Jan Theodor Schikora:

Bringing good and bad Whistle-blowers to the Lab

Munich Discussion Paper No. 2011-4

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Bringing good and bad Whistle-blowers to the Lab

Jan Theodor Schikora*

February 28, 2011

Abstract

Whistle-blowing is seen as a powerful tool in containing corruption, although theoretical findings and experimental evidence cast doubt on its effectiveness. We expand a standard corruption model by allowing both, briber and official to initiate corruption actively, in order to assess the full effect of whistle-blowing. In our laboratory experiment we find that the effect of symmetrically punished whistle-blowing is ambiguous since it reduces the impact of corruption on productive activity, but also increases its stability. We show that asymmetric leniency for the official offsets the negative effect. The results can be explained by simple arguments about belief structures within the self-interested model of payoff maximizing.

Keywords: Corruption; Experiments; Whistle-blowing; Punishment

JEL Classification Numbers: C72, C92, D73

*MGSE, University of Munich

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 (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). 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 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 measures such as applying harsh punishment and increasing detection probabilities (Becker 1968), 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'. Whistle-blowing can be seen as the act of ending a corrupt transaction and all of its consequences by incurring non-trivial personal costs (Drew 2003).[1]

In their experimental studies, using the framework of a standard corruption game, 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.[2] 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). We explain this by the fact that the standard game of corruption 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).[3] The second, indirect, consequence (which is disregarded in the standard corruption game) concerns the 'crowding out' of legal productive activities. The latter requires an actively corrupt official. The compactness of the standard game of corruption does not allow for corrupt initiation by the official and is therefore not able to capture the indirect consequence of corruption. It does

[1]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).

[2]Neither study focusses on the mechanism of whistle-blowing or on the assessment of its effectiveness.

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

not 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. We add an option to the action space of the client in form of a choice of the 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 implimentation 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) [4]. On the one hand, symmetric whistle-blowing increases the level of corruption by providing a 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.

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

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

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. [5] Most patterns of behaviour can be explained by the combination of payoff maximizing strategies and first order belief structures.[6]

The remainder of the paper is structured as follows. Section 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. In Section 4 we analyze the model in theoretical terms and derive the main hypotheses. Section 5 interprets the results and tests the hypotheses. Section 6 summarizes and concludes.

2 Model

According to the NIE of corruption, the weakest link of a corrupt transaction is its lack of legal enforceability.[7]. 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. In case corruption is successful a large negative externality is realized. (See e.g. Abbink 2006)

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 extend the standard game of corruption in two ways. First, we provide a real alternative to corruption for the client and, second, we enable both, the client

[5]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.

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

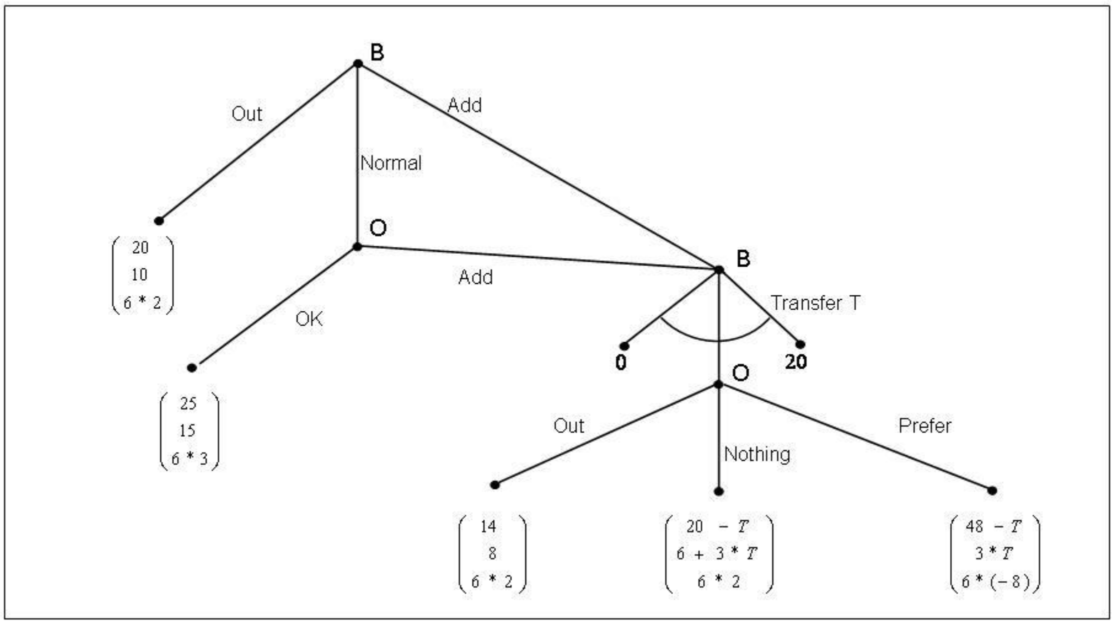
[7]Since corruption is illegal, third parties, i.e. the courts, cannot be used to enforce a corrupt transaction

