Essays in Applied and Psychological Game Theory: Cooperation, Corruption and Political Economy

Loukas Balafoutas

PhD in Economics

The University of Edinburgh

September 2008

Declaration

I hereby certify that this thesis is entirely my own work and that it has not been submitted for any other degree or professional qualification.

Loukas Balafoutas Edinburgh, 29/09/2008

Abstract

The first chapter of the thesis applies game theory in order to examine the question of income redistribution from a fresh angle. In particular, it considers a mechanism of patron-client relationships which enables a society's rich class to limit the extent of redistributive taxation. In effect, the aim of patronage is to "buy" the votes of some poor citizens and lower the demand for redistribution. Income tax rates are further shown to depend negatively on government corruption in the form of fund capture, provided that a democratic regime is in place and the government cares about reelection. This link is tested empirically using cross country data and the evidence is consistent with the predictions of the model.

The second and third chapters shift the focus of attention towards the process of decision making in games and the role of emotions in this process. Corruption, which is taken as exogenous in the first chapter, is now considered in detail as the outcome of a cooperation game between two players, with a third player (or "third party") having a stake in the outcome of the game but no opportunity to take any direct action. This situation is analysed using psychological game theory. Players' utility functions are extended to include beliefs and the emotions that these generate. In the theoretical model of the second chapter the emotion of interest is guilt and this is conditioned on the perceived beliefs of the third party. The two players are then less likely to collude if they believe that the third party expects a favourable outcome for herself. The model solves for the conditions under which collusion emerges in equilibrium. The main assumption of the model (i.e. the role of beliefs in decision making) as well as some of its predictions are then tested using an economic experiment in the third and final chapter of the thesis. The experimental findings strongly support the impact of beliefs on the incidence of collusion: Perceived expectations of the third party about the outcome of the game appear to be the most significant factor that determines the outcome itself. There seems to be a mechanism of self-fulfilling expectations which can be applicable to a number of economic situations, including corruption in public administration.

Table of contents

Introduction 5

Part I

Chapter 1: How much income redistribution? An explanation based on patronage and corruption 9

Part II

Chapter 2: Third party beliefs and collusion in a repeated psychological game 54

Chapter 3: Third party beliefs and collusion: An experimental study 82

Introduction

This work applies the intellectual tools of game theory and psychological game theory in order to analyse certain economic and social issues. The theoretical analysis is supplemented and supported by empirical evidence using both real world data and data from a laboratory experiment.

The thesis is structured in two parts and three chapters. The first part (Chapter 1) asks a fundamental socioeconomic question: What are the determinants of and the mechanisms behind income redistribution? Although this question is not a new one and has been dealt with in the literature, I believe that there may be room for "an explanation based on patronage and corruption" -as the title of Chapter 1 suggests. I present a theoretical model which views income tax rates as the equilibrium of the strategic interaction between rich voters, poor voters, and a purely self interested government. The main novelty of the model and driver of the results is the idea of *patronage*: The rich can pay a fraction of their after tax income to some of the poor, with the purpose of influencing their voting behaviour. By means of this mechanism the rich succeed in limiting the extent of income redistribution in equilibrium. Both the setup of the model and the specifics of the patronage story require a number of assumptions which inevitably affect the generality of the results, as is true of any economic model. Keeping that in mind, the purpose of the model is to convey the general idea that the rich classes in a society may be able to "put a brake" on redistribution by exercising their influence not only on politicians, but also on poor voters. While the median voter theorem, common throughout the relevant literature, still holds, the interests of the rich minority can ultimately outweigh those of the majority. The rich achieve this by "buying off" some of the poor voters, who are in effect hurting their own class with their behaviour. I motivate the existence of such a scenario with examples from the literature.

Beyond these insights, a second contribution of Chapter 1 is the investigation of the relationship between income tax rates and corruption, which to the best of my knowledge has not been studied in the literature. Equilibrium tax rates are shown to be a negative

function of government corruption in the form of fund capture. This relationship is robust to alternative theoretical specifications which correspond to different assumptions regarding key elements of the model. The empirical section which follows the model consists of the following elements: Collection of cross country data on top marginal income tax rates, corruption and a number of control variables, along with a discussion on the nature and limitations of the data and especially the measurement of corruption; a number of linear regressions with tax rates as the dependent variable; and the fitting of alternative specifications including instrumental variables estimation in order to check the robustness of the results. The key econometric finding of the chapter is the discovery of a seemingly robust negative impact of corruption on income tax rates in democratic countries. A number of econometric issues relating primarily to measurement may affect the strength and generality of these results. Nevertheless, one cannot dismiss the fact that the evidence is in line with the main testable prediction of the theoretical model.

Part II of the thesis (Chapters 2 and 3) revolves around the role of psychology in games. This part of my research originally focused on corruption in public administration and later developed into a more general framework for the analysis of the effect of emotions on decision making and on collusion between players. Hence, the intellectual thread that brings parts I and II together is more evident in the sections which discuss the implications of the findings of part II for economics, and in particular for corruption in bureaucracies.

Chapter 2 initially sets the stage. A sequential cooperation game with three players and a twist: only two of them take actions in the game, and if they achieve the cooperative outcome they increase their own payoffs but lower the payoff of the third player, or "third party". Abstracting from psychological factors, cooperation -or, as it is called throughout Chapters 2 and 3, collusion- can only be sustained in the repeated version of the game assuming a plausible trigger strategy. Now, the crucial psychological assumption of the model is that the utility function of one of the players includes his beliefs about the beliefs of the third party regarding the outcome of the game. To be more precise, the player suffers an emotional cost whenever he colludes because he is averse to letting the

third party down; this emotional cost is proportional to the player's 2nd order beliefs (i.e. what he thinks the third party thinks). Taking this "guilt aversion" assumption as the departure point, the paper reaches the following key results for the repeated game:

(i) When beliefs are constained to be constant over time, there are two possible equilibria. Collusion always occurs in one of them and never in the other. The effect of guilt aversion is to make collusion less likely in equilibrium compared to a standard, non-psychological game.

(ii) When beliefs are allowed to vary over time, it is appropriate to model them as the product of an updating process based on the history of play. Again, there is a collusion and a no-collusion equilibrium, and guilt aversion makes collusion relatively less likely to emerge. In this context beliefs no longer necessarily correspond to actual play; this contradicts the concept of a Psychological Nash Equilibrium and is discussed in some detail in Chapter 2.

In my view, the contribution of this analysis to the literature is twofold. First, it examines the emergence of collusion in equilibrium and conditions it on third party beliefs as well as on a number of other parameters (players' relative payoffs and the horizon of the game), in the context of the three-player game described above. The chapter discusses real world situations where the model and its predictions are relevant; the lion's share in this discussion belongs to corruption in public administration. Second, this model is an attempt to further the very limited literature on dynamic psychological games, time-varying beliefs and psychological games with asymmetric information. The model could be seen as incompatible with the standard theory on psychological games, because it generates beliefs which are not only time-varying, but also potentially inaccurate even in equilibrium. Yet, I believe that such beliefs should be incorporated and cannot be ruled out -as a matter of fact they are more realistic in many settings. Far from offering a complete theoretical framework for the study of these matters, Chapter 2 aims to give some useful insights that emerge from the analysis of a game with a specific structure.

The empirical investigation of some of these insights is left to Chapter 3. This last chapter of my thesis reports the findings of a laboratory experiment on decision making with

human subjects (who, in this particular case, were students at the University of Aberdeen). The experimental design was intended to follow the theoretical model: Subjects were randomly assigned to groups of three, given detailed instructions and asked to play the collusion game of Chapter 2 with two "active" players and a third player who only reported his (her) beliefs. The setup also incorporated the features of repeated interaction and asymmetric information.

The analysis of the data generated from the experiment ranges from simple statistical tests to the estimation of probit models of choice. I consider the single most important finding of this chapter to be the following: *2nd order beliefs matter for players' choices and they strongly affect the likelihood of collusion.* In fact, beliefs appear to have the strongest influence on decision making among all explanatory variables, in terms of magnitude and statistical significance. Hence, the experiment provides solid support for the impact of third party beliefs on the outcome of the specific game which is the workhorse in the second part of the thesis. In broader terms, it adds to the stock of experimental evidence in favour of psychological game theory and its implications for the role of beliefs and emotions in games.

A number of people have contributed to this work in various ways. Their inputs are gratefully acknowledged in the corresponding chapters of the thesis.

PART I

Chapter 1

How much income redistribution? An explanation based on patronage and corruption

*How much income redistribution? An explanation based on patronage and corruption**

Abstract

This paper adds to the political economy literature on income redistribution by studying how income tax rates are determined and how they are related to government corruption in the form of fund capture. A simple model is presented where a lobby formed by the rich voters tries to block redistribution by forming patron-client ties with some of the poor voters. It is shown that in equilibrium there is some redistribution and income tax rates are a negative function of government corruption. When we allow the lobby to make bribe payments to the government, an additional equilibrium is possible where patronage implies a gap between statutory income tax rates and real tax rates which are zero. The link between corruption and tax rates operates through the electoral motive; one should therefore expect it to depend on the existence of a democratic regime. This link is tested against recent data on corruption, income tax rates, type of political regime and a set of controls for a cross section of countries. The empirical evidence is fully consistent with the predictions of the model.

* I am thankful to Stephane Straub for his guidance and advice. I would also like to thank Andy Snell, Joseph Sakovics, Jonathan Thomas, Santiago Sanchez-Pages, Julia Darby, Mark Schaffer, Pablo Zoido, Juan De Laiglesia, Lee Alston, Yannick Perez, and participants at the 5th conference of ESNIE for useful discussions and comments.

1. Introduction

Redistribution of income can take various forms, the most common of which is direct transfers from the government to certain groups or individuals. Although the identity of these groups can vary, the direction of redistribution is usually from the rich to the poor classes. Thus, this paper restricts attention to the use of taxation as a means of redistributing income from the top towards the bottom of the income distribution within a country, and tries to tackle the question of why and how societies choose to redistribute - or not- income, and to what extent. On a related issue, it is generally the case that tax revenues for the government are negatively associated with corruption; however, most of the literature has identified tax evasion as responsible for this fact, while neglecting the stage that comes prior to tax (compliance or) evasion, namely the choice of tax rates by the government.[1] This stage is the focus of interest here.

A model is presented where politicians are self-interested and corrupt and where two social classes (rich and poor) compete with each other over fixed resources. Accordingly, the tax rate[2] is not mechanically set at the level preferred by the median voter, but depends on the political environment and on the strategic options and choices of the players in a game theoretic setting.

The purpose of the paper is twofold. First, it offers an original explanation for why tax rates are relatively low in most countries. Consider a country with an unequal income distribution: both the electoral motive and the desire to raise a lot of revenue and divert part of it imply that income tax rates -especially for the richer parts of the population- should be much higher than the ones observed in practice.[3] The main explanation offered here is that rich voters can form a coherent group and lobby for their

---

[1] For example, Friedman et al (2000) observe a negative correlation between corruption and tax revenue and attribute it to the fact that "weak institutions undermine the government's ability to collect tax revenue" (p. 480).

[2] Throughout the paper, the only tax rates considered are income tax rates, therefore the two terms refer to the same thing and are interchangeable.

[3] Benabou and Tirole (2005, p.2) ask themselves the question: "What are the forces that limit the extent of redistribution in a democracy, preventing the poor from 'soaking the rich'?"

interests, whereas poor voters cannot. In particular, the paper identifies the existence of patron-client relationships between the rich and some of the poor as the mechanism which reduces the demand for -and consequently the extent of- income redistribution. The next section reviews some relevant examples and insights from the literature, which provide support for such a mechanism.

Second, the paper predicts a negative effect of the extent of government corruption in a country on statutory income tax rates. This effect is driven by patronage and by the electoral motive of the government. It is therefore only in democratic countries that more corruption leads to lower taxation. In the empirical section, this hypothesis is tested against cross country data on income tax rates, corruption and a set of controls. The evidence indeed suggests a negative and significant relationship between corruption and top marginal income tax rates in democracies. Identifying the existence of such a relationship is important from a positive point of view, as it allows us to say something about the effect of restraining government corruption on the extent of income redistribution. An important point to keep in mind is that this link depends on the political system, and more specifically on whether this is democratic; then so do the policy implications. In fact, the evidence in the empirical section suggests that in non-democratic countries less corruption might actually pull tax rates down and inhibit income redistribution.

