for the resulting outcomes in terms of efficiency under the respective institution. Second, we consider the fraction of successful 'corrupt' transactions ('Prefer' in Stage 4/5 and 'Ok' in Stage 6), which is directly proportional to the sum of the negative external effects (direct consequence of corruption). Third, we measure social welfare by considering the average payoff level and its distribution across participants. Taking average payoffs (alone) as a measure of welfare is problematic, since they are strictly increasing in the total amount of bribes paid during the experiment (which is due to the tripling of the transfer stemming from the assumed difference in marginal utilities between B and O). This may lead to relatively high levels of payoffs in treatments with high corrupt activity. Using the amount of average payoffs only in combination with the two other performance variables and its distribution across participants (within a treatment) as a measure of welfare presents a more comprehensive picture.

With respect to these measures, our experimental study concentrates on two main motivations for blowing the whistle. Whistle-blowing can be used to avoid (forced) participation in a corrupt transaction. This may have several reasons. Participants (both

B and O) may want to signal unwillingness to cooperate in a corrupt transaction in order to maximize their own expected payoff, save the public from the damage (altruism, see Andreoni and Miller 2002), or promote socially efficient 'legal' transactions in future periods. The motivations may be dominated by altruism, efficiency seeking or inequality aversion (Engelmann and Strobel 2006, Fehr and Schmidt 1999, 2003). However, whistle-blowing can be intentionally misused by B as a tool for stabilization. Although it is part of a strictly dominated strategy, the deployment of costly punishment as a form of negative reciprocity in case O does not deliver the corrupt service (chooses 'Nothing' in Stage 5), and the anticipation thereof, may increase the level of corruption (Abbink 2004, Lambsdorff and Frank 2010).

In this section we discuss possible systematic effects of whistle-blowing. A strict separation of the different motivations is neither possible nor intended in our experiment, since we are ultimately interested in the total effect of the introduction of whistle-blowing. To form hypotheses on the differences in subjects' behaviour between the treatments we restrict ourselves to the use of simple arguments based on individual expected payoff maximization under varying sets of beliefs, relaxing the assumption of common knowledge of rationality (Kreps et al. 1982).[21]

3.4.1 Number of legal transactions

One argument in favour of the introduction of whistle-blowing is that it will hinder corruption from discouraging the realization of 'legal' transactions (Klitgaard 1988). Before taking the effect of whistle-blowing into account, regard the mechanism of the indirect negative effect of corruption in the control treatment. Consider a client B in the CT who would prefer to choose (at least in the last period in which there are no strategic considerations stemming from the repetition of the game) the 'legal' procedure ('Normal')[22]. Not believing in the rationality of O, B may fear that O chooses 'Add' in Stage 2 (with some probability) to force her to enter the corrupt transaction if she

[21] We refrain from the analysis of models of strong reciprocity (Fehr and Schmidt 2003). Note that the following analysis has the sole purpose of formalizing simple arguments on the mechanism of whistle-blowing within the self interested model of payoff maximization. It does not claim to provide a full description of strategies that may be rationalized (off the equilibrium path) once the assumption of common knowledge of rationality is relaxed.

[22] Note that this is consistent with the prediction of the SNE under the assumption of common knowledge of rationality.

chooses 'Normal' in Stage 1 (O may want to force B to enter the corrupt transaction hoping for a (large) bribe). Depending on B's belief about the probability of O to initiate the corrupt transaction in Stage 2 and her expected payoff conditional on this initiation, she may, in expectation, be better off to choose the outside option ('Out') in Stage 1, staying away from the productive market (see Appendix 3B for a details on the set of beliefs rationalizing such behaviour). Hence, the fear of encountering a corrupt official may, in CT, lead to a rational choice of the outside option by participants holding a certain set of (off-equilibrium) beliefs.

The introduction of whistle-blowing may change B's belief about O's Stage 2 behaviour as well as her expected payoff conditional on being forced into the corrupt transaction (O choosing 'Add' in Stage 2). This may in turn change B's decision between 'Out' and 'Normal' in Stage 1 and hence the magnitude of the indirect negative consequence of corruption. While B's payoff after a request for a bribe in Stage 2 (by O) is uncertain and can be as low as 14 EMU in the CT (see Appendix 3B), it is at least 19 EMU (in expectation) in the whistle-blowing (WB) treatments.[23] The option of whistle-blowing in Stage 3 can reduce the expected loss B is risking if she chooses 'Normal' in case she encounters a corrupt official (in Stage 2). Hence whistle-blowing serves as a safeguard against the exploitation of an 'honest' client by a corrupt official. For a relevant set of beliefs (see Appendix 3B), a client who would choose 'Out' in CT (securing 20 EMU) would choose 'Normal' in the whistle-blowing treatments (risking only 1 instead of 6 EMU).[24] This yields the following hypothesis.[25]

Hypothesis 1: "The relative number of observations in which type B participants choose 'Normal' will be higher in both WB treatments compared to the CT, while the number of Stage 1-failures ('Out' choices) will be lower."

[23] B can always blow the whistle in Stage 3 and get 19 EMU and therefore will never choose 'Add' in Stage 3, if the expected payoff does not exceed 19 EMU.

[24] The anticipation of this behaviour by the official and feedback effects (on the beliefs of the client) may even strengthen this argument.

[25] The formal argument established in Appendix 3B is valid for the last period and should, by backward induction feed back to all previous periods.

3.4.2 Stabilizing corruption

The opportunity of direct punishment as a 'norm' enforcement device has been found to be effective in situations in which subjects act in a selfish way or are insufficiently motivated by positive reciprocity alone (Fehr et al. 1997, Fehr and Gächter 1998). Lambsdorff and Frank (2010) as well as Abbink et al. (2000) show in an experiment using a one shot version of the corruption game that the presence of whistle-blowing, as a specific form of costly punishment, increases the stability of a corrupt transaction (although corrupt reciprocity is questionable as a social 'norm'). B may exploit the (threat of) whistle-blowing in Stage 6 to punish defective[26] (opportunistic) behaviour in order to stabilize the corrupt transaction. O may adhere to a set of beliefs assigning a positive probability for B retaliating defection with whistle-blowing despite the non-trivial personal costs.[27]

Consider an official who maximizes her expected payoff in the last period.[28] Once O believes that B will blow the whistle with a certain probability $r(T)$ (which may depend on the size of the bribe) after defection (choosing 'Nothing' in Stage 5), O may be, in expected terms, better off when choosing 'Prefer' (see Appendix 3B for explications on critical beliefs). If the official believes the threat of costly punishment to be credible[29], whistle-blowing can serve as a tool of retaliation which stabilizes the corrupt transaction.

Hypothesis 2: "The relative number of successful deals ('Prefer' in Stage 5 and 'Ok' in Stage 6) will be higher in SWT than in CT."

Hypothesis 3: "The relative number of O's defection will be lower in SWT than in CT."

The dependency of $r^*(T)$ on the level of transfer may not be the only reason for individuals to condition their reciprocal behaviour on the size of T. Reciprocation after a high transfer can be interpreted as part of a strategy aimed at triggering high transfers

[26] O chooses 'Nothing' after receiving a bribe

[27] Alternatively O may adhere to beliefs assigning positive probabilities to the existence of certain types of B who will (always) negatively retaliate opportunistic behaviour.

[28] We consider the last period in order to shut off all effects stemming from repetition, ignoring the consequences of O's decisions on B's response in future periods.

[29] Note that whistle-blowing in Stage 6 is costly in both WB treatments and hence belongs to strictly dominated strategies.

in future periods. While the argument of strategic reciprocity holds for all treatments, the additional effect stemming from the threat of retaliation is to be expected only in SWT, since it is absent in CT and avoidable in LT.

3.4.3 Gender effects

The belief on the punishment probability $r(T)$ is likely to differ substantially across individuals. Lambsdorff and Frank (2010) find that female participants are less likely to engage in costly (hence off-equilibrium) punishment and, as a direct consequence, anticipate (irrational) negative reciprocity less often than male participants. These results provide experimental evidence and an explanation for the empirical findings of lower corruptibility of women in Dollar et al. 2001 and Swamy et al. 2001, while they do not contradict diverse evidence on women's attitude towards corruption (Alatas et al. 2006).

The last three stages (Stage 4, 5 and 6) of the set-up in SWT are close to the experiment of Lambsdorff and Frank (2010) with regard to methodology as well as contents. However, in contrast to their one shot situation, whistle-blowing in our set-up is not the only tool of negative retaliation for the client. In our repeated game, B can also (negatively) reciprocate by choosing less cooperative actions in future periods. (Only in the last period, direct punishment is B's only tool of negative retaliation.) Hence we investigate whether the gender effect found in Lambsdorff and Frank (forthcoming) survives the repetition of the game and explore the role of the direct punishment device (i.e. whistle-blowing in Stage 6) with respect to its interaction with cross-period reciprocity.

Hypothesis 4: "Female participants (in the role of O) will show different levels of defection compared to their male counterparts only in the WB treatments if the gender effect depends on the direct punishment device of whistle-blowing (alone)."

This hypothesis is supported by the results of several experimental studies. The lack of significant differences in the behaviour between female and male participants with respect to social preferences (Croson and Gneezy 2009), risk attitudes (Dekel and Scotchmer 1999), or trusting behaviour (Croson and Buchan 1999) suggests that the gender effect in corruption may be solely due to differences in the anticipation of 'irrational' and costly direct punishment (Eckel and Grossman 1996). According to the findings of Lambsdorff and Frank (forthcoming), the gender effect is likely to be found not only in the anticipation but also in the direct application of punishment behaviour.

Hypothesis 5: "Female participants in the role of B will show lower levels of punishment in case of O's defection in Stage 5 compared to their male counterparts."

3.4.4 Officials' whistle-blowing

In SWT, O's whistle-blowing in Stage 5 leads to the loss of the transfer-related benefit from the corrupt transaction ($3T$). Although it is unattractive (yielding the same payoff as 'Out') for O, she may blow the whistle in order to signal her unwillingness to engage in corrupt reciprocity (which is relevant for the situations in which B has chosen 'Add' in Stage 1 and for early periods) or to punish B in case of a low transfer (which is relevant in situations in which O has chosen 'Add' in Stage 2, see Section 3.4.1). For low transfer levels, the options 'Whistle' and 'Out' yield similar outcomes for O and only small differences (i.e. the amount of a low transfer) for B. Hence, we do not expect O's possibility to blow the whistle in Stage 5 to be of any relevance in this treatment.

In LT, however, O can keep the benefit from the bribe ($3 * T$) at a relatively low cost, i.e. 3 EMU, if she blows the whistle in Stage 5, while she remains safe from the possible negative retaliation through B's whistle-blowing in Stage 6. Thus, the opportunity of whistle-blowing provides a powerful incentive to defect for those officials whose corrupt reciprocity is based on a belief that B will punish defection with a high probability (i.e. a high $r(T)$, see Section 3.4.2). Again, we use an argument of backward induction ignoring effects stemming from interactions of behaviour across periods (starting with the last period).

Once reached Stage 5, payoff maximizing officials would always prefer 'Whistle' over 'Prefer' if $EP(\text{`}Whistle'|H_{I_5}) = 3 + 3 * T > EP(\text{`}Prefer'|H_{I_5}) = 3 * T$ for any T and 'Whistle' over 'Nothing' if $EP(\text{`}Whistle'|H_{I_5}) = 3 + 3 * T > EP(\text{`}Nothing'|H_{I_5}) = (1-r(T)) * (6+3T) + r(T) * 8$. This is the case if $r(T) \geq \frac{3}{3T-2} = r^*_{LT}$. Hence those officials who hold a belief $r(T)$ that would make it optimal for them to reciprocate the bribe in Stage 5 of SWT (All officials holding $r(T) \geq \frac{6}{3T-2} = r^*_{SWT}$), would blow the whistle in LT since $r(T)$ must then be larger than the critical bribe r^*_{LT} $(= \frac{3}{3T-2} < r^*_{SWT})$[30]. Though irrelevant for those participants who reciprocate for strategic reasons, providing leniency of O's Stage 5 whistle-blowing may reduce the stability of the reciprocal corrupt transaction by giving an incentive to defect to those who reciprocate for fear of B's direct negative retaliation.

Hypothesis 6: "In Stage 5 of LT, the relative number of 'Prefer' choices will be lower and the relative number of 'Whistle' choices higher than in SWT."

While we expect the negative effect of whistle-blowing, i.e. the stabilization of the corrupt transaction through the threat of retaliation (by the client), to be substantially diminished through asymmetric leniency, we expect the positive effects (i.e. higher numbers of 'Normal' choices in Stage 1) to remain. The total effectiveness of the introduction of whistle-blowing depends on the magnitude of the positive effect, the magnitude of the negative effect and the extent to which the latter can be contained by providing asymmetric leniency.

3.5 Results and Interpretation

3.5.1 Descriptive statistics

The main questions addressed in this paper can be essentially answered by considering the differences in the main performance variables, the average relative number of successful 'corrupt' transactions (Corrupt Deals), the average relative number of successful 'legal' transactions (Legal Deals, out of ten situations/periods), the average payoff (Payoff, total income in Euros) and the average transfer level (Transfer, over all ten periods and pairs of participants, in EMU) shown in Table 3.1.

[30] Note that this holds only for the last period and for situations in which the concerns about immediate direct punishment dominate strategic considerations.

Table 3.1: **Main Results**

Treatment	CT		SWT		LT	
	mean	std. dev.	mean	std. dev.	mean	std. dev.
Legal Deals	1.75	1.11	3.48	1.67	3.39	2.19
Corrupt Deals	2.83	1.24	3.57	1.58	2.37	0.99
Payoff	13.04	2.44	14.83	4.19	16.08	3.45
Transfer	4.18	1.40	4.67	2.27	4.74	2.36

All figures are averages across individuals for the relevant types and over all 10 periods within a treatment.

As we have expected (**Hypothesis 1**), we find a large positive effect of whistle-blowing on the relative number of legal successful deals. On average 1.75 out of 10 potential transactions in CT end in the (socially) first best option, compared to 3.48/3.39 in SWT/LT (differences: CT vs. SWT/LT tested with pair-wise, two-sided Mann-Whitney U-tests[31]: $p = 0.002/0.001$; $N = 33/30$).

Corruption levels are by 0.74 $(3.57 - 2.83)$ significantly ($p = 0.089$; $N = 33$) higher in SWT than in CT, confirming the effect of symmetric whistle-blowing on corrupt stability (**Hypothesis 2**). The low level of corruption in LT (2.37 in LT vs. 3.57 in SWT: $p = 0.064$; $N = 39$) indicates the effectiveness of asymmetric punishment of whistle-blowing (leniency) with respect to the incentives for defection (**Hypothesis 6**).

There is no significant difference in the total payoff levels between CT and SWT (13.04 Euros compared to 14.83 Euros: $p = 0.23$; $N = 33$), while the differences in payoff levels between LT (16.08 Euros) and both other treatments are highly significant ($p < 0.001$; $N \geq 30$). Not only does the value of total payoffs differ substantially across treatments but also their composition. The difference in average payoffs between type O and type B participants is lowest in LT at 1.52 $(16.84 - 15.32)$ Euros, 2.18 $(14.18 - 11.90)$ Euros in CT and highest in SWT at 5.88 $(16.77 - 12.89)$ Euros.

LT not only yields the lowest rate of corruption and a high number of successful legal transactions (see Table 3.1), it also generates the highest per capita payoff level over all 10 periods with the most desirable outcomes in terms of the distribution of

[31] If not stated otherwise, reported p-values are all for two-sided Mann-Whitney U-tests.

income. When we compare the Gini-coefficients[32], we obtain the lowest level for LT (0.12) compared to 0.15 for SWT and 0.14 for CT. This distributional effect stems from the fact that corruption increases inequality by allowing the corrupt partners to extract large benefits while causing a negative externality to the public. Hence participants who do not engage in corruption receive only moderate payoffs (engaging in legal transactions or falling back to the outside option) which may be further reduced by the negative external effect caused by their fellow participants (of their session) who are corrupt. A large number of legal and a low number of corrupt transactions not only maximize the sum of payoffs but also flatten their distribution.

Efficiency

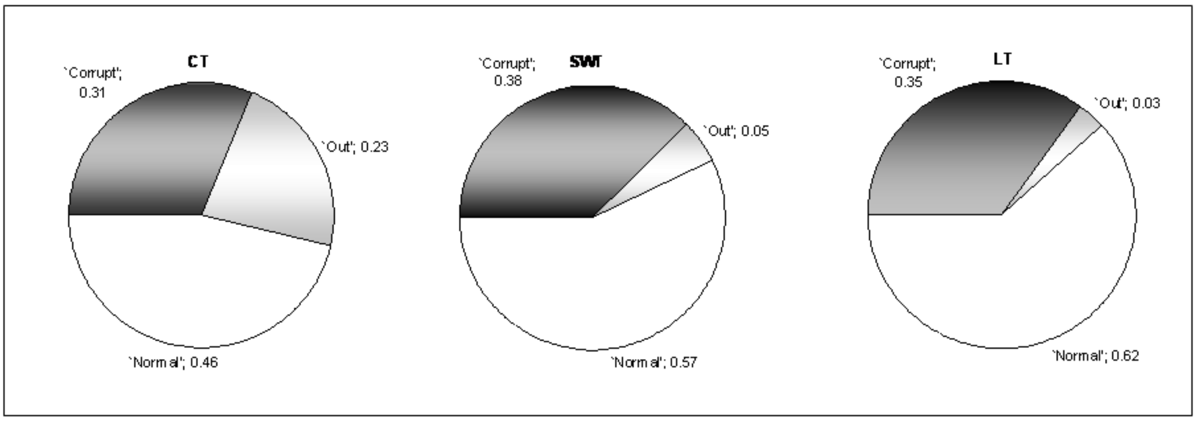
As explicated in Section 3.4, corruption indirectly reduces productive activity and thereby efficiency. The risk of being forced into a corrupt transaction may hinder honest clients to enter productive markets. In our set-up, the presence of corruption causes participants to choose the sub-optimal outside option (‘Out’) in Stage 1, although any strategy set that includes this choice is strictly Pareto-dominated (with respect to outcomes) by the strategy set predicted by the SNE ([‘Normal’; ‘Ok’; ‘T=1’; ‘Nothing’] in CT and [‘Normal’; ‘Ok’; ‘Add’; ‘T=1’; ‘Nothing’; ‘Ok’] in both WB treatments, see Sections 3.2.1 and 3.2.2). We show that the opportunity to blow the whistle offers a mechanism that is suitable to decrease this effect by providing the client with a safeguard against the exploitation by a corrupt official (**Hypothesis 1**). There are two conditions for a ‘legal’ transaction to be successful.

First, B has to choose ‘Normal’ in Stage 1. Figure 3.3 shows that, while clients seem to be equally willing to initiate the corruption transaction across all three treatments (31% ‘Add’ choices in CT, 38% in SWT and 35% in LT, differences between treatments, CT vs. SWT, CT vs. LT, and SWT vs. LT, are not significant: $p \geq 0.312$; $N \geq 30$), they choose to enter the legal procedure significantly more often in the WB treatments, 62% in SWT and 57% in LT compared to 46% in CT (SWT vs. CT and LT vs. CT: $p < 0.01$; $N \geq 30$). The fraction of cases in which the outside option is taken in Stage 1 is drastically reduced through whistle-blowing, from 23% in CT to 5% in SWT and

[32] The Gini-coefficient measures the distribution of wealth within a certain population on a scale between 0 and 1, where a value of 0 corresponds to complete equality and a value of 1 signifies complete inequality.

3% in LT (supporting **Hypothesis 1**). The differences between CT and SWT as well as between CT and LT are significant on the 1%-level ($p < 0.001$; $N \geq 30$).

Figure 3.3: Shares of choices in Stage 1



Second, O has to choose 'Ok' in Stage 2. The rates of acceptance ('Ok' in Stage 2) differ significantly between WB treatments, at 61% in SWT and 55% in LT, and CT (38%). Pair-wise U-tests for SWT vs. CT and LT vs. CT yield $p < 0.001$ with $N \geq 30$. Officials seem to anticipate their clients to blow the whistle in Stage 3 (which they do in 22% of possible cases in SWT and in 21.5% in LT) after being forced to enter the corrupt transaction and therefore choose the legal procedure more often. According to the argumentation of Section 3.4.1 (and Appendix 3B), the choices in Stage 1 depend critically on the expected payoff which are determined by B's beliefs on O's behaviour in Stage 2 and Stage 4/5 (CT/WB treatments).

The high frequency (18%) of O choosing 'Out' (given that B has chosen 'Normal' in Stage 1) in Stage 4 of CT, signifies the relevance of this option as a (costly) punishment device. The rational expectation of this behaviour determines B's expected payoff conditional on being forced by O to enter the corrupt transaction in Stage 2. Even if we assume that B's belief on the probability of O choosing 'Ok' in Stage 2 is constant over the treatments, this creates a gap in B's expected payoff when choosing 'Normal' between CT and the WB treatments.[33] It may be just this expected gap (which is

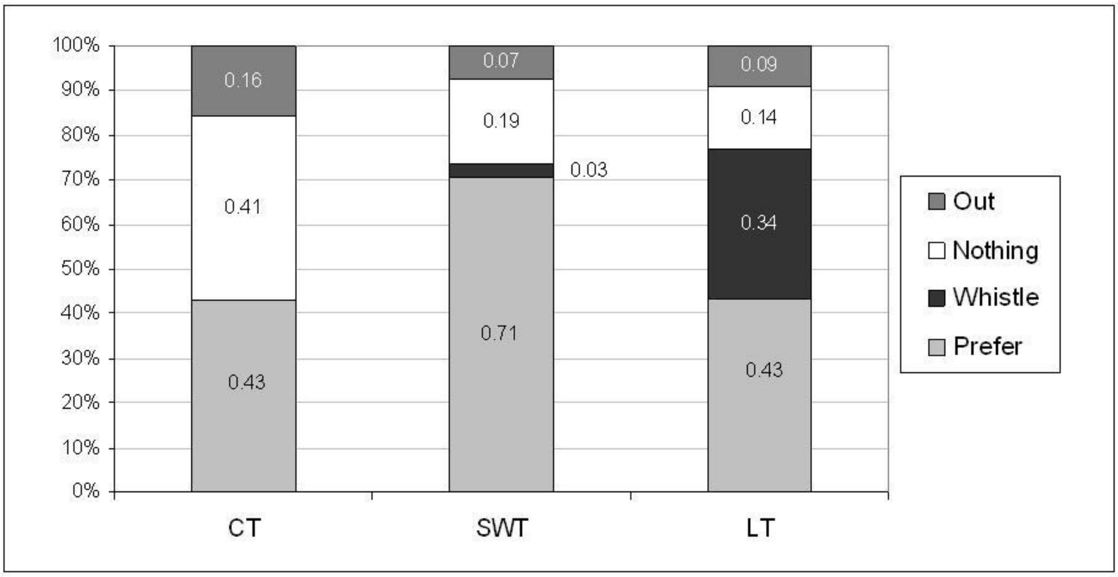
[33]To put it differently, given $1 - q = 0.18$, B's belief on O choosing 'Out' in Stage 4, the gap in B's critical beliefs on O choosing 'Ok' in Stage 2 (p) between CT and the WB treatments is substantial, see Figure 3.8 in Appendix 3B. Using the argumentation of Section 3.4, the critical belief in CT is $p^*_{CT} = \frac{6-5*0.18}{11-5*0.18} = 0.51$ while it would only be $p^*_{WB} = \frac{1}{6} = 0.17$ in the WB treatments. This means that B has to expect O to choose 'Ok' in Stage 2 with a probability of at least 51% to make her choose 'Normal' in Stage 1 of CT while she would already do so in the WB treatments as long as she expects O to choose 'Ok' with a probability of more than 17%. Note that the actual 'probability' of O choosing

based on realistic beliefs) that leads to the large differences in the fractions of 'Out' and 'Normal' choices in Stage 1 between the treatments (see Figure 3.3). The gap in outcomes (successful legal choices, see Table 3.1) is even wider due to the differences in behaviour of O in Stage 2. Our results suggest that whistle-blowing equips the client with a useful safeguard against the official's threat of exploitation, allowing her to initiate a 'legal' transaction at a low(er) risk. This makes it, in expectation, more attractive for the client to enter (or stay) in the productive market.

Corruption levels

Figure 3.4 depicts the composition of O's choices in Stage 4/5 (CT/WB treatments). Having received a bribe (reaching Stage 4/5), officials in LT as well as in CT choose to be corrupt ('Prefer') in 43% of cases, which is substantially less than the percentage of conditional 'Prefer' choices in SWT, 71% (U-tests; CT vs. SWT: $p = 0.071$; $N = 33$, LT vs. SWT: $p = 0.002$; $N = 39$).

Figure 3.4: Shares of choices in Stage 4/5



In contrast to the case of CT, defection ('Nothing') is risky in terms of expected payoffs in both WB treatments and therefore occurs (significantly) less often (U-tests: CT vs. SWT and CT vs. LT: $p < 0.001$; $N \geq 30$). The high fraction of the (conditional) success of corruption in SWT (71%) can be explained by payoff maximizing motives of officials. In SWT, O can only benefit with certainty from a received bribe (get $3*T$) if she

'OK' in Stage 2 is 0.38 in CT and thereby lies in the relevant range to expect a treatment effect.

cooperates in the corrupt transaction. If she does not, she faces the risk of punishment (**Hypothesis 3**). Officials adhering to a belief structure that makes them defect in CT may reciprocate in SWT rather than face the threat of punishment (which results in the loss of the benefit from the bribe). In LT, most officials (60%)[34] who decide not to deliver the corrupt service, 'insure' themselves against the possibility of retaliation by the client and substitute risky defection by whistle-blowing. The difference in O's payoffs between the options 'Whistle' and 'Nothing' (which amounts to 3 EMU) can be interpreted as an insurance fee.

While whistle-blowing does not provide any real benefit for the official (in direct comparison to the option 'Out') in SWT, it enables officials in LT to benefit from the transfer without risk and without collaborating in the corrupt transaction.[35] The results suggest that the institution of asymmetric leniency for O's whistle-blowing is successful in the sense that a considerable number of officials can be incentivized to blow the whistle rather than reciprocate in the corrupt transaction (**Hypothesis 6**).

Transfer levels

The differences in the average transfer levels between the treatments (see Table 3.1) can be explained by path-dependency of B's behaviour with respect to reaching the bribing stage (Stage 3 in CT and Stage 4 in the WB treatments). Table 3.2 shows four different types of average transfer levels. Transfer_{mean} measures the average transfer level across periods (and thereby treats transfers in situations for which the bribing stage has not been reached as 0 values). Transfer_{cond} only considers average transfers of situations for which the bribing stage has been reached. These situations are further divided into those for which B has reached the relevant stage voluntarily (as her first choice i.e. by choosing 'Add' in Stage 1) and those for which B has revealed to prefer the legal option (by choosing 'Normal' in Stage 1) and hence has entered the bribing stage 'involuntarily'. Transfer_{vol} is relevant for the former set of situations, Transfer_{invol} for the latter.

[34]The fraction is calculated by $\frac{0.34}{0.34+0.14+0.09} * 100\%$, see Figure 3.4.

[35]This assumes that choosing 'Prefer' also yields a riskless return to O since B has no reason to blow the whistle in this case.

Table 3.2: **Transfer levels**

Treatment	CT		SWT		LT	
	mean	std. dev.	mean	std. dev.	mean	std. dev
Transfer_{mean}	4.18	1.40	4.67	2.27	4.74	2.36
Transfer_{cond}	6.77	0.27	8.61	1.10	8.71	1.21
Transfer_{vol}	9.07	1.04	9.11	1.18	9.44	1.15
Transfer_{invol}	5.36	1.13	5.92	1.07	5.61	1.22

All figures are averages across subjects in the relevant periods.