and the official, to activate corrupt transactions. Therefore we add a third option in Stage 1 and a fourth stage to the standard 3-stage corruption game.

2.1 Representation of the basic game

Figure 1 shows the extensive form of the game used as a basis for our analysis (control treatment).

Figure 1: Extensive form representation of the basic set-up



The first line stands for the payoff of the potentially corrupt client (B), the second linefor that of official (O) and the third for the monetary 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.

8

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

In Stage 1, client B chooses between three alternatives. The choice 'Out' leads to the end

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

of the game and represents a second best outside option in terms of total payoffs.
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 B chooses 'Normal', Stage 2 is entered and O decides between providing the licence ('Ok'), which leads to a 'legal' transaction, and demanding a bribe ('Add'). If 'Add' is chosen by, either B in Stage 1, or 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.
Having received T, O decides again 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).
If 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.[9]
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)[10] 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

[9]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).

[10]Note that the costs of 6 EMU represent the total costs of reciprocation (delivering the corrupt service). We may interpret these costs as effort-costs 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.)

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 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'].

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).

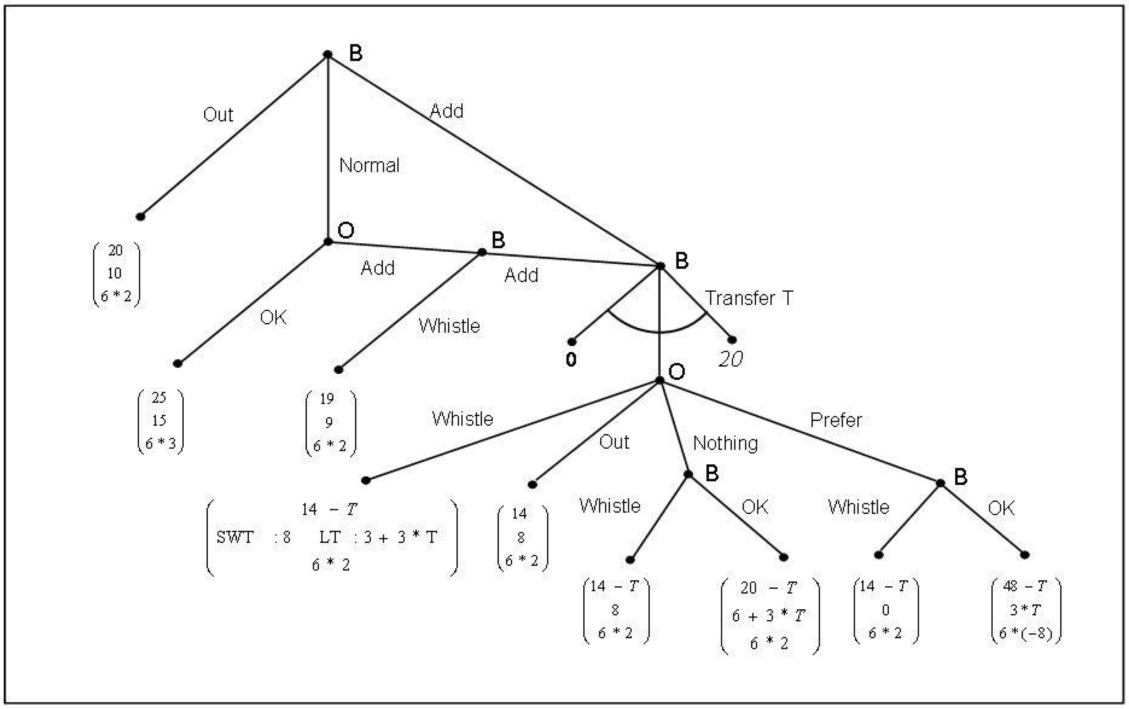
2.2 Treatment specifications

To our knowledge this paper provides 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 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 1, for ten consecutive periods (repeated game in partner design).