The next section reviews some relevant literature. Section 3 presents a simple game and derives the central results on the determination of income tax rates, the role of patronage and the relationship between tax rates and corruption. Section 4 adds a new feature to the basic model by allowing the government to take bribes, while section 5 modifies one of the assumptions regarding the rules of the political environment. Section 6 discusses the data that are used along with their limitations and section 7 tests empirically the implications of the model. Section 8 concludes.

2. Related literature

There are several references in the literature that lend credibility to the patronage scenario in this paper. Baland and Robinson (2006) offer a detailed account and a number of references on mechanisms through which landowners have kept the votes of their workers under control, across centuries and in different parts of the world. Their case study of Chile reveals that in the past landowners had an absolute control over the votes of peasants through the use of coercion and the threat of expulsion, and that this control depended critically on the absence of a secret ballot. They claim, however, that even in the presence of a secret ballot *"strategies were found to keep voting under control" [p.4]*. And then, that these strategies *"were used and remain up to the present day in democratic third world countries. Nowhere is the evidence about landlord control of elections so conclusive as in Latin America" [p.5]*. The mechanism of control is slightly different, as it relies on coercion and threat rather than on an incentive mechanism that would align the interests of the peasant with those of the landowner. It has nevertheless a very similar flavour to the story modelled in this paper. Ravallion and Lokshin (2000) also offer some support for the patronage scenario presented here by arguing that *"demand for redistribution will be lower in a socially cohesive setting in which reciprocal relationships (though possibly unequal ones, such as patronage) offer security" [p.89]*. A well-documented discussion on the evolution of patronage can be found in Platteau (1995). Finally, on a more general and intuitive level, and in the context of feudal societies of past centuries, Moore (1966) argues that *"where the links arising out of [the] relationship between overlord and peasant community are strong, the tendency toward peasant rebellion (and later revolution) is feeble" [p.469]*.

This paper also relates closely to the literature on the determination of the level of redistributive taxation. One strand of this literature uses the median voter theorem to predict demand for redistribution and equilibrium tax rates under majority voting. In general, poor voters prefer high tax rates, while the opposite is true of rich voters. The unique tax rate actually chosen is the one preferred by the voter with median income (or productivity). The tax rate increases with inequality, as measured by the distance between

the median and mean income.[4] The most influential paper in this literature is probably Meltzer and Richard (1981). They use a general equilibrium model to show that individual demand for redistribution is determined by each voter's pre-tax income in comparison to the median income, while the voter with median income is decisive over the tax rate. Similar works include Romer (1975), Roberts (1977), Persson and Tabellini (2000).

In a somewhat different spirit, a number of papers emphasize social mobility and expectations of future income as critical factors that shape preferences for redistribution, although they do share the insights of Meltzer and Richard (1981) on the role of relative current income. In Piketty (1995) individual social mobility experiences shape attitudes towards redistribution through learning. Alesina and La Ferrara (2002) suggest that expected future income must be taken into account in so far as some of today's poor may –or expect to- be tomorrow's rich. Higher social mobility enhances this possibility and therefore reduces demand for redistribution.[5] Other factors that are identified in that paper are risk aversion, altruism, and perceptions about equal opportunities. In Ravallion and Lokshin (1999) uncertainty about the tax rate and about future income is what drives the main result, which is that the poor always favor redistribution and the rich have mixed attitudes towards it. Here the authors focus on downward social mobility as a possible explanation of why some rich voters may support redistributive policies. Benabou and Ok (2001) formalize what they call the "Prospect Of Upward Mobility" (POUM) hypothesis. They show that, under certain assumptions, it is possible and consistent with rational expectations to have some voters with below mean incomes who oppose redistribution, as long as tomorrow's expected income is an increasing and concave function of today's income.

Alesina and Angeletos (2005) build on the Meltzer and Richards framework and incorporate fairness considerations into the demand for redistribution: Societies which

[4] This result is also found in the literature that relates inequality and redistribution to economic growth: Alesina and Rodrik (1994) and Persson and Tabellini (1994) show that a more unequal distribution of income leads to a higher level of tax on capital (through stronger demand for redistribution) and subsequently to lower growth rates.

[5] Only upward social mobility is considered in this context.

perceive differences in income as arising primarily from luck or socially unworthy activities (such as corruption) choose higher levels of redistribution to correct for "unfair" outcomes. [6]

All these models provide interesting insights into voters' demand for income redistribution. However, they do not take into account government incentives other than the desire to be reelected, nor do they consider the behavior of the rich class; one could say that they restrict themselves to the demand side of redistribution and completely ignore the supply side. To give an example of the literature that studies "supply-side" redistributive politics, Persson and Tabellini (1999) present a model where self-interested politicians tax the voters and then allocate the tax revenue among public goods, targeted redistributive transfers,[7] and rents for themselves. The choice of tax rate depends on the intensity of competition between politicians and between voters, as well as on the nature of the political system and the electoral rules. The crucial feature here is that politicians are self-interested and divert resources for their own purposes; in other words corruption enters the model and, along with the electoral motive, influences the economic outcomes. No explicit attempt is made, however, to study the relationship between the extent of corruption the choice of tax rate.

Finally, there are some very interesting insights concerning redistributive taxation in Fernandez and Levy (2006) and Levy (2005). The first of these papers uses a framework with two social classes (rich and poor) and preference heterogeneity within each group. It arrives at two political equilibria, one of which implements extensive redistribution, while in the other a fraction of the poor joins forces with the rich leading to less redistribution. Although this result is entirely driven by taste heterogeneity, it is similar to the results in my paper in the sense that extensive redistribution is inhibited by the existence of some poor voters who hurt the general interests of their class in order to advance their own. In Levy (2005) coalitions between rich and poor voters are also possible and reduce the extent of redistribution. That model considers two redistributive

[6] See also Benabou and Tirole (2005), who attempt a broad treatment of beliefs and attitudes towards redistribution and perceived fairness, giving more weight and structure to psychological parameters.
[7] Redistribution in this setting is not from the rich to the poor, but among different electoral districts.

tools (transfers and public education) and voter heterogeneity along two dimensions (income and age) and models the interactions between those dimensions.

Turning to related empirical work, La Porta et al (1999) observe that larger governments are also higher quality ones, which means that there is a negative association between government size and corruption. This is consistent with my findings, although the implied direction of causality is the reverse. An explicit attempt to study empirically the relationship between corruption and redistribution is found in Olken (2006). That paper uses evidence from an Indonesian transfer program where subsidized rice was distributed to poor households and assesses empirically the effect of corruption on redistribution. The data suggest that at least 18 percent of the rice was captured by officials at different levels before it reached the targeted households; this implies that corruption lowers the extent of redistribution. The paper then uses a CRRA utility function to estimate that the presence of corruption lowers the welfare gain from redistribution by approximately 20 percent. Comparing the actual welfare with the program and the one without it, the paper concludes that corruption, combined with the deadweight loss of the taxation raised to fund the program, can more than offset the beneficial effects of redistribution and lead to a net welfare loss. However, the effect of corruption is simply due to the fact that part of the tax revenue is captured by officials. The decisions on whether to implement the programme in the first place and on the level of tax rates –which determines the extent of the programme- are not considered; nor are the possible links between these decisions and the extent of corruption in the government.

3. The Model

3.1. General

Consider a stationary endowment economy with a voting population of a mass of 1. There are two social classes, the rich and the poor, accounting for proportions n and $(1-n)$ of the population respectively, with $n<1/2$. In each period the rich receive an exogenous

endowment of y_r and the poor an endowment of y_p. The endowment of the poor is defined as the subsistence income and normalized to zero. Rationality and full information are assumed.

Redistribution can be achieved through a linear income tax set by the government (which is positive only for the rich since $y_p=0$), with the tax revenue being distributed evenly among the poor in the form of direct transfers. Since this is an endowment economy, and following much of the literature on redistributive taxation,[8] I assume that taxes are non-distortionary and do not impose any deadweight loss. Furthermore, the model distinguishes between statutory and real tax rates. Statutory rates are denoted by $\hat{\tau}$ and real rates by τ: having set the statutory tax rate at $\hat{\tau}$, the government can then collect taxes that correspond to a tax rate which is less or equal to $\hat{\tau}$; in other words, $\hat{\tau}$ is the maximum rate at which an individual can be taxed. But the real rate can be lower, if for instance the government gives large rebates or tax refunds or simply overlooks tax evasion.[9] The role of this distinction between real and statutory rates will become clear in the solution of the model.

There are two identical parties that are both selfish, in the sense that they only care about the maximization of their own payoffs.[10] There are no ideological preferences of governments or citizens and they all act as rational utility maximizers. There is a limit to the number of consecutive terms that an incumbent is allowed to stay in office, and this limit is taken to be two terms –this assumption is relaxed in section 5. I therefore initially consider a game that lasts only two periods, which simplifies the analysis. There is no discounting in the model and actors weigh equally their payoffs during the two periods. Elections are majoritarian and there is a single voting district, which means that the party which receives more than half of the total number of votes wins the elections. Voting is purely retrospective and a government that pleases its voters in the first term will be

[8] See, for instance, Lindbeck and Weibull (1987), Persson and Tabellini (1999).
[9] So in this model, tax evasion only occurs to the extent that the government allows it to occur.
[10] To motivate this assumption, see for instance Downs (1957, p.28): "[party members] act solely in order to attain the income, prestige and power which come from being in office", Barro (1973, p.22): "the officeholder's objective is the maximization of his own utility", or Persson and Tabellini (1999, p.702): "politicians would like to raise a lot of revenue and spend it on rents for themselves"

rewarded with reelection. The assumption of retrospective voting is often made in the literature.[11]

3.2. The actors: Government, rich voters, poor voters

The government is not only selfish, but also corrupt. Corruption here takes the form of diversion of funds:[12] The government appropriates some proportion β of the tax revenue that it has raised, $0 \leq \beta < 1$. The parameter β will therefore be used to relate corruption to redistributive taxation. Corruption is exogenous in this model. This exogeneity assumption may appear unrealistic at first; however, corruption here is envisaged to reflect the set of political and social norms that prevail in a country. For instance, the ability of politicians to divert funds without being caught and punished depends on the efficiency of the judiciary, the freedom and influence of the press etc. It is natural to think that a government cannot change these factors in the short run, or at least it is not easy to do so. In addition, the purpose of the paper is not to explain the extent of corruption, but to show how a given level of corruption in a country relates to income redistribution, and to offer some original insights on potential mechanisms that determine income tax rates. Thus, the choice variable of the government is τ, and not β.

The period payoff function of the government is given by: $U = R + \beta \tau m y_r$, where R represents an exogenous rent from being in office. The government's payoff function thus includes an exogenous rent and an endogenous one (the part of the total tax revenue that it captures).

The rich class is able to form a coherent lobby with the purpose of promoting the interests of its members.[13] Redistribution clearly opposes these interests, the more so the higher is the tax rate. In order to block redistribution, the members of the lobby can take

[11] The idea is that parties are unable to make binding and credible commitments to the public, and that one should therefore focus on the behaviour of the incumbent government rather than on pre-election politics. See, among others, Barro (1973), Persson and Tabellini (1999).

[12] Svensson (2005, p. 19) argues that "The most devastating forms of corruption include the diversion and outright theft of funds for public programs", and gives examples from Zaire, Indonesia, Philippines and Angola.

[13] Throughout the paper, the terms "the rich (class)" and "the lobby" mean exactly the same.

the following course of action: they can "buy" the alliance of some members of the poor class, by offering them some fraction of their after tax income. The objective of this strategy is to make the majority of the population prefer a zero real tax rate to any redistribution, by tying the realizations of y_p to y_r.[14] If this strategy is successful, there will be no government induced redistribution; still, some of the poor citizens will benefit from their ties to the rich, so in fact some partial redistribution will take place. This patronage story is the main novelty of the theoretical model and a key driver of its results.