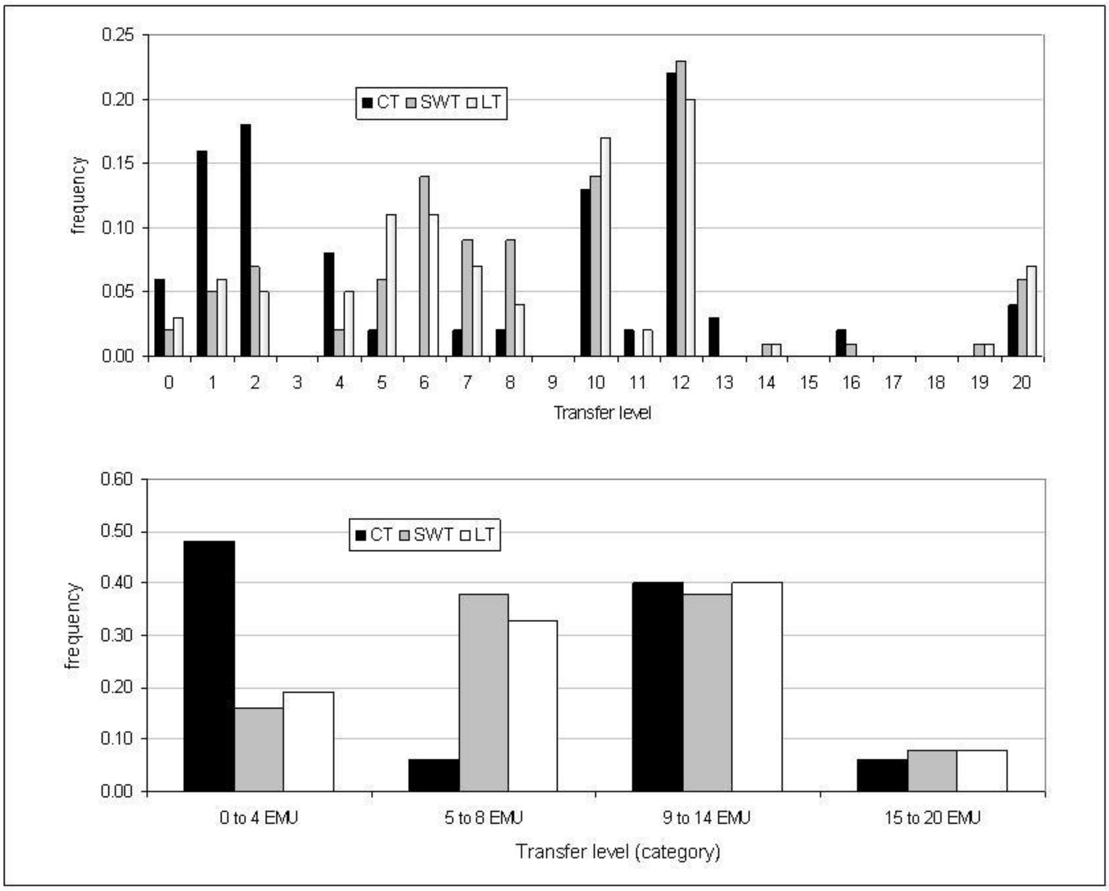
Among these four types of transfers, we find a significant difference between the CT and the WB treatments only in the average transfer level conditional on reaching Stage 4, Transfer_{cond} (CT vs. SWT: $p = 0.089$; $N = 33$ and CT vs. LT $p = 0.007$; $N = 30$). Neither of the other three levels (Transfer_{mean}, Transfer_{vol} and Transfer_{invol}) show any significant difference between the treatments ($p \geq 0.452$; $N \geq 15$). Within all three treatments, the differences between Transfer_{vol} and Transfer_{invol} are all highly significant (Wilcoxon signed ransum tests: Transfer_{vol} vs. Transfer_{invol}: $p < 0.001$; $N \geq 8$)[36]. We infer that clients who have been forced to enter the corrupt transaction seem to be less inclined to transfer high bribes.[37] The treatment difference in Transfer_{cond} can be explained by differences in the 'voluntariness' of reaching the bribing stage (shares of clients reaching the bribing stage 'voluntarily' or 'involuntarily') between CT and the WB treatments. In CT, only 52% of clients who reach the bribing stage are of the 'voluntary' type, compared to 78% in SWT and 71% in LT.

Figure 3.5 shows the distribution of transfers conditional on reaching Stage 4/5. Since the number of observations is limited for some transfer levels (e.g. transfer levels 3, 9, 17 and 18 were not chosen at all), we illustrate the frequency of transfers in two ways. The upper panel of Figure 3.5 shows the frequency of transfers separated in exact levels. For reasons of clarity we pool situations in which transfers were 'very low' (0 – 4 EMU), 'low' (5 – 9 EMU), 'medium' (10 – 14 EMU) and 'high' (15 – 20 EMU) in respective categories in the lower panel of Figure 3.5.

[36]Here we loose the observation by those pairs who have either made only voluntary or only involuntary transfers.

[37]This may be due to differences in beliefs on O's willingness to reciprocate or differences in social preferences etc.

Figure 3.5: Distribution of transfer levels and categories



In the upper panel (of Figure 3.5) we see a large mode at the transfer level of 12 EMU in all treatments. We interpret this as evidence for the presence of 'fairness' considerations (Fehr and Schmidt 1999) between B and O, since a transfer of 12 EMU followed by positive reciprocity by O equalizes payoffs between the two in all treatments (since $48 - 12$ EMU $= 36$ EMU $= 3 * 12$ EMU, see Figure 3.1 or 3.2). The second mode for all treatments at a transfer of 10 EMU may be explained by the focal character of the number 10 or its proximity to the payoff equalizing transfer level (12 EMU) and a self-serving bias. Neither of the treatments shows substantial activity at the high end of the transfer scale ($b > 12$). Considering the lower panel (of Figure 3.5), transfers are almost equally distributed in the 'medium' and 'high' range across treatments, while subjects' transfers in CT lie significantly more often in the 'very low' and significantly less often in the 'low' range than those in the WB treatments (CT vs. SWT/LT: $p < 0.01$; $N \geq 11$). We attribute this finding to differences in the fractions of situations of 'voluntary' and 'involuntary' bribing between the treatments.

3.5.2 Conditional reciprocity

In our experiment, the occurrence of corruption mainly relies on (strategic) reciprocity between B and O. Confirming results from previous experiments on corruption (e.g. Abbink 2004, Lambsdorff and Frank 2010), the spearman rank correlation coefficients of $\rho_s(CT) = 0.62$, $\rho_s(LT) = 0.60$ and $\rho_s(SWT) = 0.69$ support a strong positive correlation between the transfer levels and the probability of corrupt success ('Prefer') in all treatments. Independence can be rejected at the 1% -level ($p < 0.003$; $N \geq 61$).

Figure 3.6: Corrupt success conditional on transfer levels

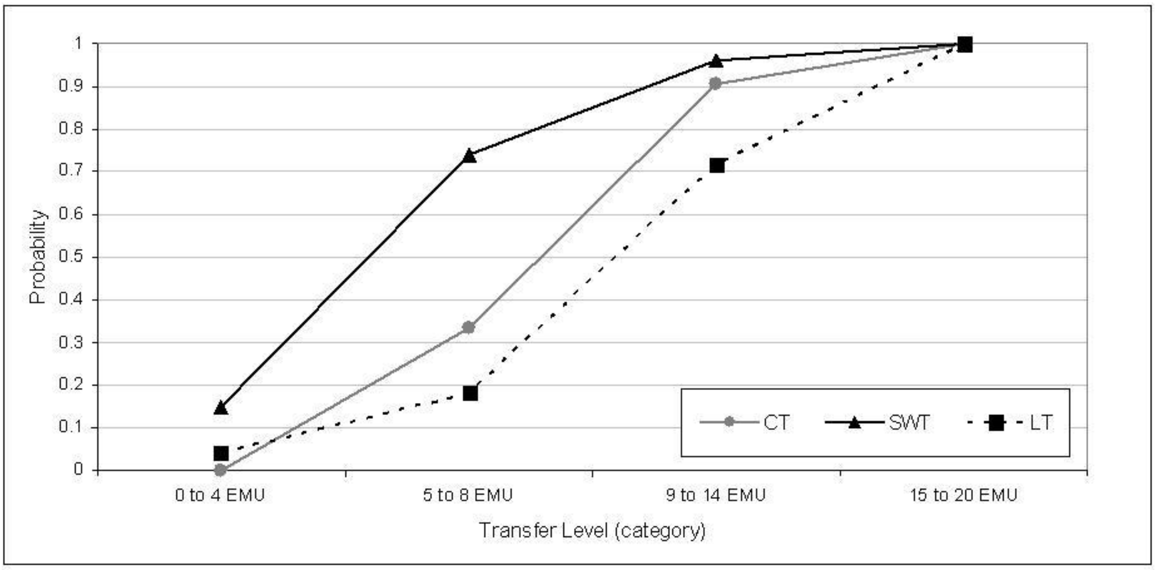


Figure 3.6 shows the probability of corrupt success conditional on transfer level categories.[38] We confine our analysis of conditional corruption to non-parametric tests.[39] The difference in the correlation coefficients between CT and SWT in combination with the differences in average levels of corrupt success conditional on transfer levels (Figure 3.6)

[38]A figure without clustering in categories shows a qualitatively similar pattern but decreases clarity because of spikes and flat spots for transfer levels with no or few observations.

[39]An identification of treatment differences in the causal relationship between the transfer level and the probability of success (or defection) in a regression is not possible. The selection process of (officials) reaching the bribing stage (only about 60% of possible cases in CT and about 50% of cases in the WB treatments reach Stage 4/5) cannot be explained by our data in a way that would satisfy the conditions of a Heckman correction process (Heckman 1979). In our experiment most B-O pairs of participants reach the bribing stage for some periods and not for others. This makes it impossible to find a set of variables that explains the selection process and, at the same time, is irrelevant for the choice of the size of the transfer.

suggests that symmetric whistle-blowing strengthens the reciprocal relationship between the briber and the official. While the conditional probability of corrupt success in SWT is not significantly different from the ones in LT and CT for very low, medium and high transfer levels (SWT vs. CT, SWT vs. LT: $p \geq 0.342$; $N \geq 33$), it is significantly higher for low transfer levels ($p \leq 0.008$; $N \geq 33$). For low transfer levels (which are relevant for about 30% of all observations) the rate of success is 74% in SWT compared to 33% in CT and 18% in LT.

Corrupt reciprocity is likely to be motivated not only by securing the present period's payoff (avoiding whistle-blowing as a form of punishment in the WB treatments) but also by strategic considerations (triggering B's positive reciprocity, i.e. high transfers in future periods). The lower the transfer levels, the lower the expected flow of future transfers (of the same magnitude) and hence the less important strategic considerations become, compared to immediate concerns about B's direct negative reciprocity. For 'medium' and 'high' levels of transfer, strategic considerations are strong enough to cause high levels of corruption in all treatments, making any other potential motivation redundant. For 'very low' transfer levels, the motivation to secure the present period's payoff is weak and hence does not lead to a treatment difference. Only for 'low' levels of transfers, for which strategic considerations are not sufficient to trigger positive reciprocity for most officials (in most situations),do participants in SWT show significantly higher success rates than those in CT and LT. We attribute this to O's fear of B's retribution (through WB in Stage 6), which is absent in CT and avoidable[40] in LT.

3.5.3 Whistle-blowing and gender

We want to test whether the gender difference in corrupt behaviour (found e.g. in Swamy et al. 2001 or Sung 2003) can be demonstrated experimentally in a repeated game. In a one shot game, Lambsdorff and Frank (forthcoming) show that gender effects in corrupt reciprocity can be attributed to differences in the tendencies to engage in costly punishment and to anticipate direct negative retaliation. In our set-up, strategic considerations may crowd-out effects stemming from differences in belief structures and

[40] O can avoid retaliation in Stage 6 by choosing 'Whistle' in Stage 5

hence cut out potential gender effect.

Gender and defection

In CT we do not find any significant difference in the rate of defection (fraction of 'Nothing') between female and male officials. Both genders defect in 44% (female: $\frac{0.24}{0.55} = 0.44$; male: $\frac{0.20}{0.45} = 0.44$) of possible cases, see Table 3.6 in Appendix 3C. In SWT, female officials choose 'Nothing' in 38% ($\frac{0.12}{0.32} = 0.0.38$) of cases (which is not significantly different from the rates of defection for both genders in CT, $p = 0.295$; $N = 15$), while male officials defect in only 10% ($\frac{0.07}{0.68} = 0.10$) of relevant situations, see Table 3.7 in Appendix 3C. The defection rates in LT show a similar pattern as in SWT, however, the gender difference is slightly smaller and statistically insignificant ($p = 0.283$; $N = 12$), see Table 3.8 in Appendix 3C. Only in SWT do female officials defect significantly more often than their male counterparts (female vs. male: $p = 0.002$; $N = 20$). The lack of a significant gender effect in CT suggests that female and male officials do not differ in the consideration of the effect of defective behaviour on the clients' future behaviour (with respect to transfers in future periods). They do not seem to follow systematically different approaches with respect to their strategies in the corrupt context. We interpret the strong gender effect in SWT as evidence for the hypothesis that male officials show a stronger reaction to the possibility of getting punished for defection (through whistle-blowing in Stage 6) than female officials. As a consequence, pairs of participants involving a male official show significantly higher success rates of corruption than those with a female official (**Hypothesis 4**).[41] This confirms the validity of the argumentation of Eckel and Grossman (1996) with respect to gender differences in anticipation behaviour.

Gender and punishment

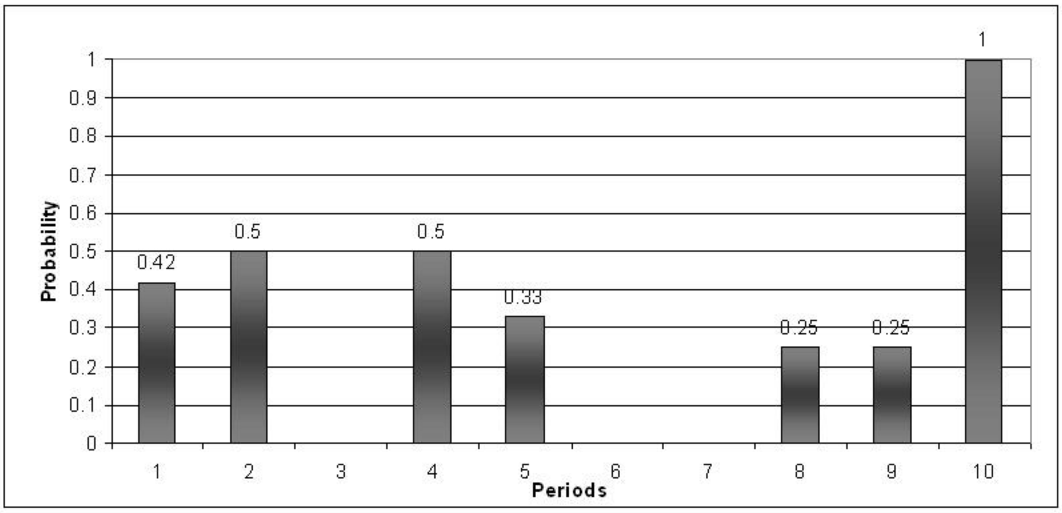
In both WB treatments a substantial number of clients are found to punish defective behaviour of the official (by choosing 'Whistle' in Stage 6). In 40% (SWT) and in 38% (LT) of all relevant situations (i.e. O chose 'Nothing' in Stage 5)[42] the client accepted substantial costs (6 EMU, independent of the size of T) to blow the whistle. In our

[41] Note that the gender of the transaction partner is neither known nor noticeable to any of the participants.

[42] As expected, not a single participant of type B chose 'Whistle' in Stage 6 after O had chosen 'Prefer' in Stage 6

experiment, whistle-blowing in Stage 6 is designed in a way that the impact of punishment depends positively on the transfer ($Pun(T) = 3*T-2$), while the cost remains constant at 6 EMU making punishment at high rates of T relatively cheap. Hence it is not surprising to find a strong correlation between the frequency of punishment and the transfer level T (Spearman Rank correlation coefficient; $\rho_{p(WB)/T} = 0.62$; $p = 0.004$; N=63: observations from both WB treatments). The pattern of whistle-blowing in both SWT and LT along all 10 periods is shown in Figure 3.7.[43] Slightly higher punishment rates in early periods may

Figure 3.7: Probability of whistle-blowing in Stage 6



be due to the (strategic) signalling purpose of negative retaliation. The high punishment rate in the last period emphasizes the existence of pure retaliation (possibly motivated by fairness considerations) as a motive for defective behaviour, since punishment in the last period is unlikely to target reputation formation (Fehr and Gächter 2000, Bochet et al. 2006). Table 3.3 shows the average conditional punishment rates in the WB treatments. Rates are presented for both genders as well as separated for female and male participants.

As expected, female clients punish defection less often than their male counterparts (25%/33% compared to 55%/46% in SWT/LT). Only in SWT is the difference in punishment behaviour between male and female clients significant (female vs. male: $p = 0.052$; $N = 16$). This result further strengthens the hypothesis that the gender effect found in O's behaviour (**Hypothesis 5**) critically depends on differences in the anticipation of immediate and costly direct punishment, which is likely to be strongly related to the 'own' (gender specific) preferences towards punishment behaviour. Our evidence confirms

[43] A separation into SWT and LT is not informative because of the low number of total observations.

the findings in Lambsdorff and Frank (forthcoming) and supports the interpretation of the gender effect being mainly determined by differences in O's belief on B's behaviour with respect to direct (off-equilibrium) punishment of defective behaviour (**Hypothesis 4**). The differences in defection rates of officials and in actual punishment behaviour of clients between genders are also in line with the experimental findings of the literature on the gift exchange and trust game (see e.g. Croson and Gneezy 2009, Andreoni and Vesterlund 2001, Cox 2002). However, we find that men tend to be more reciprocal and women act more equitably only in the situation where a direct costly punishment device is present (as in SWT).

Table 3.3: **Gender and punishment**

Treatment	SWT		LT	
Whistling after Defection	0.40		0.38	
	female	male	female	male
	0.25	0.55	0.33	0.46
Acceptance after Defection	0.60		0.62	
	female	male	female	male
	0.75	0.45	0.67	0.54

All figures show averages of shares across subjects and periods for the relevant situation (O has chosen 'Nothing' in Stage 5)

The higher level in the application of costly punishment (and consequently the anticipation thereof) by male participants may be interpreted as a tendency of men to stick to their principles (Eckel and Grossman 1996). In our experiment this would mean that men consider bilateral reciprocation as their behavioural target even though this leads to aggregate inefficiencies and potentially unfair outcomes with respect to the distribution of payoffs. The gender difference in behaviour may also be explained by the finding that men are more inclined to maximize joint payoffs than women (efficiency seeking in Andreoni and Vesterlund 2001). This argument may only be conceded if we consider joint payoffs on (O-B) 'pair'-level alone, but not if we regard aggregated payoffs on session or treatment levels. Ignoring the behaviour of other pairs, punishment and the anticipation thereof leads to higher levels of corruption yielding larger payoffs for B

and O but reducing aggregated payoffs within a session (and thereby a treatment) by the resulting negative externality. In this sense, the argument of joint payoff maximization cannot be extended to the social dilemma of inefficient corruption. We leave the question on the particular reasons for the gender differences to future research.

3.5.4 Path-dependent behaviour

Decisions made in later stages of the game may depend on whether it was B or O who has initiated the corrupt transaction. For example B's behaviour of punishing defective behaviour (choose 'Whistle' in Stage 6 of the WB treatments) may depend on the 'voluntariness' of corrupt participation. Considering the rate of punishment, we do not find any indication of a difference between 'voluntary' and 'involuntary' participation when we control for transfer levels (separate U-tests in the four transfer categories, averaging all relevant observations on the individual subject level; 'voluntary' vs. 'involuntary': $p \geq 0.225$; $N \geq 12$). This contradicts the presumption that clients who reveal to be committed to choose the corrupt path, may be more inclined to punish defection. In line with our argumentation concerning the gender effect in corupt behaviour, our data provides clear evidence that the voluntariness of participation in corruption is independent of B's gender (the Spearman correlation coefficient between gender and voluntariness $\rho_{vol/gender} = 0.21$ is not significantly different from 0; $p = 0.543$, using observations of both WB treatments).

Those officials who have been able to choose whether or not to enter the corrupt transaction (passing Stage 2) may systematically differ in their willingness to reciprocate from those who are sent straight into Stage 4/5 (by B choosing 'Add' in Stage 1). Again we test for the difference in the mean probability of defection and reciprocation between the two sub-samples separately for the four transfer categories. The rates of defection and reciprocity do not differ significantly in either of the three treatments ('voluntary' vs. 'involuntary' within transfer categories: $p_{CT} \geq 0.311$; $p_{SWT} \geq 0.218$; $p_{LT} \geq 0.342$; $N \geq 12$). We conclude that O's reciprocity is not affected by the way Stage 4/5 has been reached other than (indirectly) by the effect on the size (category) of the transfer.

In the following we assess the effectiveness of corrupt and non-corrupt strategies for both types of players in terms of payoff maximization.

Payoff maximizing strategies (client)

While it is (by design) never socially optimal for the participants to engage in corruption, it may be individually optimal for B to initiate a corrupt transaction in Stage 1 (assuming self interested payoff maximization). To determine the effect of all relevant decisions on individual payoffs we run three linear random effects panel regressions separately for observations for the three treatments. For all three estimations we use the following specification.

$$\text{(M1)} \quad PP_{it} = \beta_0 + \beta_1 Add_{it} + \beta_2 Legal_{it} + \beta_3 Bribe_{it} + \beta_4 Add_{it} * Bribe_{it} + \epsilon_{it}$$

Index i stands for type B participants, index t for periods (t=1,2,...,10). We use individual 'period payoffs' (PP, measured in EMU) as the dependent variable. *Add* represents a dummy variable for B's decision whether or not to activate a corrupt transaction in Stage 1. *Legal* is a dummy variable for choosing 'Normal' in Stage 1 and *Bribe* is a variable measuring the amount of bribe (in EMU) transferred in Stage 3/4. The interaction term $Add * Bribe$ measures the difference in the marginal effect of an additional unit of bribe on the period payoff between situations in which bribing has been chosen by B in Stage 1 and situations in which B has been forced into the corrupt transaction (passing Stage 2). Table 3.4 reports the output of the three regressions.

Choosing the legal path ('Normal' in Stage 1) significantly increases period payoffs compared to the outside option (choosing 'Out' in Stage 1) only in the WB treatments (H_0: $\beta_2 = 0$: $p_{SWT} = 0.071$, $p_{LT} = 0.065$). The lack of a significant positive effect of a 'legal' choice in CT explains the high percentage of situations in which the (Pareto-dominated) outside option is chosen in this treatment (see Figure 3.3 in Section 3.5.2). In all treatments, initiating a corrupt transaction is worthwhile only in combination with a relatively large transfer. On average, it needs a bribe larger than 4.55 ($\frac{6.28}{1.38}$) EMU in CT, 2.32 ($\frac{2.69}{1.16}$) EMU in SWT and 4.41 ($\frac{3.00}{0.68}$) EMU in LT to yield an expected period payoff greater than the one obtained by choosing the outside option.

Table 3.4: **Output for random effects estimation (M1)**

	Dependent variable: PP					
	CT		SWT		LT	
	Coeffs	Std. dev	Coeffs	Std. dev	Coeffs	Std. dev
Constant	20.16***	1.47	20.00***	2.37	20.00***	4.07
Add	−6.28**	2.49	−2.69***	0.57	−3.00**	1.45
Legal	0.21	1.69	2.95*	1.59	3.04*	1.85
Bribe	1.38***	0.20	1.16***	0.21	0.68***	0.14
$Add * Bribe$	0.09	0.31	0.03	0.31	0.15	0.21
Overall R^2	$R^2 = 0.53$		$R^2 = 0.47$		$R^2 = 0.34$	
Number of Subjects	$N_i = 12$		$N_i = 21$		$N_i = 18$	

*** denotes significance at the 1%-level, ** denotes significance at the 5%-level, and * denotes significance at the 10%-level. Number of periods: 10

In terms of period payoff maximizing, bribing is not significantly more effective in situations in which the bribing stage has been reached voluntarily (β_4 ($Add * Bribe$) is not significantly different from 0 in any of the three treatments; $p \geq 0.284$). To check for robustness we proceeded in two dimensions. First, we included several combinations of variables of individual characteristics from the demographic data obtained in the questionnaire (e.g. dummy variables for B's and O's gender) and the variable $Bribe^2$ (the squared transfer level) in order to control for the possibility of diminishing marginal returns of the transfer on period payoffs. Second, we used two alternative models to estimate our regression. In addition to the linear random effects model we applied pooled OLS with dummy variables for periods (to account for dynamic effects) and a Tobit model to account for the censored structure of the dependent variable (PP is restricted to integers between 0 and 48 EMU). None of the alternative models, nor the inclusion of additional explanatory variables, yields qualitatively different results with respect to direction and significance of the coefficients reported in Table 3.4. We interpret the lack of a significant effect of the variable $Bribe^2$ in all specifications ($p > 0.378$) as strong evidence for the absence of a decreasing marginal effectiveness of the amount of transfer.[44]

[44] For reasons of parsimony we do not report detailed results of (any of) these estimations.

Payoff maximizing strategies (official)

If B chooses the legal path, a situation reaches Stage 2 and O has to decide between agreeing to the legal proposal ('Ok') and demanding a bribe ('Add'). In order to obtain a crude measure of the effectiveness of O initiating a corrupt transaction with respect to payoff maximization we calculate the average payoff generated by O's respective decisions in Stage 2. Using observations for which Stage 2 is reached (B has chosen 'Normal' in Stage 1), we run OLS regressions with period payoff (PP) as the dependent variable and a dummy variable which takes the value 1 if O initiates a corrupt transaction ('Add') and 0 for O's acceptance of the legal procedure ('Ok') as the only explanatory variable, separately for all three treatments.[45] The results for the OLS regressions (M2) to (M4) for observations in the respective treatments are shown in the following equations.

$$\text{CT (M2):} \quad \hat{PP_i} = 18.65^{***} + 5.00^{**} Add_i \quad (R^2 = 0.53;\ N_{obs} = 55;\ \text{Clusters} = 12)$$

$$\text{SWT (M3):} \quad \hat{PP_i} = 20.59^{***} - 4.89^{**} Add_i \quad (R^2 = 0.45;\ N_{obs} = 113;\ \text{Clusters} = 21)$$

$$\text{LT (M4):} \quad \hat{PP_i} = 20.13^{***} - 0.54 Add_i \quad (R^2 = 0.47;\ N_{obs} = 111;\ \text{Clusters} = 18)$$

*** stands for significance on the 1%−level, ** for significance on the 5%−level.

In CT, it pays off for O to force B into a corrupt transaction in Stage 2. On average, this yields additional 5 EMU per period compared to accepting the legal proposal ($\beta_1^{CT} = 5.00$, $p < 0.001$). In contrast, it is counterproductive in SWT (significantly negative effect of corrupt initiation) for O to initiate a corrupt deal in Stage 2 ($\beta_1^{SWT} = -4.89$, $p < 0.001$) and does not make any significant difference for O in LT ($\beta_1^{LT} = -0.51$, $p = 0.328$) with respect to payoff levels. We conclude that in our model of corruption, whistle-blowing (of both types) is effective in extinguishing the incentive for O to engage in corruption under rational expectations as to her monetary payoff.

[45]We use robust standard errors, clustered on subject level. As a robustness check we also ran respective specifications of the Tobit model to account for the censored dependent variable, which did not yield qualitative different results. So we did not report the results of these regressions.

3.6 Conclusion

The main objective of this paper is to evaluate the introduction of (various set-ups of) whistle-blowing with respect to its effect on the two negative consequences of corruption. We expand the standard game of corruption, which is often used as a vehicle to study individual behaviour in the experimental corruption research, in two ways. First, we provide subjects with a true 'legal' alternative to corruption and second, we allow both deciding agents, the client and the official, to choose to actively initiate a corrupt transaction. This allows us to consider not only the ex-post problem of a realized corrupt transaction causing damage to members of the public (who are not involved in the corrupt transaction), but also the ex-ante problem of corruption keeping 'honest' clients away from productive markets. Since whistle-blowing may have (opposing) effects on the two negative consequences of corruption, an analysis focusing on only one aspect (as in the standard game of corruption) may lead to the wrong conclusions. We show in a controlled laboratory experiment that our model is able to capture both consequences, enabling us to study the total effect of two institutions of whistle-blowing differing with respect to the symmetry of leniency.