In the symmetric whistle-blowing treatment (SWT), see Figure 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 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[11]) outside option less the transfer T.

The leniency treatment (LT), see Figure 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 only costs of 3 EMU, 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 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.

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 A.

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 (Fehr et al. 1993, Berg et al. 1995).

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 None of the respective models seems to be able to provide a consistent and convincing theoretical explanation of experimental findings.

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

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 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 the standard game of corruption, Abbink et al. (2002), Abbink (2004, 2006), Abbink and Hennig-Schmidt (2006) and Lambsdorff and Frank (2007, 2010) 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.

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, 2000) 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 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 1 and 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 Schikora (2010).

The instructions were partly read aloud (See Appendix E 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. [12]

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, 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.

4 Analysis and hypotheses

To evaluate the effectiveness of whistle-blowing in general and different institutions of whistle-blowing in particular we compare our three treatments with respect to the following variables.

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

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. 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. Third, we measure social welfare by considering the average payoff level and its distribution across participants.

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. Whistle-blowing can also 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.

In this section we discuss possible systematic effects of whistle-blowing. A strict separation of the different motivations is not 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 use 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).

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. 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')[13]. 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 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 B

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

for 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 her expected payoff of choosing the legal path. This may change B's decision between 'Out' and 'Normal' in Stage 1 and hence the magnitude of the indirect 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 B), it is at least 19 EMU (in expectation) in the whistle-blowing (WB) treatments.[14] 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 B), 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).[15] This yields the following hypothesis.[16]

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."

4.2 Stabilizing corruption

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 (2004) 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[17] (opportunistic) behaviour in order to stabilize the corrupt transaction. O may adhere to a

[14] 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.

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

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

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

set of beliefs assigning a positive probability for B retaliating defection with whistle-blowing despite the non-trivial personal costs.

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."

4.3 Gender effects

The belief on the punishment probability 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 (forthcoming) 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). 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."

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 ($3T$). Since it is not more attractive than the option 'Out' we do not expect it to be of any relevance in this treatment. 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 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. 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 $EP(`Whistle'|H_{I_5}) = 3 + 3 * T > EP(`Prefer'|H_{I_5}) = 3 * T$ for any T and 'Whistle' over 'Nothing' if $EP(`Whistle'|H_{I_5}) = 3 + 3 * T > EP(`Nothing'|H_{I_5}) = (1 - r(T)) * (6 + 3T) + r(T) * 8$. This is the case if $r(T) \geq \frac{3}{3T-2} = r^*_{LT}$. Hence those officials who hold a set of beliefs $r(T)$ 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} ($= \frac{3}{3T-2} < r^*_{SWT}$) [18]. Though

[18]Note that this holds only for the last period and for situations in which the concerns about immediate

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.

5 Results and Interpretation

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 1.

direct punishment dominate strategic considerations.

Table 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[19]: $p = 0.002/0.001$; $N = 33/30$).

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 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 income. When we compare the Gini-coefficients[20], 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

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

[20] 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.

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.