Poor voters adopt a strategic voting rule, which specifies the conditions under which they will reelect an incumbent government[15] and which is shown in the solution of the game in the next subsection.

Timing of events:

1. At the beginning of the first period an incumbent government is exogenously in office and poor voters determine their voting rule.
2. The government sets the statutory income tax rate.
3. The lobby decides upon a course of action: it can either stay inactive or try to block redistribution through patronage. If it opts for patronage, the clients are selected at this stage.
4. After observing the action taken by the lobby, the government "sets" the real tax rate, collects the tax revenue and distributes it.
5. At the end of the first period, elections take place and the government is either reelected or replaced by another party. In addition, one must note that transfers from the patrons to the clients are made after the outcome of the elections is known.[16]

---

[14] Individual votes are not observable. Hence, the lobby cannot offer a contract according to which a client is paid only if (s)he votes in a particular way; they have to give the share of their after tax income that *induces* the clients to oppose redistribution.

[15] Such voting rules are common in the literature. In Barro (1973) voters adopt a strategic voting rule such that the government is reelected if and only if it provides a public good above some level. Similarly, in Persson and Tabellini (1999) voters set strategically their reservation utilities.

[16] See the Appendix (A.4) for a discussion on the issue of credible commitment. I discuss in particular the ability of the lobby to commit to make its payments, both in the simple version of this section and in the one of section 4.

3.3. Solution of the game

To arrive at the solution of the game, I begin by looking at the strategies adopted by the lobby and by the poor class:

<u>*Lemma 1*</u>

If the lobby uses patronage to block redistribution, each rich voter pays a fraction k of his (her) after tax income to each of (1/2n -1) poor voters, where* $k^* = \frac{n}{1-n}(1-\beta)$

Proof: In order to achieve a majority which wants to preserve the status quo with zero redistribution, the lobby must ally with *(1/2-n)* randomly selected poor citizens. This means that each member of the lobby will have to "buy" the alliance of *(1/2n-1)* members of the poor class, by paying them fraction k of his (or her) after-tax income. This part of the poor population will prefer a zero tax rate to any positive tax rate if: $ky_r \geq k(1-\tau)y_r + \frac{(1-\beta)\varpi y_r}{1-n}$, or $k \geq \frac{n}{1-n}(1-\beta) = k^*$. The left hand side of the first of the above inequalities is the payoff to the selected poor (the "clients") of maintaining the status quo with no government induced redistribution at all. The right hand side is the payoff from redistribution: the clients still receive fraction k of the income of their patrons, but this income is now $(1-\tau)y_r$. In addition they receive fraction *(1-β)* of the tax revenue. For k high enough, the clients prefer zero taxes to redistribution, and the government knows this. Hence, when patron-client relationships prevail, the preferred policy is zero taxes and only this policy will lead to reelection of the government.

I call this the patronage outcome, or *PAT*. In addition, I call *RED* the redistribution outcome in which the government collects taxes, captures part of the revenue and distributes the rest among the poor.

The voting rule adopted by the poor class is the following:

$$p^{R/NC} = \begin{cases} 1, & if \quad \tau \geq \underline{\tau} \\ 0, & if \quad \tau < \underline{\tau} \end{cases} \quad , \quad p^{R/C} = \begin{cases} 1, & if \quad \tau = 0 \\ 0, & if \quad \tau > 0 \end{cases} \qquad (1)$$

This rule specifies the voting behaviour of each poor voter in each of the two states that can occur: the two states are *C* (the voter has been selected as a client) and *NC* (the voter has not been selected as a client),[17] and the strategies conditional on these states are $p^{R/C}$ and $p^{R/NC}$ respectively, where p^R is the probability of reelecting the incumbent government. $\underline{\tau}$ is a reservation tax rate set strategically by the poor voters: Voters observe both statutory and real tax rates (the latter can be inferred from the transfers that take place), but voting strategies in *(1)* are conditional on real rates, because these are the ones that determine individual income.

<u>*Lemma 2*</u>

The voting rule given in (1) is optimal for the poor voters in both states, C and NC.

Proof: Since redistributive taxation is the only policy dimension and clients are always better off with zero than with any positive tax rates, it is optimal for them to set $p^{R/C}$ as given in *(1)*, which will indeed lead to $\tau=0$ given the electoral rule. The rest of the poor set a positive reservation real tax rate, the value of which under different assumptions is calculated in the Appendix (A1, A2).

I have ignored the voting strategies of the lobby, largely because it is the poor voters who form the majority and determine the electoral outcome. Only if the outcome is patronage is it necessary to know that the lobby will reelect the incumbent government if and only if $\tau=0$ (i.e. their voting rule is the same as $p^{R/C}$ in *(1)*).

In order to solve this game I use backwards induction and start from the second period. Provided the government has been reelected, in its second term it will set the statutory and real tax rate at 1. Since the government does not care about reelection, allying with the poor does not serve the interests of the lobby and total expropriation cannot be avoided.

[17] Note that the state *NC* occurs not only when a voter is not among the selected clients, but also when there is no patronage at all.

In the first period, however, the patronage strategy is relevant, since the government values reelection. Compared to the second period, the government now has the constraint that the rich might choose to ally with a fraction of the poor class, which would lead to a policy of zero real taxes ensuring reelection and potentially being optimal for the government. In that respect, the "softness" of the government's tax decision (distinction between statutory and real tax rates) is crucial for the relevance and effectiveness of the patronage strategy. This is because it allows the government to collect no revenue in the patronage outcome and thus to be reelected for a second term (given the voting rule), even if tax rates are officially positive.

We can now look at the equilibrium of this game.

<u>*Proposition 1*</u>

During its second term in office the government sets the statutory and real tax rate at $\tau = \hat{\tau} = 1$ *and the equilibrium is RED.*
In the first period the equilibrium is also RED but the tax rate is lower: The government sets the statutory and real tax rate at $\tau^* = \frac{1-2n}{2-2n}(1-\beta)$ *and is reelected.*

The proof is given in Appendix A1.

<u>*Corollary:*</u> *Income tax rates are a negative function of government corruption.*

This follows directly from the value of τ^* in Proposition 1.

In the first period redistribution is achieved through a positive income tax rate, the level of which is a negative function of government corruption, β. This negative relationship is a central result of the model. The intuition for this result is the following:

The higher is β, the cheaper is patronage for the lobby, because the clients do not expect much from the government anyway. But the government prefers redistribution to patronage[18] and so will set τ low enough to ensure that the lobby has the same preference. Hence, the assumption that the rich can ally with the poor by engaging in some form of patronage relationships drives the result concerning the equilibrium tax rate. If patronage did not exist as a possibility, the unique equilibrium would always see the government setting $\tau=1$ and getting reelected. Total expropriation of the rich is only prevented here because the government sets the tax rate to ensure that the lobby does not prefer *PAT* over *RED*.

This observation highlights the point made in the introduction: It is partly the ability of the rich for collective action that prevents more extensive income redistribution; but it is also the inability of the poor voters to "stick together" as a group that leads to this result. When offered a client's income, it is subgame optimal for each poor voter to accept, but by doing so he allows the lobby to extract a lower real tax rate from the government and hurts his own class. This argument echoes Moore (1966), who recognises the absence of solidarity and cohesiveness among peasants as a main obstacle to political action and revolution: *"The type of weak solidarity that inhibits political action of any variety is mainly a modern phenomenon" [p.477]*. It is also similar to the preference heterogeneity story in Fernandez and Levy (2005), discussed in the review of the literature in section 2.

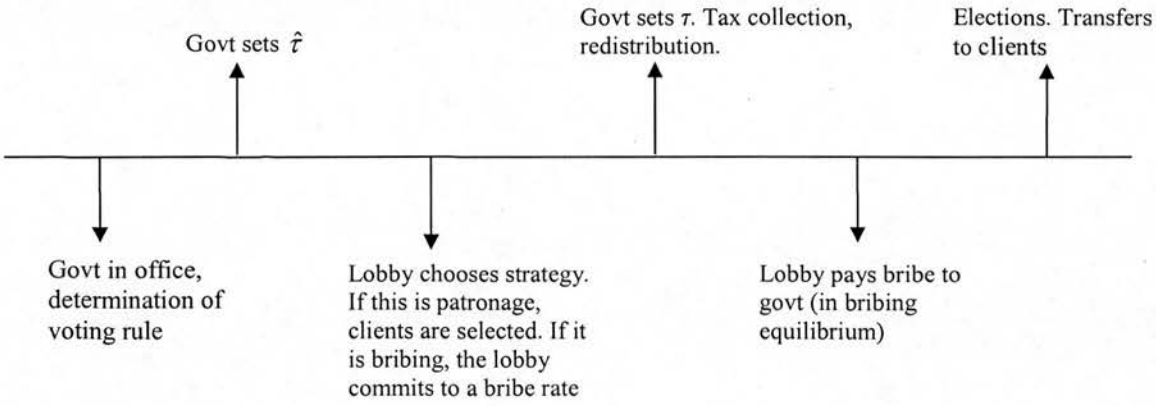
4. Bribes to the government

I will now give a richer structure to the model by allowing for bribe payments from the lobby to the government. This gives the lobby an alternative means of influencing government decisions; it can now either turn towards the government or towards poor voters, and it will choose the cheaper course of action. Bribes from the lobby to the

[18] This follows from the comparison of the government's utility under different outcomes. See the proof of Proposition 1 in the Appendix for more details.

government introduce an additional form of corruption. This is endogenous, in contrast to the share of funds captured by the government. However, the main insight of the model that links corruption to redistributive taxation only refers to fund capture, or government embezzlement, as measured by the parameter β. Therefore, throughout the paper, the term (government) corruption will be used to describe government embezzlement.

The features of the political system are the same (finite number of terms in office, majoritarian elections, backward-looking voting), and we still have the voting strategies given in *(1)*. The timing of the model is the same as before, with two additions (see also timeline below): first, when the lobby decides on its strategy (step 3) it has the option of committing to pay a bribe to the government; and second, bribe payments from the lobby to the government are made after the real tax rate has been set.



Compared to the previous section, there is now one more possible outcome, namely the bribing outcome, henceforth *LOB*: The lobby bribes the government and impedes redistribution. Following the announcement of the statutory tax rate, each member can pay fraction α of his (her) income as a bribe to the government in order to persuade it to set the real tax rate to zero.[19] The period payoff function of the government then becomes: $U = R + n\alpha y_r + \beta\tau n y_r$

[19] It can be checked that when the lobby bribes the government, the target will not simply be to lower the tax rate, but actually to push it all the way down to zero.

The lobby must offer a bribe rate sufficiently high to stop the government from redistributing income; the values of α that ensure this in each of the two periods are calculated in what follows.

Solution of the game

In the second (and last) term of the government, the lobby bribes and achieves a zero real rate. The bribe rate in this case is β,[20] and it is always an optimal strategy for the lobby to pay it. Compared to the simpler model of section 3 where the tax rate was at the highest possible level $(\tau=1)$, the rich class here is able to avoid total expropriation. This is because bribes are allowed, and because the government cannot capture the whole of the tax revenue $(\beta<1)$.[21]

Lemma 3

In the first period, if the lobby bribes the government the bribe rate is $\alpha^* = \frac{R}{ny_r} + \beta(1+\tau)$.

Proof: In the first period, if the lobby chooses to bribe, it has to make the government indifferent between *LOB* and *RED*. The respective total expected payoffs for the government are: $U^{LOB}=R+n\alpha y_r$, $U^{RED} = 2R + \beta\tau n y_r + n\beta y_r$

Comparing the two gives the value of α^*.

The first term in the expression for α^* represents compensation for the loss of the second-period exogenous rent from office. The second term represents compensation for the fact that, by not redistributing in the first period, the government foregoes the part of the tax revenue that it would capture as well as the second period bribe.

Proposition 2 outlines the outcome of the game.

---

[20] It is easy to check this: a bribe rate of β makes the government indifferent between redistributing and accepting the bribe.

[21] In Persson and Tabellini (1999) the same intuition applies: an incumbent would like to maximize the tax rate and consequently his revenue. In that case, the force that disciplines the incumbent and prevents this outcome is political competition.