Using three different treatments of a repeated version of an extended multi-stage corruption game we are able to show that there are two opposing effects of the introduction of symmetrically punished whistle-blowing. First, giving the client an opportunity to blow the whistle on an official who attempts to force her into a corrupt transaction leads to a higher number of legally proceeded successful transactions (the 'true' alternative to corruption which is first best in terms of aggregated payoffs) compared to the control treatment where whistle-blowing is not possible. This behaviour can be explained by arguments based on the structure of a payoff maximizing client's set of (realistic) first order beliefs on the (off-equilibrium) behaviour of the official. Moreover, we can show that whistle-blowing helps to decrease the attractiveness of the outside option. In this respect whistle-blowing serves as a client's safeguard against the exploitation through a corrupt official and thereby enables productive activity. Second, we find that the possibility of 'post-transaction' whistle-blowing by the briber (occurring after the decision of the official) can be misused as a tool to stabilize the corrupt transaction. Despite the personal costs, clients show a substantial amount of whistle-blowing in order to punish the official for not delivering the (implicitly) demanded corrupt service. The (realistic) anticipation of this behaviour leads to an increase in the success-rate of corrupt transactions (both conditional and un-conditional on reaching the relevant stage of the game). The second effect confirms experimental evidence from Lambsdorff and Frank (2010, forthcoming) and Abbink (2006). For a relevant range of transactions (i.e. for relatively small transfer levels) the possibility of direct retaliation is shown to matter despite the obvious relevance of strategic considerations.

We are able to show that the effect of stabilization is stronger for transactions involving a male official. Men punish defection more often than their female counterparts in the role of the client and act accordingly with respect to (revealed) anticipation of this behaviour in the role of the official. In the control treatment we do not find any significant gender effects. This provides further support for the hypothesis that the gender effect in corrupt behaviour found in a series of experimental and empirical studies (Lambsdorff and Frank 2010, Croson and Gneezy 2009) can be attributed to the reactions to the existence and anticipation of direct and costly negative retaliation (and not to patterns of general differences in preferences, e.g. risk aversion, Dekel and Scotchmer 1999). We demonstrate that asymmetric leniency can at least partly offset the stabilization effect, giving the official an opportunity to defect on the client while still benefiting from the bribe and being safe from the client's negative retaliation. Leniency for whistle-blowing can be used by the official as an 'insurance' against punishment after defection and undercuts the stabilization effect of the anticipation of punishment found under symmetric whistle-blowing.

Only under asymmetric leniency for the official does the introduction of whistle-blowing yield consistently positive effects with respect to all performance variables, the average level of payoff, the relative number of corrupt transactions and the relative number of successful legal transactions. Asymmetric leniency policies, providing incentives for defection while protecting the official from retaliation, can enhance the effectiveness of the institution of whistle-blowing, offering a strong argument for its use and consideration in the design of anti corruption laws. Our results relate strongly to the theoretical findings of Lambsdorff and Nell (2007) and Buccirossi and Spagnolo (2006), who advocate the effectiveness of well designed asymmetric punishment policies with respect to illegal activities. The findings with respect to behavioural gender differences in the reactions to whistle-blowing may serve to fine-tune these policies. In particular, asymmetric punishment for whistle-blowing may be especially effective for male officials which is in line with the argumentation of Lambsdorff and Frank (forthcoming) and Krajcova and Ortmann (2008).

Future theoretical as well as experimental corruption research with respect to the optimal design of institutions of whistle-blowing may be directed at taking heterogeneity of decision-makers into account. Especially for large private or public entities, decision-making agents involved in corruption may consist of a group rather than individuals. Exploiting the heterogeneity of individuals, asymmetrically punished whistle-blowing may not only serve as a tool to incentivize and secure defection between briber and official (as shown in our experiment) but also as a tool to take advantage of weaknesses in the corrupt complicity *between* individuals within a decision-making group.

3.7 Appendix 3

Appendix 3A: Proof of Equilibrium in the 6-Stage Game

We include the proof of **Proposition 2** only for the specification of LT. The proofs for the respective propositions in the treatment specifications of SWT and CT (e.g. **Proposition 1**) are similar in structure. The proof is by backward induction.

Denote by $I_{i,n}$ the information set in stage i of period n, where $i \epsilon \{1, 2, ..., 6\}$ and $n \epsilon \{1, 2, .., 10\}$. Let $p(I_{i,n})$ be the probability of reaching the respective stage and $q(\text{‘}s\text{’}|I_{i,n})$ the conditional probability of the relevant agent choosing ‘s’ once she has reached stage i, n having access to the respective information set.
An information set contains all relevant information about ego's and alter's decisions in all relevant stages up to the respective stage (i, n). Information sets $I_{1,n}$ $I_{3,n}$ $I_{4,n}$ and $I_{6,n}$ are relevant for type B, $I_{2,n}$ and $I_{5,n}$ for type O participants. Moreover, let $PO(\text{‘}Set\text{’})$ be the payoff by realizing the strategy set ‘Set’.
First, we show that there cannot be an equilibrium in which B chooses ‘Whistle’ in Stage 6 of the last (10^{th}) period.
Consider a strategy set $Set1 = [s_{1,1}, s_{2,1}, ..., s_{6,1}, s_{1,2}..., s_{6,10}]$ in which $p(I_{6,10}) > 0$ and $q_w = q(\text{‘}Whistle\text{’}|I_{6,10}) > 0$. Compare the expected payoff, resulting from this set ($PO(\text{‘}Set1\text{’})$) to that of an alternative which consists of equal strategies up to $I_{6,10}$ but for which $q(\text{‘}Whistle\text{’}|I_{6,10}) = 0$ and call it $PO(\text{‘}Set1_0\text{’})$. Irrespective of the history up to $I_{6,10}$, it is better for B not to blow the whistle since $q_w(14-T)+(1-q_w)(21-T) < 21-T$ (for the case O has chosen ‘Nothing’) and $q_w(14 - T) + (1 - q_w)(48 - T) < 48 - T$ (for the case O has chosen ‘Prefer’), hence $PO(Set1) < PO(\text{‘}Set1_0\text{’})$, so that $Set1$ cannot constitute an equilibrium (Sub game perfect Nash Equilibrium). A rational B will never play ‘Whistle’ in Stage 6.
Second, we show that, in the last (10^{th}) period, O will never choose ‘Prefer’ in Stage 5.
Consider a strategy set $Set2 = [s_{1,1}, s_{2,1}, ..., s_{6,9}, s_{1,10}..., s_{6,10}]$ in which $p(I_{5,10}) > 0$, $q(\text{‘}Whistle\text{’}|I_{6,10}) = 0$ (as shown above) and $q(\text{‘}Prefer\text{’}|I_{5,10}) > 0$.
Again, compare $PO(Set2)$ to $PO(Set2_0)$ which differs from the former only in $q(\text{‘}Prefer\text{’}|I_{5,10}) = 0$. Since $3*T < 6+3*T$, the expected payoff $PO(Set2) < PO(Set2_0)$,

if probability mass $q(\text{'}Prefer\text{'}|I_{5,10})$ is shifted to $q(\text{'}Nothing\text{'}|I_{5,10}) > 0$.[46] Hence $Set2$ cannot be an equilibrium.

It is easy to show by similar argumentation that any strategy set exhibiting $q(\text{'}T = 0\text{'}|I_{4,10}) = 0$, $q(\text{'}Prefer\text{'}|I_{5,10}) = 0$, $q(\text{'}Whistle\text{'}|I_{5,10}) > 0$ or $q(\text{'}Out\text{'}|I_{5,10}) > 0$ cannot constitute an equilibrium when compared to the expected payoff of a strategy set that is similar except that it excludes the possibilities of 'Whistle' and 'Out' choices in Stage 5 of the 10^{th} period. If T=0, the expected payoff from any of these strategy sets is strictly smaller than that of a strategy set being similar but exhibiting $q(\text{'}Nothing\text{'}|I_{5,10}) = 1$

Third, it is straightforward to see that any strategy set for which $q(\text{'}T = 1\text{'}|I_{4,10}) < 1$, $(q(\text{'}Whistle\text{'}|I_{5,10}) = 0, q(\text{'}Out\text{'}|I_{5,10}) = 0, q(\text{'}Whistle\text{'}|I_{6,10}) = 0)$ is strictly dominated by a strategy set with the same characteristics except $q(\text{'}T = 1\text{'}|I_{4,10}) = 1$. This is because setting $T > 1$ wastes B's payoff as it leads to the same behaviour of O. $T = 0$ leads to a lower expected payoff since in this case $q(\text{'}Out\text{'}|I_{5,10}) = 1$ and $14 < 20 - T = 19$.

In Stage 3, we cannot exclude any strategy set given the above argumentation since B will be indifferent between strategy sets that differ only in $q(\text{'}Add\text{'}|I_{3,10})$ (if $p(I_{3,10}) > 0$) since both actions 'Add' or 'Whistle'lead to an expected payoff of 19 EMU.

It is clear that an equilibrium exhibiting $p(I_{2,10}) > 0$ has to include the acceptance of O, $q(\text{'}Ok\text{'}|I_{2,10}) = 1$ since this yields strictly greater payoffs for O under the equilibrium properties derived above $(15 > 9)$.

Choosing 'Normal' for any information set $I_{1,10}$, setting $q(\text{'}Normal\text{'}|I_{1,10}) = 1$, leads to the maximum payoff for B.

An equilibrium Strategy must therefore fulfill the following characteristics of actions in the last period: $q(\text{'}Whistle\text{'}|I_{6,10}) = 0$ if $p(I_{6,10}) > 0$, $q(\text{'}Nothing\text{'}|I_{5,10}) = 1$ if $p(I_{5,10}) > 0$, $q(\text{'}T = 1\text{'}|I_{4,10}) = 1$ if $p(I_{4,10}) > 0$, $q(\text{'}Add\text{'}|I_{3,10}) = [0, 1]$ if $p(I_{3,10}) > 0$, $q(\text{'}Ok\text{'}|I_{2,10}) = 1$ and $q(\text{'}Normal\text{'}|I_{1,10}) = 1$. We call this set of actions the Stage Game Equilibrium. Hence $Set_{equ} = [s_{1,1}, s_{2,1}, ..., s_{6,9}, \text{'}Normal\text{'}, \text{'}Ok\text{'}, \text{'}Whistle\text{'}/\text{'}Ok\text{'}, \text{'}T = 1\text{'}, \text{'}Nothing\text{'}, \text{'}Ok\text{'}]$.

Consider a period-set $PS = \{k, ...9\}$ of the last $10 - k$ consecutive periods for which the Stage Game Equilibrium of the last (10^{th}) period is played and regard period $k - 1$. By the same line of arguments for the equilibrium properties of the Stage Game Equilibrium,

[46]This is because $q(\text{'}Whistle\text{'}|I_{5,10}) * 8 + q(\text{'}Out\text{'}|I_{5,10}) * 8 + q(\text{'}Nothing\text{'}|I_{5,10}) * (6 + 3T) + (1 - q(\text{'}Whistle\text{'}|I_{5,10}) - q(\text{'}Out\text{'}|I_{5,10}) - q(\text{'}Nothing\text{'}|I_{5,10})) * 3T < q(\text{'}Whistle\text{'}|I_{5,10}) * 8 + q(\text{'}Out\text{'}|I_{5,10}) * 8 + (1 - q(\text{'}Whistle\text{'}|I_{5,10}) - q(\text{'}Out\text{'}|I_{5,10})) * (6 + 3T)$.

we can repeat excluding all sets of strategies in period $k-1$ that do not exhibit the strategy characteristics of the Stage Game Equilibrium.
Letting k decrease from 9 one by one until it reaches 1, it is easy to see that the Stage Game Equilibrium of the last period remains the only Sub game perfect Nash Equilibrium (SNE) in the finitely repeated game.

Appendix 3B: Clarifications to hypotheses

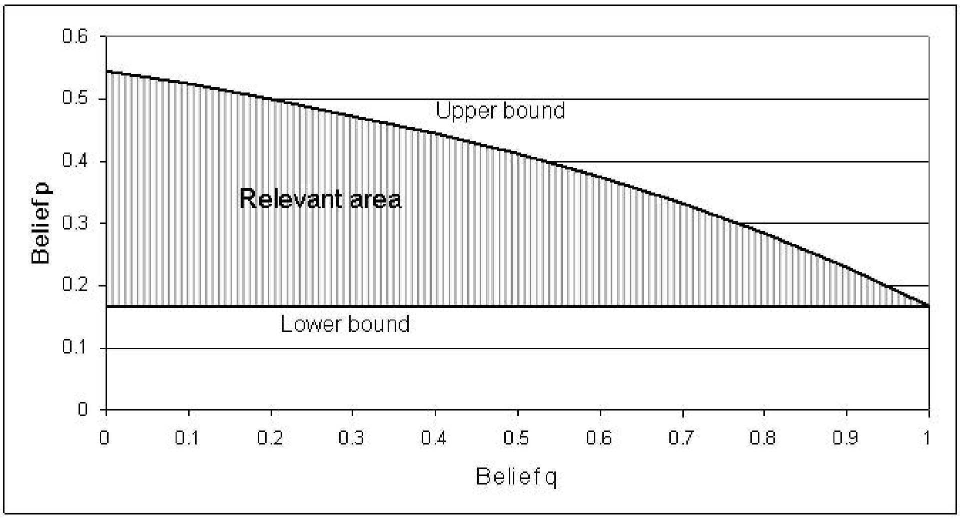
Hypothesis 1

The following argument in favour of a positive effect of giving B an opportunity to blow the whistle builds on the consideration of payoff maximizing strategies under a reasonable set of first order beliefs derived by induction. Consider first the considerations of B in the CT. Departing from the actions predicted in the SNE, a risk neutral payoff maximizing client B will (in the last period of the repeated game) choose 'Out' in Stage 1 if her beliefs about O's behaviour are structured such that her expected payoff from choosing 'Normal' is smaller than 20 EMU, the value of the outside option.
If she chooses 'Normal' in Stage 1 O may her (by choosing 'Add') to enter Stage 3. Not aiming at initiating corrupt reciprocity, B would either choose '$T=0$', in which case she would (most likely) end up with 14 EMU, or '$T=1$', in which case she would expect to get 19 EMU with some probability (q) and 14 EMU with the residual probability $(1-q)$. A belief of $q<1$ is justified by B expecting O to punish a low transfer at relatively low costs. We assume that B's belief on the probability of O choosing 'Prefer' after receiving a low transfer is 0. Hence, B's expected payoff from choosing 'Normal' in Stage 1 (when choosing a low transfer in Stage 3) is $EP(\text{`}Normal'|T \leq 1) = 25p + (19q + 14(1-q)) * (1-p) = 25p + EP(Bribe) * (1-p)$, where p is B's belief on the probability that O will choose 'Ok' in Stage 2 and $EP(Bribe)$ is B's expected payoff after leaving Stage 2.
The higher the probability of O choosing 'Out' in Stage 4 $(1-q)$, the lower $EP(Bribe)$ and hence the lower the belief of the probability of O choosing 'Add' $(1-p)$ in Stage 2 needs to be in order to cause B to choose 'Out' in Stage 1 of CT, leaving $EP(\text{`}Normal'|T \leq 1) < 20$ EMU. In CT B's belief on O's Stage 2 honesty ('Ok') must be $p_{CT} \leq \frac{6-5q}{11-5q}$. Hence those B participants characterized by the belief structure defined above will choose 'Out' instead

of 'Normal' in Stage 1.

When introducing whistle-blowing (SWT or LT), the expected payoff from choosing 'Out' remains 20 EMU, while the expected payoff after leaving Stage 2 is at least 19 EMU, $EP(Bribe)_{WB} = min[EP(Bribe); 19] = 19$ EMU, since B can always blow the whistle in Stage 3. Hence participants holding the same belief structures as described above, will choose 'Out' in the WB treatments only if $20 > 25p + 19(1-p)$, hence $p_{WB} \leq \frac{1}{6}$. The larger the belief on the probability of O punishing a low T by choosing 'Out' $(1-q)$, the larger the difference between the expected payoff after leaving Stage 2 in the CT and the WB treatments. Hence participants who hold belief structures about O's Stage 2 honesty that satisfy $P^*_{WB} = \frac{1}{6} < p < \frac{6-5q}{11-5q} = p^*_{CT}$ will choose 'Out' in CT but 'Normal' in the WB treatments. The relationship between the believed probabilities p and q is illustrated in Figure 3.8. Especially for low values of q (B expecting O to punish a low transfer with a high probability) there is a wide range of values for p ('relevant area') for which we would expect a treatment difference according to Hypothesis 1.[47]

Figure 3.8: Relation between p and q



The opportunity to blow the whistle in Stage 3 allows B to bail on a corrupt transaction initiated by O at a small cost. By the same line of arguments, O may, in the WB

[47]Note that the arguments is at least valid for the last period as it implicitly assumes a restrictive set of first order beliefs that prevents the individual from considering future payoff through corrupt reciprocity and reputation-building. According to the argument of backward induction, however the argument should go through to earlier periods.

treatments, anticipate B's willingness to blow the whistle in Stage 3 (if forced to enter the corrupt transaction). The higher the believed probability of such behaviour, the less likely it is that O will try to initiate a corrupt transaction in Stage 2. Hence O will choose 'Ok' with a higher probability in Stage 2 of the WB treatments ($p_{WB} > p_{CT}$). A cascade of belief-anticipation and updating would yield expectations of even stronger treatment effects in O's and B's Stage 1 and 2 behaviour.

Hypothesis 2

We can calculate the critical belief (conditional on the level of transfer) that causes O to choose 'Prefer' rather than 'Normal' for a bribe $T > 1$ in Stage 5 (For $T = 0$ the choice of 'Out' would be always preferable to O).

Let $EP(\text{`}Prefer'|H_{I_5}) = 3T$ and $EP(\text{`}Nothing'|H_{I_5}) = (1 - r(T)) * (6 + 3T) + r(T) * 8$ be the expected payoffs resulting from O's Stage 5 choices, given a certain history H_{I_5} which contains the information of all relevant choices up to Stage 5. An official will deliver the corrupt task (choose 'Prefer') if:

$EP(\text{`}Prefer'|H_{I_5}) > EP(\text{`}Nothing'|H_{I_5})$, hence $3T > (1 - r(T)) * (6 + 3T) + r(T) * 8$.

Solving for $r^*(T)$, the critical belief above for which O will deliver, yields $r(T)^* \geq \frac{6}{3T-2}$, which is strictly decreasing in T ($\frac{\partial r(T)^*}{\partial T} = -\frac{18}{(3T-2)^2} > 0$).

This means, the larger the transfer T, the lower the critical belief $r^*(T)$ and hence the more likely O will reciprocate in the transaction because of the immediate threat of retaliation.

For bribes $T \geq 3$ EMU, ('Out' and 'Whistle' are strictly dominated by 'Prefer'), those subjects of type O who hold respective beliefs with respect to the punishment probability will choose 'Prefer' instead of 'Nothing' in the SWT (fearing B's retaliation). By the argument of backward induction this argument is valid for all earlier periods as well.

Appendix 3C: General Outcomes

Table 3.5: **All relevant Variables**

	CT	SWT	LT
‘Out’ in Stage 1	2.25	0.53	0.33
‘Normal’ in Stage 1	4.63	5.71	6.17
Conditional ‘Ok’ in Stage 2	38%	61%	55%
Conditional ‘Whistle’ in Stage 3	-	22%	22%
Reaching Stage 3/4	60%	49%	48%
‘T’	4.18	4.67	4.74
Conditional ‘T’	6.77	8.61	8.71
Whistling after Defection	-	40%	38%
Whistling after Success	-	0	0
Total Corrupt Success	2.83	3.57	2.37
Total Legal Success	1.75	3.48	3.39
Total Failures	5.42	2.95	4.24
Payoff	13.04 Euros	14.83 Euros	16.08 Euros

All figures are averages across subjects and periods within treatments

Appendix 3D: Defection across genders

The following tables show gender differences with respect to defection in CT, SWT and LT. Rates of defection are conditional on reaching the bribing stage (Stage 4/5).

Defection in CT

	Defection	No Defection	
Female	0.24	0.31	0.55
Male	0.20	0.25	0.45
	0.44	0.56	1

In CT male as well as female officials defect in 44% of cases.

Defection in SWT

	Defection	No Defection	
Female	0.12	0.20	0.32
Male	0.07	0.61	0.68
	0.19	0.81	1

In SWT female officials defect in 38% of cases compared to 10% in the male population of officials.

Defection in LT

	Defection	No Defection	
Female	0.11	0.29	0.40
Male	0.04	0.56	0.60
	0.15	0.85	1

In LT female officials defect in 28% of cases compared to 7% in the male population of officials.

Appendix 3E: Instructions for the SWT treatment (translated from German)

Thank you very much for your appearance. In the next 90 minutes you will take part in an experiment in the laboratory of MELESSA.

If you read the following instructions carefully, you can (depending on your decisions and the decisions of other participants of the experiment) earn money, additional to the show-up fee of 4 Euros. The money you will earn during the experiment will be added to the show-up fee and paid out in cash at the end of the experiment.

During the experiment you are not allowed to communicate with the other participants. If you have questions, please approach one of the experimenters by raising your hand. In the case of violation of this rule we have to exclude you from any payments.

During the experiment we will refer to Experimental Monetary Units (EMU) instead of Euros. Your income will be calculated in EMU. At the end of the experiment the total amount will be exchanged in Euros.

The Exchange rate is **1 EMU = 6 Eurocents**.

All 24 participants are distributed into **groups of two**. **Neither** the experimenters **nor** the other participants **know** which group you will be in. Your decisions remain **completely anonymous**.

The Decision Situation

There are two types in this experiment: type A and type B. The types play different roles and make decisions that affect their own income and potentially the income of the other participants of the experiment. There will be as many type A participants as type B participants. The type of a participant is allocated **randomly**. The probability to play the role of B is therefore equal to the probability of being an A type.

A group of two consists of one type A and one type B participant. The members of a group stay together for the entire duration of the experiment. The experiment consists of 10 periods.

Procedure:

Each of the 10 periods has at most 6 Stages:

Stage 1: In the first Stage each participant of type A has 3 Alternatives.

If he/she chooses End the period ends. In this case, type A gets 20 EMU, and type B gets 10 EMU, and 6 randomly chosen participants of the experiment get 2 EMU each.

If he/she chooses Add, Stages 2 and 3 are skipped.

If he/she chooses Normal, Stage 2 is reached.

Stage 2: In Stage 2 (which is only reached if type A has chosen Normal in Stage 2) type B can choose between the alternatives OK and Add.
If type B chooses OK, the period ends and type A receives 25 EMU, type B 15 EMU, 6 randomly chosen participants of the experiment get 3 EMU each.
If type B chooses Add, Stage 3 is reached.

Stage 3: In Stage 3 (which is only reached in case type B has chosen Add in Stage 2) type A can choose between Report und Add.
If he chooses Report, the period ends, and type A receives 19 EMU and type B 9 EMU.
If he chooses Add, Stage 4 is reached.

Stage 4: In Stage 4 (which is only reached in case type A has chosen Add in Stage 3) type A chooses the level of Transfer T (which has to be an integer between 0 and 20 EMU).

Stage 5:In Stage 5 type B learns the level of transfer T of type A and has four alternatives:
Alternative 1: Type B chooses End. In this case type A receives 14 EMU and type B 8 EMU for this period.
Alternative 2: Type B chooses Report. In this case type A receives $14 - T$ and type B: 8 EMU for this period.
Alternative 3: Type B decides for Nothing. In this case Stage 6 is reached.
Alternative 4: Type B chooses Prefer. In this case Stage 6 is reached.

Stage 6: In Stage 6 it is again type A who decides. Note that this stage is only reached in case type B has chosen Nothing or Prefer in Stage 5. In both cases type A has 2 options: Report or OK.

1^{st} **Case**: Type B has chosen alternative **Nothing** in Stage 5:
A) If type A chooses Report, he receives 14 EMU less the transfer T: ($14 - T$ EMU), type B receives 6 EMU, and 6 randomly chosen participants of the experiment get 2 EMU each.
B) If type A chooses OK, type A receives 21 EMU less the transfers T: ($21 - T$ EMU), type B receives 6 EMU in addition to the tripled value of T ($6 + 3 * T$ EMU). 6 randomly chosen participants of the experiment get 2 EMU each.

2^{nd} **Case**: Type B has chosen alternative **Prefer** in Stage 5:
A) If type A chooses Report, he receives 14 EMU less the transfer T: ($14 - T$ EMU), type B receives 6 EMU, and 6 randomly chosen participants of the experiment get 2 EMU each.
B) If type A chooses OK, type A receives 48 EMU less the transfers T: ($48 - T$ EMU), type B receives the tripled value of T ($3 * T$ EMU). The payment of 6 randomly chosen participants of the experiment is reduced by 8 EMU.

Example

The following example will help you to better understand the situation.

Assume type A chooses Normal in Stage 1 of the experiment. This means we consider the middle cell of the first line in the payment table (see the last page of the instructions).

Now we consider those cells that are directly beneath the chosen cell: In the second Stage (represented by the second line) type B decides. We see that he has 2 alternatives: OK and Add. Assume that he chooses Add.

The cells lying directly beneath this choice (in line 3) are the alternatives Report and Add for type A. Assume again, type A chooses Add. This means that we reach the fourth Stage in which type A decides about the Transfers T. Assume type A chooses a transfer of $T = 10$ EMU.

In Stage 5 type B learns about the transfer (sees it on his screen) and decides between four alternatives: End, Report, Nothing und Prefer. Assume he chooses Prefer which means we reach Stage 6.

In Stage 6 type A decides between Report and OK. Assume he/she chooses OK.

Then the period ends and we can infer from the payment table what the payment of the respective participant is:

Type A receives: $48 - T$. Since T is 10, his income is $48 - 10 = 38$ EMU

Type B receives: $3 * T$. Since T is 10 EMU his income amounts to: $3 * 10 = 30$ EMU

The last line contains the income generated through the situation for 6 randomly chosen participants of the experiment. In our example their total income is **reduced** by 8 EMU each.

Timing

Stage 1: Type A chooses between End, Normal and Add.

Stage 2: Is only reached if type A has chosen Normal in Stage 1. Type B chooses between OK and Add.

Stage 3: Is only reached if type B has chosen Add in Stage 2: Type A decides between Report und Add.

Stage 4: Is only reached if type A has chosen Add in Stage 1 or in Stage 3. Type A chooses Transfer T (between 0 and 20 EMU).

Stage 5: Type B learns the value of Transfer T and decides between End, Report, Nothing and Prefer.

Stage 6: Is only reached if type B has chosen Nothing or Prefer in Stage 5. In these cases type A decides between Report and OK

At the end of each period every participant gets to know his/her income in this period. Note that type A as well as type B can infer from this information what the decision of the other participant was. Note that this income does not include the additional payments or reductions potentially caused by the decisions of other participants of the experiment.

The following **control questions** will help you to better understand the situation. Please answer all control questions and raise your hand when you have finished. An experimenter will come to your place and check your solutions.

Question 1

Assume that **you are type A**. In Stage 1 you have chosen **Add**. Therefore **Stage 2 and 3 were skipped**, and you proceeded with Stage 4, where you chose a transfer of **3 EMU** (T is 3). In Stage 5 type B has chosen **Nothing** and you chose **OK** in Stage 6.

a) What is **your income** (type A) in this period (without additional payments or reductions of income caused by other participants of the experiment that you cannot know)?

Your answer:___________________________

b) What is the **sum of additional payments or reductions** of income caused by your (and your partner's of type B) decisions for 6 randomly chosen participants of the experiment?