5.1.1 Efficiency

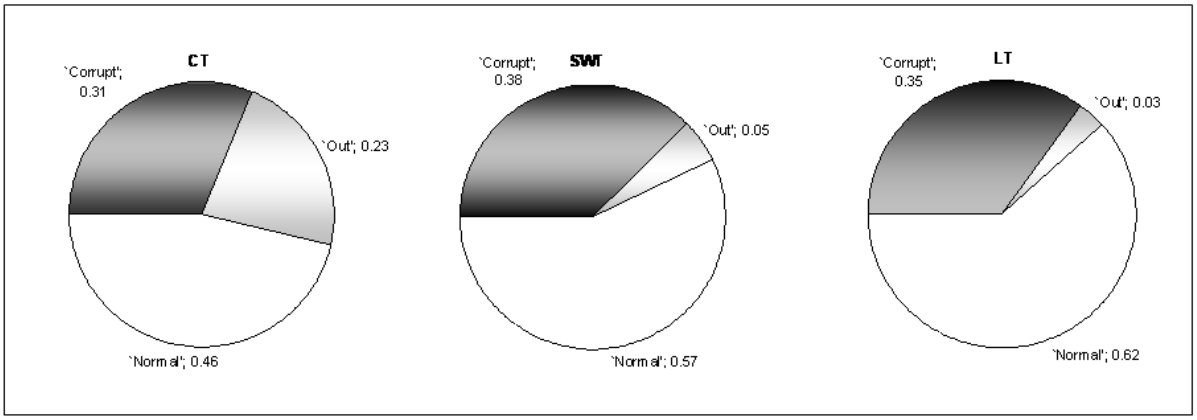
As explicated in Section 4, corruption indirectly reduces productive activity and thereby efficiency. The risk of being forced into a corrupt transaction may hinder honest clients to enter productive markets. In our set-up, the presence of corruption causes participants to choose the sub-optimal outside 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 Section 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 shows that, while clients seem to be equally willing to initiate the corruption transaction across all three treatments (31%/38%/35% 'Add' choices in CT/SWT/LT), 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 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 ($p < 0.001$; $N \geq 30$).

Second, O has to choose 'Ok' in Stage 2. The rates of acceptance ('Ok' in Stage 2) of a legal proposal differ significantly between WB treatments at 61%/55% in SWT/LT and CT at 38%. (SWT vs. CT and LT vs. CT: $p < 0.001$ with $N \geq 30$.) Officials seem to anticipate their clients to blow the whistle in Stage 3 (which they do in 22%/21.5% of possible cases in SWT/LT) and therefore choose the legal procedure more often.

We explain this behaviour according to the argumentation of Section 4. The choices in

Figure 3: Shares of choices in Stage 1



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 and the difference in acceptance rates in Stage 2 explains the differences in B's Stage 1-behaviour across treatments. Our results suggest that whistle-blowing equips the client with a useful safeguard against the threat of exploitation, allowing her to initiate a 'legal' transaction at a lower risk. This makes it, in expectation, more attractive for the client to enter (or stay) in the productive market.

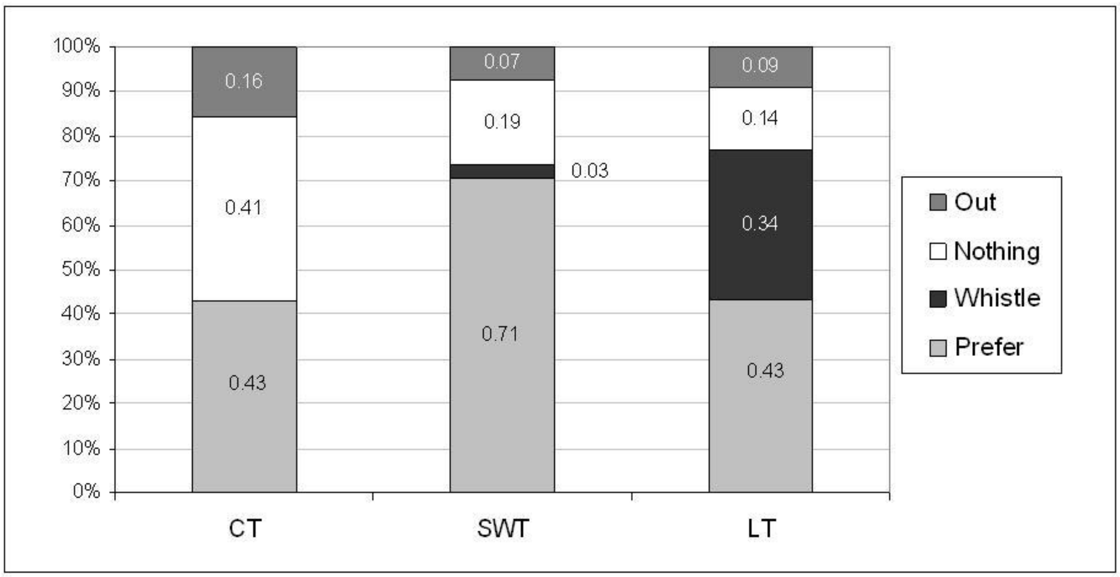
Even if we assume that B's belief on the probability of O choosing 'Ok' in Stage 2 is constant over the treatments, this creates a gap in B's expected payoff when choosing 'Normal' between CT and the WB treatments.[21] It may be just this expected gap (which is based on realistic beliefs) that leads to the large differences in the fractions of 'Out' and 'Normal' choices in Stage 1 between the treatments (see Figure 3). The gap in outcomes (successful legal choices, see Table 1) is even wider due to the differences in behaviour of O in Stage 2.