Proposition 2

When the lobby can make bribe payments to the government the solution of the game is modified as follows:
During its second term in office the government sets the statutory rate to $\hat{\tau} = 1$ *and the real rate to* $\tau=0$*. The equilibrium is LOB with* $\alpha=\beta$*.*

In the first term there are two cases:

(i) *If* $\frac{1-2n}{2-2n} \leq \frac{2R/ny_r + \beta}{(1-\beta)(1-2\beta)}$, *(2)*

the first period equilibrium is RED: The government sets the statutory and real tax rate at $\tau^* = \frac{1-2n}{2-2n}(1-\beta)$ *and is reelected. The exact same outcome obtains if (2) does not hold but* $\alpha^*>1$*.*

(ii) *If (2) does not hold and* $\alpha^* \leq 1$*, the first period equilibrium is PAT: The statutory tax rate is between* τ^* *and 1, the real tax rate is zero and the government is reelected.*

The proof is given in Appendix A2.

In the first of the above cases the outcome is the same as in the simple game of section 3, with redistribution taking place, its extent nevertheless being limited due to the existence of the patronage option for the lobby.

In case *(ii)* the equilibrium outcome is patronage. Thus, allowing for bribe payments from the lobby to the government has the effect of making patronage a candidate equilibrium! An interesting aspect of this case is the multiplicity of statutory tax rates in equilibrium. A way out of this multiplicity is the following proposed refinement: Let there be a cost c to the government of deviating from the announced tax rates and collecting less revenue in the patronage equilibrium. This cost can come, for example, from international monitoring bodies which may impose sanctions on the country. Alternatively, it can be thought of as a decline in the government's overall

credibility which may affect its ability to implement other policies in the next period. The cost of deviation is positive (although it can be arbitrarily small) and a strictly increasing function of the distance d between statutory and real tax rates: $c(\hat{\tau}-\tau)=c(d)$, $d \in [0,1]$, with $c>0$, $\frac{dc}{dd}>0$, $c(0)=0$. In addition, I assume an upper bound on c: $c \leq R$, which ensures that the patronage strategy remains effective.

Introducing this cost of deviation refines the equilibrium in Proposition 2(ii) above and pins down the statutory tax rate at τ^*. Simply put, the idea is that since the government plans to set the real tax rate at τ^*, there is no reason why it should choose any other statutory rate when τ^* equally serves its interests.

Taking the equilibrium refinement into account, the statutory tax rate is τ^* in every case; the proposed negative relationship between corruption and tax rates holds. Real tax rates, on the other hand, are zero if the outcome of the game is *PAT*, which implies a distance of $d=\tau^*$ between statutory and real rates.

Whether an economy ends up in case *(i)* or *(ii)* depends on the realization of the exogenous parameters n, R, y_r, β in *(2)*. More specifically, an economy is more likely to be in the redistribution rather than the patronage equilibrium the higher the level of government corruption β and the ratio of the rent from office to the initial income of the rich, (R/y_r), and the lower n, the proportion of rich voters in the population. The result that the likelihood of redistribution is increasing in β is interesting and somewhat surprising. Intuitively it is because a more corrupt government that plans to appropriate a large part of the tax revenue would require a relatively high bribe rate, thus making *RED* the more likely equilibrium as the lobby would prefer paying the taxes to paying the bribe. In addition to this, high values of β and of (R/y_r) may cause α^* to exceed 1, in which case the bribing strategy is no longer relevant and we end up in the redistribution equilibrium of the baseline model. On the contrary, for low values of β, *LOB* could be cheaper for the lobby, in which case poor voters will set a relatively high reservation tax

rate in order to obtain *PAT* rather than *LOB*.[22] An implication of this result is that, when bribes to the government are part of the setting, reducing corruption only leads to higher tax rates conditional on *(2)* being satisfied; if β is so low that this is no longer the case, then the equilibrium will be *PAT*, the partial redistribution equilibrium with zero real taxes.

5. Unlimited consecutive terms in office

In Propositions 1 and 2 a distinction was made between a government's first and second term in office, with the tax rates varying between the two. This was due to the assumption that parties are not allowed to stay in office for more than two consecutive terms. This assumption is now relaxed.

In the simple model of section 3 the effect of this modification will be to render the result identical in every period; the government will always set the statutory and real tax rate at τ^* in order to prevent patronage. This eliminates the discrepancy in tax rates between periods.

When we allow for bribes from the lobby to the government, the analysis is modified as follows. First, if *LOB* is the equilibrium in one period, then –having assumed identical parties– it will also be the equilibrium in the next period with the other party in office, and so after the second period we are faced with the same problem again. Moreover, the redistribution and patronage equilibria are stationary. This allows us to consider only two periods in order to solve the game with a potentially infinite number of consecutive terms in office. Second, the bribe rate is now different. Finally, the effectiveness of the patronage strategy requires a slightly stricter condition than in section 4 (with bribes but only two periods in office). This is calculated in Appendix A3.

[22] See the proof of Proposition 2 in the Appendix (A2) for details.

Lemma 4

In any period, if the lobby bribes the government the bribe rate is $\tilde{\alpha} = \frac{R}{ny_r} + 2\beta\tau$.

Proof: In any period the lobby has to make the government indifferent between *LOB* and *RED*. The respective payoffs for the government in that period and the next are: $U^{LOB}=R+n\alpha y_r$, $U^{RED} = 2R + 2\beta\tau n y_r$. Comparing the two gives the value of $\tilde{\alpha}$.

Proposition 3

When no upper limit is imposed on the number of consecutive terms that a government can stay in office and bribes from the lobby are allowed, there are two possible equilibria:

(i) If $\frac{1-2n}{2-2n} \leq \frac{R}{(1-\beta)(1-2\beta)ny_r}$, *(3)*

the equilibrium is RED in every period: The government sets the statutory and real tax rate at $\tau^* = \frac{1-2n}{2-2n}(1-\beta)$ *and is reelected. The same result is obtained if (3) does not hold but* $\tilde{\alpha} > 1$.

(ii) If (3) does not hold and $\tilde{\alpha} \leq 1$, *the equilibrium is PAT in every period: The statutory tax rate is* τ^*, *the real tax rate is zero and the government is reelected.*

As in the previous section, the unique determination of the statutory tax rate at $\hat{\tau} = \tau^*$ in case *(ii)* relies on the existence of the cost of deviation, *c*.

The proof is given in Appendix A3.

From *(3)*, which is the analogue of *(2)* in Proposition 2, the likelihood of being in the redistribution equilibrium is increasing in the extent of government corruption, β.

As in the simple case, relaxing the assumption of no more than two consecutive terms in office has the effect of eliminating the difference between first and second period tax rates. The statutory tax rate is always τ^*, while the real rate is either τ^* or zero.

6. Data

The testable hypothesis of the model is that statutory income tax rates are a negative function of government corruption in the form of embezzlement in democratic countries. This result follows from the value τ^* that was calculated in the solution of the game, both for the simple model of section 3 and for those of sections 4 and 5 (see Propositions 1, 2, 3).

I used two alternative sources to obtain statutory income tax rates, which serve as the dependent variable in the regressions. One is the Index of Economic Freedom published by the Heritage Foundation and the other the Annual Report on the Economic Freedom of the World, published by the Fraser Institute (Gwartney and Lawson, 2005). They both provide top marginal income tax rates (expressed in percentages), but there are some differences in their data arising from the fact that they use different sources. There are two reasons why the empirical section uses statutory and not real tax rates as the dependent variable. The first has to do with data availability, as I know of no reliable cross country dataset on real tax rates, i.e. the actual extent to which individuals pay income taxes relative to their income. The second reason relates to the theoretical model, which predicts a unique statutory tax rate of τ^* in every period (section 5) or in all but the last period (sections 3, 4). On the contrary, the result on the real tax rate is not clear cut, as it is predicted to be either τ^* or zero (in the event of patronage). Therefore, although an econometric investigation into the mechanism behind the determination of real tax rates is equally -if not more- important than that of statutory tax rates, the latter are a better fit for this particular model and for the available data.

In order to distinguish between democratic and non-democratic countries I used the Political Rights (Gastil) index, which is published by the Freedom House. Countries with a score of 1-4 are classified as democratic, whereas those with scores of 5-7 as non-democratic. To check robustness, an alternative classification is used, with countries scoring strictly less than 4 classified as democracies.

The way to measure government corruption is less straightforward due to the nature of the phenomenon. I believe that the best way to capture corruption in the way it is modeled in this paper is by means of Public Expenditure Tracking Surveys (PETS).[23] These follow the path of public funds from the central government to the end users in order to determine the extent to which these funds are captured and the stage where the capture occurs. Although they are typically used to track leakage of funds meant for public goods (education, healthcare) and not for direct redistribution, they are a good estimate of a government's proneness and ability to capture resources. However, the PETS that have been carried out up to this date are very limited in number and exhibit methodological differences; therefore they are not suited for cross country analysis.

The readily available alternative is found in aggregate subjective indices, three of which are used in this paper.[24] Although far from perfect, these are arguably the best available tools for measuring corruption across countries. The first is an index of "Diversion of Public Funds" (which I will call *DIV*), published in the World Economic Forum's 2004-2005 *Global Competitiveness Report*. This index corresponds quite closely to the particular form of corruption discussed in this paper, as it is based on a survey question on the frequency of "diversion of public funds to companies, individuals or groups due to corruption". The second index is the Transparency International Corruption Perceptions Index (*CPI*), a composite indicator that uses surveys from various sources. For the 2004 *CPI*, Transparency International used sources from 12 independent institutions to collect data on corruption for 146 countries. The data are collected by

[23] For a comprehensive survey on PETS and other micro level measures of corruption see Reinikka and Svensson (2006).
[24] For a rigorous discussion on governance indicators see Arndt and Oman (2006).

means of surveys of businessmen and country analysts, and the questions asked typically refer to the frequency of irregular payments to government officials. This is quite different from the definition of corruption in this paper; nevertheless one expects different aspects of corruption to be highly correlated. The main regressions performed in the paper use the *DIV* and *CPI* indices, while a third index is used to check robustness: The Control of Corruption (*CC*) index, which is constructed by the World Bank (Kaufmann et al, 2005) and uses a variety of sources (up to 14 surveys/polls for each country, including *DIV*) and a broad definition of corruption.[25] The *CPI* assumes values from 0 to 10, the *DIV* index ranges from 1 to 7 and the *CC* index from -2.5 to 2.5. In order to make their coefficients comparable I rescaled all the indices so that they measure from 0 to 10, and transformed them such that higher values always correspond to more corruption.

The controls that are used in the basic regressions are log GDP per capita and log population. GDP per capita captures a country's level of development, which should affect both corruption and tax rates. Wagner's Law states that the share of public expenditure in national income increases with the level of development, so one should expect a positive relationship between log GDP per capita and tax rates. Log population is used to account for the fact that countries with higher population tend to rely more on income taxes and less on trade taxes. Easterly and Rebelo (1993) observe empirically this scale effect and attribute it to the fact that income taxes are associated with high set-up costs but low marginal administrative costs, so they are more likely to be used extensively in countries with high populations. We must therefore expect a positive sign for log population in the regressions.

Table 1 provides the summary statistics of the dependent and independent variables in the regressions, including the instruments that are used to check robustness.

[25] According to its official definition the Control of Corruption Index measures "the exercise of public power for private gain, including both petty and grand corruption and state capture".

Charts 1 and 2 give a rough impression of the relationship between corruption, tax rates and regime type. Chart 1 is a scatterplot of tax rates (from the Fraser Institute) against corruption (the *DIV* index) for the subsample of democratic countries, whereas Chart 2 includes only non-democratic ones. The linear regression lines are also fitted: There is a clear negative relationship in the first chart, whereas in the second chart the direction of this relationship is inverted. The same pattern emerges if one uses alternative measures of tax rates or corruption.