Your answer:___________________________

c) What is the **income of type B** in this period (without additional payment or reductions of income caused by other participants of the experiment that you cannot know)?

Your answer:___________________________

d) What is the **sum of all payments** caused by your (and your partner's of type B) decisions?

Your answer:___________________________

Question 2

Assume that you are **type B**. In Stage 1 type A has chosen **Normal**. In Stage 2 you have chosen **Add**. Type A chose **Add** in Stage 3 and a transfer of **0**. (T is 0). Then you chose **End** in Stage 5. a) What is **your income** (type B) in this period (without additional payment or reductions of income caused by other participants of the experiment that you cannot know)?

Your answer:___________________________

b) What is the **sum of additional payments or reductions** of income caused by your (and your partner's of type A) decisions for 6 randomly chosen participants of the experiment?

Your answer:___________________________

c) What is the **income of type A** in this period (without additional payment or reductions of income caused by other participants of the experiment that you cannot know)?

Your answer:___________________________

d) What is the **sum of all payments** caused by your (and your partner's of type A) decisions?

Your answer:___________________________

Question 3

Assume that you are **type B**. In Stage 1 type A has chosen **Add**. **Stage 2 and 3** are therefore **skipped**. In Stage 4 type A chooses a transfer of **12 EMU** (T is 12). You chose **Prefer** in Stage 5. Type A chose **OK** in Stage 6.

a) What is **your income** (type B) in this period (without additional payment or reductions of income caused by other participants of the experiment that you cannot know)?

Your answer:___________________________

b) What is the **sum of additional payments or reductions** of income caused by your (and your partner's of type A) decisions for 6 randomly chosen participants of the experiment?

Your answer:___________________________

c) What is the **income of type A** in this period (without additional payment or reductions of income caused by other participants of the experiment that you cannot know)?

Your answer:___________________________

d) What is the **sum of all payments** caused by your (and your partner's of type A) decisions?

Your answer:___________________________

Question 4

Assume that you are **type A**. You chose **Normal** in Stage 1. In Stage 2 type B decides for **OK**. **Stage 3, 4, 5 and 6** are therefore **not reached**.

a) What is **your income** (type A) in this period (without additional payment or reductions of income caused by other participants of the experiment that you cannot know)?

Your answer:___________________________

b) What is the **sum of additional payments or reductions** of income caused by your (and your partner's of type B) decisions for 6 randomly chosen participants of the experiment?

Your answer:___________________________

c) What is the **income of type B** in this period (without additional payment or reductions of income caused by other participants of the experiment that you cannot know)?

Your answer:___________________________

d) What is the **sum of all payments** caused by your (and your partner's of type B) decisions?

Your answer:___________________________

Payment table

This table can be read as follows. Generally we always start from the top and proceed downwards, cell by cell. If a participant chooses a specific alternative, only those cells that are located directly beneath this cell are relevant in the next stage (represented by the next line of cells).

<table>
<tr><td>1. Stage</td><td>Type A chooses End</td><td colspan="3">Type A chooses Normal</td><td colspan="5">Type A chooses Add</td></tr>
<tr><td>2. Stage</td><td>Stage 2 is not reached</td><td>Type B chooses OK</td><td colspan="2">Type B chooses Add</td><td colspan="5">Stage 2 is not played, go to Stage 4</td></tr>
<tr><td>3. Stage</td><td colspan="2">Stage 3 is not reached</td><td>Type A chooses Report</td><td>Type A chooses Add</td><td colspan="5">Stage 3 is not played, go to Stage 4</td></tr>
<tr><td>4. Stage</td><td colspan="3">Stage 4 is not reached</td><td colspan="6">Type A chooses the amount of Transfer T</td></tr>
<tr><td>5. Stage</td><td colspan="3">Stage 5 is not reached</td><td>Type B chooses End</td><td>Type B chooses Report</td><td colspan="2">Type B chooses Nothing</td><td colspan="2">Type B chooses Add</td></tr>
<tr><td>6. Stage</td><td colspan="5">Stage 6 is not reached</td><td>Type A chooses Report</td><td>Type A chooses OK</td><td>Type A chooses Report</td><td>Type A chooses OK</td></tr>
<tr><td>Income of Type A</td><td>20</td><td>25</td><td>19</td><td>14</td><td>14 - T</td><td>14 - T</td><td>20 - T</td><td>14 - T</td><td>48 - T</td></tr>
<tr><td>Income of Type B</td><td>10</td><td>15</td><td>9</td><td>8</td><td>8</td><td>6</td><td>6 + 3*T</td><td>0</td><td>3*T</td></tr>
<tr><td>Additional income/reduction in income of 6 randomly chosen participants of the experiment</td><td>6*2</td><td>6*3</td><td>6*2</td><td>6*2</td><td>6*2</td><td>6*2</td><td>6*2</td><td>6*2</td><td>6*(-8) (reduction)</td></tr>
</table>

Chapter 4

Cooperation with Uncertain Endowments

4.1 Introduction

It has long been the goal of numerous public goods experiments to identify factors that increase or decrease cooperation among individuals. It is important to know these factors in order to set up institutions that can enhance cooperative behaviour. Several factors have been analyzed that try to capture a more realistic environment of situations where cooperation is important, like heterogeneous endowments, varying group size, communication, or punishment. Surprisingly, there is little evidence how uncertainty about endowments influences cooperative behaviour. One can think of many situations where cooperation is important and where individuals usually do not know the endowment of others. Charitable giving, for example, usually occurs without any knowledge about the income of those who donated, but donations become known.[1] Team work often allows the observation of contributions by individual team members while these do not know how much disposable time each of the other team members actually has.

Making endowments uncertain necessarily involves the possibility of heterogeneous in addition to homogeneous endowments. In contrast to uncertain endowments, the literature on heterogeneous endowments is relatively large.[2] Besides investigating the effects

[1] Gächter (2007) lists campaigns where it was practice to reveal the donated amounts of all donators.

[2] Studies investigating the effects of heterogeneous endowments include Marwell and Ames (1979), Aquino et al. (1992), Rapoport and Suleiman (1993), van Dijk and Wilke (1994), Ledyard (1995), Carde-

of heterogeneous *and* uncertain endowments on contributions in a repeated linear public goods game, we further explain these effects with conditional cooperation preferences[3] elicited separately. Our experimental design builds upon the design of Fischbacher and Gächter (2010). They used conditional cooperation preferences elicited in the so-called P-Experiment to explain declining cooperation in a repeated linear public goods game, the so-called C-Experiment. We take their setup to a more realistic environment by investigating and explaining cooperative behaviour under heterogeneous and uncertain endowments. Our C-Experiment consists of two treatments, the certainty and the uncertainty treatment. In both treatments subjects played a repeated linear public goods game. There were homogeneous as well as heterogeneous groups. While subjects know others' endowments in the certainty treatment they know neither the others' endowments nor the total group endowment in the uncertainty treatment. In our P-Experiment we ask subjects (using the strategy method) to indicate their own contribution for all combinations of others' contributions and endowments. Besides explaining our results from the C-Experiment, our P-Experiment allows us to investigate open questions concerning conditional cooperation preferences.[4]

Overall, we find only a small effect of endowment uncertainty on cooperation in the C-Experiment. Contributions decrease slightly when endowments are uncertain. However, we further show that this is due to two opposing effects that partly cancel each other. Low endowed individuals (L types) contribute substantially less to the public good and high endowed individuals (H types) slightly more under endowment uncertainty. This behaviour causes groups with a majority of L types to lose and groups with a majority of H types to gain under uncertainty. While homogeneous groups are more efficient than heterogeneous groups under certainty (which is consistent with the existent literature), under uncertainty this relationship breaks down, and we find that a group becomes more

nas (2003), Zelmer (2003), Cherry et al. (2005), Buckley and Croson (2006), and Anderson et al. (2008). Most of these studies indicate that making endowments heterogeneous decreases cooperation. The only studies we are aware of that consider the effect of uncertain endowments (although with a focus different from ours) are Isaac and Walker (1988), van Dijk and Grodzka (1992), Chan et al. (1999), and Levati et al. (2007). We discuss how these studies relate to ours at the end of the paper.

[3]Conditional cooperation describes the preference to contribute to the public good *if* others contribute as well. Gächter (2007) provides an overview to this literature.

[4]With the design of our P-Experiment we can first answer the question whether the assumption made in all other experiments that investigate conditional cooperation preferences, i.e. the assumption that subjects condition their own contribution on the *average* of others' contribution, can be justified. The second question we investigate (which is more important for our results) is whether subjects condition their own contribution on others' relative or absolute contributions.

efficient the richer it is. Although there is almost no effect of uncertainty on the aggregate level, uncertainty causes the inequality of the income distribution to increase.

We further explain these treatment effects with individual preferences elicited in the P-Experiment. The existing literature on conditional cooperation preferences showed that subjects are on average imperfect conditional cooperators who match others' contributions with a self-serving bias. Nothing has been said on whether they match absolute or relative contributions. With homogeneous endowments these two forms of conditional cooperation coincide. However, with heterogeneous endowments it becomes important whether subjects are on average *absolute* or *relative* conditional cooperators. We show that subjects are imperfect *relative* conditional cooperators who match others' *relative* contributions with a self-serving bias. It is exactly this preference that can explain our treatment effects in the repeated public goods game played in the C-Experiment in the following two ways.

A direct effect of a preference for relative conditional cooperation is that it implies that subjects contribute more in absolute terms the poorer their group members are (holding their absolute contributions constant). Thus, under uncertainty it becomes important whether subjects *believe* their group members are poorer or richer than they actually are. In the C-Experiment we elicited beliefs about both others' contributions (in both treatments) and others' endowments (only in the uncertainty treatment). Subjects in poor groups think they are in a richer group than they actually are (since half the subjects were L types and the other half H types) and therefore give less than their counterparts under certainty. Subjects in rich groups believe to be in a poorer group than they actually are and therefore give more than under certainty. This explains why poor groups lose and rich groups gain through uncertainty. Moreover, in rich groups subjects can better update their beliefs over time since they get informative signals. This is one reason why poor groups suffer more than rich groups benefit.

Similar to previous studies we observe (strategic) over-contribution[5] in the repeated public goods game with homogeneous endowments under certainty. There is, however,

[5]Over-contribution occurs when the actual contribution in the public goods experiment (the C-Experiment) exceeds the predicted contribution that is derived through the individual preferences (elicited in the P-Experiment) and beliefs about others' contributions and endowments (elicited in the C-Experiment). Similar to Fischbacher and Gächter (2010), we observe over-contribution under certainty for homogeneous groups.

no over-contribution observed in heterogeneous groups, neither under certainty (which explains why homogeneous groups are more efficient than heterogeneous groups) nor under uncertainty. For homogeneous groups under uncertainty we find that only rich groups still over-contribute. Poor groups do not over-contribute under uncertainty. A possible reason for this observation and indirect effect of a relative conditional cooperation preference is that L types may anticipate that others are relative conditional cooperators. Given this anticipation they fear sending signals that may misidentify them as H types and hence drive down contributions by others. So, L types will not over-contribute under uncertainty because they anticipate that others are relative conditional cooperators as well. H types do not stop over-contributing under uncertainty in order to send L type signals. Perhaps surprisingly, this suggests that H types do not mimic being L types under uncertainty and thus do not exploit their private information. Since poorer groups consist of more L types, this is another reason why only poor groups lose through the uncertainty.

Both effects of a preference for relative conditional cooperation, the direct effect caused by the deviation of beliefs from reality and the indirect effect due to the absence of strategic over-contribution in poor groups, indicate why poor groups lose more than rich groups gain through endowment uncertainty (leading to the small negative overall treatment effect). We further show that these two effects can explain a substantial part of our quantitative treatment effects from the C-Experiment.

The paper is structured as follows. Section 4.2 describes our experimental design. In Section 4.3 we discuss our experimental results. In Section 4.3.1 we present the results of the C-Experiment showing what our treatment effects are and in Section 4.3.2 we explain these treatment effects and present the results from the P-Experiment. In Section 4.4 we relate our results to the relevant literature and conclude.

4.2 The Experiment

The experiment was computer-based and was conducted at the experimental laboratory MELESSA of the University of Munich. It used the experimental software z-Tree (Fischbacher 2007) and the organizational software Orsee (Greiner 2004). 180 subjects (graduate students were excluded) participated in 8 sessions and earned 14.76 Euros (including 4 Euros show-up fee) on average (with a minimum of 10.14 Euros and a maximum of 22.56 Euros). The experiment took approximately 90 minutes. At the beginning of the experiment subjects received written instructions that were read privately by them.[6] At the end of these instructions they had to answer test questions that showed whether everything was understood. There was no time limit for the instructions, and subjects had the opportunity to ask questions in private. The experiment started on the computer screen only after everybody had answered the test questions correctly and there were no further questions.

Contributions in the public goods game were modeled through a standard linear voluntary contribution mechanism (VCM). In each period, the payoff of player i was given by

$$\Pi_i = E_i^k - g_i + 0.6 \sum_{j=1}^{n=3} g_j \tag{4.1}$$

where E_i^k is player i's endowment, g_i her individual contribution, n is the number of group members (with $i, j = \{1, 2, 3\}$), and 0.6 is the efficiency factor that determines how valuable the public good is. Under standard selfish preferences subjects would choose $g_i = 0$, whereas $g_i = E_i^k$ is the social optimum.

In each session of the experiment, half of the subjects had a low endowment of 4 EMU (Experimental Monetary Units, where the exchange rate was such that 1 EMU corresponded to 0.10 Euros) and the other half of the subjects was high endowed with 8 EMU. Hence, $k = \{L, H\}$ with $E_i^L = 4$ and $E_i^H = 8$. All groups in the experiment consisted of three subjects and thus $n = 3$. Endowments were randomly assigned to subjects (via the instructions[7]) and everybody knew that each subject had the same chance to be low or high endowed, since they knew that half of the subjects in their

[6]The instructions of the experiment can be found in Appendix 4B.

[7]The instructions were given in three sets. The first set of instructions at the beginning of the experiment gave information about the general setup and how the VCM worked. The second set of instructions was handed out after everybody had correctly answered the test questions from the first set of instructions. In the second set subjects received information about their type (H or L) and about the first part of the experiment (i.e. the P-Experiment). At this point they only knew that there would be a second part (i.e. the C-Experiment), but did not know what this second part would look like. After the first part of the experiment was completed, subjects finally received the third set of instructions, which gave them information about their treatment and the second part that followed.

session was low and the other half was high endowed.

Subjects kept their endowment type in both parts of the experiment. The first part of the experiment, the P-Experiment, adopts a design first proposed by Fischbacher et al. (2001). Here, subjects stated their contributions conditional on others' contributions separately for all possible group structures. Since they only knew their own type but neither the type of the other two group members nor the group structure and thus the total group endowment, we used strategy tables where they stated their contribution given every combination of others' contributions and others' endowments (LL, LH, and HH). Table 4.1 illustrates our elicitation method for the case where one other group member was high and the other low endowed.

Table 4.1: **Strategy Table**

Example for group structure LH		
Contribution of Type L	Contribution of Type H	Own Contribution
...	...	...
0	4	?
1	3	?
2	2	?
2	3	?
...	...	...

For a complete version of the table refer to the instructions in Appendix 4B.

In order to make these conditional contributions incentive compatible we also asked for their unconditional contribution in each of the three possible group structures.[8] The stated contribution preferences from the P-Experiment are used later in order to explain the treatment effects from the second part of the experiment. While the seminal P-Experiment of Fischbacher et al. (2001) and other studies that used it elicit cooperation preferences conditional on the *average* of others' contributions in *homogeneous*

[8] For two randomly selected subjects in each group the unconditional contribution was payoff relevant and for one randomly selected subject the conditional contribution was payoff relevant. Payoffs were then realized according to (4.1) but feedback about the P-Experiment was only given at the end of the session.

groups, our P-Experiment elicits cooperation preferences conditional on every possible combination of others' contributions in homogeneous and heterogeneous groups. This modification allows us to investigate two open questions concerning conditional cooperation preferences. First, whether subjects condition their contributions on the averages of others' contributions or whether inequality in contributions among others matters. And second, whether subjects condition on relative or absolute contributions. While answering the former question shows whether the assumption of averages is justified in general, answering the latter is especially important under heterogeneous endowments and allows us to explain our results.

In the second part, the C-Experiment, subjects played the standard public goods game as described above repeatedly for ten rounds. We used a partner design where groups stayed together for all rounds.[9] Subjects were randomly allocated to groups with the structures LLL, LLH, LHH, and HHH.[10] These group structures appeared equally often in each of the following two treatments.

In the *certainty* treatment, subjects learned before the first period, whether the other two group members were low or high endowed. They further received information after each period how much each of the other group members contributed (together with the information about their types). In each period, they first stated their belief about others' contributions (knowing others' type) before we asked them to state their own contribution.

In the *uncertainty* treatment, subjects neither received information about the types of the other two group members nor about the total group endowment.[11] They only knew their own endowment. After each period they also received information about others' contribution (but without type information). Within each period, we first asked each subject to state her probabilistic belief about the group structure, then her belief about others' contributions, and finally asked for her own contribution.[12]

[9] Using two parts and connecting stated preferences from the P-Experiment to observed behaviour in the C-Experiment, we adopt a design similar to Fischbacher and Gächter (2010). The difference between their and our P-Experiment is outlined above since they used the original design of Fischbacher et al. (2001). Moreover, their C-Experiment consists of ten one-shot linear public goods games with random matching and a group size of four. The design of our C-Experiment is further explained below.

[10] Although these group structures were also used to determine the payoffs from the P-Experiment, feedback about the P-Experiment was only given at the end and after the C-Experiment.

[11] Note that all subjects knew that half of the them was high and the other half was low endowed and therefore knew the endowment distribution in the experiment. They did, however, not know the frequency of each group structure.

[12] In both treatments, certainty and uncertainty, beliefs about others' contributions were incentivized

The experiment consisted of eight sessions in total. In seven of those sessions, 24 subjects participated (12 in the certainty and 12 in the uncertainty treatment) and in one session only 12 subjects participated (in the uncertainty treatment). So, we had each possible group structure (LLL, LLH, LHH, and HHH) eight times in the uncertainty treatment and seven times in the certainty treatment. Having finished both parts of the experiment, first the P-Experiment and then the C-Experiment,[13] subjects answered a short questionnaire about socio-economic characteristics and then received feedback about their total earnings from the experiment. Before they left, each subject was paid in private and by a person that was *not* the experimenter.

4.3 Results

We organize our results in the following way. First, we report the results from the C-Experiment where we investigate what the treatment effects (i.e. the effects of uncertain endowments) are. Additionally, we compare heterogeneous to homogeneous endowments. In a second step, we will then report the results from the P-Experiment and use them to explain our observed treatment effects from the C-Experiment.

4.3.1 Treatment Effects

In the C-Experiment, subjects played the public goods game as outlined in Section 4.2 repeatedly for ten rounds. In the certainty treatment, contrary to the uncertainty treatment, subjects knew the endowment of their group members. In this section we compare these two treatments and investigate treatment effects on the overall, type, and group level.

Figure 4.1 shows the average contributions in the ten periods for the uncertainty and the certainty treatment. The mean contribution over all ten periods is 2.00 EMU in the

in the following way. If their belief was correct, subjects received 2 EMU, if their belief deviated from the actual contribution by only 1 EMU they received 1 EMU, and if their belief deviated by 2 EMU or more, they received nothing. Again, feedback was only given at the end of the experiment. We did not incentivize the stated probabilistic beliefs about others' endowment in the uncertainty treatment, since the only existing method we are aware of (i.e. quadratic scoring rule) requires the assumption of risk neutrality.

[13] We did not reverse the order of conducting the P- and C-Experiment since Fischbacher and Gächter (2010) already showed that it has no effect on outcomes.

uncertainty treatment and 2.21 EMU in the certainty treatment. This difference (9.5% less contribution in the uncertainty treatment) is, however, not significant ($p = 0.1891$; $N = 60$).[14]

Figure 4.1: Contributions under Certainty and Uncertainty

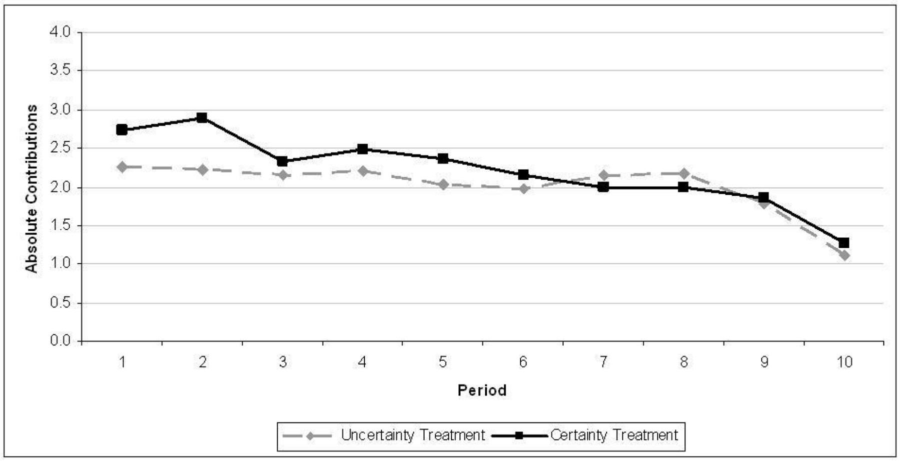
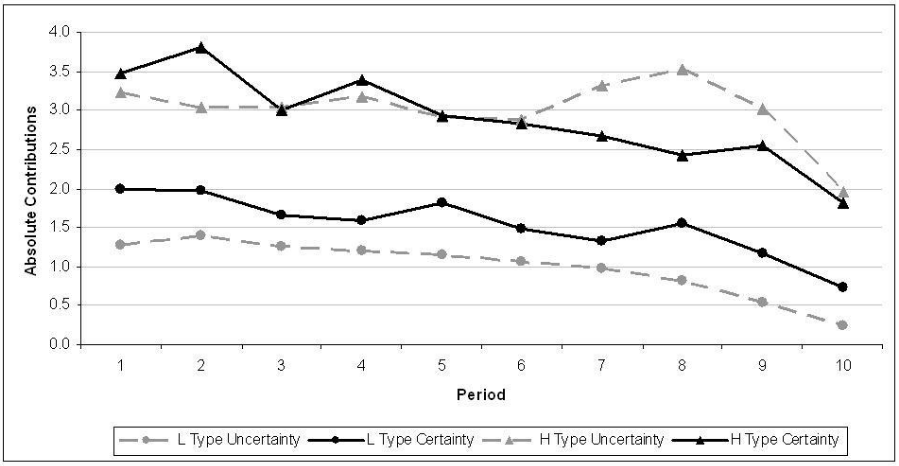


Figure 4.1 suggests that uncertainty about others' endowment levels does not have much of an effect on cooperation. In fact, this is what the sparse literature touching endowment uncertainty concludes. However, although there is only a slight negative effect of uncertainty on the overall level, there may well be a larger effect of uncertainty when types (L vs. H) or groups (poor vs. rich) are considered instead.

Figure 4.2 shows type-specific absolute contributions in the ten periods for both treatments. We find that L types contribute substantially less under uncertainty. The mean absolute contribution over all ten periods is 0.99 EMU in the uncertainty treatment and 1.53 EMU in the certainty treatment. This difference (35.3% less contribution in the uncertainty treatment) is significant at the 1%-level ($p = 0.0002$; $N = 45$). By contrast, H types contribute slightly more under uncertainty. Here, the mean absolute contribution over all ten periods is 3.01 EMU in the uncertainty treatment and 2.89 EMU in the certainty treatment. This difference (4.2% more contribution in uncertainty treatment), however, is not significant ($p = 0.6186$; $N = 45$). Nevertheless, it suggests that the small overall negative effect of uncertainty in Figure 4.1 may be due to two opposing effects that partly offset each other. L types contribute substantially less and H types slightly more under uncertainty.

[14]If not stated otherwise, all reported p-values are those of a two-sided Mann Whitney U test. The number of observations is denoted by N.

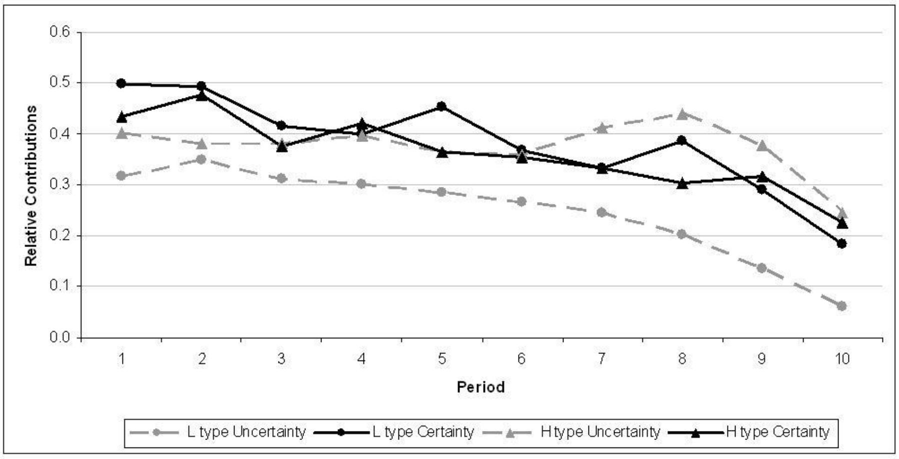
Figure 4.2: Type-Specific Absolute Contributions under Certainty and Uncertainty



The results under certainty replicate existing findings on the effect of heterogeneous endowments. There are several studies (see e.g. Marwell and Ames 1979 or Wit et al. 1992) that show that high endowed individuals contribute more in absolute terms than low endowed individuals, and many studies even indicate that subjects contribute approximately the same in relative terms (see e.g. Rapoport 1988, Rapoport and Suleiman 1993, van Dijk and Grodzka 1992, van Dijk and Wilke 1994). A notable exception is the study by Buckley and Croson (2006), which finds that L types contribute relatively more than H types. However, when looking at field evidence, Buckley and Croson (2006, p 937) note the following: "Data from the Social Welfare Research Institute at Boston College shows that in 2000, U.S. families with incomes under \$125,000 (91.39% of families in the U.S.) gave an average of 2.34% of their income to charity. There was little variation of giving across incomes with the poorest families, those with incomes under \$10,000 giving 2.25% (Giving USA 2003). Thus low-income families give the same percentage of their income as high-income families."

Figure 4.3 shows type-specific relative contributions in the ten periods for both treatments. We find that under certainty L and H types contribute similar amounts relative to their endowment. Over the ten periods L types contribute on average 38% and H types 36% of their endowment. Thus, there is no significant difference in relative contributions between types under certainty ($p = 0.3212$; $N = 28$). This picture changes however under uncertainty. Average relative contributions of L types over the ten periods drop to 25%, whereas the average relative contributions of H types slightly increase to 38%. When testing for differences in these relative contributions, we now find a significant difference in relative contributions between types under uncertainty. L types contribute significantly less at the 5%-level relative to their endowment than H types ($p = 0.0232$; $N = 32$). Moreover, between treatments there is no significant change in relative contributions among H types (testing the difference between certainty and uncertainty yields $p = 0.6184$; $N = 45$), but L types contribute significantly less at the 1%-level under uncertainty ($p = 0.0000$; $N = 45$).