[21] 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 8 in Appendix B. Using the argumentation of Section 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 'OK' in Stage 2 is 0.38 in CT and thereby lies in the relevant range to expect a treatment effect.

5.1.2 Corruption levels

Figure 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 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 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%)[22] 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 com-

[22]The fraction is calculated by $\frac{0.34}{0.34+0.14+0.09} * 100\%$, see Figure 4.

parison 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.[23] 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**).

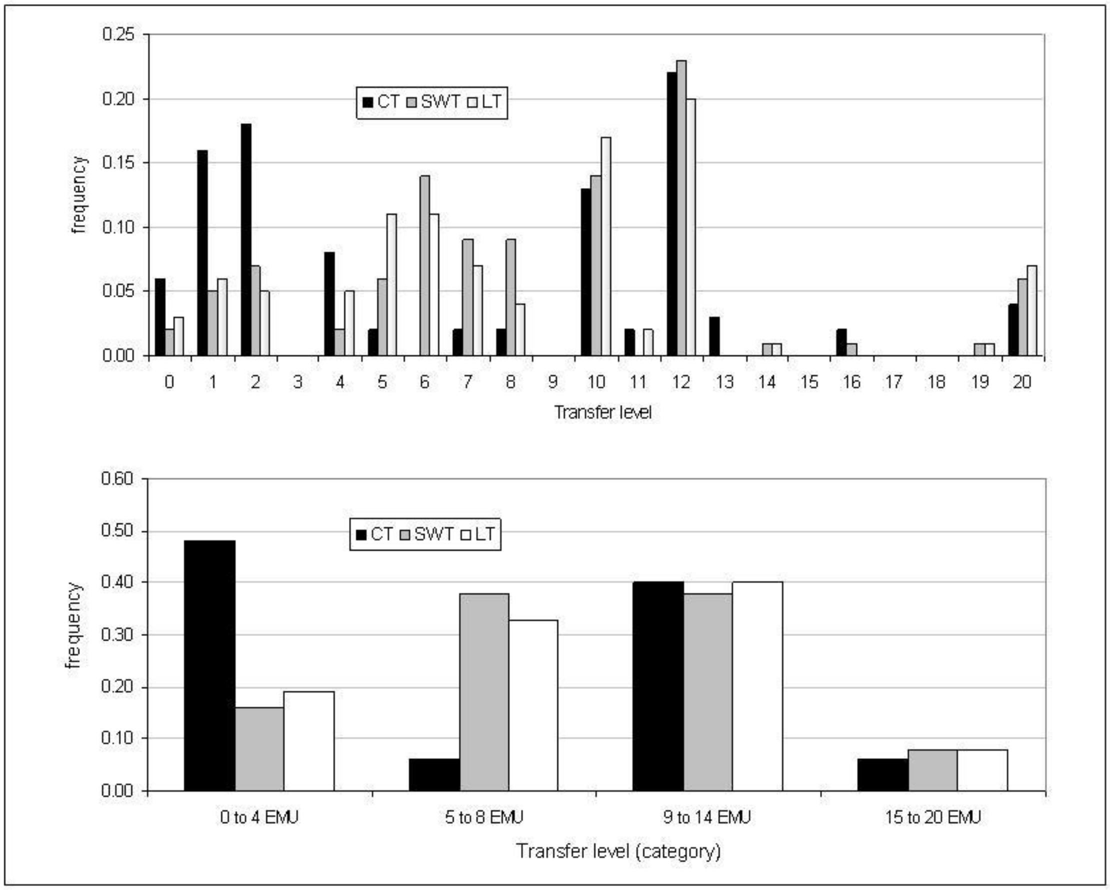
5.1.3 Transfer levels

Figure 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 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 5.