Table 1: Summary statistics

All countries	*CPI*	*DIV*	*HER*	*FRA*	*LGDPPC*	*LPOP*	*ELF*	*ANT*	*LAT*
Mean	5.73	5.13	33.53	36.81	3.41	7.10	0.35	0.43	0.28
Median	6.55	5.67	33.50	35.00	3.35	7.03	0.27	0.38	0.22
Max	8.50	8.67	60.00	65.00	4.68	9.11	0.86	1.00	0.72
Min	0.30	0.67	12.00	12.00	2.05	5.45	0.00	0.07	0.01
St. deviation	2.30	2.33	9.78	10.76	0.68	0.69	0.30	0.26	0.19
No. of observations	118	82	118	85	118	118	92	92	92
Democracies	*CPI*	*DIV*	*HER*	*FRA*	*LGDPPC*	*LPOP*	*ELF*	*ANT*	*LAT*
Mean	5.33	5.11	34.15	36.26	3.53	7.03	0.30	0.43	0.30
Median	6.10	5.67	34.00	35.00	3.57	7.02	0.19	0.37	0.22
Max	8.50	8.67	60.00	59.00	4.68	9.03	0.86	0.98	0.72
Min	0.30	0.67	13.00	12.00	2.22	5.45	0.00	0.07	0.01
St. deviation	2.37	2.37	9.29	9.75	0.69	0.71	0.28	0.26	0.20
No. of observations	89	72	89	71	89	89	67	67	67
Non-democracies	*CPI*	*DIV*	*HER*	*FRA*	*LGDPPC*	*LPOP*	*ELF*	*ANT*	*LAT*
Mean	6.95	5.28	31.64	39.57	3.02	7.31	0.47	0.44	0.20
Median	7.30	5.92	30.00	35.00	3.06	7.25	0.61	0.39	0.21
Max	8.50	7.67	60.00	65.00	4.37	9.11	0.86	1.00	0.39
Min	0.70	1.17	12.00	13.00	2.05	6.14	0.00	0.07	0.01
St. deviation	1.52	2.11	11.13	15.03	0.48	0.60	0.31	0.26	0.14
No. of observations	29	10	29	14	29	29	25	25	25

CPI: Corruption Perceptions Index, DIV: Diversion of Public Funds Index, HER: Tax rates (Heritage Foundation), FRA: Tax rates (Fraser Institute), LGDPPC: Log GDP per capita, LPOP: Log population, ELF: Ethnolinguistic fractionalisation, ANT: State antiquity, LAT: Latitude

Chart 1

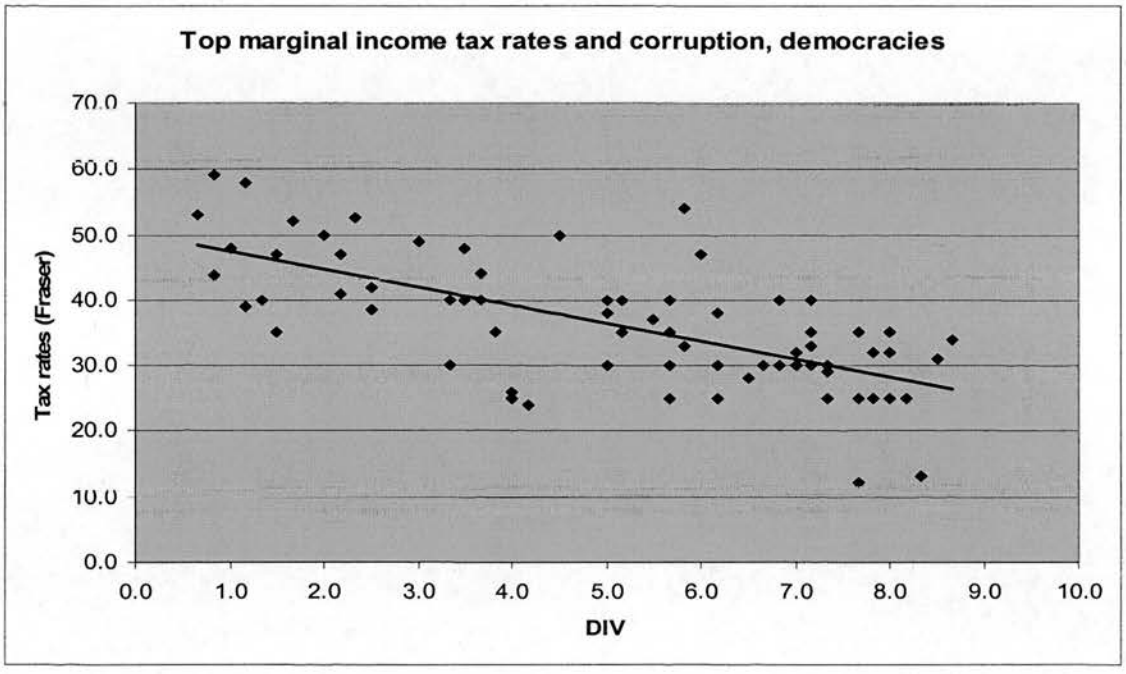
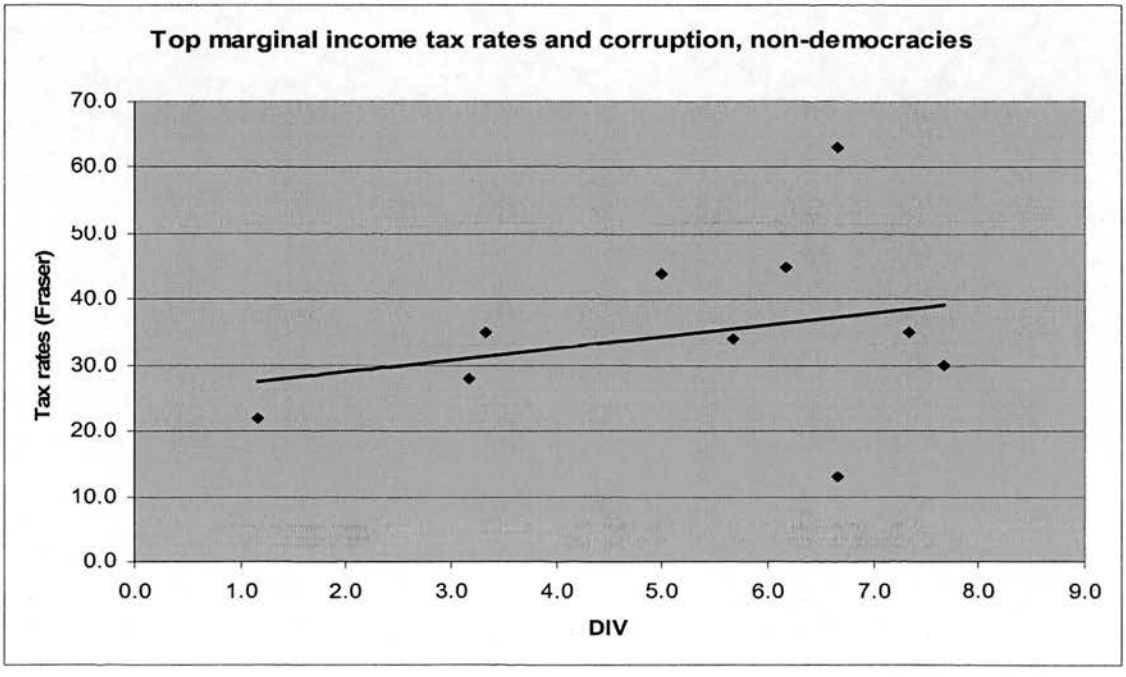


Chart 2



Specification

The regression specification is:

$$TAX_i = \alpha + \beta_1 * DEM_i + \beta_2 * CORRUPTION_i + \beta_3 * (CORRUPTION_i * DEM_i) + \beta_4 * LGDPPC_i + \beta_5 * LPOP_i + u_i$$

where the index i denotes different countries, *TAX* is the top marginal income tax rate, *DEM* is a dummy variable that takes the value 1 if a country is classified as democratic and 0 otherwise, *CORRUPTION* is one of the corruption indices, *LGDPPC* is log GDP per capita and *LPOP* is log population.

The regressions use data for 2004. The number of countries included in each specification is constrained by data availability; in total, the sample comprises 118 countries. 38 of those are located in Europe, 24 in Asia, 23 in Sub-Saharan Africa, 20 in Latin America and the Caribbean, 9 in North Africa and the Middle East, 2 in North America and 2 in Oceania.

The main hypothesis is that the effect of corruption on tax rates in a democracy should be negative and significant. This effect is captured by the combined coefficient $(\beta_2 + \beta_3)$, for which one should expect a negative sign. On the other hand, the effect of corruption on tax rates in non-democracies is given by β_2. This paper does not predict anything about that effect, but one would ideally see a clear distinction between democracies and non-democracies. Such a distinction would be captured by a significant coefficient for the interaction variable.

7. Regression results

Table 2 shows the results of running the above regression for a cross-country sample using Ordinary Least Squares. Specifications 1-2 use the tax rates obtained from the

Heritage Foundation as the dependent variable, while specifications 3-4 those obtained from the Fraser Institute.

The key thing to note is that all specifications confirm the main result that corruption has a negative impact on income tax rates in democracies. Calculating the combined coefficient $(\beta_2 + \beta_3)$ allows us to estimate the magnitude of this impact: a 1-unit increase in the corruption index leads to a fall in income tax rates which ranges from 1.39 to 2.19 percentage points. Running a test on the restriction $\beta_2 + \beta_3 = 0$ reveals that this impact is always statistically significant at the 1% level.

The interaction variable is highly significant in all four specifications, thereby supporting the hypothesis that the relationship between corruption and tax rates is affected by the nature of the political system; it is in democracies only that a negative empirical relationship is observed. Hence, the paper advocates the relevance of voting models by providing evidence that democracy matters.

The coefficient on the corruption variable, which corresponds to the effect of corruption on tax rates in non-democracies, is significant in only one specification. But it is very interesting to see that this coefficient is positive in all four specifications, and in the one where it is significant a 1-unit increase in the corruption index is associated with an increase in income tax rates of 4.14 percent. This points to a positive relationship between corruption and tax rates in non-democracies; intuitively, one could think that a more corrupt government will have stronger incentives to raise tax rates and appropriate part of the tax revenue, unconstrained by electoral considerations.

The democracy dummy captures the part of the effect of democracy on tax rates which cannot be attributed to corruption and to the motive of maximising the diverted tax revenue. In other words, the coefficient on *DEM* measures the effect of democracy on tax rates in the absence of corruption. Accountability to the electoral body implies a positive effect. Indeed, *DEM* enters all regressions highly significant and with a positive sign: democracies tend to have higher income tax rates on average. Log GDP per capita has the

expected positive sign in the three specifications where it is significant; log population is always positive and it is also significant in three out of four specifications, lending support to the intuition given above concerning scale effects.

Table 2: Basic Regression Results

OLS estimation. Dependent variable is: Top marginal income tax rates				
	1	**2**	**3**	**4**
Intercept	-24.1 (14.902)	3.849 (12.559)	-15.129 (15.083)	-13.655 (15.982)
DEM	18.481 *** (4.897)	19.63 *** (5.266)	22.313 *** (6.773)	31.92 *** (5.181)
DIV	1.681 (1.29)		2.168 (1.776)	
CPI		0.628 (0.93)		4.141 *** (1.306)
*DIV*DEM*	-3.073 *** (1.142)		-4.142 ** (1.685)	
*CPI*DEM*		-2.82 *** (0.865)		-6.125 *** (1.166)
LGDPPC	4.983 ** (1.955)	-0.286 (2.173)	4.871 *** (1.826)	4.43 * (2.627)
LPOP	4.097 *** (1.203)	3.324 *** (1.139)	2.993 *** (1.131)	1.662 (1.316)
R^2 *adj.*	0.346	0.205	0.405	0.400
SAMPLE SIZE	82	118	82	85
*Robust standard errors in brackets. *, **, *** denotes significance at the 10%, 5%, 1% level respectively.*				
DIV and CPI stand for the corruption indices from the World Economic Forum and Transparency International respectively. DEM is a democracy dummy, LGDPPC is log GDP per capita and LPOP is log population.				