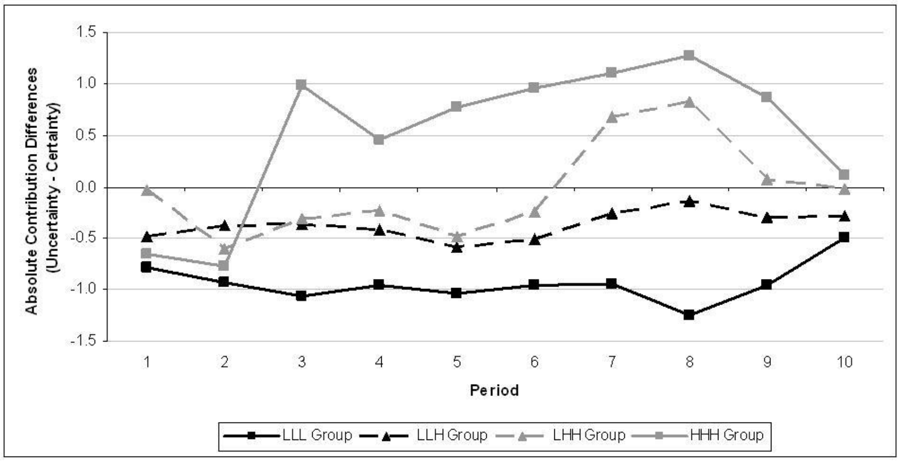
Figure 4.3: Type-Specific Relative Contributions under Certainty and Uncertainty



Knowing that L types contribute substantially less and H types slightly more under uncertainty raises the question about the effect of uncertainty on different group structures. Intuitively, we should expect that the more L types a group consists of the more this group suffers from uncertainty. Moreover, since the decrease in contributions of L types is higher than the increase in contributions of H types, we should expect that poor groups lose more than rich groups gain through the uncertainty. This is indeed what we observe. Figure 4.4 shows the treatment differences (uncertainty minus certainty) of group-specific absolute contributions over the ten periods. Groups with a majority of H types (LHH and HHH) gain from the uncertainty, whereas groups with a majority of L types (LLL and LLH) lose from making endowments uncertain. The average treatment difference over the ten periods is -0.94 EMU for LLL, -0.37 EMU for LLH, 0.35 EMU for LHH, and 0.51 EMU for HHH. These treatment differences deviate significantly from zero only for homogeneous groups.[15] We can further test whether these treatment differences differ between the various group structures. Here we find that treatment differences are significantly different between all group structures except between LLH vs. LHH and

[15] $p = 0.0000$ for LLL, $p = 0.1536$ for LLH, $p = 0.1204$ for LHH, and $p = 0.0622$ for HHH; $N = 15$.

Figure 4.4: Group-Specific Contribution Differences between Certainty and Uncertainty



LHH vs. HHH.[16]

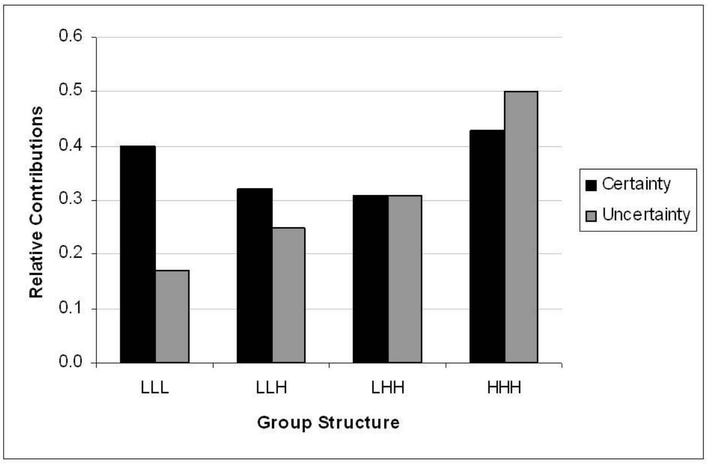
Having shown that poor groups contribute less and rich groups more under uncertainty we can now further analyze the efficiency of groups in the two treatments. Figure 4.5 shows group-specific mean contributions over the ten periods relative to the total group endowment. This determines how efficient a group is in providing the public good in the two treatments.

Under certainty, there is a U-shaped relationship between the number of H types in a group and the efficiency of the group. LLL groups contribute on average 40% of their endowment, LLH groups 32%, LHH groups 31%, and HHH groups contribute 43% of their endowment. Thus, heterogeneous groups are less efficient than homogeneous groups. When testing these differences we find that they are significant between homogeneous and heterogeneous groups but not within these two categories.[17] This is consistent with the existing literature. Most studies show that endowment heterogeneity decreases cooperation (for an overview see Ledyard 1995 or Zelmer 2003). For instance, Cherry et al. (2005) also use a linear public goods game and find that contributions are significantly lower in heterogeneous (33.1% of endowment) than in homogeneous (42.1% of endowment) groups (note that we find 31.4% and 41.8%, respectively). Anderson et al. (2008) vary the distribution of a fixed payment for participating in the public goods

[16] $p = 0.0782$ for LLL vs. LLH, $p = 0.0176$ for LLL vs. LHH, $p = 0.0000$ for LLL vs. HHH, $p = 0.3378$ for LLH vs. LHH, $p = 0.0646$ for LLH vs. HHH, and $p = 0.1740$ for LHH vs. HHH; $N = 30$.

[17] Note that this holds with the exception of LLL vs. LLH where the difference is marginally insignificant. For the comparisons of homogeneous vs. heterogeneous groups we find $p = 0.1182$ for LLL vs. LLH, $p = 0.0395$ for LLL vs. LHH, $p = 0.0822$ for HHH vs. LLH, and $p = 0.0379$ for HHH vs. LHH; $N = 14$. Within these two categories we get $p = 0.9750$ for LLL vs. HHH and $p = 0.9483$ for LLH vs. LHH; $N = 14$.

Figure 4.5: Group Efficiency under Certainty and Uncertainty



experiment and also find that inequality reduces contributions. In a field experiment, Cardenas (2003) examines the effect of real wealth on cooperation. Villagers in rural Columbia knew each other and others' wealth that they brought into the public goods experiment. Again, increased inequality decreased contributions.

Under uncertainty, the relationship between the number of H types and group efficiency is not U-shaped anymore, but becomes increasing in the number of H types. The effect of heterogeneous endowments (less efficiency than with homogeneous endowments) breaks down under uncertainty. Instead, it becomes important how rich a group is. LLL groups now contribute on average only 17% of their endowment, LLH groups 25%, LHH groups 31%, and HHH groups contribute 50% of their endowment. The efficiency differences between groups under uncertainty are all significant with two marginally insignificant exceptions.[18] Clearly, the richer a group the more efficient it is in providing the public good. Again, Figure 4.5 further shows that under uncertainty poorer groups are much less efficient and richer groups are slightly more efficient than under certainty. The difference between certainty and uncertainty is however significant only for homogeneous groups (LLL and HHH) but not for heterogeneous groups (LLH and LHH).[19]

Overall, we saw that uncertainty has only a small negative effect on cooperation. However, we showed that this is due to two opposing effects that partly cancel each other. Poor groups lose substantially and rich groups gain slightly under uncertainty,

[18] $p = 0.1181$ for LLL vs. LLH, $p = 0.0001$ for LLL vs. LHH, $p = 0.0004$ for LLL vs. HHH, $p = 0.1429$ for LLH vs. LHH, $p = 0.0003$ for LLH vs. HHH, and $p = 0.0004$ for LHH vs. HHH; $N = 16$.

[19] Testing the differences in efficiency between certainty and uncertainty we find $p = 0.0000$ for LLL, $p = 0.2406$ for LLH, $p = 0.3561$ for LHH, and $p = 0.0563$ for HHH; $N = 15$.

since L types contribute less and H types tend to contribute more than under certainty. Although the overall effect is almost neutralized, uncertainty causes inequality to increase compared to certainty. When measuring inequality in the income distribution via the Gini coefficient[20] we find that cooperation causes inequality to decrease in both treatments, but much less so under uncertainty. Without any cooperation, the Gini coefficient is 0.167. It decreases in both treatments, to 0.154 in the certainty treatment and to 0.161 in the uncertainty treatment.

4.3.2 Explaining Treatment Effects

In this section we provide an explanation for the treatment effects that were reported in the preceding section. In the P-experiment subjects stated their contributions conditional on every combination of others' contributions for each possible group structure. The main purpose of the P-Experiment was to elicit individual contribution preferences that can be used to explain cooperation behaviour in the C-Experiment. Before turning to the explanation of our treatment effects from Section 4.2, we will first discuss the underlying preference that can generate these findings.

Relative Conditional Cooperation

It has been argued before that a preference for conditional cooperation can explain why people contribute in public goods games. While earlier studies used more indirect approaches (see e.g. Sonnemans et al. 1999; Keser and van Winden 2000 or Brandts and Schram 2001), Fischbacher et al. (2001) proposed a direct mechanism (based on the strategy method of Selten 1967) to elicit individuals' willingness to cooperate given that others cooperate as well (see Ockenfels (1999) for a similar approach). This elicitation method asks subjects to choose their own contribution for every possible average of others' contributions. It has been used in other experimental studies in order to investigate the robustness (see Kocher 2008) or cultural differences (see Kocher et al. 2008 or Herrmann

[20]The Gini coefficient is based on the Lorenz curve which plots the proportion of the total income (on the vertical axis) that is cumulatively earned by the bottom $x\%$ of the population (where x goes from 0 to 100 on the horizontal axis). The Gini coefficient is then the ratio of the area that lies between the line of equality (i.e. the 45 degree line) and the Lorenz curve over the total area under the line of equality. The Gini coefficient lies always between 0 and 1 and the closer its value is to 0 the more equal is the income distribution.

and Thöni 2009) of conditional cooperation. Fischbacher and Gächter (2010) used it in order to explain the decline of cooperation in public goods games. The findings of these studies support conditional cooperation with a self-serving bias (i.e. imperfect conditional cooperation) both as the prevalent individual preference and as the average preference of all individuals.

As mentioned above, subjects in these studies are asked to indicate their own contributions for every possible *average* of others' contributions. In our P-Experiment we also used the elicitation method of Fischbacher et al. (2001) but asked for own contributions given others' *individual* contributions. We can therefore test whether the implied assumption that subjects condition their own contribution on the *average* of others' contributions is in fact reasonable. In Appendix 4A we show that subjects indeed condition their contributions on the average of others' contributions. This finding supports the assumption made in Fischbacher et al. (2001) and can be seen as another robustness result on the preference for conditional cooperation.

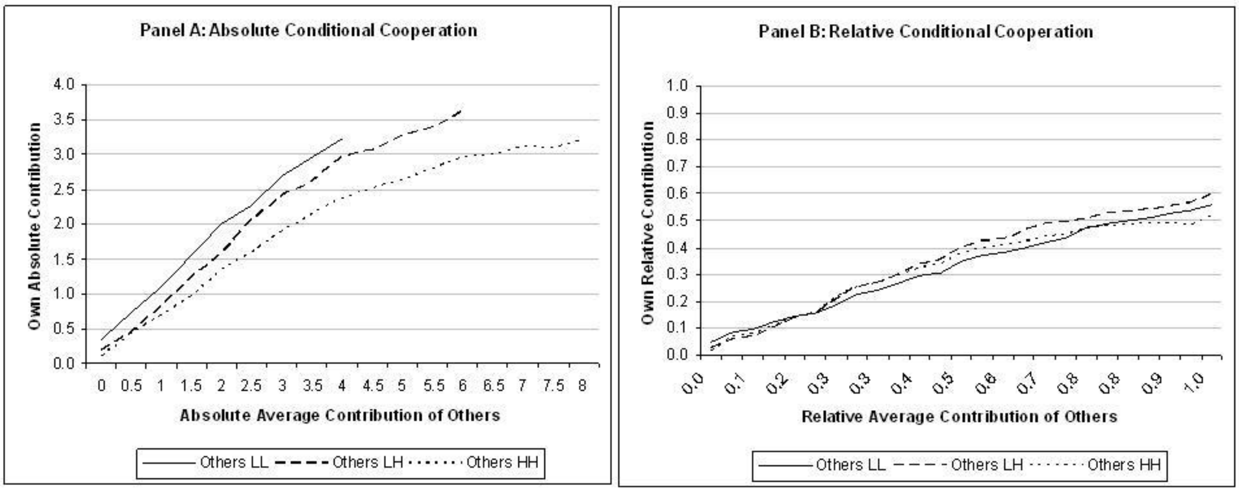
In all previous studies on conditional cooperation individuals had the same endowment. Under homogeneous endowments, there is no difference whether subjects condition on absolute or rather on relative contribution levels. Under heterogeneous endowments it becomes, however, important whether subjects (imperfectly) choose to match the absolute amount of the average contribution of other group members or whether they choose contributions similar to the others' average contribution relative to their endowments. In our experiment this conditional behaviour determines whether subjects contribute more the poorer the group is. On a broader scale, it determines whether and how subjects condition their own contributions on contribution levels *and* endowments of others.[21]

Figure 4.6 shows the average contribution schedule separately for each combination of others' endowments.[22] Figure 4.6A shows own absolute contributions (on the vertical axis) given the absolute average level of others' contributions (on the horizontal axis) separately for the cases where the other group members are LL, LH, and HH. Subjects

[21]Note that it is not our intention to test any theoretical foundations of *why* people may want to condition their contributions on relative rather than on absolute contributions of others. We are rather interested in disclosing the systematic behavioural patterns that can be empirically observed in public goods games.

[22]Figure 4.6 as well as the regression analysis are based on all subjects and does not separate into specific *preference types* like free-riders, conditional or triangle cooperators. We therefore investigate the average preference of all individuals.

Figure 4.6: Conditional Cooperation Preferences: Absolute and Relative



contribute more in poorer groups, holding the absolute level of others' average contribution constant. The slopes of the average contribution schedules in Figure 4.6A are less than one, and they are the higher the poorer the group members are. While positive slopes of less than one just confirm the stylized fact of a self-serving bias in conditional cooperation, the relation between slopes and others' endowments, i.e. greater slopes in poorer groups, remains to be explained.

That subjects give more in poorer groups may be due to the fact that relative rather than absolute contributions matter. Indeed, that this is the case can be seen in Figure 4.6B. Figure 4.6B plots own relative contributions (on the vertical axis) given others' relative average contributions (on the horizontal axis) separately for the cases when others are LL, LH, and HH. Since subjects condition their contributions on others' relative rather than absolute contributions, we do not observe a difference in contribution schedules between the combinations of others' endowments in Figure 4.6B. The differences in the slopes between the situations in which group members are LL, LH, or HH in Figure 4.6A disappear almost completely when we control for their endowment level in Figure 4.6B. We interpret this observation as subjects matching relative rather than absolute average contributions of others. Still, the slopes are less than one, so there is a self-serving bias in relative conditional cooperation as well.

There are two ways subjects can conditionally cooperate on relative levels. They can condition their own contribution either on others' relative average contributions (as in Figure 4.6B) or on the average of others' relative contributions. In the first specification they would take the average of others' individual absolute contributions and then divide it by the average of others' absolute endowment. In the second specification subjects would take others' individual absolute contributions, divide them by others' individual absolute endowments, and then take the average of these relative individual contributions. Both ways of relative conditional cooperation coincide when the endowments of other group members are the same. But they do not when others are differently endowed. Consider an example where one group member is low and the other is high endowed. Suppose the L type contributes 4 EMU and the H type contributes 0 EMU. In the first case a subject would (imperfectly) match a contribution of 33.3% ($= \frac{4+0}{4+8} * 100$) of her endowment, and in the second case she would (imperfectly) match a contribution of 50% ($= [\frac{4}{4} + \frac{0}{8}] * 100$) of her endowment. So we need to find out whether the first specification, which we used in Figure 4.6B, can explain our data better than the second specification. In the following quantitative analysis we will detect which of the specifications yields more reasonable results.

To quantify conditional cooperation in absolute as well as in relative terms we estimate OLS regressions[23] of the following specification.

$$OwnCon_k = \beta_0 + \beta_1 \; OthCon_g + \beta_2 \; OthCon_g * D_{LH} + \beta_3 \; OthCon_g * D_{LL} + \epsilon. \tag{4.2}$$

The dependent variable $OwnCon_k$ is an individual's own contribution where $k = \{A, R\}$. $OwnCon_A$ stands for own absolute contribution and $OwnCon_R$ for own relative contribution (i.e. own absolute contribution relative to own endowment). $OthCon_g$ is the average of others' contributions where $g = \{A, R1, R2\}$. $OthCon_A$ represents others' absolute average contribution, $OthCon_{R1}$ is others' relative average contribution (first specification from above), and $OthCon_{R2}$ is the average of others' relative contributions (second specification from above). D_{LL} is a dummy variable that takes the value one for being in a group where the others are LL, D_{LH} is a dummy variable for others being LH, and ϵ is the individual error term. For relative conditional cooperation under the first specification (with $k = R$ and $g = R1$) we ran (4.2) with all subjects (M2; Table 4.2), only with H types (M4; Table 4.2), and only with L types (M5; Table 4.2). For relative

[23] We additionally ran a Tobit regression, which disregards the clustering of the data but takes the censoring of the dependent variable into account. Since we received similar results we only report those of the OLS regression.

conditional cooperation under the second specification (with $k = R$ and $g = R2$) and absolute conditional cooperation (with $k = A$ and $g = A$) we ran (4.2) with all subjects in (M3) and (M1) in Table 4.2, respectively.

Table 4.2 reports regression results for all five specifications.[24] Let us first look at (M1) in Table 4.2. The marginal effect of an additional unit of others' absolute average contribution is 0.4177 ($= \beta_1$) when the others are HH, 0.5831 ($= \beta_1 + \beta_2$) when the others are LH, and 0.7234 ($= \beta_1 + \beta_3$) when the others are LL. So, individuals contribute more the more others contribute. The marginal effect is strictly lower than one, indicating a slope of absolute conditional cooperation that is smaller than one. Moreover, holding others' absolute average contribution constant, individuals contribute significantly more the poorer their group members are. The slope of the contribution schedule significantly increases by 0.1654 when going from HH to LH and by another 0.1403 when going from LH to LL, since the difference between β_2 and β_3 is also significant at the 1%-level (F-test; $p = 0.0000$).

These differences (almost) disappear when we consider own relative contribution given the relative average contribution of others in (M2) in Table 4.2. The corresponding marginal effects are 0.5258 ($= \beta_1$) when the others are HH, 0.5764 ($= \beta_1 + \beta_2$) when the others are LH, and 0.5088 ($= \beta_1 + \beta_3$) with the others being LL. Now, individuals increase their relative contribution by roughly half of the increase of others' relative average contribution. Interestingly, they make almost no difference between groups anymore. The difference in the slopes between HH and LL is not significant. However, despite being very small there is a significant difference both between HH and LH and between LL and LH, since the difference between β_2 and β_3 is significant at the 5%-level (F-test; $p = 0.0122$). Before investigating the reasons for the small but significant differences in these slopes we first consider the second specification of relative conditional cooperation below.

[24] The results of Table 4.2 are qualitatively similar when we include dummy variables for the different group structures (LL and LH) in our regression.

Table 4.2: **Absolute and Relative Conditional Cooperation**

	Dependent Variables				
	(M1) Abs.: $OwnCon_A$ and $g = A$	(M2) Rel.1: $OwnCon_R$ and $g = R1$	(M3) Rel.2: $OwnCon_R$ and $g = R2$	(M4) Rel.H: $OwnCon_R$ and $g = R1$	(M5) Rel.L: $OwnCon_R$ and $g = R1$
$OthCon_g$	0.4177***	0.5258***	0.4600***	0.6192***	0.4324***
	(0.0329)	(0.0380)	(0.0335)	(0.0520)	(0.0538)
$OthCon_g * D_{LH}$	0.1654***	0.0506**	0.1443***	-0.0227	0.1238***
	(0.0197)	(0.0231)	(0.0225)	(0.0298)	(0.0338)
$OthCon_g * D_{LL}$	0.3057***	-0.0170	0.0698**	-0.1900***	0.1560***
	(0.0299)	(0.0322)	(0.0305)	(0.0359)	(0.0471)
Constant	0.4646***	0.0773***	0.0878***	0.0778***	0.0767***
	(0.0778)	(0.0122)	(0.0119)	(0.0172)	(0.0175)
Observations	18900	18900	18900	9450	9450
Clusters	180	180	180	90	90
R^2	0.1433	0.1650	0.1441	0.2183	0.1417

OLS Regression with data from the P-Experiment. Robust standard errors (clustered in subjects) are in parentheses. *** denotes significance at the 1%-level, ** denotes significance at the 5%-level.

In (M3) in Table 4.2 we consider own relative contribution given the average of others' relative contributions. We find that the differences in the slopes of the absolute contribution schedules between different endowments of others (LL, LH, or HH) do not disappear when we consider our second alternative relative specification. The marginal effect is 0.4600 ($= \beta_1$) when others are HH, 0.6043 ($= \beta_1 + \beta_2$) when others are LH, and 0.5298 ($= \beta_1 + \beta_3$) when others are LL. Therefore, all differences in the slopes between group structures are larger and remain significant when the second specification of relative conditional cooperation is used. Note that the difference between LH (β_2) and LL (β_3) is also significant at the 1%-level (F-test; $p = 0.0092$). In addition, the coefficient of determination (i.e. R^2) of (M2) is larger than that of both (M3) and (M1). We therefore choose (M2), the first specification of relative conditional cooperation, as our preferred model.

As mentioned above, the preference of absolute conditional cooperation and relative conditional cooperation coincide with homogeneous endowments. Our results from Figure 4.6 and Table 4.2 therefore extend existing results of conditional cooperation preferences to heterogeneous endowments. On average, individuals condition their own contribution on others' relative average contribution but not on the average of others' relative contribution. So far, our analysis used the data of all subjects, and we did not investigate whether there are differences between H and L types. The last two columns (M4 and

M5) of Table 4.2 estimate regression (4.2) separately for H and L types (using the first specification of relative conditional cooperation).

In (M4) in Table 4.2 we see that for H types the marginal effect of others' relative average contribution on own relative contribution is 0.6192 (= β_1) when others are HH, 0.5965 (= $\beta_1 + \beta_2$) when others are LH, and 0.4292 (= $\beta_1 + \beta_3$) when others are LL. The difference between HH and LH is very small and not significant, but the differences between HH and LL and also between LL and LH are significant, since the difference between β_2 and β_3 is significant at the 1%-level as well (F-test; $p = 0.0000$). This shows that H types contribute the same in relative terms when others are HH and LH, but they contribute significantly less when others are LL.

In (M5) in Table 4.2 we consider own relative contribution for L types given the relative average contribution of others. The slopes of their contribution schedule are 0.4324 (= β_1) when others are HH, 0.5562 (= $\beta_1 + \beta_2$) when others are LH, and 0.5884 (= $\beta_1 + \beta_3$) when others are LL. The differences between HH and LH and between HH and LL are significant, while the difference between LH and LL is not (F-test; $p = 0.4166$ for the difference between β_2 and β_3). L types contribute the same in relative terms when others are LL and LH, but they contribute significantly less when others are HH. This is the reverse pattern of H types. Both types seem to have a preference for relative conditional cooperation, but with a negative bias if their type is the minority in the group. Note, however, that both types still contribute more in absolute terms for a given absolute average contribution of others the poorer their group members are.

In order to investigate whether both L and H types show the same relative conditional cooperation behaviour in situations without negative minority bias, we ran additional OLS regressions of the following form.

$$OwnCon_R = \beta_0 + \beta_1 \; OthCon_{R1} + \beta_2 \; OthCon_{R1} * D_H + \epsilon. \qquad (4.3)$$

Our dependent variable remains $OwnCon_R$. D_H is a dummy variable that is one for being an H type (and zero for being an L type). We ran (4.3) when others are LL (M6; Table 4.3), LH (M7; Table 4.3), and HH (M8; Table 4.3).

Table 4.3: **Type-Specific Relative Conditional Cooperation**

	Dependent Variables		
	(M6)	(M7)	(M8)
	$OwnCon_R$ when others LL	$OwnCon_R$ when others LH	$OwnCon_R$ when others HH
$OthCon_{R1}$	0.6041***	0.5707***	0.4085***
	(0.0471)	(0.0528)	(0.0524)
$OthCon_{R1} * D_H$	-0.1576***	0.0419	0.1885**
	(0.0551)	(0.0674)	(0.0738)
Constant	0.0666***	0.0677***	0.0914***
	(0.0150)	(0.0143)	(0.0142)
Observations	2700	8100	8100
Clusters	180	180	180
R^2	0.2414	0.1903	0.1488

OLS Regression with data from the P-Experiment. Robust standard errors (clustered in subjects) are in parentheses. *** denotes significance at the 1%-level, ** denotes significance at the 5%-level.

First consider (M7) in Table 4.3. The slopes of the relative contribution schedule are not statistically different between L types ($\beta_1 = 0.5707$) and H types ($\beta_1 + \beta_2 = 0.6126$) when others are LH. This picture changes, however, when we consider (M6) and (M8) in Table 4.3. In (M6) we see that H types ($\beta_1 + \beta_2 = 0.4465$) contribute significantly less than L types ($\beta_1 = 0.6041$) when the others are LL. L types still contribute the relative amount they contributed when others were LH (this can be seen from (M5) in Table 4.2), but H types contribute less now. When looking at (M8) in Table 4.3 we see the reverse picture. Now, L types ($\beta_1 = 0.4085$) contribute significantly less than H types ($\beta_1 + \beta_2 = 0.5970$) when others are HH. H types still contribute at the relative level when others were LH (again this can be seen from (M4) in Table 4.2), but L types contribute less.

Combining the analysis of both tables yields the complete pattern of conditional cooperation preferences. We first showed that on average subjects are imperfect relative conditional cooperators who condition their own contribution on others' relative average contributions rather than on the average of others' relative contributions (this can be seen from (M1) to (M3) in Table 4.2). Instead of looking at the average subject we then considered L and H types separately in a second step. Here, we showed that both types, despite having a relative conditional cooperation preference, have a negative bias for being the minority type. H types contribute less in relative terms when others are LL than they do when others are LH or HH. L types contribute less in relative terms when others are HH than what they contribute when others are LH or LL (this can be seen from (M4) and (M5) in Table 4.2 and (M6) to (M8) in Table 4.3).

The observed minority bias may capture the effect of group identity on social preferences which has recently been analyzed by Chen and Li (2009). They show that out-group (as opposed to in-group) identity decreases charity concerns, reciprocity, and the likelihood of choosing the social-welfare-maximizing action. In our experiment, subjects may identify themselves as outside the group if they are the only ones with a certain endowment. When considering the average subject the minority bias causes relative contributions to shift downward in cases where others are LL and HH, but not in cases where others are LH. This explains why we found (in (M2) in Table 4.2) that on average subjects contribute slightly (but significantly) more when others are LH than when others are LL or HH. Nevertheless, despite having a negative minority bias in addition to the self-serving bias in the preference of relative conditional cooperation, the clear pattern of giving more in absolute terms the poorer the group members are, still holds. This is because the minority bias is rather small.

However, for explaining our results on heterogeneous endowments from the C-Experiment the minority bias may still be relevant. In contrast to heterogeneous groups, no subject is the minority type in homogeneous groups. This suggests that homogeneous groups are more efficient (i.e. contribute more in relative terms) than heterogeneous groups. Having outlined how preferences for relative conditional cooperation can explain our results from the certainty treatment in the C-Experiment, the question is now why this preference would explain our results from the uncertainty treatment as well. This is what we focus on in the next two sections. First, we explore the direct effect of relative conditional cooperation that is based on first-order beliefs, and second, we consider an indirect effect of relative conditional cooperation preferences based on second-order beliefs.

Both these effects are caused by the mere fact that subjects contribute more (for given contributions of others) the poorer their group members are. As a robustness check for the regression analysis of this section, we can additionally perform non-parametric tests. For any given combination of others' individual contributions (not averages) we can test whether subjects' own contributions are higher or lower for changes in the endowment of other group members. For instance, assume that one group member contributes 2

EMU and the other group member contributes 4 EMU. We can then compare a subject's own contribution for these given contributions of others in the situations where the other group members are LL, LH, and HH.

Since we use Wilcoxon signed rank sum tests we make pair-wise comparisons for every possible combination of others' contributions between the situations where the other group members are LL vs. LH and LH vs. HH.[25] Our results strongly support the regression analysis. For the comparison of LL vs. LH we find that out of 15 combinations of others' contributions 14 are significant ($0.0000 \leq p \leq 0.0151$; $N = 180$). The only one that is not significant in the comparison of LL vs. LH is when others both contribute nothing ($p = 0.3210$; $N = 180$), where we indeed should not expect any difference. If others contribute a given individual amount that is positive for at least one of them, subjects contribute significantly more when others are LL than what they do when others are LH. For the comparison of LH vs. HH we get similar results. Out of 25 combinations of others' contributions 23 are significant at the 1%-level ($0.0000 \leq p \leq 0.0017$; $N = 180$) in the comparison of LH vs. HH. Again, when both others contribute nothing, as expected, we do not observe a difference ($p = 0.4624$; $N = 180$). We also do not observe a significant difference between LH vs. HH for the combination where one of the others contributes nothing and the other contributes 1 EMU ($p = 0.5146$; $N = 180$).[26] However, for all other combinations subjects contribute significantly more when others are LH instead of HH.