In the upper panel (of Figure 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 1 or 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 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.

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

Figure 5: Distribution of transfer levels and categories



5.2 Conditional reciprocity

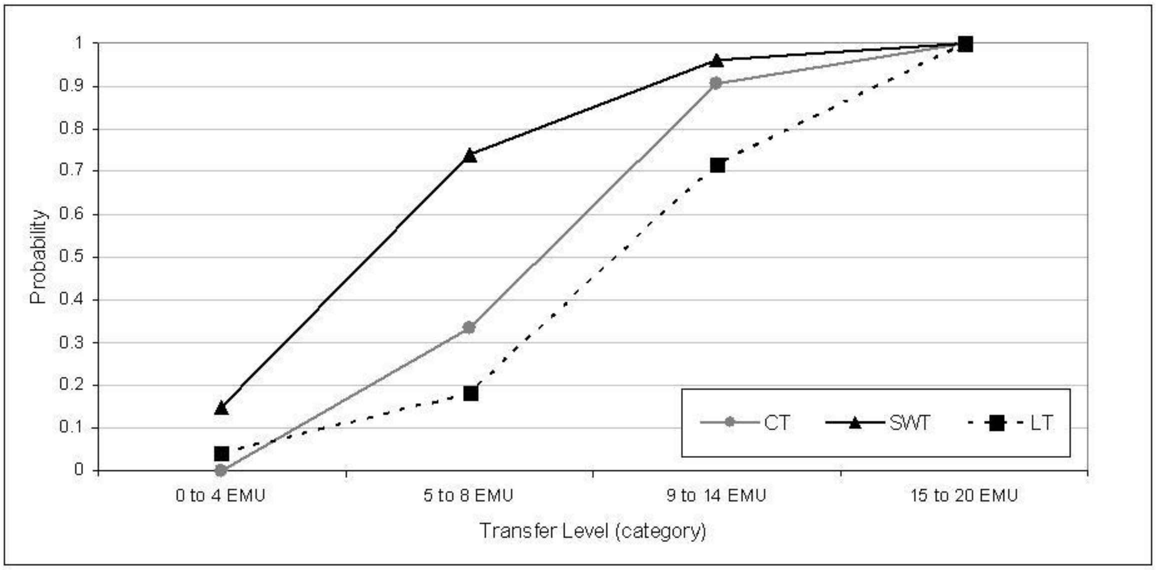
In our experiment, the occurrence of corruption mainly relies on (strategic) reciprocity between B and O. Confirming results from previous experiments on corruption (e.g. Abbink 2004, Lambsdorff and Frank 2010), the spearman rank correlation coefficients of $\rho_s(CT) = 0.62$, $\rho_s(LT) = 0.60$ and $\rho_s(SWT) = 0.69$ 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 6 shows the probability of corrupt success conditional on transfer level categories.[24] We confine our analysis of conditional corruption to non-parametric tests.[25] The selection process of (officials) reaching the bribing stage (only about 60% of possible cases in CT and

[24] 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.

[25] 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 because of the character of the selection process reachung the bribing stage.

Figure 6: Corrupt success conditional on transfer levels



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. 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 6) 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[26] in LT.

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 hence cut out potential gender effect.

5.3.1 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 6 in Appendix C. 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 7 in Appendix C. 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 8 in Appendix C. 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

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

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 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**).[27] This confirms the validity of the argumentation of Eckel and Grossman (1996) with respect to gender differences in anticipation behaviour.

5.3.2 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) [28] the client accepted substantial costs (6 EMU, independent of the size of T) to blow the whistle.

In our 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).

Table 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 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

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

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

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 2: **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 of 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.

5.4 Path-dependent behaviour

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

5.4.1 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.

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

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 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, Section 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: **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 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.[29]

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

5.4.2 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.[30] The results for the OLS regressions (M2) to (M4) for observations in the respective treatments are shown in the following equations.

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

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

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

*** 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.

[30] 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.