IV estimation and robustness checks

Endogeneity is an issue commonly raised by researchers when some aspect of economic performance (typically growth) is regressed on an institutional measure (such as corruption or other indicators of governance). In this case where the dependent variable is the income tax rate, I believe it is relatively harder to come up with potential channels through which the dependent variable determines the right hand side variable, namely corruption.[26] In any case, there might be an endogeneity issue that needs to be adressed.

In order to account for such a possibility, I ran two regressions using instrumental variables. Finding plausible instruments for corruption is an empirical challenge, and various solutions have been proposed in the literature.[27] Since IV is only performed here to check robustness, I have simply used three of the variables commonly employed as instruments. Those are: *(i)* state antiquity, from Bockstette et al (2002), to account for the possibility that countries with a longer history of state-level institutions produce better outcomes in terms of governance, *(ii)* latitude, from La Porta et al (1999), as a proxy for the negative effect of tropical -as opposed to temperate- climates on development and institutions, and *(iii)* ethnolinguistic fractionalisation, also from La Porta et al (1999), which is generally associated with worse governance and higher corruption.[28] The signs from the first stage regression of corruption on the instruments, reported in Appendix A5, are consistent with these presuppositions in the cases of state antiquity and ethnolinguistic fractionalisation. However, latitude has a counterintuitive positive coefficient. The tests of overidentifying restrictions suggest that the instruments are valid.

[26] One such channel could work in the following way: Higher tax rates induce individuals to invest more effort in tax evasion, thus leading to more corruption. It must be noted, though, that such a link implies a positive correlation between tax rates and corruption, thereby making it likely that the magnitude of the effect of corruption on tax rates is in fact even greater than the one identified in regressions 1-4. Of course, other stories could be told that would imply a negative impact of tax rates on corruption. For example, high taxation increases state revenue which can be used to improve institutions and fight corruption.

[27] A survey and classification of relevant instruments can be found in Pande and Udry (2005).

[28] See also Mauro (1995), Hall and Jones (1999), Easterly and Levine (2002) for discussions on the effects of ethnolinguistic fractionalisation and latitude on institutions and on their use as instruments.

IV estimation yields generally robust results, which are shown in columns 5 and 6 of table 3. The negative effect of corruption on income tax rates in democracies remains significant at the 5% level, with a 1 unit increase in the *CPI* corruption index leading to a fall of 2.45 percentage points in the income tax rate (column 5). In non-democracies the same effect is positive and insignificant, while the interaction effect is negative and significant at the 1% level. When the instrumented corruption variable is *DIV* (column 6), the sample size is considerably smaller and the results are weaker: the effect of corruption on tax rates in democracies is found to be smaller (1.84 percentage point) and no longer significant.

Table 3 presents a number of further robustness checks. I tried out alternative specifications using a more extended set of controls. Column 7 adds openness of the economy (exports plus imports as a percentage of GDP) and column 8 adds a Scandinavia and a sub-Saharan Africa dummy. These variables are insignificant and do not change any of the main results. Similarly, robustness checks were performed including in the set of controls the first lag of the present value of debt as a percentage of Gross National Income and the first lag of government final consumption expenditure as a percentage of GDP. Again, no changes were observed, and these regressions are not reported in the tables. However, one must point out the possibility that the results suffer from omitted variable bias, to the extent that there may be more factors that influence the government's fiscal policy and have not been accounted for.

Specification (9) is similar to (1), with the difference that a country is classified as a democracy if its score in the Political Rights index is not higher than 3. This "stricter" definition of democracy is adopted here to ensure that the results do not depend on the somewhat arbitrary choice of a democracy threshold. Finally, the OLS regressions were repeated using *CC* as the measure of corruption. All aforementioned results were found to hold, with highly significant coefficients on the index and on its interaction with democracy.

Table 3: Robustness checks

Dependent variable is: Top marginal income tax rates (Heritage Foundation)					
	IV	*IV*	*OLS*	*OLS*	*OLS*
	5	**6**	**7**	**8**	**9**
Intercept	2.66 (25.00)	-24.582 (21.046)	-25.06 (17.832)	-29.602 * (16.245)	-8.49 (14.3)
DEM	37.972 *** (9.924)	21.028 *** (5.827)	18.956 *** (7.059)	16.501 *** (5.298)	16.028 ** (7.641)
DIV		2.129 (1.586)	1.751 (1.44)	1.774 (1.347)	0.156 (1.095)
CPI	3.692 ** (1.641)				
*DIV*DEM*		-3.971 *** (1.214)	-3.143 ** (1.293)	-2.83 ** (1.203)	-2.28 ** (1.145)
*CPI*DEM*	-6.146 *** (1.402)				
LGDPPC	-1.529 (3.812)	3.984 (3.932)	4.964 ** (1.988)	6.098 *** (2.215)	3.877 ** (1.87)
LPOP	1.829 (1.204)	4.578 *** (1.29)	4.155 *** (1.316)	4.258 *** (1.222)	3.595 *** (1.28)
OPEN			0.002 (0.019)		
SCAND				5.558 (6.227)	
SUB				2.135 (2.311)	
R^2 *adj.*	0.274	0.47	0.337	0.344	0.409
SAMPLE SIZE	92	64	82	82	82

*Robust standard errors in brackets. *, **, *** denotes significance at the 10%, 5%, 1% level respectively.*

DIV and CPI stand for the corruption indices from the World Economic Forum and Transparency International respectively. DEM is a democracy dummy, LGDPPC is log GDP per capita, LPOP is log population, OPEN is openness of the economy (exports plus imports, as % of GDP), SCAND is a dummy for Scandinavian countries and SUB is a dummy for sub-Saharan African countries.

8. Concluding remarks

Income taxation is the primary tool that most governments emply to influence the distribution of income among citizens. The traditional political economy literature explains the level of income tax rates by means of an almost mechanical application of the median voter theorem. This paper suggests that a somewhat different mechanism may underlie the determination of redistributive taxation. Governments are not only concerned about elections, but have a stake in the level of taxation because they divert part of the revenue. Moreover, the interaction between social classes shapes voting preferences and consequently equilibrium tax rates. More specifically to this last point, the paper identifies patron-client ties and the lack of coherence among poor voters' strategies as responsible for the relatively limited extent of income redistribution and persistent income inequality. A further result of the paper is the negative impact of government corruption on income tax rates, conditional on the existence of a democratic regime. This impact was derived theoretically and tested empirically, with the evidence corroborating the theory.

On a theoretical level, possible next steps would be to consider party heterogeneity, and to endogenize government corruption in order to look at the joint determination of corruption and redistribution. Another interesting direction would be the use of the model's predictions about statutory and real tax rates in order to analyse and predict levels of tax evasion. Finally, the analysis could be extended to include public goods provision and a broader set of redistributive policies. The latter is also an empirical challenge, along with the use of micro level data on corruption and with the measurement of the occurrence and real extent of patronage.

References

Alesina, A., AngeletosG.M., 2005. "Fairness and redistribution", *American Economic Review* 95, 960-980

Alesina, A., La Ferrara, E., 2002. "Preferences for redistribution in the land of opportunities", *CEPR Discussion Paper* no. 3155

Alesina, A., Rodrik, D., 1994. "Distributive politics and economic growth", *The Quarterly Journal of Economics* 109, 465-490

Arndt, C., Oman, C., 2006. "Uses and abuses of governance indicators", *OECD Development Centre Studies*

Baland, J-M., Robinson, J., 2006. "Land and Power: Theory and Evidence from Chile", *NBER Working Paper* no. 12517

Barro, J., 1973. "The control of politicians: An economic model", *Public Choice* 14, 19-42

Benabou, R., Ok, E.A., 2001. "Social mobility and the demand for redistribution: The POUM hypothesis", *The Quarterly journal of economics* 116, 447-487

Benabou, R., Tirole, J., 2005. "Belief in a just world and redistributive politics", *NBER Working Paper* no. 11208

Bockstette, V., Chanda, A., Putterman, L., 2002. "States and markets: The advantage of an early start", *Journal of Economic Growth* 7, 347-369

Downs, A., 1957. "An economic theory of democracy", New York: Harper and Row

Easterly, W., Levine, R., 2002. "Tropics, germs, and crops: How endowments influence economic development", *NBER Working Paper* no. 9106

Easterly, W., Rebelo, S., 1993. "Fiscal policy and economic growth: An empirical investigation", *Journal of Monetary Economics* 32, 417-458

Fernandez, R., Levy, G., 2005. "Diversity and redistribution", *NBER Working Paper* no. 11570

Friedman, E., Johnson, S., Kaufmann, D., Zoido-Lobaton, P., 2000. "Dodging the grabbing hand: The determinants of unofficial activity in 69 countries", *Journal of Public Economics* 76, 459-493

Gwartney, J., Lawson, R., 2005. "Economic Freedom of the World: 2005 Annual Report". Vancouver: *The Fraser Institute*. Data retrieved from www.freetheworld.com.

Hall, R., Jones, C., 1999. "Why do some countries produce so much more output per worker than others?", *The Quarterly journal of Economics* 114, 83-116

Kaufmann, D., Kraay, A., Mastruzzi, M., 2005. "Governance matters IV: Governance indicators for 1996-2004", World Bank

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R.W., 1999. "The quality of government", *The Journal of Law, Economics, and Organization* 15, 222-279

Levy, G., 2005. "The politics of public provision of education", *The Quarterly Journal of Economics* 120, 1507-1534

Lindbeck, A., Weibull, J., 1987. "Balanced-budget redistribution as the outcome of political competition", *Public Choice* 52, 273-297

Mauro, P., 1995. "Corruption and growth", *The Quarterly Journal of Economics* 110, 681-712

Meltzer, A.H., Richard, S., 1981. "A rational theory of the size of government", *Journal of Political Economy* 89, 914-927

Moore, B., 1966. "The social origins of dictatorship and democracy: Lord and Peasant in the Making of the Modern World", *Boston: Beacon Press*

Olken, B.A., 2006. "Corruption and the costs of redistribution: Micro evidence from Indonesia", *Journal of Public Economics* 90, 853-870

Pande, R., Udry, C., 2005. "Institutions and development: A view from below", *Yale University Working Paper*

Persson, T., Tabellini, G., 1994. "Is inequality harmful for growth?", *American Economic Review* 84, 600-621

Persson, T., Tabellini, G., 1999. "The size and scope of government: Comparative politics with rational politicians. 1998 Alfred Marshall Lecture", *European Economic Review* 43, 699-735

Persson, T., Tabellini, G., 2000. "Political economics: Explaining economic policy", MIT Press

Piketty, T., 1995. "Social mobility and redistributive politics", *The Quarterly Journal of Economics* 110, 551-584

Ravallion, M., Lokshin, M., 2000. "Who wants to redistribute? Russia's tunnel effect in the 1990s", *Journal of Public Economics* 76, 87-104

Platteau, J.P., 1995. "A Framework for the analysis of evolving patron-client ties in agrarian economies", *World Development* 23, 767-786

Reinikka, R., Svensson. J., 2006. "Using Micro Surveys to Measure and Explain Corruption", *World Development* 34, 359-370

Roberts, K., 1977. "Voting over income tax schedules", *Journal of Public Economics* 8, 329-340

Romer, T., 1975. "Individual welfare, majority voting and the properties of a linear income tax", *Journal of Public Economics* 7, 163-168

Svensson, J., 2005. "Eight questions about corruption", *Journal of Economic Perspectives* 19, 19-42

World Economic Forum, 2004. "The Global Competitiveness Report 2004-2005"

Appendix

A1. Proof of Proposition 1

It has already been shown that in the second period the electoral motive disappears and the government will set the tax rate at 1 so as to maximize the diverted tax revenue. The government's payoff in the second period is therefore $R + n\beta y_r$. We can then write out the total ex ante payoffs of the government in period 1, i.e. when it decides on its strategy. These are:

$$U^{PAT} = 2R + n\beta y_r$$
$$U^{RED} = 2R + \beta\tau n y_r + n\beta y_r$$

For *PAT* to be sustainable in equilibrium, the patronage strategy on part of the lobby must be effective and lead to zero taxes and reelection. Indeed, if the lobby opts for patronage, then the government should either set $\tau=0$ (if it wants to be reelected) or $\tau>0$ (if it chooses to implement the statutory rate and forgo reelection). Even if $\tau=1$, the respective payoffs are: $U^{PAT} = 2R + n\beta y_r$, $U^{l}=R + n\beta y_r$, so that a real tax rate of zero is always the best response by the government to patronage.