Deviation of Beliefs

A preference for relative conditional cooperation causes subjects to contribute more in absolute terms the poorer their group members are. Therefore, under uncertainty it is important what subjects believe about their group members' endowments. Figure 4.7 shows how correct subjects' beliefs are in the C-Experiment. The variable 'deviation of beliefs' measures how realistic subjects' beliefs on the endowment of their group members are. For instance, consider a subject who has group members that are both L types. If

[25] We could also test LL vs. HH. However, if we can show a significant difference in a consistent direction for the comparisons of LL vs. LH *and* LH vs. HH, testing LL vs. HH will be redundant.

[26] The non-significance of this combination disappears for the comparison of LL vs. HH. Here, subjects contribute significantly more at the 1%-level if one of the others contributes nothing and the other contributes 1 EMU ($p = 0.0000$; $N = 180$). Obviously, for all other combinations the comparison of LL vs. HH also yields significant results.

this subject believed to be in HH with probability 1, her belief would deviate by 100% from reality. If she believed to be in LH with probability 1, her belief would deviate by 50%, and if she believed to be in LL with probability 1, her belief would not deviate (i.e. deviate by 0%).[27] Since subjects know that half of the subjects in the experiment are L types and the other half are H types, they mostly start out with the belief to be in LH. Hence, subjects who actually are in LH have realistic beliefs and subjects who actually are in LL or HH start out with beliefs that deviate by about 50% from reality. Over the ten periods, there is not much updating. Subjects in LH keep their realistic beliefs, and subjects in LL and HH mostly keep their deviating beliefs. However, subjects in HH update better than subjects in LL, since they get informative signals of contributions exceeding 4 EMU. Figure 4.7 implies that subjects in LL think they are in a richer group than they actually are.[28] Because of their relative conditional cooperation preference they therefore contribute less under uncertainty. Likewise, subjects in HH think they are in a poorer group than they actually are, and since they prefer to give more in poorer groups, they contribute more under uncertainty. This effect is strongest in LLL and HHH since individual effects of all group members go in the same direction. Moreover, the fact that subjects with other group members being HH exhibit more realistic beliefs than subjects who have LL group members, explains why poor groups lose more than rich groups gain through the uncertainty.

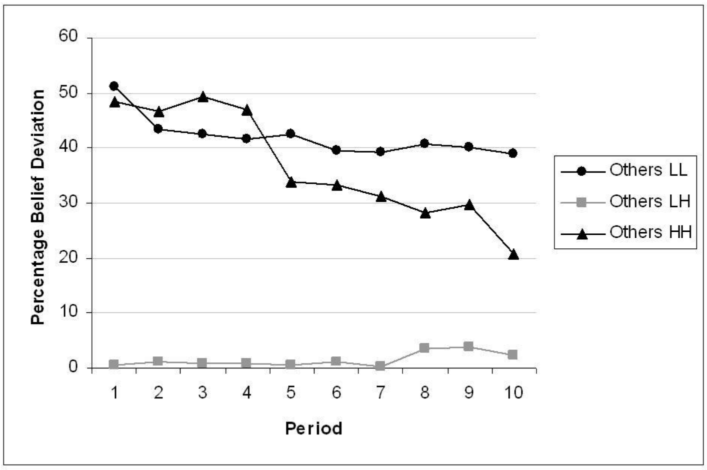
In order to quantify this explanation we have to compare the *predicted* contributions that are based on actual and often deviating beliefs to predicted contributions that would have occurred if subjects had held the correct beliefs about their group structure. We will call the latter type of predicted contributions *hypothetical* contributions.

Hypothetical contributions are calculated from the individual contribution preferences (elicited in the P-Experiment), the beliefs about others' contributions in each period (elicited in the C-Experiment), and under the assumption of correct beliefs about the

[27]We restricted beliefs to be certain in this example for illustration only. Since subjects in the C-Experiment stated their *probabilistic* beliefs about their group members' endowments, the variable 'deviation of beliefs' is calculated as a linear combination of the different group structures that give 100%, 50%, or 0% deviation. For instance, if this subject believed to be in LL with 33% probability, in LH with 33% probability, and in HH with 34% probability, her belief would also deviate by 50%.

[28]Note that the 'deviation of beliefs' can be negative or positive. Under the assumption that the deviation is positive for thinking incorrectly to be in a richer group and negative for thinking incorrectly to be in a poorer group we would observe the HH curve in Figure 4.7 to be mirrored in the negative on the horizontal axis. Although our analysis fully accounts for this, we chose to omit this complication in Figure 4.7 in order to compare the LL and HH curve directly.

Figure 4.7: Deviation of Beliefs about Group Structure from Reality



group structure. Predicted contributions are calculated from the individual contribution preferences, the beliefs about others' contributions, and the actual probabilistic beliefs about the group structure. Both predicted and hypothetical contributions are the point predictions. Although Section 4.3.2 showed that individuals have on average a preference for relative conditional cooperation, the computations of predicted and hypothetical contributions do not impose any assumptions. The calculation is based on the exact stated preference of each individual given others' individual contributions (and not the average of their contributions) in each specific group structure. Nevertheless, the findings in Section 4.3.2 are extremely useful in order to understand the mechanisms driving our results from the C-Experiment.

The difference between predicted (X_u^p) and hypothetical (X_u^h) contributions gives the net amount that subjects contributed more or less under uncertainty because of their deviating beliefs. We refer to this difference as $DOB_u = X_u^p - X_u^h$, where u stands for uncertainty. For L types in LL we expect $DOB_u < 0$ since they would give more if they knew others' real endowments. For H types in HH it should hold that $DOB_u > 0$, since they would give less if they knew others' real endowments. We expect L types and H types in LH to have $DOB_u = 0$ since their belief is (more or less) correct. L types in HH will have $DOB_u > 0$ since they would give more, and H types in LL will have $DOB_u < 0$ since they would give less, if they knew the real endowments of others. Under certainty, beliefs cannot be wrong, $X_c^p = X_c^h$, and therefore $DOB_c = 0$, where c stands for certainty. So, the difference in contributions between uncertainty and certainty, which is caused by deviating beliefs, is captured by $\Delta DOB = DOB_u - DOB_c = DOB_u$.

We find that poor groups contribute less under uncertainty and rich groups contribute more because of their deviating beliefs. The ΔDOB is -0.24 EMU for LLL, -0.03 EMU for LLH, 0.11 EMU for LHH, and 0.27 EMU for HHH.[29] These findings are in line with our predictions. That this is the case can be easily seen in homogeneous groups. In LLL, all subjects have $DOB_u < 0$. Likewise, all subjects in HHH have $DOB_u > 0$. So, all individual effects go in the same direction in homogeneous groups. By contrast, in heterogeneous groups individual effects do not go in the same direction. To see this, consider first an LLH group. L types in this group are in LH leading to $DOB_u = 0$, and the H type in this group is in LL leading to $DOB_u < 0$. So the total effect for the LLH group should be slightly negative. Second, consider the LHH group. Here, the L type is in HH leading to $DOB_u > 0$, and both H types are in LH leading to $DOB_u = 0$. Thus, the total effect in LHH should be slightly positive.

When considering types we find that the ΔDOB for L types is -0.28 EMU and for H types it is 0.34 EMU.[30] For both types these differences are significant at the 1%-level ($p = 0.0032$ for L types, and $p = 0.0012$ for H types; $N = 16$).

(Strategic) Over-Contribution

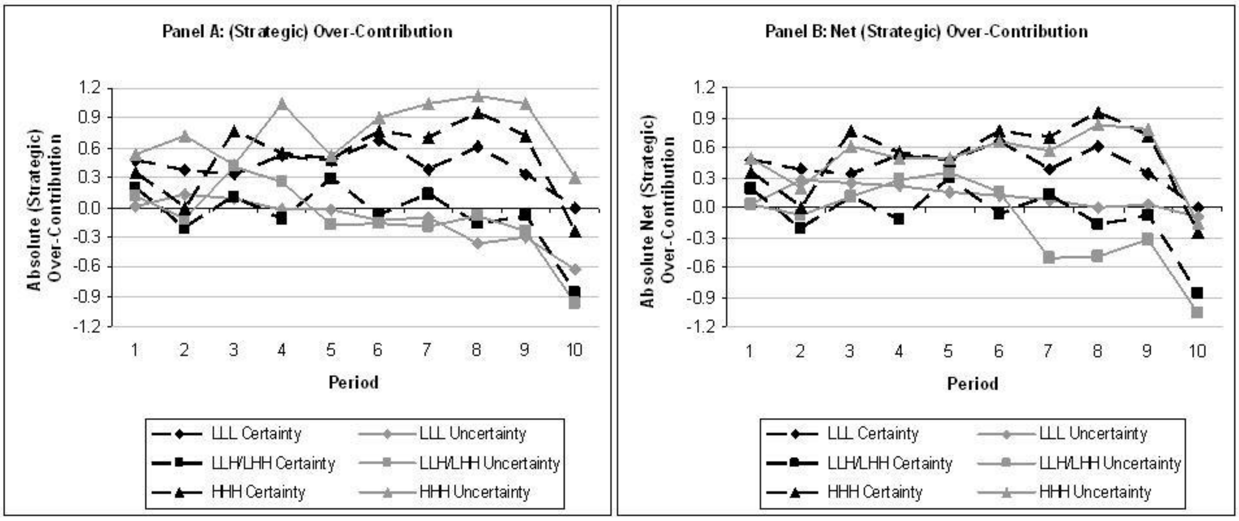
Botelho et al. (2009) show that subjects free-ride more in a perfect stranger design where they only meet each other once. As soon as there is the possibility to meet more than once (random stranger design), subjects contribute more. Fischbacher and Gächter (2010) found in their random stranger design that subjects contribute more in the C-Experiment than what is predicted from their preferences in P-Experiment. In our partner design, such over-contribution in the C-Experiment should be even more pronounced since subjects may strategically want to sustain high cooperation as long as possible.

(Strategic) over-contribution is calculated as the difference between predicted contributions and *actual* contributions (X_j^a), where the latter are the observed contribu-

[29]While the differences for homogeneous groups are significant, they are not for heterogeneous groups ($p = 0.0240$ for LLL, $p = 0.6432$ for LLH, $p = 0.2196$ for LHH, and $p = 0.0176$ for HHH; $N = 8$).

[30]Note that the amount that L types contribute less because of their deviating beliefs is smaller than the amount H types contribute more despite the fact that L types' beliefs deviate more than those of H types. The reason is that both types are relative conditional cooperators who give more in absolute terms the poorer their group members are. H types are more often in richer groups where the average contribution of others is higher. The vertical distance (caused by the different slopes) between the curves of Figure 4.6A is larger at higher average contributions of others. The impact of the deviating belief is therefore stronger than for L types who are more often in poorer groups who contribute less on average.

Figure 4.8: (Strategic) Over-Contribution under Certainty and Uncertainty



tions from the C-Experiment with $j = \{c, u\}$. We define (strategic) over-contribution as $SOC_j = X_j^a - X_j^p$. Figure 4.8A shows the SOC_j for homogeneous and heterogeneous groups under certainty and uncertainty over all ten periods. In contrast to heterogeneous groups, homogeneous groups strategically over-contribute under certainty. (Strategic) over-contribution is significantly different from zero for both LLL and HHH, but not for LLH and LHH. Under certainty, the SOC_c is 0.42 EMU ($p = 0.0788$; $N = 7$) for LLL and 0.51 EMU ($p = 0.0424$; $N = 7$) for HHH. For LLH and LHH the SOC_c is -0.02 EMU ($p = 0.9321$; $N = 7$) and -0.14 EMU ($p = 0.3720$; $N = 7$), respectively. This again indicates why homogeneous groups are more efficient than heterogeneous groups under certainty. Regarding the quantitative effects of our two identified explanations for the efficiency loss of heterogeneous endowments under certainty, the minority bias and (strategic) over-contribution, we find that the latter almost completely drives the result. Recall that the difference in group efficiency between homogeneous and heterogeneous groups is 10.4 percentage points (homogeneous groups contribute 41.8% of their endowment, whereas heterogeneous groups contribute only 31.4% of their endowment). The difference in relative (strategic) over-contribution between the two groups is 9.5 percentage points. Thus, 93% of the efficiency loss caused by heterogeneous endowments can be explained by the absence of (strategic) over-contribution in heterogeneous groups. The minority bias is therefore only marginally relevant.

Under uncertainty, heterogeneous groups still show no (strategic) over-contribution. The SOC_u is -0.13 EMU ($p = 0.1610$; $N = 8$) for LLH and -0.10 ($p = 0.6688$; $N =$

8) for LHH. These differences still do not deviate significantly from zero. However, rich homogeneous groups still over-contribute under uncertainty, but poor homogeneous groups do not. The SOC_u drops to -0.14 EMU ($p = 0.5584$; $N = 8$) for LLL but stays at 0.77 EMU ($p = 0.0200$; $N = 8$) for HHH. (Strategic) over-contribution under uncertainty is not significantly different from zero for LLL, but for HHH it is significant at the 5%-level. We denote the difference in (strategic) over-contribution between the uncertainty and certainty treatment as $\Delta SOC = SOC_u - SOC_c$. The ΔSOC is -0.56 EMU ($p = 0.0178$; $N = 15$) for LLL, -0.11 EMU ($p = 0.0526$; $N = 15$) for LLH, 0.04 EMU ($p = 0.2640$; $N = 15$) for LHH, and 0.26 EMU ($p = 0.0874$; $N = 15$) for HHH. Note that all differences are significant except for LHH.

(Strategic) over-contribution captures the amount subjects contribute more in a repeated set-up given their preferences and actual beliefs. We showed in the previous section that poor groups would contribute more under uncertainty if they had correct beliefs. Thus, the SOC_u of poor groups would be higher and therefore ΔSOC would be less negative if subjects had correct beliefs. So, ΔSOC includes the effect caused by deviating beliefs, which is ΔDOB. For rich groups we showed that they would contribute less if they had correct beliefs. Hence, the SOC_u of rich groups would be lower and the ΔSOC therefore less positive. Again, ΔSOC includes the effect caused by deviating beliefs.

The net (strategic) over-contribution ($NSOC_j$) is the amount subjects would over-contribute even if they had correct beliefs, namely $NSOC_j = X_j^{a'} - X_j^p$, where $X_j^{a'} = X_j^a + (X_j^h - X_j^p)$. The term in brackets is the amount subjects would contribute more (poor groups) or less (rich groups) if they had correct beliefs.[31] Figure 4.8B shows the $NSOC_j$ for homogeneous and heterogeneous groups under certainty and uncertainty over all ten periods. Under certainty, $NSOC_c = SOC_c$ since $DOB_c = 0$. Under uncertainty, heterogeneous groups still show no net (strategic) over-contribution. The $NSOC_u$ is -0.09 EMU ($p = 0.4234$; $N = 8$) for LLH and -0.22 ($p = 0.2104$; $N = 8$) for LHH. These differences still do not significantly deviate from zero. However, rich homogeneous groups still over-contribute under uncertainty, but poor homogeneous groups do not. The $NSOC_u$ drops to 0.11 EMU ($p = 0.3188$; $N = 8$) for LLL but stays at 0.50 EMU ($p =$

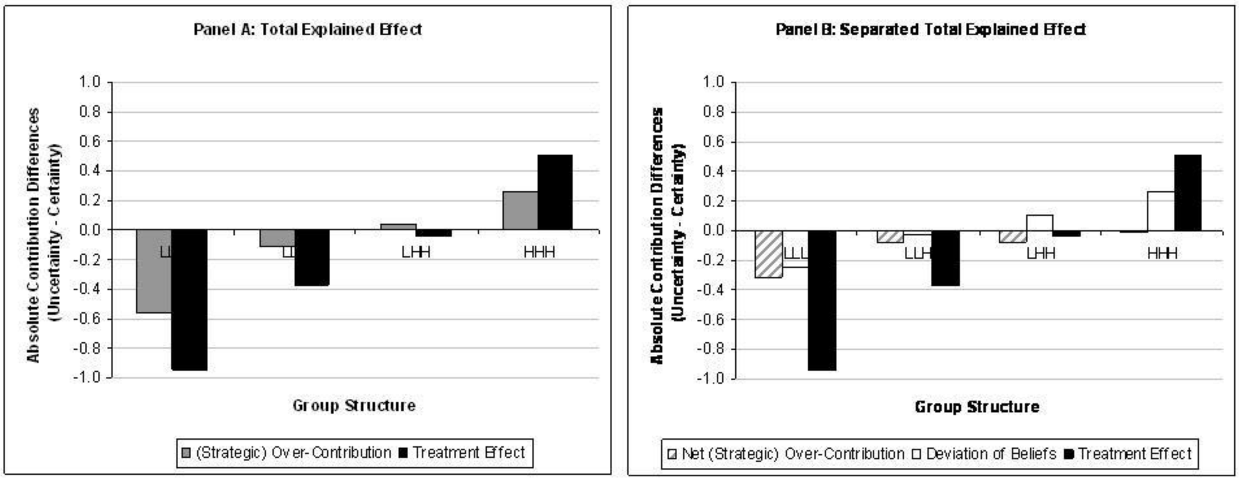
[31] Note that $SOC_j = NSOC_j + DOB_j \Leftrightarrow DOB_j = SOC_j - NSOC_j = [X_j^a - X_j^p] - [X_j^a + (X_j^h - X_j^p) - X_j^p] = X_j^p - X_j^h$. Under certainty, $X_c^p - X_c^h \Leftrightarrow X_c^a = X_c^{a'}$ and hence $SOC_c = NSOC_c$.

0.0002; $N = 8$) for HHH. Net (Strategic) over-contribution for LLL is not significantly different from zero, but for HHH it is significant at the 1%-level. We denote the difference in net (strategic) over-contribution between the uncertainty and certainty treatment as $\Delta NSOC = NSOC_u - NSOC_c$. The $\Delta NSOC$ is -0.31 EMU ($p = 0.0732$; $N = 15$) for LLL, -0.07 EMU ($p = 0.3092$; $N = 15$) for LLH, -0.08 EMU ($p = 0.4514$; $N = 15$) for LHH, and -0.01 EMU ($p = 0.7894$; $N = 15$) for HHH. Note that the difference is only significant for LLL.

A possible explanation for the results of $\Delta NSOC$ is that under uncertainty L types may fear to send H type signals when they strategically over-contribute. If they anticipate that others give less in richer groups, i.e. if they know about the relative conditional cooperation preference of others, it may strategically make sense not to over-contribute under uncertainty even though they want to sustain cooperation. That such an explanation may be plausible can be seen from type-specific net (strategic) over-contributions. Under certainty, L types' $NSOC_c$ is 0.38 EMU ($p = 0.0228$; $N = 21$) and under uncertainty their $NSOC_u$ is 0.18 EMU ($p = 0.2108$; $N = 24$). The $NSOC_c$ of H types is 0.30 EMU ($p = 0.0362$; $N = 21$) and their $NSOC_u$ is 0.26 EMU ($p = 0.0429$; $N = 24$). While under certainty both L and H types show significant net (strategic) over-contribution, under uncertainty only H types still over-contribute significantly. In other words, uncertainty causes only L types to reduce their net (strategic) over-contribution. The $\Delta NSOC$ for L types is -0.20 EMU ($p = 0.0936$; $N = 45$), but for H types it is only -0.04 EMU ($p = 0.8666$; $N = 45$). The difference of net (strategic) over-contribution between uncertainty and certainty is therefore only significant for L types but not for H types. It is interesting that H types do not try to mimic L types in order to increase their payoff. This suggests that subjects follow their elicited preference of relative conditional cooperation.

Figure 4.8A showed SOC_c and SOC_u for homogeneous and heterogeneous groups over the ten periods. We further outlined that ΔSOC captures two effects (i.e. $\Delta SOC = \Delta NSOC + \Delta DOB$), the effect of deviating beliefs (as discussed in Section 4.3.2) and the effect of net (strategic) over-contribution (as shown in Figure 4.8B). Both these effects are ultimately due to relative conditional cooperation preferences based on first- and second-order beliefs, respectively. Since our objective is to explain the results from the C-Experiment, we now have to compare ΔSOC to the treatment effects.

Figure 4.9: Figure 9: Explaining Treatment Effects



The gray bars in Figure 4.9A illustrate ΔSOC graphically for all group structures. The black bars in Figure 4.9A and 4.9B depict our treatment differences from the C-Experiment, namely $X_u^a - X_c^a$. Roughly speaking, they show that poor groups contribute less under uncertainty and rich groups contribute more than under certainty. Figure 4.9A shows that the ΔSOC captures the observed treatment effects reasonably well. In Figure 4.9B ΔSOC is split graphically into its components, $\Delta NSOC$ and ΔDOB. The shaded bars capture the amount explained by net (strategic) over-contribution and the white bars capture the effect of deviating beliefs.[32]

4.4 Conclusion

In this paper we investigated the effect of uncertain endowments on cooperation behaviour in a linear public goods game. Making endowments uncertain necessarily involves the possibility of heterogeneous in addition to homogeneous endowments. In our C-Experiment we found that results for heterogeneous endowments are in line with the existing literature. Homogeneous groups are more efficient in providing the public good than heterogeneous groups. We further showed that this is almost entirely due to the absence of (strategic) over-contribution in heterogeneous groups. Concerning uncertain endowments, our C-Experiment showed that uncertainty has only a slight negative effect (9.5% less) on cooperation. However, we found that this is caused by two effects

[32]Results are similar when separating into types. We observe a ΔSOC of -0.48 EMU ($p = 0.0134$; $N = 45$) for L types and 0.30 EMU ($p = 0.1924$; $N = 45$) for H types. ΔSOC is significantly different from zero only for L types.

that partly cancel each other. L types contribute substantially less and H types slightly more under uncertainty. Poor groups (with a majority of L types) therefore suffer more than rich groups (with a majority of H types) benefit from uncertainty. As a result, the pattern that prevailed under certainty, i.e. that homogeneous groups are more efficient than heterogeneous groups, breaks down under uncertainty. Here we find that the poorer a group the less efficient it is under uncertainty. Thus, despite the small overall effect, uncertainty causes the inequality of the income distribution to increase.

There are only a few other studies that investigate the effect of uncertain endowments. Isaac and Walker (1988) use a linear public goods game and find that mean contributions under certainty are 8.2% higher (though not significantly) than under uncertainty. Their interest in the effects of communication, however, may have confounded these results.[33] Chan et al. (1999) are also interested in the effects of communication. They use a non-linear public goods setting and find that uncertainty has a small but significant negative effect (7.4% less) on mean contributions.[34] Levati et al. (2007) use a linear public goods game to study the effects of leadership. They do not find any effect of uncertainty on contributions without a leader. However, in their uncertainty setting subjects knew the total group endowment and, since we find that subjects condition their own contribution on others' relative average contribution, we should not see any difference in contributions between certainty and uncertainty based on our results. Van Dijk and Grodzka (1992) use a one-shot step-level public goods setting where they are interested in contribution rules. All groups in their experiment consisted of two H and two L types. They find no effect of uncertainty on actual contributions. However, in their uncertainty treatment subjects were not informed about any endowment asymmetry. So, subjects probably believed that endowments were symmetrically distributed. This suggests that their uncertainty treatment rather reflects a certainty treatment with homogeneous endowments. We showed that the efficiency difference between homogeneous and heterogeneous endowments under certainty is only due to the absence of (strategic) over-contribution in heterogeneous groups. Since there is no (strategic) over-contribution in a one-shot public goods game, it is not surprising that van Dijk and Grodzka (1992) do not find an effect.

[33] Isaac and Walker (1988) find 18% higher levels of contributions with homogeneous than with heterogeneous endowments. These findings are also similar to our results under certainty.

[34] In their setup without communication and without preference heterogeneity, they also find 16% higher mean contributions with homogeneous than with heterogeneous endowments.

None of these studies has attempted to examine the opposing effects of uncertainty on groups and types. Moreover, we are not aware of any studies explaining the effects of uncertain and heterogeneous endowments in a way similar to our approach. We explain behaviour observed in the C-Experiment with *relative* conditional cooperation preferences elicited in the P-Experiment. A preference for relative conditional cooperation causes subjects in poor groups to contribute less and subjects in rich groups to contribute more under uncertainty than under certainty, since subjects in poor groups believe to be in a richer group and subjects in rich groups believe to be in a poorer group than they actually are. Moreover, L types may fear sending 'H-type signals' when they anticipate that others are relative conditional cooperators as well. As a result, they do not strategically over-contribute under uncertainty which causes poor groups to lose. In contrast to homogeneous groups, heterogeneous groups do not over-contribute under certainty which further explains their relative inefficiency.

Besides explaining behaviour that is observed in the C-Experiment, our results from the P-Experiment extend the growing literature on conditional cooperation. We found that subjects are on average *relative* conditional cooperators, and ascertained that the implicitly made assumption in the literature on conditional cooperation, i.e. that subjects condition their own contribution on the *averages* of others' contributions, is indeed reasonable. As noted for instance by Kocher (2008), the theoretical foundation for conditional cooperation is, however, still unclear. Future research should therefore try to incorporate existing findings on conditional cooperation and develop a theoretical foundation for a preference that seems to be empirically highly relevant in public goods games.

4.5 Appendix 4

Appendix 4A: Do Subjects Condition on Averages?

The seminal work by Fischbacher et al. (2001) that proposed the method we used in our P-Experiment in order to elicit conditional cooperation preferences, as well as other papers that used it (e.g. Kocher et al. 2008, Herrmann and Thöni 2009 or Fischbacher and Gächter 2010), ask subjects to state their own contributions conditional on the *averages* of others' contributions. However, it may well be possible that inequality in the individual contributions of others matters for own contribution levels.

For instance, suppose imperfect conditional cooperation behaviour was based on the model of inequity aversion by Fehr and Schmidt (1999). This would imply that own contributions are sensitive to the inequality of others' contributions. That this is the case, can already be seen in the linear specification of Fehr and Schmidt (1999). For simplicity, suppose groups have homogeneous endowments. Then, one's own predicted contribution is either zero or equal to the lowest of the others' individual contributions. This could reason *imperfect* conditional cooperation behaviour in cases where others contributed unequal amounts, since then one's own contribution would match the lowest of the other contributions, which is lower than the average of others' contributions by definition. In these cases the model of Fehr and Schmidt (1999) would predict a change in own contributions for increases in the inequality of others' contributions (holding their average contribution constant) since increases in inequality imply that the lowest of the other contributions changes.[35] A similar prediction could be made in the model of Charness and Rabin (2002) for situations where their model can generate (perfect or imperfect) conditional cooperation. The reason is that increasing inequality of others' contributions always decreases the income of *another worst off* person. The ERC model of Bolton and Ockenfels (2000) can also generate conditional cooperation. However, in their model inequality of others' contributions has no effect on own contributions, since it is only important how one's own payoff compares to the *average* payoff. Note that it is not our intention to test models of social preferences in the context of conditional cooperation.

[35]Note that own contributions would also change for increases in the inequality of others' contributions in cases where their model would predict *perfect* conditional cooperation, i.e. when others give the same amounts.

We only use these models as inspiration for the tests we perform in this appendix.