6 Conclusion

The main objective of this paper is to evaluate the introduction of (two distinct 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, but also the ex-ante problem of corruption keeping 'honest' clients away from productive markets. 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 (first best outcome) 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 behaviour of the official. 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 whistle-blowing by the briber occurring after the decision of the official can be misused as a tool to stabilize the corrupt transaction by the threat of punishing opportunism. The (realistic) anticipation of this behaviour leads to an increase in the success-rate of corrupt transactions. The second effect confirms experimental evidence from Lambsdorff and Frank (2010, forthcoming) and Abbink (2006).

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.

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.

Appendix

Appendix A: 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 $q(\text{`}Nothing\text{'}|I_{5,10}) > 0$.[31] Hence $Set2$ cannot be an equilibrium.

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

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, 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 B: 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.

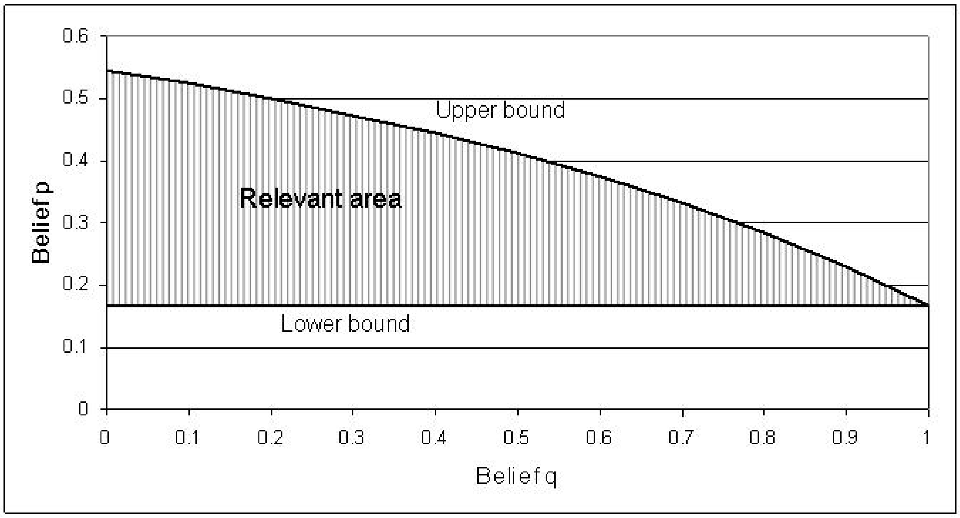
If she chooses 'Normal' in Stage 1 O may her (by choosing 'Add') to enter Stage 3. Not aiming at initiating corrupt reciprocity, B would either choose '$T = 0$', in which case she would (most likely) end up with 14 EMU, or '$T = 1$', in which case she would expect to get 19 EMU with some probability (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.[32]

Figure 7: 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 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

[32]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.

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:
$EP(`Prefer'|H_{I_5}) > EP(`Nothing'|H_{I_5})$, 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 D: 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 E: 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>

References

Abbink, K. (2004), “Staff Rotation as an Anti−corruption Policy: an Experimental Study”, European Journal of Political Economy VOL 20, pp 887−906

Abbin k, 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

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

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

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

Buccirossi, P., G. Spagnolo (2006), “Leniency Policies and Illegal Transactions”, Journal of Public Economics VOL 90/6;7, pp 1281−97

Charness, G., M. Rabin (2002), “Understanding Social Preferences with Simple Tests”, Quarterly Journal of Economics VOL 117, pp 817−69.

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

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

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

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

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

Fischbacher, U. (2007), "Z-Tree: Zurich Toolbox for Ready-made Economic Experiments", Experimental Economics VOL 10/2, pp 171−8

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

Hall, D., S. Davis (1999), "Corruption and Whistle Blower − a background note for TUAC", PSIRU, London, Available at www.psiru.org/reportsindex.htm

Klitgaard, R. (1988), "Controlling Corruption", University of California Press

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

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

Mauro, P. (1995), "Corruption and Growth", Quarterly Journal of Economics VOL 110, pp 681−712

Rabin, M. (1993), "Incorporating Fairness into Game Theory and Economics", American Economic Review VOL 83, pp 1281−1302

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

Spagnolo, G. (2006), "Leniency and Whistleblowers in Antitrust", CEPR Discussion Paper 5794

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 (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