Turning to the lobby, we can write out the total ex ante cost of the two strategies (which correspond to the two possible equilibria) available to them in period 1 (costs as fractions of y_r):

$$C^{PAT} = \frac{1-2n}{2-2n}(1-\beta)+\beta$$

$$C^{RED} = \tau + \beta$$

Comparing these costs we have the following condition for the lobby:

RED is preferred to *PAT* when $\tau \leq \frac{1-2n}{2-2n}(1-\beta)$

Now, from an inspection of the government's payoffs it is immediately obvious that *PAT* is dominated by *RED*, and so the government will set the statutory rate in order to ensure that the lobby never chooses *PAT* over *RED*, that is: $\hat{\tau} = \frac{1-2n}{2-2n}(1-\beta)$. I call this value of the tax rate τ^*.

Let us now consider the strategy of the poor class. It was already explained in the proof of Lemma 2 that it is optimal for clients to set $p^{R/C}$ as given in *(1)*. As for $p^{R/NC}$, note the following: Given the rationality and full information assumptions, the poor that have not been selected as clients face the following problem: they can set $\underline{\tau} \leq \tau^*$, in which case the equilibrium will be *RED* and the real tax rate will be τ^*. This gives us the tax rate in Proposition 1. Or they can set $\underline{\tau} > \tau^*$. In this case redistribution is no longer feasible: to

see this, note that redistribution would require the government to set a statutory tax rate greater or equal to $\underline{\tau}$, and therefore strictly greater than τ^*. However, for such a tax rate, the lobby strictly prefers *PAT* to *RED*. In fact, when $\underline{\tau} > \tau^*$ it is optimal for the government to set the statutory rate greater or equal to τ^*. This will trigger the patronage strategy on part of the lobby and the equilibrium will be *PAT*. If, instead, the government were to set some $\hat{\tau} < \tau^*$, its payoff, say U', would be lower: $U' = R + n\beta\tau y_r < U^{PAT}$.

From the point of view of the poor class *RED* is preferred to *PAT*, since it yields a higher expected payoff. They will therefore set $\underline{\tau} \leq \tau^*$ (any $\tau \in [0, \tau^*]$ will do) and the equilibrium will always be *RED* with $\tau = \tau^*$.

This completes the proof.

A2. Proof of Proposition 2

In the second period the equilibrium is *LOB*, the lobby pays fraction $\alpha=\beta$ of its income and the real tax rate is zero. The government's payoff in the second period is therefore $R + n\beta y_r$.

In period 1, if $\alpha^*>1$, the bribing strategy is not relevant and the analysis is the same as in the basic model, leading to the same first period equilibrium *(RED)* as in Proposition 1.
If $\alpha^* \leq 1$, the possible total ex ante payoffs of the government in period 1 are:

$$U^{PAT} = 2R + n\beta y_r$$
$$U^{LOB} = R + n\alpha^* y_r$$
$$U^{RED} = 2R + \beta\tau n y_r + n\beta y_r$$

The total ex ante costs of the three different strategies (which correspond to the three different equilibria) available to the lobby in period 1 are (as fractions of y_r):

$$C^{PAT} = \frac{1-2n}{2-2n}(1-\beta)+\beta$$

$$C^{LOB} = 2\left[\frac{R}{ny_r}+\beta(1+\tau)\right]$$

If *LOB* is the equilibrium in the first period, it will be the equilibrium with a new government in the following period assuming a stationary environment and identical parties; this is why both terms are multiplied by 2.

$$C^{RED} = \tau+\beta$$

Comparing these costs pairwise we have the following conditions for the lobby:

(i) *RED* is preferred to *PAT* when $\tau \leq \frac{1-2n}{2-2n}(1-\beta)$

(ii) *RED* is preferred to *LOB* when $\tau \leq 2\left(\frac{R}{ny_r}+\beta\tau\right)+\beta$ *(A1)*

(iii) *PAT* is preferred to *LOB* when $\frac{1-2n}{2-2n}(1+\beta)+\beta \leq 2\left[\frac{R}{ny_r}+\beta(1+\tau)\right]$ *(A2)*

As in the simple model, $U^{PAT}<U^{RED}$, and the government will set the statutory rate in order to ensure that the lobby never chooses *PAT* over *RED*, that is: $\hat{\tau} = \tau^* = \frac{1-2n}{2-2n}(1-\beta)$

Let us assume for now that $\underline{\tau} \leq \tau^*$. Upon the announcement of $\hat{\tau} = \tau^*$, the lobby prefers *RED* to *PAT*. The equilibrium then depends only on *(A1)*, while *(A2)* is redundant. Indeed, if from *(A1) RED* is preferred to *LOB*, then *RED* is preferred to both alternatives and is the choice of the lobby, with the real rate also at the level of τ^*. In the opposite case that *LOB* is preferred to *RED*, then by transitivity it is also preferred to *PAT*. Thus, given $\hat{\tau}$, the decision of the lobby is only based on *(A1)*. According to the direction of this inequality the first period equilibrium will either be *LOB* or *RED*.

Finally, substituting $\tau=\tau^*$ in *(A1)* we get the result that the outcome is redistribution if and only if $\frac{1-2n}{2-2n} \leq \frac{2R/ny_r+\beta}{(1-\beta)(1-2\beta)}$, *(A3)*

which is condition *(2)*.

The analysis for the strategy of the poor class is the following (for $p^{R/NC}$): If the poor set $\underline{\tau} \leq \tau^*$, the equilibrium is either *RED* or *LOB* depending on *(A3)*. If they set $\underline{\tau} > \tau^*$, redistribution is not feasible. It is then optimal for the government to set the statutory tax rate in the interval $\hat{\tau} \in [\tau^*,1]$ in order to trigger patronage. This is because, absent the redistribution option, patronage can easily be shown to be the best possible outcome for the government.

From the point of view of the poor class, the best equilibrium is *RED*, since it yields a higher expected payoff than *PAT*. The worst equilibrium is *LOB*. If they know that *(A3)* holds, their choice is between *RED* and *PAT*. They will then set $\underline{\tau} \leq \tau^*$ (any $\tau \in [0,\tau^*]$ will do) and the equilibrium will be *RED* with $\tau = \tau^*$. If, however, they know that *(A3)* does not hold, then their choice is between *LOB* and *PAT*, in which case they will set $\underline{\tau} > \tau^*$ (any $\tau \in [\tau^*,1]$ will do) and the equilibrium will be *PAT* with $\hat{\tau} \in [\tau^*,1]$.

This completes the proof.

A3. Proof of Proposition 3

In this case, the various payoffs to the government and costs to the lobby become (it has been shown in the text that we can restrict our attention to two periods only):

$$U^{PAT} = 2R$$

$$U^{LOB} = R + n\tilde{a}y_r$$

$$U^{RED} = 2R + 2\beta\tau ny_r$$

$C^{PAT} = \frac{1-2n}{2-2n}(1-\beta)$, each period

$C^{LOB} = \frac{R}{ny_r} + 2\beta\tau$, »»

$C^{RED} = \tau$, »»

Proceeding as in the proof of Proposition 2 and using the cost of deviation c to get rid of the multiplicity of statutory tax rates in the patronage equilibrium, we arrive at the result that the equilibrium is either *RED* with $\tau = \hat{\tau} = \tau^*$, or *PAT* with $\hat{\tau} = \tau^*$ and $\tau=0$.
The condition that determines whether the equilibrium is *RED* or *PAT* is derived in the same manner as in Proposition 2, and it is the following:
The outcome is redistribution if and only if $\frac{1-2n}{2-2n} \leq \frac{R}{(1-\beta)(1-2\beta)ny_r}$,
which is condition *(3)*.

Finally, the necessary condition for patronage to be effective is now $c \leq R - n\beta y_r$.
Indeed, if in any given period the lobby opts for patronage, the alternative government utility levels for that period and the next are: $U^{PAT} = 2R - c(1)$, $U^1 = R + n\beta y_r$.
Comparing the two gives the above condition. Note that, if it is optimal for a party to set $\tau=1$ when faced with patronage, this will also be the case for the other party in the next period. So after one period the first party will be in power again and the problem will be the same. This allows us to consider only two periods, even though there is now no limit in the number of consecutive terms in office. This relies crucially on the assumption that the parties are identical.

This completes the proof.

A4. Issues of commitment and discounting

The results obtained in this paper rely on the implicit assumption that the lobby is able to commit credibly to make the payments to the other agents, whether that be the government or the clients (the timing is such that the government and clients would have no incentive to renege on their agreements with the lobby).

One way to motivate this assumption is to invoke trigger strategies in a repeated game. Consider the basic model in section 3 and the following idea: If the lobby does not make the promised patronage payments to the clients, the clients do not trust the lobby in any subsequent period. In other words the lobby loses its credibility and therefore the ability to use the patronage strategy in the future. This means that we move from the equilibrium in Proposition 1 to the equilibrium with redistribution and $\tau=1$ in every period. In the second term of every government the two equilibria are the same. However, in the first term the equilibrium with $\tau=1$ is more costly to the lobby, and the difference in costs as a share of y_r is (as a share of y_r): $1-\frac{1-2n}{2-2n}(1-\beta)=\frac{1+\beta(1-2n)}{2(1-n)}$. Call this $C1$; this cost is borne every second period. The benefit from not making the patronage payment in some period is $k^*=\frac{n}{1-n}(1-\beta)$.

Given that there is no discounting, if we extend the horizon of the game and take all future periods into account the total cost is bound to be greater than the benefit and the lobby will never break its promise to pay the bribe, leading to the result that commitment is in fact credible.

The reason why there is no discounting in the model is that it would add nothing substantial to the analysis, in the light of the fact that we can restrict attention to just two periods to define the equilibrium of the game. However, in the context of the present discussion on commitment in a repeated game, it makes sense to introduce a discount

factor and derive a condition for credible commitment. Let δ be the discount factor. Then the future cost to the lobby of losing credibility is: $\delta^2 C1+\delta^4 C1+\delta^6 C1+... = \frac{\delta^2 C1}{1-\delta^2}$

Comparing this with the one-off benefit, it follows that the lobby has no incentive to break its commitment as long as $\frac{2n(1-\beta)}{1+\beta(1-2n)} \leq \frac{\delta^2}{1-\delta^2}$, i.e. when the lobby cares sufficiently about the future.

The same argument can be applied to payments from the lobby to the government. Suppose that, in the bribing equilibrium of the second period in Proposition 2, if the lobby does not pay the promised bribe the government never trusts it again. This means that the lobby loses the bribing option and we move from the equilibrium in Proposition 2 to the one in Proposition 1. Comparing the two equilibria, we can verify that in the first period the cost to the lobby is the same in any equilibrium (*RED* in Proposition 1, *RED* or *PAT* in Proposition 2), and equal (as a function of y_r) to $\frac{1-2n}{2-2n}(1-\beta)$. We can therefore restrict the comparison to the second period, when the equilibrium in Proposition 2 comes at a cost of β, whereas that in Proposition 1 comes at a cost of 1. Hence the cost to the lobby from the loss of credibility is $1-\beta$, and is borne every second period. On the other hand, the one-off benefit from not paying the bribe in any period is β.

Then the future cost to the lobby of of losing credibility is:

$$(1-\beta)\delta^2+(1-\beta)\delta^4+(1-\beta)\delta^2+... = \frac{\delta^2(1-\beta)}{1-\delta^2}$$

It follows that the lobby has no incentive to break its commitment as long as $\frac{\beta}{1-\beta} \leq \frac{\delta^2}{1-\delta^2}$, i.e. when the horizon is sufficiently long relative to the immediate benefit of cheating today.

A5. First stage regression results (from specification 5)

First stage regression. Dependent variable is: *CPI*			
	coefficient	*st. error*	*p-value*
Intercept	14.346	2.156	0.000
ELF	2.228	1.294	0.089
LAT	5.861	3.116	0.064
ANT	-3.912	1.341	0.005
R^2 adj. = 0.77, N=92			
ELF=ethnoliguistic fractionalisation, from La Porta et al (1999). LAT=latitude: distance from the equator scaled between 0 and 1, from La Porta et al (1999). ANT=state antiquity, from Bockstette et al (2002). The list of independent variables further included DEM, LGDPPC, LPOP, and the interactions of the three instruments with the democracy dummy.			