Instead of asking for subjects' own contributions conditional on the averages of others' contributions, in our P-Experiment we asked for own contributions conditional on every combination of others' individual contributions. This allows us to test whether the implied assumption of previous studies, namely that subjects condition their own contributions on the *averages* of others' contributions, can indeed be justified. For empirical research on conditional cooperation it is important to know whether the elicitation method of Fischbacher et al. (2001) is robust toward our design variation. Moreover, for future research on the theoretical foundations of conditional cooperation it seems important to know not only whether conditional cooperation is relative or absolute (which we showed in Section 4.3.2), but also, whether inequality increases in others' contributions matter for own contributions. Since all previous studies on conditional cooperation endowed all subjects with the same income, we focus in this appendix on homogeneous groups but additionally report the results for heterogeneous groups.[36]

From the data of our P-Experiment we know how much every subject contributes for any given combination of others' individual contributions for all combinations of others' endowments. We can therefore make multiple pairwise comparisons (i.e. matched pairs) of own contributions where we increase the inequality of others' individual contributions and, at the same time, hold their average contribution constant. As an example, consider a subject that contributes a certain amount given the other group members contribute 2 and 3 units. We then compare this amount to her contribution when the other group members contribute 1 and 4 units. We can make 7 such pairwise comparisons when the other group members are LL and 50 comparisons when they are HH.

In homogeneous poor groups (LLL) we find that none of the pairwise comparisons is significant.[37] Likewise, in homogeneous rich groups (HHH) none of the pairwise comparisons is significant.[38] Moreover, there is no tendency of subjects contributing more or less for increases in inequality since the z-values of the Wilcoxon signed rank sum tests

[36] Note that our explanation for the effects of heterogeneous and/or uncertain endowments in Section 4.3.2 is not based on the averages but on individual contributions of others.

[37] Using two-sided Wilcoxon signed rank sum tests we find $p = 0.182$ for $c(1,1)$ vs. $c(0,2)$, $p = 0.738$ for $c(1,2)$ vs. $c(0,3)$, $p = 0.658$ for $c(1,3)$ vs. $c(0,4)$, $p = 0.248$ for $c(2,2)$ vs. $c(0,4)$, $p = 0.680$ for $c(2,2)$ vs. $c(1,3)$, $p = 0.808$ for $c(2,3)$ vs. $c(1,4)$, and $p = 0.860$ for $c(3,3)$ vs. $c(2,4)$ with $N = 90$ and $c(x,y)$ being own contribution given the other group members contribute x and y.

[38] Using two-sided Wilcoxon signed rank sum tests we find $0.111 \leq p \leq 0.988$ with $N = 90$ for the 50 comparisons.

are negative *and* positive.[39] This suggests that the implicitly made assumption in the elicitation method of Fischbacher et al. (2001) can be justified. Making these multiple pairwise comparisons, we do not find any evidence that subjects change their contributions for increases in the inequality of others' contributions when holding the average contribution of others constant. As stated above, a model of inequity aversion with a linear specification would rather predict changes in own contributions. One can, however, construct a model of inequity aversion with a non-linear specification that would predict imperfect conditional cooperation where certain inequality increases would not have an effect. In such a model, there would be no effect of inequality increases as long as they do not lead to rank changes in the income distribution. For inequality increases leading to a rank change, the model would still predict changes in own contributions.

We therefore separate the pairwise comparisons into three classes. The first class consists of those comparisons that do not lead to a rank change. For instance, suppose a subject contributes 3 units, when others contribute 2 and 6 units. Then, this subject's rank does not change in the comparison of others contributing 1 and 7 units.[40] The second class consists of comparisons leading to negative rank changes. As an example, consider a subject contributing 1 unit given others contribute both 2 units. Then, there is a negative rank change for the comparison where others contribute 0 and 4 units. If this subject had contributed 3 instead of 1 units, there would have been a positive rank change. This is the third class of pairwise comparisons that we consider.

In homogeneous poor groups (LLL) there is only one potential situation (out of 7) which can generate a rank change. This is the comparison of $c(2,2)$ vs. $c(0,4)$. This comparison generates a negative rank change if $c(2,2) = 1$ and a positive rank change if $c(2,2) = 3$. Out of 90 subjects, 67 subjects contributed amounts that did not generate a rank change in this comparison, 9 subjects had a positive rank change, and 14 subjects had a negative rank change. Those with a positive rank change contribute significantly more and those with a negative rank change contribute significantly less (as predicted by the above mentioned model of inequity aversion).[41] However, since there is only one

[39]In LLL 2 are positive and 5 are negative. In HHH 12 are positive and 38 are negative.

[40]Note that we defined rank changes to be strict. For instance, there is no rank change in the comparison of $c(2,2) = 2$ vs. $c(1,3)$, but there is a rank change when comparing $c(2,2) = 1$ vs. $c(0,4)$.

[41]Using two-sided Wilcoxon signed rank sum tests we find for the comparison of $c(2,2)$ vs. $c(0,4)$ $p = 0.467$ with $N = 67$ if there is no rank change, $p = 0.083$ with $N = 9$ if there is a positive rank change, and $p = 0.046$ with $N = 14$ if there is a negative rank change.

potential situation generating a rank change in homogeneous poor groups and only few observations that in fact experience rank changes in this situation, we should not derive any conclusions from this result before having analyzed homogeneous rich groups.

In homogeneous rich groups (HHH) there are 22 pairwise comparisons (out of 50) that can possibly generate rank changes. Of those 22 situations that induce positive rank changes only one is significant.[42] Of the 22 situations inducing negative rank changes only two are significant.[43] Although the direction of change in the three significant situations is consistent with the predictions of the model, in the remaining 39 comparisons rank changes do not have a significant effect.[44] Again, even the z-values of the Wilcoxon signed rank sum tests are both negative and positive suggesting that there is not even a tendency that can be observed in these 39 non-significant cases.[45]

Moreover, in none of the 57 comparisons inducing no rank change, neither in poor (7 comparisons) nor in rich groups (50 comparisons), do we observe any significant effect of inequality increases. Overall, distinguishing between rank changes and no rank changes does not lead to any results different from those for the pooled data. Rank changes do not have an effect in virtually all cases. In those few cases where they have an effect, the direction of change is such that own contributions increase for higher ranks and decrease for lower ranks in the income distribution.

So far, our analysis used data which was aggregated over subjects and tested whether there is a prevalent pattern that can be observed. From this analysis we obtain the following picture. Holding the average contribution constant, increases in the inequality of others' contributions has on average no effect on own contributions, not even if these inequality increases lead to rank changes in the income distribution. However, since each subject makes multiple decisions where the average but not the inequality of others' con-

[42]Using two-sided Wilcoxon signed rank sum tests we find for the comparison of $c(2,2)$ vs. $c(0,4)$ $p = 0.189$ with $N = 68$ if there is no rank change, $p = 0.083$ with $N = 7$ if there is a positive rank change, and $p = 0.317$ with $N = 15$ if there is a negative rank change.

[43]Using two-sided Wilcoxon signed rank sum tests we find for the comparison of $c(3,4)$ vs. $c(0,7)$ $p = 0.240$ with $N = 60$ if there is no rank change, $p = 0.934$ with $N = 9$ if there is a positive rank change, and $p = 0.046$ with $N = 21$ if there is a negative rank change. For the comparison of $c(4,4)$ vs. $c(1,7)$ we find $p = 0.290$ with $N = 67$ if there is no rank change, $p = 1.000$ with $N = 4$ if there is a positive rank change, and $p = 0.080$ with $N = 19$ if there is a negative rank change

[44]Using two-sided Wilcoxon signed rank sum tests we find $0.317 \leq p \leq 1.000$ with $0 \leq N \leq 13$ in 21 situations inducing positive rank changes and $0.159 \leq p \leq 0.934$ with $6 \leq N \leq 27$ in 20 situation inducing negative rank changes.

[45]In situations inducing positive rank changes 5 are negative, 9 are positive, and 7 have $p = 1.000$. Concerning negative rank changes 15 are negative and 5 are positive.

tributions remains the same, we can further use Binomial tests to classify subjects on the individual level. Subjects are classified if their behaviour can be significantly distinguished from random behaviour based on the 5%- (1%-) significance level. Otherwise, they remain unclassified.

Our within-subject analysis yields a result similar to our across-subject analysis. In homogeneous poor groups (LLL), 77 (62) out of 90 subjects can be classified as contributing the same for inequality increases in others' contributions. No subject can be classified as contributing differently and 13 (28) subjects remain unclassified. In homogeneous rich groups (HHH), 79 (78) out of 90 subjects are classified as contributing the same, none are classified as changing their contribution, and 11 (12) remain unclassified.

Our results do not change when we consider heterogeneous instead of homogeneous groups. Out of 90 H types 69 (56) can be classified as contributing the same when the other group members are LL, none can be classified as contributing differently, and 21 (34) remain unclassified. With the other group members being LH 65 (53) H types contribute the same, none contributes differently, and 25 (37) remain unclassified. Out of 90 L types 87 (87) can be classified as contributing the same when the other group members are HH, none can be classified as contributing differently, and only 3 (3) remain unclassified. If the other group members are LH instead, 73 (68) L types contribute the same, none differently, and 17 (22) remain unclassified.

Neither in homogeneous nor in heterogeneous groups can we classify a single subject as changing her contribution for increases in the inequality of others' contributions. Our within subject analysis therefore supports the results from our across subject analysis. To conclude, this appendix showed that subjects condition their contribution on the *average* of others' contribution. Inequality in others' individual contributions has no effect on own contributions.

Appendix 4B: Instructions (translated from German)

General Instructions[46]

Thank you very much for your appearance. In the next 90 minutes you are going to take part in an experiment at the laboratory of MELESSA.
If you read the following instructions carefully, you can earn money depending on your decisions. The money earned during the experiment will be added to the show-up fee and paid out in cash at the end of the experiment.

During the experiment you are not allowed to communicate with the other participants. If you have any questions, please raise your arm. One of the experimenters will then come to your seat to answer your questions. You are not allowed to use your mobile phones or to open any programs on the computer. In the case of violation of these rules we have to exclude you from the experiment including all payments.

During the experiment we will refer to Experimental Monetary Units (EMU) and not to Euros. Your income will be calculated in EMU. At the end of the experiment, the total sum of earnings in EMU will be exchanged into Euros.

The exchange rate is: **1EMU = 10 Eurocents**.

In this experiment there are **two different types**: **L** and **H**. Type L and type H differ in their initial endowment. The total number of both types in the experiment is equal, hence there are as many L as there are H types.
At the beginning of the experiment you are randomly assigned a type. All participants have an equal chance to be of either type (L or H).
Then all participants are, again randomly, distributed into groups of three. None of the participants knows the identity of his or her group partners. Thus, all your decisions remain completely anonymous.

The Decision Situation:

We will explain the details of the experiment later. First, we want you to get familiar with the decision problem.

You are in a group together with two other participants (hence there are three group members in total). Every group member receives an initial endowment which depends on the type (L or H). Remember: the type is assigned randomly. Type L and type H have the following **initial endowments**.

Type L has an **initial endowment** of **4 EMU**.

[46] All participants; both types; both treatments.

Type H has an **initial endowment** of **8 EMU**.

All group members have to decide how many EMU of their initial endowment they want to transfer to a project. All EMU that are not transferred to the project are kept in private (by the respective participant). In the following we will describe how your total income is calculated.

Income from Private Property:

Every unit (EMU) of the initial endowment that has not been transferred to the project is kept in private property. For instance, if you are an L type (you have an initial endowment of 4 EMU) and decide to transfer 3 EMU to the project, 1 EMU will be kept as your private property. If you decide, for example, to transfer nothing (0 EMU) to the project, your private property will contain 4 EMU.

Income from the Project:

All group members benefit equally from the contributions made toward the project. This means that your income from the project is equal to that of the two other group members. Only the sum of contributions determines this income. Your income (and that of the two other group members) is calculated as follows.

Income from the project = Sum of all 3 contributions ∗ 0.6.

Example:
If you contribute 3 EMU to the project and the other group members contribute 1 EMU and 6 EMU, the sum of contributions is 10 EMU ($3 + 1 + 6 = 10$). Then, all group members receive 6 EMU ($10 * 0.6$) as their income from the project.

Total income:

Your total income is calculated as the sum of the income from private property and the income from the project:

Your private income (= Your endowment - Your contribution to the project)
+ Your income from the project (= 0.6 ∗ Sum of all contributions)
= Your total income.

The following **control questions** will help you to better understand the calculation of your total income.

Control Questions:

Please answer the following questions. They will help you to better understand the calculation of your income. Please fill in the solutions of all questions and raise your arm as soon as you have finished this task. Then, one of the experimenters will come to your seat in order to check your answers.

1. Assume that all three group members have an initial endowment of 4 EMU. Assume further that neither you nor any of the other group members has contributed anything to the project.

a) What will your total income be?_____________

b) What will the total income of each of the other two group members be? _____________

2. Assume that all three group members have an initial endowment of 8 EMU. Assume further that you have contributed 8 EMU and the two other group members have also contributed 8 EMU.

a) What will your total income be? _____________

b) What will the total income of each of the other two group members be? _____________

3. Assume that you have an initial endowment of 4 EMU. The two other group members have an initial endowment of 8 EMU. The two other group members contribute 5 EMU each.

a) What will your total income be if you contribute nothing (0 EMU)? _____________

b) What will your total income be if you contribute 2 EMU?_____________

c) What will your total income be if you contribute 4 EMU?_____________

4. Assume that all group members have an initial endowment of 4 EMU. Assume further that you contribute 2 EMU.

a) What will your total income be if the two other group members contribute the sum of 6 EMU?_____________

b) What will your total income be if the two other group members contribute the sum of 8 EMU?_____________

c) What will your total income be if the two other group members contribute the sum of 2 EMU?_____________

After you have answered all four control questions or in case you have any other questions, please raise your arm. Then, one of the experimenters will check your answers of the control questions and will answer further questions.

Instructions Part 1[47]

The Experiment

The experiment includes the decision situations that you have just become acquainted with. At the end of the experiment you will receive a payment according to your decisions.

The experiment consists of **two parts**. These two parts are completely independent from each other. Your decisions in the first part of the experiment have no influence on the process or the income of the second part of the experiment, and vice versa. During the experiment you have to confirm your decisions

[47]P-Experiment; H type; both treatments.

by pressing the “OK" button. Please note that decisions that have been confirmed are final and cannot be reversed.

You have been randomly assigned the type H. This means you are **type H for the entire duration of the experiment** and you can decide over the **initial endowment** of **8 EMU**.

The **procedure** of the experiment is the following:

1. Please read the instructions for the first part of the experiment carefully and put them on your desk in front of you (up-side down) when you have finished reading.

2. Next, you will exercise Part 1 of the experiment on the computer.

3. As soon as all participants have completed Part 1, you will receive the instructions for Part 2 of the experiment. Again, you should read them carefully and put them up-side down on your desk when you have understood the set-up and in case you have no further questions.

4. Then, you will complete Part 2 of the experiment on the computer.

5. As soon as all participants have finished Part 2, you will be asked to complete a questionnaire on the computer. Your answers to this questionnaire have no influence on your income.

6. At the end of the experiment you will get information about your income of both parts of the experiment.

In the following we describe the first part of the experiment in detail.

Instructions for Part 1

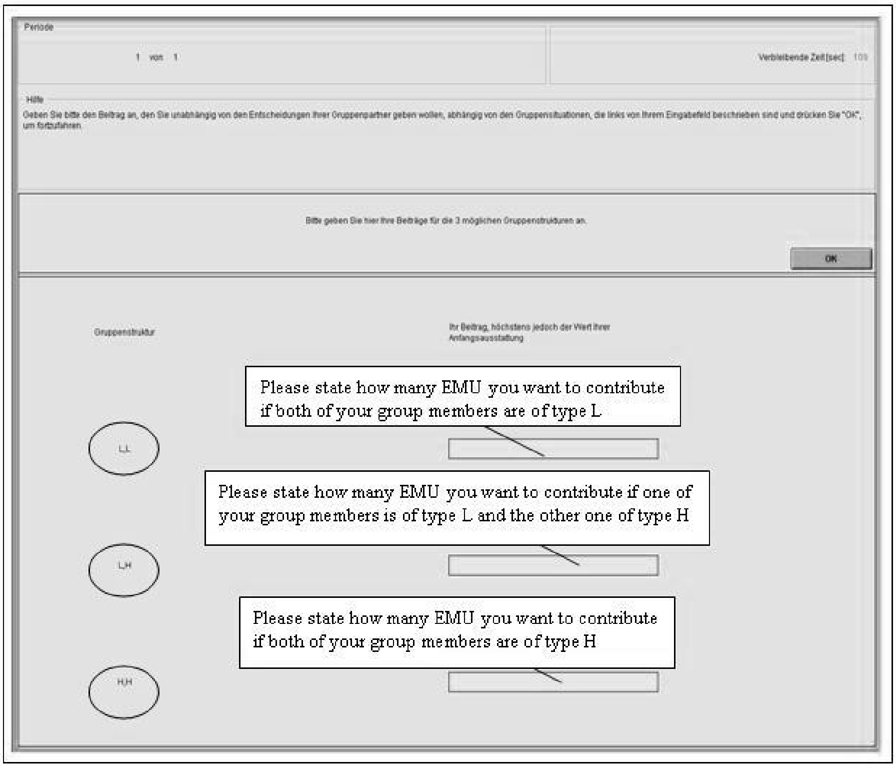
Since you do not know the types (L or H) of the other two members of your group, you may find yourself in one of three possible situations. The two other group members may be both of type L, they may both be of type H or one group member may be a L type and the other may be an H type.

The first part of the experiment consists of **two stages**:

Stage 1: unconditional decisions
Stage 2: conditional decisions

1. Unconditional Decisions:

Here you decide about your contribution to the project **independently** from the decisions of the other group members. In this table you have to state your contribution to the project for all three possible group structures in which you may find yourself (see above). Since all participants of the experiment decide simultaneously, neither you nor the two other participants in your group know the level of contributions of the other group members.



2. Conditional Decisions:

Here you decide about your contribution, **conditional** on the decisions of your group members. To do this you have to fill in the following three tables. For each of the three possible group structures you get one table. The rows in the tables tell you what the contributions of your group members are. You have to state in each row and column how much you want to contribute given the contributions of the other group members. This means you know how much the other group members have contributed before you decide about your own contributions in the different situations.

In the following example you see the table for the case where both other group members are of type L:

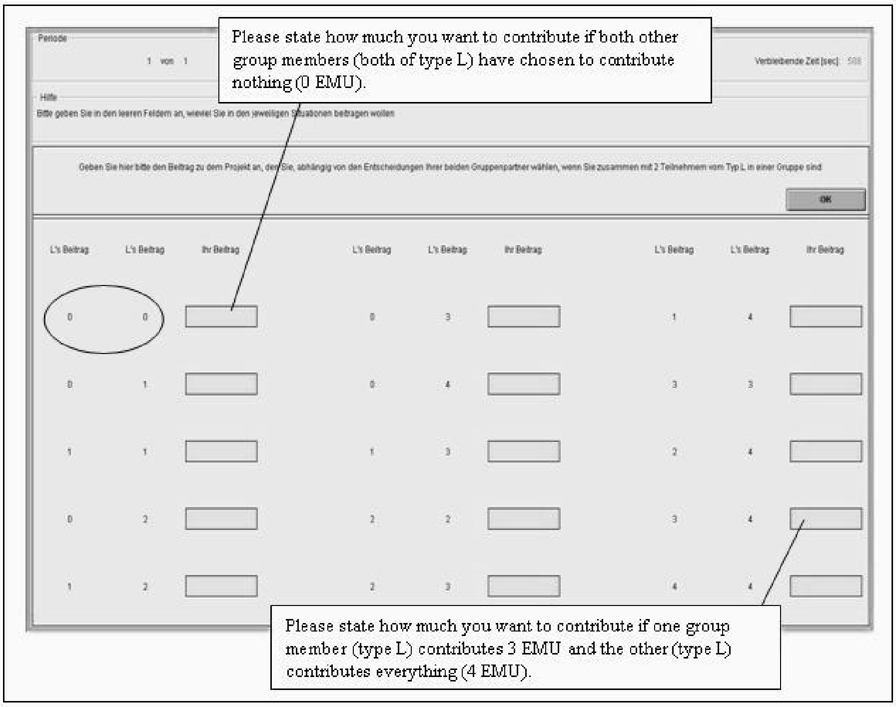
Two **Examples**:

1. In the first situation (upper left side of the screenshot) the two other group members (in this table both are of type L) have decided to contribute 0 EMU. Hence you have to decide how many EMU you want to contribute to the project (between 0 and 8, since you are of type H) and enter the amount into the respective cell.

2. In the second situation (lower right side) one of the group members has contributed 3 EMU, and the other has contributed 4 EMU. Again, you have to decide about your own contribution and enter the number into the respective cell.

These decisions have to be made for all presented situations.

Note that every single situation may be relevant for your income.



After all the participants have made all the decisions, conditional and unconditional, they will be allocated randomly into groups of three. As mentioned above, you may be matched with two group members of type L, both group members of type H, or one group member of type L and the other one of type H. Then, one member of the group is chosen randomly. For this participant, the three tables of the conditional contribution decisions are relevant (which of these three tables is relevant, depends on the randomly determined group structure). Which particular situation (entry) of the respective table is in fact relevant, depends now on the unconditional contributions of the other two group members for the randomly determined group structure.

So, the income of all three group members depends on the conditional contribution of the randomly chosen participant and on the unconditional contributions of the other two group members.

At the time when you make your decisions, you will not know, whether you will be one of the participants for whom the unconditional contribution is relevant or one for whom the conditional contributions are relevant. Therefore you should think carefully about every single decision (in both stages), since every one of them may determine your income (of this part of the experiment).

For an illustration consider the following **example**:

Assume you are randomly allocated to a group in which the other group members are of type L. Assume further that you are randomly chosen to be the participant for whom the conditional contribution is relevant. Consequently, the unconditional contribution is relevant for the other two group members. Assume that the unconditional contributions of the other two group members (both type L) are 2 EMU and 3 EMU for the relevant situation (for both, the situation is relevant in which one group member is of type L and the other is of type H). For yourself, the table for the group structure with two type L participants is relevant. Your income is now determined by the decision on your contribution in case one group member has contributed 2 EMU and the other has contributed 3 EMU. Assume you have chosen to contribute 5 EMU in this situation.

In this case, the sum of contributions is $2+3+5 = 10$ EMU. Hence, all group members receive $0.6*10 = 6$ EMU from the project. In total you receive 3 EMU from your private property ($8-5 = 3$ EMU) and 6 EMU from the project leading to the sum of 9 EMU.

The group member (type L) who has contributed 2 EMU receives 2 EMU from his/her private property ($4-2 = 2$ EMU) and 6 EMU from the project and hence has a total income of 8 EMU.

The other group member (type L), who has contributed 3 EMU, receives 1 EMU from his/her private property ($4-3 = 1$ EMU) and 6 EMU from the project, making it a total income of 7 EMU.

If you have any further questions, please raise your arm. An experimenter will then come to your seat in order to answer the questions. If you have no further questions, please put the instructions up-side down on your desk.

Instructions for Part 2[48]

In the second part of the experiment, the now familiar decision situation is repeated 10 times (structured in 10 periods).

You are still **type H** and have an initial endowment of **8 EMU** (you will receive these 8 EMU in each period).

You are still together with the two other group members from the first part of the experiment. You still do not know the types of the other group members. You will stay together with these two participants over all 10 periods.

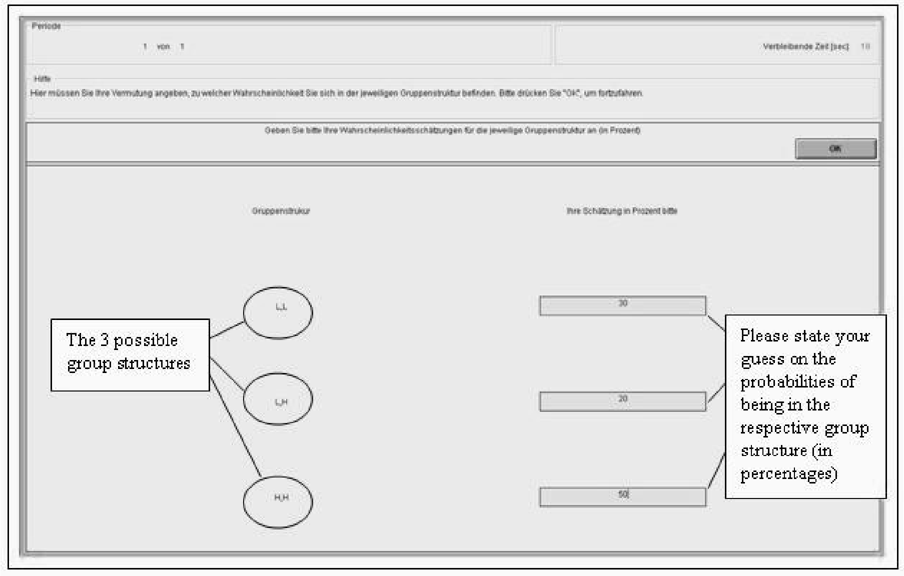
The decision situation remains unchanged. You decide in each period about the allocation of your initial endowment of 8 EMU. The calculation of your income remains also unchanged and is repeated in the following:

Private income (= 8 - contribution to the project)
+ Income from the project (= $0.6 *$ Sum of all contributions)
= Total income.

At the beginning of each period, we will ask you to indicate your expectations about the initial endowments of the other two group members. So, you are required to make a guess about the type of the other participants in your group. There are three possible compositions of your group: you are together with two participants of type L, with two participants of type H, or with one participant of type H and one of type L. Please allocate probabilities (in percentage points) to each of these group structures.

For example, if you are 100% sure that you are in a group in which both other group members are of type L, you will fill in 100 in the first row and 0 in each of the other two rows.

[48] C-Experiment; H type; uncertainty treatment

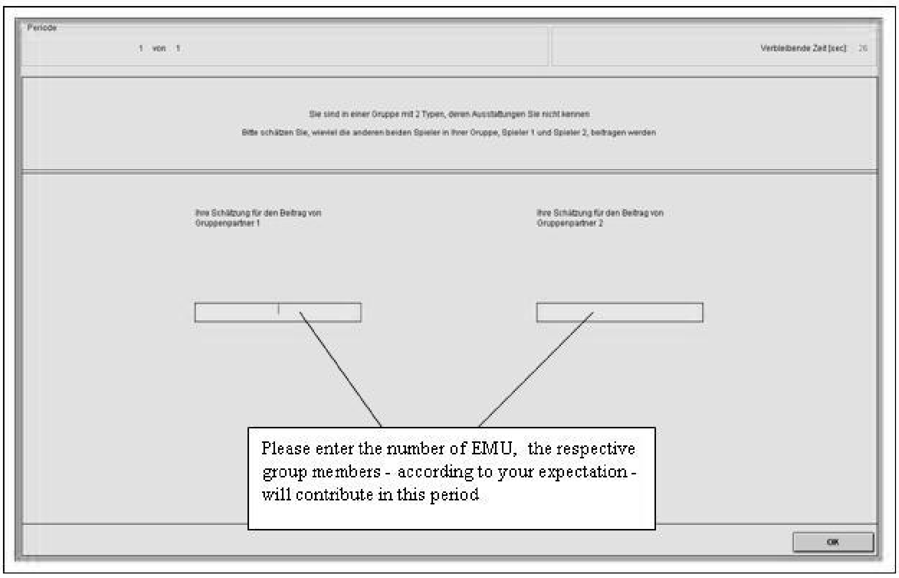


If you think the chances are 50% to be in a group in which both other group members are of type L, 25% to be in a group in which one group member is of type L and the other is of type H, and 25% to be in a group with two participants of type H, you will fill in 50 in the first, 25 in the second and 25 in the third row.

Note that the probabilities of your expectations (the numbers you fill in) have to add up to 100.

If you believe that the probability to be in each of the three group structures is equal, please fill in 34, 33, and 33 in the three rows.

Moreover, you are asked how many EMU you think the other two group members will contribute in the respective period. So, you have to guess how many EMU each of your group members will contribute.



You will be paid for the accuracy of this guess in the following way:

- If your expectation is completely correct, you will receive 2 EMU (for each group member).
- If your expectation deviates by not more than 1 EMU from the real value, you will receive 1 EMU (for each group member).
- If your expectation deviates by 2 EMU or more, you will receive nothing.

In a last step you have to decide how many EMU of your initial endowment of 8 EMU (which you receive at the beginning of each of the 10 periods) you actually want to contribute to the project in the respective period.

As soon as all group members have stated their decisions, you will be shown a table that states how many EMU all other group members have contributed and how many EMU you have earned in the respective period (excluding the income from correct expectations about the group structure). Then, a new period starts.

All group members get the same information about the contributions of the other group members.
None of the group members receives information about the type (L or H) of any of the other group members.

The income of all periods is summed up. After the 10^{th} period you receive the payment in private and separately from other participants.

The **final income** then consists of the following:

1. Show-up fee of 4 Euros.
2. The income from the first part of the experiment.
3. The sum of incomes from the 10 periods of the second part of the experiment, including the payments for correct expectations.

If you have any questions, please raise your arm. An experimenter will come to your seat in order to answer your questions. If you do not have any further questions, please put the instructions up-side down on your desk, so that we can start with the second part of the experiment on the computer.

Bibliography

Abbink, K. (2004), "Staff Rotation as an Anti−corruption Policy: an Experimental Study", European Journal of Political Economy VOL 20, pp 887−906

Abbink, K. (2006), "Laboratory Experiments on Corruption" in Rose-Ackerman, S. v. 'International Handbook on the Economics of Corruption', Cheltenham, pp 418−37

Abbink, K., H. Hennig-Schmidt (2006), "Neutral versus Loaded Instructions in a Bribery Experiment", Experimental Economics VOL 9/2, pp 103−21

Abbink, K., B. Irlenbusch, E. Renner (2000), "The Moonlighting Game", Journal of Economic Behavior and Organization VOL 22, pp 265−77

Abbink, K., B. Irlenbusch, E. Renner (2002), "An Experimental Bribery Game", Journal of Law, Economics and Organization VOL 18/2, pp 428−54

Aidt, T. (2003), "Economic Analysis of Corruption: A Survey", The Economic Journal VOL 113, pp 632−52

Alatas, V., L. Cameron, A. Chaudhuri, N. Erkal, L. Gangadharan (2006), "Gender and Corruption: Insights from an Experimental Analysis", Department of Economics − Working Papers Series 975, The University of Melbourne

Alesina, A., E. La Ferrara (2000), "Participation in Heterogeneous Communities", Quarterly Journal of Economics VOL 115/3, pp 847−904

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, R. Wacziarg (2003), "Fractionalization", Journal of Economic Growth VOL 8/2, pp 155−94

Anderson, L., C. Holt (1997), "Information Cascades in the Laboratory", American Economic Review VOL 87, pp 847−62

Anderson, L., J. Mellor, J. Milyo (2008), "Inequality and Public Good Provision: An Experimental Analysis", Journal of Socio-Economics VOL 37/3, pp 1010−28

Andreoni, J., J. Miller (2002), "Giving according to GARP: An experimental Test of the Consistency of Preferences for Altruism", Econometrica VOL 70, pp 737−53

Andreoni, J., L. Vesterlund (2001), "Which is the Fair Sex? Gender Differences in Altruism", Quarterly Journal of Economics VOL 116, pp 293−312

Armantier, O., A. Boly (2008), "Can Corruption Be Studied in the Lab? Comparing a Field and a Lab Experiment", Cirano Working Paper Series, Quebec

Aquino, K., V. Steisel, A. Kay (1992), "The Effects of Resource Distribution, Voice, and Decision Framing on the Provision of Public Goods", Journal of Conflict Resolution VOL 36/4, pp 665−87

Axelrod, R. (1984), "The Evolution of Cooperation", Basic Books, New York

Bardhan, P. (1997), "Corruption and Development: A Review of Issues", Journal of Economic Literature VOL 35/3, pp 1320−46

Becker, G. (1968), "Crime and Punishment: an Economic Approach", Journal of Political Economy VOL 78, pp 169−217

Berg, J., J. Dickhaut, K. McCabe (1995), "Trust, Reciprocity, and Social History", Games and Economic Behavior VOL 10, pp 122−42

Bertrand, M., S. Djankov, R. Hanna, S. Mullainathan (2007), "Obtaining a Driver's License in India: An Experimental Approach to Studying Corruption", The Quarterly Journal of Economics VOL 122/4, pp 1639−76

Bishop G., D. Myers (1974), "Informational Influences in Group Discussion", Organizational Behavior and Human Performance VOL 12, pp 92−104

Blinder, A., J. Morgan (2005), "Are Two Heads Better than One? An Experimental Analysis of Group vs. Individual Decision Making", Journal of Money Credit and Banking VOL 37, pp 789−812

Bochet, O., T. Pagea, L. Puttermana (2006), “Communication and Punishment in Voluntary Contribution Experiments”, Journal of Economic Behavior & Organization VOL 60/1, pp 11−26

Bohn, F. (2004), “White Elephants and the Limits to Efficient Investment”, Centre for Economic Research Working paper Series, WP04/13

Bolton, G., A. Ockenfels (2000), “ERC: A Theory of Equity, Reciprocity, and Competition”, American Economic Review VOL 90/1, pp 166−93

Bone J., J. Hey, J. Suckling (1999), “Are Groups more (or less) Consistent than Individuals?”, Journal of Risk and Uncertainty VOL 8, pp 63−81

Bonein, A., D. Serra (2007), “Another Experimental look at Reciprocal Behavior: indirect reciprocity”, MPRA Paper No. 3257

Bornstein G., I. Yaniv (1998), “Individual and Group Behavior the Ultimatum Game: are Groups more ‘Rational’ Players?”, Experimental Economics VOL 1, pp 101−08

Bossert, W., C. D’Ambrosio, E. Laferrara (2006), “A Generalized Index of Fractionalization”, CEPR, NBER Working Paper Series

Botelho, A., G. Harrison, L. Costa Pinto, E. Rutström (2009), “Testing Static Game Theory with Dynamic Experiments: A Case Study of Public Goods”, Games and Economic Behavior VOL 67/1, pp 253−65

Brandts, J., A. Schramm (2001), “Cooperation and Noise in Public Goods Experiments: Applying the Contribution Function Approach”, Journal of Public Economics VOL 79/2, pp 399−427

Buccirossi, P., G. Spagnolo (2006), “Leniency Policies and Illegal Transactions”, Journal of Public Economics VOL 90/6;7, pp 1281−97

Büchner S., A. Freytag, L. G. González, W. Güth (2008), “Bribery and Public Procurement: an Experimental Study”, Public Choice VOL 137, pp 103−11

Burnstein E., A. Vinokur, Y. Trope (1973), “Interpersonal Comparison versus Persuasive Argument: a more Direct Test of Alternative Explanations for Group-induced Shifts in Individual Choices”, Journal of Social Psychology VOL 9, pp 236−45

Buckley, E., R. Croson (2006), "Income and Wealth Heterogeneity in the Voluntary Provision of Linear Public Goods", Journal of Public Economics VOL 90/4;5, pp 935−55

Cardenas, J.-C. (2003), "Real Wealth and Experimental Cooperation: Experiments in the Field Lab", Journal of Development Economics VOL 70/2, pp 263−89

Cason, T., V. Mui (1997), "A Laboratory Study of Group Polarization in the Team Dictator Game", Economic Journal VOL 107, pp 1465−83

Chalos P., S. Pickard (1985), "Information Choice and Cue use: An experiment in Group Information Processing", Journal of Applied Psychology VOL 70, pp 634−41

Chan, S., S. Mestelman, R. Moir, R. Muller (1999), "Heterogeneity and the Voluntary Provision of Public Goods", Experimental Economics VOL 2/1, pp 5−30

Charness, G., M. Rabin (2002), "Understanding Social Preferences with Simple Tests", Quarterly Journal of Economics VOL 117, pp 817−69.

Chen, Y., S. Xin Li (2009), "Group identity and Social Preferences", American Economic Review VOL 99/1, pp 171−78

Cherry, T., S. Kroll, J. Shogren (2005), "The Impact of Endowment Heterogeneity and Origin on Public Good Contributions: Evidence from the Lab", Journal of Economic Behavior and Organization VOL 57/3, pp 357−65

Cox, J. (2002), "Trust, Reciprocity, and Other−Regarding Preferences: Groups vs. Individuals and Males vs Females" in Zwick R., A. Rapoport 'Experimental Business Research', Kluwer Academic Publishers, pp 331−49

Cox, J., C. Friedman, D. Friedman, S. Gjerstad, (2007), "A Tractable Model of Reciprocity and Fairness", Games and Economic Behavior VOL 59/1, pp 17−45

Croson, R., N. Buchan (1999), "Gender and Culture: International Experimental Evidence from Trust Games", Gender and Economic Transactions VOL 89/2, pp 386−91

Croson, R., U. Gneezy (2009), "Gender Differences in Preferences", Journal of Economic Literature VOL 47/2, pp 448−74

Dekel, E., S. Scotchmer (1999), "On the Evolution of Attitudes towards Risk in Winner−Take−All Games", Journal of Economic Theory VOL 87, pp 125−43 della Porta, D., A. Vanucci (1999), “Corrupt Exchanges, Actors, Resources and Mechanisms of Political Corruption”, de Gruyter, New York

Dincer, O. (2008), “Ethnic and Religious Diversity and Corruption”, Economic Letters VOL 99/1, pp 98−102

Dollar, D., R. Fisman , R. Gatti (2001), “Are Women Really the Fairer Sex? Corruption and Women in Government”, Journal of Economic Behavior and Organization VOL 46, pp 423−29

Drew, K. (2003), “Whistle-blowing and Corruption, an Initial and Comparative Review”, PSIRU, London, available at: www.psiru.org

Dufwenberg, M., U. Gneezy (2000), “Measuring Beliefs in an Experimental Lost Wallet Game”, Games and Economic Behavior VOL 30, pp 163−82

Dufwenberg, M., G. Kirchsteiger (2004), “A Theory of Sequential Reciprocity”, Games and Economic Behaviour VOL 47, pp 268−98

Dusek, L., A. Ortmann, L. Lizal (2004), “Understanding Corruption and Corruptibility through Experiments”, Prague Economic Papers VOL 2, pp 147−62

Eckel, C., P. Grossmann (1996), “The Relative Price of Fairness: Gender Differences in a Punishment Game”, Journal of Economic Behavior and Organization VOL 30/2, pp 143−58

Egger P., H. Winner (2006), “How Corruption Influences Foreign Direct Investment: A Panel Data Study”, Economic Development and Cultural Change VOL 54/2, pp. 459−86

Engelmann, D., M. Strobel (2006), “Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments”, American Economic Review VOL 94/4, pp 857−69

Falk, A., U. Fischbacher (2006), “A Theory of Reciprocity”, Games and Economic Behavior VOL 54/2, pp 293−315

Fehr, E., U. Fischbacher (2003), “The Nature of Human Altruism”, Nature VOL 425, pp 785−91

Fehr, E., S. Gächter (1998), "Reciprocity and Economics: The Economic Emplications of Homo Reciprocans", European Economic Review VOL 42, pp 845−59

Fehr, E., S. Gächter (2000), "Cooperation and Punishment in Public Goods Experiments", American Economic Review VOL 90/4, pp 980−94

Fehr, E., S. Gächter (2000), "Fairness and Retaliation: The Economics of Reciprocity", Journal of Economic Perspectives VOL 14/3, pp 159−81

Fehr, E., S. Gächter, G. Kirchsteiger (1997), "Reciprocity as a Contract Enforcement Device: Experimental Evidence", Econometrica VOL 65/4, pp 833−60

Fehr, E., G. Kirchsteiger, A. Riedl (1993), "Does Fairness Prevent Market Clearing? An Experimental Investigation", Quarterly Journal of Economics VOL 108, pp 437−59

Fehr, E., K. Schmidt (1999), "Theory of Fairness, Competition and Cooperation", Quarterly Journal of Economics VOL 114, pp 817−68

Fehr, E., K. Schmidt (2003), "Theories of Fairness and Reciprocity−Evidence and Economic Applications", in Dewatripont, M., L.P. Hansen, S.J. Turnovsky, 'Advances in economics and econometrics', 8th World Congress, Economic Society Monographs, pp 208−57

Fehr, E., K. Schmidt (2006), "The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories" in 'Handbook on the Economics of Giving, Reciprocity and Altruism' VOL 1, pp 615−91

Fischbacher, U. (2007), "Z-Tree: Zurich Toolbox for Ready-made Economic Experiments", Experimental Economics VOL 10/2, pp 171−8

Fischbacher, U., S. Gächter (2010), "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments", American Economic Review VOL 100/1, pp 541−56

Fischbacher, U., S. Gächter, E. Fehr (2001), "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment", Economics Letters VOL 71/3, pp 397−404

Fudenberg, D., J. Tirole (1991), "Game Theory", The MIT Press, Cambridge Massachusetts/London England

Fukuyama, F. (1995), "Trust: the Social Virtues and the Creation of Prosperity", Free Press, New York

Fukuyama, F. (1999), "Social Capital and Civil Society", IMF Conference on Second Generation Reforms, www.imf.org

Gächter, S. (2007), "Conditional Cooperation: Behavioral Regularities from the Lab and the Field and their Policy Implications" in Frey, B.S., A. Stutzer, 'Economics and Psychology. A Promising New Cross-Disciplinary Field', MIT Press, Cambridge, MA, pp 19−50

Gächter, S., A Falk (2002), "Reputation and Reciprocity: Consequences for the Labour Relation", Scandinavian Journal of Economics VOL 104, pp 1−26

Gneezy, U., J. A. List (2006), "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments", Econometrica VOL 74, pp 1365−84

Greif, A. (1993), "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition", American Economic Review VOL 83/3, pp 525−48

Greif, A. (2006), "Institutions and the Path to the Modern Economy: Lessons from Medieval Trade", Cambridge University Press

Greiner, B. (2004), "An Online Recruitment System for Economic Experiments" in Kremer, K., V. Macho, 'Forschung und wissenschaftliches Rechnen', GWDG Bericht 63, Ges. für Wiss. Datenverarbeitung, Göttingen, pp 79−93

Groenendijk, N. (2004), "A Principal-agent Model of Corruption", Crime, Law and Social Change VOL 27/3-4, pp 207−29

Güth, W., B. Schwarze (1982), "An Experimental Analysis of Ultimatum Bargaining", Journal of Economic Behavior and Organization VOL 3/4, pp 367−88

Gunasekara, V.L (2008), "The Political Economy of Ethnic Fragmentation: A Preliminary Study that Explores the Dynamic Interactions Among Ethnic Fragmentation, Trust, Institutions, and Corruption", Paper presented at the annual meeting of the MPSA Annual National Conference, Chicago

Guriev, S. (2004), “Red Tape and Corruption”, Journal of Development Economics VOL 73/2, pp 489−544

Hall, D., S. Davis (1999), “Corruption and Whistle Blower − a background note for TUAC”, PSIRU, London, Available at www.psiru.org/reportsindex.htm

Harris, D. (2007), “Bonding Social Capital and Corruption: A Cross-National Empirical Analysis”, Working Paper at the Department of Land Economy, University of Cambridge

Heckman, J. (1979), “Sample Selection Bias as a Specification Error”, Econometrica VOL 47, pp 153−61

Herrmann, B., Chr. Thöni (2009). “Measuring Conditional Cooperation: A Replication Study in Russia”, Experimental Economics VOL 12/1, pp 87−92

Hung, A., C. Plott (2001), “Information Cascades: Replication and an Extension to Majority Rule and Conformity−Rewarding Institutions”, American Economic Review VOL 91/5, pp 1508−20

Huntington, S., (1968), “Political Order in Changing Societies”, Yale University Press, New Haven

Hussein, A. (2005), “How the New Law on Public Procurement Implements the Anti-corruption Recommendations of the Supreme Chamber of Control” in ‘Fighting Corruption and Promoting Integrity in Public Procurement’, OECD Publishing, Paris

Isaac, M., J. Walker (1988), “Communication and Free-Riding Behavior: The Voluntary Contribution Mechanism”, Economic Inquiry VOL 26/4, pp 585−608

Jacquemet, N. (2005), “Corruption as Betrayal: Experimental Evidence”, GATE Working Paper Nr 2005-06

Kandori, M. (1992), “Social Norms and Community Enforcement”, The Review of Economic Studies VOL 59/1, pp 63−80

Keser, C., F. van Winden (2000), “Conditional Cooperation and Voluntary Contributions to Public Goods”, Scandinavian Journal of Economics VOL 102/1, pp 23−39

Kingston, C. (2007), “Parochial Corruption”, Journal of Economic Behavior and Organization VOL 63/ 1, pp 73−87

Klitgaard, R. (1988), "Controlling Corruption", University of California Press

Knapp, W. (1986), "Sittenwidrige Vermittlung von Regierungsaufträgen durch Bestechungsgelder", Recht der Internationalen Wirtschaft, RIW, pp 999−1001

Kocher, M. (2008), "Robustness of Conditional Cooperation in Public Goods Experiments", Mimeo

Kocher M., M. Sutter (2005), "The Decision Maker Matters. Individual versus Team Behavior in Experimental Beauty-contest Games", Economic Journal VOL 115, pp 200−23

Kocher M., S. Strauss, M. Sutter (2006), "Individual or team decision making − Causes and consequences of self-selection", Games and Economic Behavior VOL 56, pp 259−70

Kocher, M., T. Cherry, S. Kroll, R. Netzer, M. Sutter (2008), "Conditional Cooperation on Three Continents", Economics Letters VOL 101/3, pp 175−8

Kocher M., M. Sutter (2007), "Individual versus Group Behavior and the Role of the Decision making Procedure in Gift−Exchange Experiments", Empirica VOL 34, pp 63−88

Krajcova, J., A. Ortmann (2008), "Testing Leniency Programs Experimentally: The Impact of 'Natural' Framing", Mimeo, CERGE-EI, Prague

Kreps D, P. Milgrom, J. Roberts, R. Wilson (1982), "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma", Journal of Economic Theory VOL 27, pp 245−52

Kugler, T., G. Bornstein, M. Kocher, M. Sutter (2007), "Trust Between Individuals and Groups: Groups are Less Trusting than Individuals but just as Trustworthy", Journal of Economic Psychology VOL 28, pp 646−57

Lambert-Mogiliansky, A., M. Majumdar, R. Radner (2006), "Petty Corruption: a Game−theoretic Approach", International Journal of Economic Theory VOL 4/2, pp 273−97

Lambsdorff, J. Graf (2002), "Corruption and Rent−Seeking", Public Choice, VOL 113/2, pp 97−125

Lambsdorff, J. Graf (2007), "The Institutional Economics of Corruption and Reform, Theory, Evidence and Policy", Cambridge University Press

Lambsdorff, J. Graf, B. Frank (2007), "Corrupt Reciprocity − an Experiment", Diskussionsbeitrag Nr. V−51−07, Volkswirtschaftliche Reihe ISSN 1435−3520

Lambsdorff, J. Graf, B. Frank (2010), "Bribing versus Gift-giving − An Experiment", Journal of Economic Psychology VOL 31/3, pp 347−57

Lambsdorff, J. Graf, B. Frank (forthcoming), "Gender Effects in Laboratory Corruption Experiments", Forthcoming in European Journal of Development Research

Lambsdorff, J. Graf, M. Nell (2007), "Fighting Corruption with Asymmetric Penalties and Leniency", CEGE Discussion Paper ISSN 1439−2305

Lambsdorff, J. Graf, M. Taube, M. Schramm (2005), "The New Institutional Economics of Corruption", Routledge, London and New York

La Porta R., F. Lopez de Silanes, A. Shleifer, R. Vishny (1999), "The Quality of Government", Journal of Law, Economics and Organization VOL 15/1, pp 222−79

Ledyard, J. (1995), "Public Goods: A Survey of Experimental Research" in Kagel, J., A. Roth, 'The Handbook of Experimental Economics', Princeton University Press, Princeton, NJ, pp 111−94

Leff, N. (1964), "Economic Development through Bureaucratic Corruption", American Behavioral Scientist VOL 8/3, pp 8−14

Leigh, A. (2006), "Trust, Inequality and Ethnic Heterogeneity", Discussion paper No. 511, The Australian National University, ISSN: 1442-8636

Levati, V., M. Sutter, E. van der Heijden (2007), "Leading by Example in a Public Goods Experiment with Heterogeneity and Income Information", Journal of Conflict Resolution VOL 51/5, pp 793−818

Levine, J., R. Moreland (1998), "Small groups" in 'Handbook of social psychology' VOL 2/4, pp 415−69

Lipset, S. M., G. S. Lenz (2000), "Corruption, Culture, and Markets", in Harrison, L., S. Huntington 'Culture Matters. How Values shape Human Progress', Basic Books, New York, pp 112−24

Marwell, G., R. Ames (1979), "Experiments on the Provision of Public Goods. Resources, Interest, Group Size, and the Free-Rider Problem", American Journal of Sociology VOL 84/6, pp 1335−60

Mauro, P. (1995), "Corruption and Growth", Quarterly Journal of Economics VOL 110, pp 681−712

Nowak, M., K. Sigmund (1998), "Evolution of Indirect Reciprocity by Image Scoring", Nature VOL 393, pp 573−77

Ockenfels, A. (1999), "Fairness, Reziprozität und Eigennutz – Ökonomische Theorie und Experimentelle Evidenz. Die Einheit der Gesellschaftswissenschaften", Bd. 108, Mohr Siebeck, Tübingen

OECD (2010), "Guidelines for Managing Conflict of Interest in the Public Service", http://www.oecd.org/ dataoecd/ 31/15/ 36587312.pdf

Ogilvie, S. (2004), "The Use and Abuse of Trust: Social Capital and Its Deployment by Early Modern Guilds", CESifo Working Paper No. 1302, Munich

Pechlivanos, L. (1997), "Self-Enforcing Corruption and Optimal Deterrence", GREMAQ Working Paper Series 97.18.460

Pereira, P.T., N. Silva, J. Andrade e Silva (2006), "Positive and Negative Reciprocity in the Labor Market", Journal of Economic Behavior & Organization VOL 59, pp 406−22

Pörting, P., W. Vahlenkamp (1998), "Internal Strategies Against Corruption: Guidelines for Preventing and Combatting Corruption in Police Authorities", Crime Law and Social Change VOL 29, pp 225−49

Pruitt, D.G. (1971), "Choice Shifts in Group Discussion: An Introductory Review", Journal of Personality and Social Psychology VOL 20/3, pp 339−60

Putnam, R.D. (2000), "Bowling Alone: The Collapse and Revival of American Community", New York: Simon and Schuster

Rabin, M. (1993), "Incorporating Fairness into Game Theory and Economics", American Economic Review VOL 83, pp 1281−1302

Rapoport, A. (1988), "Provision of Step-Level Goods: Effects of Inequality in Resources", Journal of Personality and Social Psychology VOL 54/3, pp 432−40

Rapoport, A., R. Suleiman (1993), "Incremental Contribution in Step-Level Public Goods Games with Asymmetric Players", Organizational Behavior and Human Decision Processes VOL 55/2, pp 171−94

Realo, A., J. Allik, B. Greenfield (2008), "Radius of Trust: Social Capital in Relation to Familism and Institutional Collectivism", Journal of Cross-Cultural Psychology VOL 39, pp 446−62

Rieger, H.-J. (2005), "Prevention − a Key Factor in Fighting Corruption", United Nations Public Administration Network

Rivas, M. (2007), "An Experiment on Corruption and Gender", Mimeo

Rose-Ackerman, S. v. (1999), "Corruption and Government, Causes, Consequences and Reform" in Lambsdorff, J. Graf 'The New Institutional Economics of Corruption', Cambridge University Press 1999

Rose-Ackerman, S. v. (2006), "International Handbook on the Economics of Corruption", Marston Book Services Limited, Abingdon, UK

Schramm, M., M. Taube (2001), "Institutionenökonomische Anmerkungen zur Einbettung von Korruption in das Ordnungssystem Chinesischer Guanxi−Netzwerke", Duisburg Working Papers on East Asian Economic Studies

Schubert, R., M. Brown, M. Gysler, H.W. Brachinger (2009), "Financial Decision-Making: Are Women Really More Risk-averse?", The American Economic Review VOL 89/2, pp 381−5

Schulze, G., B. Frank (2003), "Deterrence versus Intrinsic Motivation: Experimental Evidence on the Determinants of Corruptibility", Economics of Governance VOL 4/2, pp 143−60

Seinen, I., A. Schram (2006), "Social Status and Group Norms: Indirect Reciprocity in a Repeated Helping Experiment", European Economic Review VOL 50/3, pp 581−602

Selten, R. (1967), "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments" in Sauermann H. 'Beiträge zur experimentellen Wirtschaftsforschung' VOL 1, pp 136−68

Shleifer, A., R. Vishny (1993), "Corruption", Quarterly Journal of Economics VOL 108/3, pp 599−617

Sonnemans, J., A. Schramm, T. Offerman (1999), "Strategic Behavior in Public Goods Games: When Partners Drift Apart", Economics Letters VOL 62/1, pp 35−41

Spagnolo, G. (2006), "Leniency and Whistleblowers in Antitrust", CEPR Discussion Paper 5794

Strassmair, C. (2009), "Can intentions spoil the kindness of a gift? − An experimental study", Munich Discussion Paper No. 2009−4 Department of Economics

Sung, H.E. (2003), "Fairer Sex or Fairer System? Gender and Corruption Revisited", Social Forces VOL 82, pp 703−23

Sutter, M., S. Haigner, M. Kocher (2010, forthcoming), "Choosing the Stick or the Carrot? Endogenous Institutional Choice in Social Dilemma Situations", Review of Economic Studies

Swamy, A., S. Knack, Y.Lee, O. Azfar (2001), "Gender and Corruption", Journal of Development Economics VOL 64, pp 25−55

Tajfel, H., J.C. Turner (1986), "The Social Identity Theory of Intergroup Behavior", in Worchel, S., W.G. Austin 'Psychology of intergroup relations', Chicago, pp 7−24

The World Bank (2008), www.worldbank.org/publicsector/anticorrupt

Tonoyan, V. (2005), "The Dark Side of Trust: Corruption and Entrepreneurship" in Höhmann, H., F. Welter, 'Trust and Entrepreneurship: A West-East-Perspective', Cheltenham, Edward Elgar, pp 39−58

Transparency International (2005), "Transparency International Global Corruption Report 2005", www.transparency.de

Van Dijk, E., M. Grodzka (1992), "The Influence of Endowments Asymmetry and Information Level on the Contribution to a Public Step Good", Journal of Economic Psychology VOL 13/2, pp 329−42

Van Dijk, E., H. Wilke (1994), "Asymmetry of Wealth and Public Good Provision", Social Psychology Quarterly VOL 57/4, pp 352−9

Wiehen, M. (2005), "Global Corruption Report 2005", Transparency International, Pluto Press, London and Ann Harbor

Wit, A., H. Wilke, H. Oppewal (1992), "Fairness in Asymmetric Social Dilemmas" in Liebrand, W., D. Messick, H. Wilke, 'Social Dilemmas: Theoretical Issues and Research Findings', Pergamon, New York, pp 183−97

World Values Survey (2008), www.worldvaluessurvey.org

Zelmer, J. (2003), "Linear Public Goods Experiments: A Meta-Analysis", Experimental Economics VOL 6/3, pp 299−310

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Datum: 18.6.2010

Jan Schikora

Curriculum Vitae

	Lehrassistent Volkswirtschaftliche Fakultät, Ludwig-Maximilians-Universität München
Oktober 2006 - September 2010	Doktorandenstudium in Ökonomie Munich Graduate School of Economics
September 2007 - Januar 2008	Visiting Ph.D. Student in Economics Cambridge University, Cambridge, England
März 2003- Juli 2006	Diplom in Volkswirtschaftslehre Ludwig-Maximilians-Universität München
März 2004- August 2004	Austauschprogramm International Business Administration Udayana University Denpasar, Indonesien
Oktober 2001- März 2003	Vordiplom in Betriebswirtschaftslehre Ludwig-Maximilians-Universität München
Juni 2000	Abitur St.-Anna-Gymnasium, München
Januar 1981	Geboren in München, Deutschland