Sargan chi-square test of overidentifying restrictions= 7.81 (p=0.099)

Basmann chi-square test of overidentifying restrictions= 7.61 (p=0.11)

PART I I

Chapter 2

Third party beliefs and collusion in a repeated psychological game

*Third party beliefs and collusion in a repeated psychological game**

Abstract

Psychological games include players' beliefs directly into their utility functions, thus allowing emotions to affect strategies and the outcome of the game. The paper uses this framework and studies the effect of guilt on decision making in the context of a repeated collusion game with three players. The distinctive feature of the model is that it conditions guilt on the perceived beliefs of a "third party": This is one of the players, who does not move in the game but is affected by its outcome. Two equilibria are possible, one with collusion in every period and one in which collusion never occurs. When I allow for updated beliefs, collusion is more likely to emerge in equilibrium when the horizon of the game is relatively long and when the third party holds low initial expectations and updates his (her) beliefs fast. Compared to the standard, non-psychological version of the game, the inclusion of a psychological component (guilt) in utility has the effect of making collusion less likely to be the equilibrium. The results highlight the self-fulfilling nature of expectations in psychological games. A real world example of this game is given, where it is translated into a corruption game played between a bureaucrat and a lobby, and with the public in the role of the third party.

---

* The ideas presented in this paper originated from discussions with Louka Katseli and Yiannis Varoufakis of the University of Athens, and from some of their earlier work on the topics of corruption and public administration. I also want to thank Ed Hopkins, Santiago Sanchez-Pages, Stephane Straub and Kohei Kawamura for their very useful comments.

1. Introduction

Conventional game theory, based on strict rationality principles, uses equilibrium concepts -such as (Bayesian) Nash equilibrium, sequential equilibrium etc- which do not allow for any role of emotions, and is therefore often criticised as inadequate to describe human behaviour in many real world situations. This scepticism has led in recent years to extensive research in the field of experimental economics, and has in turn gained ground owing to the findings of this research. One approach to the modelling of strategic choices which would be considered irrational in the standard game theoretic framework is the use of psychological game theory.[29] [30] The basic premise here is that players' payoffs depend not only on their actions, but also on their beliefs. To be more specific, in a typical psychological game higher order beliefs directly enter the players' payoff functions, so that a player's actions are influenced by what (s)he thinks other people think (s)he will do.

This paper uses the framework of psychological game theory and adds to the literature that studies the impact of beliefs and emotions on decision making. This is done in the context of an applied collusion game which features three players: a decision maker (i.e. a player who in effect determines the final payoffs), an active stakeholder who can make a strategic transfer to the decision maker, and a passive stakeholder who does not move in the game but whose payoff depends on its outcome. The setup resembles a trust game between two players which generates an externality for a third player.

The first significant contribution of the paper is that it analyses the effect of the perceived beliefs of the passive stakeholder (to whom I will also refer as the third party) on the

---

[29] For a broad and interesting discussion on rationality, game theory and psychological game theory see Colman (2003).

[30] The term behavioral game theory (Camerer, 1997) overlaps with psychological game theory in the sense that it abstracts from the assumption of fully rational and self-interested agents and invokes nonstandard reasoning processes in order to explain real world behavior and experimental evidence. But behavioral game theory is more of an umbrella term for all theories that deviate from strict rationality, and therefore it incorporates ideas such as fairness discussed later on. Psychological game theory on the other hand uses a given framework and focuses explicitly on the role of beliefs, expectations, and the emotions that these generate.

outcome of the game played between the other two players. In particular, a term is included in the decision maker's utility function to account for the fact that he is unwilling to let the third party down.[31] Hence, the more he believes that the third party expects a favourable outcome, the more likely he is to deliver it. This specification leads to a self-sulfilling mechanism, in which expectations tend to be confirmed through their impact on the decision maker's psychological payoff. To be more precise, the paper studies the effect of 2nd-order beliefs -that is, what the decision maker thinks the third party expects of him- and guilt aversion on the probability that the outcome of the game is collusion. The term "collusion" refers to the outcome where the decision maker receives a transfer from the active stakeholder, acting in her favour in return. Because this is costly for the decision maker, and given the additional feature that the game is sequential with the transfer being made in advance, collusion can only be sustained in a repeated game. I therefore consider a game with a finite but indefinite horizon, in which collusion indeed turns out to be a candidate equilibrium. In this game, the effect of guilt aversion is that it makes collusion less likely to emerge in equilibrium, compared to the standard repeated cooperation game with no psychological dimension.

The paper's second contribution is that it takes a closer look at the dynamic properties of beliefs and at how these affect the outcome of the game. Beliefs are initially constrained to be constant over time; I then relax this assumption and move to a version of the model in which players' beliefs are allowed to vary over time. In such a setting, players must take into account the effect of their actions on others' beliefs in the future. This introduces an additional element of strategic behaviour. I solve for equilibrium strategies and analyse the factors that make collusion easier or harder to establish and sustain in equilibrium. It turns out that collusion is more likely when the psychological component in the decision maker's utility is strong relative to the monetary incentives (i.e. when he

[31] The disutility from letting others down is called guilt aversion in the literature (Charness and Dufwenberg, 2006); what is new here is that guilt is conditioned on the expectations of a third party who does not take any action in the game.

is very guilt averse), when beliefs[32] are initially low and are updated fast, and when the expected horizon of the game is relatively long.

In all of the one-shot psychological games discussed in the review of the literature that follows, beliefs (even higher-order ones) correspond to reality in equilibrium -this is required by the definition of a Psychological Nash Equilibrium (henceforth PNE). In this paper this is no longer the case, because the game is repeated and the model assumes that beliefs follow a particular updating rule based on the history of play. Accordingly, beliefs in this model not only change every period, but they also do not necessarily match actual behaviour even in equilibrium (indeed, they only do so in special cases). This slightly "odd" notion of equilibrium is discussed later on in the paper.

Outsiders' beliefs matter in games where players care about how others view them, a well-known example being the bravery game studied in Geanakoplos et al (1989). The difference here is that the payoff of the outsider is affected by the decision maker's actions.[33] Such a setting can be relevant in certain cases; a potential application that is discussed in some detail is corruption in public administration with a bureaucrat in the role of the decision maker, a lobby in the role of the active stakeholder and the public in that of the third party.

The rest of the paper is organised as follows: Section 2 reviews the literature on psychological games and related concepts such as guilt and reciprocity. Section 3 introduces the game and its one-shot solution, while section 4 solves the repeated version. Section 5 adds dynamic beliefs to the model and discusses its main findings. Section 6 links the game to real world examples. Finally, section 7 takes a look at the implications of allowing for a richer psychological structure, relaxing the assumption that guilt is only directed towards the third party.

---

[32] In this case, as well as throughout most of the paper, by "beliefs" I mean the 2nd order beliefs of the decision maker, i.e. the extent to which he believes that the third party expects him to act in their favour.
[33] A further difference between my collusion game and the bravery game -and other similar games- is that in the bravery game there are only two players, while I describe a situation where two players interact and the outcome of their interaction affects a third party.

2. Review of the relevant literature

In their seminal paper, Geanakoplos et al. (1989) underline the role of beliefs and define the concepts of a psychological game *("players' payoffs depend not only on what everybody does but also on what everybody thinks", p.61)* and Psychological Nash Equilibrium (players choose their optimal strategies and beliefs correspond to reality). In such a framework, players' payoffs incorporate higher-order beliefs, so that expectation-dependent emotions are taken into account.[34] Geanakoplos et al. (1989) initially consider static games, and then generalize to extensive games. However, they only allow initial beliefs to enter the players' utility functions. Battigalli and Dufwenberg (2005) generalize the theory of Geanakoplos et al to include a number of features such as dynamic beliefs,[35] sequential equilibrium analysis, and imperfectly observable outcomes. They also briefly discuss the implications of incomplete information in psychological games.

An explicit example of the effect of expectation-dependent emotions is *guilt aversion*. The term is introduced in Charness and Dufwenberg (2006), the idea being that a player suffers a cost whenever he believes he is letting others down;[36] thus, in order to understand the player's strategic choices, one must know what he thinks others expect of him. Introducing guilt aversion transforms a standard game into a psychological one, by making utility dependent on beliefs about beliefs. Although the paper considers a principal-agent setting with a hidden action, it recognizes that guilt aversion may be relevant in several forms of strategic interaction. It is important to note that the authors only consider a one-shot game, while suggesting that allowing for repeated interaction should have a significant impact on the game if players can update their beliefs based on observation of the history of play.

---

[34] Expectation-dependent emotions include for example guilt, embarassment, disappointment, anger, surprise, gratitude etc.

[35] The authors use what they call a "psychological forward induction" argument and allow players to update their beliefs as events unfold.

[36] The paper defines guilt as the emotion that arises when a player *"believes he hurts others relative to what they believe they will get" [p.1583].*

Huang and Wu (1994) look at a psychological game of trust, where a player's decision on whether or not to honour another player's trust depends on his expectations of the other player's expectations of him. Remorse from betraying someone's trust is proportional to their expectations.[37] The main insight of the paper is that the expectation-dependent moral cost of betrayal leads to a multiplicity of equilibria, some of which feature (at least some degree of) honest behaviour. Interestingly, the authors use corruption in bureaucratic organizations as their central example: bureaucrats may act less corruptly if they believe that the public has high expectations of them. Then, different equilibria with their corresponding sets of beliefs can be viewed as alternative social norms. Another paper which assumes that bureaucrats are averse to letting the public down is Varoufakis (2006); the author studies the evolution of corruption in bureaucracies, with the added twist of linking it to political participation.

Although in the same spirit, Rabin's (1993) theory of *reciprocity* is somewhat different in that it stresses the role of beliefs about intentions, instead of beliefs about beliefs. To be more precise, a player is kind to those whom she considers to be kind and well-intentioned, but she is unkind to those whom she perceives as ill-intentioned. A subtle but significant difference between Rabin's paper and that of Geanakoplos et al (1989) is that Rabin does not start out with beliefs included in the payoffs of players. Instead, he derives a psychological game, starting from what he calls a basic material game and then introducing a "kindness function" which incorporates beliefs. Once beliefs have made their way into the game, the solution concept employed is the Psychological Nash Equilibrium. In another formal treatment of the theory of reciprocity, Falk et al (2008) disentangle the role of intentions from that of outcomes in the shaping of reciprocal behaviour. In the same manner as Rabin, they transform a -finite horizon, complete information- material game into a psychological one through the use of kindness and reciprocity terms. They also apply their framework to a number of well-known games (such as the ultimatum, gift-exchange and public goods game and the prisoner's dilemma).

[37] Hence, remorse here is the same idea as that of guilt aversion in Charness and Dufwenberg (2006).

The idea of *fairness* or *inequity aversion* is also quite similar, since it leads to choices that are not consistent with the mere maximisation of the material benefits enjoyed by players.[38] Inequity aversion means that players care not only about their absolute, but also about their relative payoffs, and that they are to some extent averse to outcomes that are very unequal. The main references in this line of the literature are Fehr and Schmidt (1999) and Bolton and Ockenfels (2000).

Theories based on reciprocity and/or fairness are often quite successful in predicting individual behaviour, as documented by experimental evidence.[39] The experimental literature on psychological games is far more limited. An experimental test of a psychological game is found in Dufwenberg et al (2006) who look at the effect of 2nd-order beliefs on strategic choices, as well as at the effect of framing which is transmitted through changes in those beliefs. Their workhorse is a public good game, and the main result which is of interest here is that guilt aversion is shown to be relevant: people tend to contribute more towards the public good when they think that others expect them to do so. Experimental evidence in favour of a relationship between 2nd-order beliefs beliefs and actions can also be found in Dufwenberg and Gneezy (2000), Bacharach et al. (2007).

3. The basic one-shot game

3.1. Players, structure and beliefs

Consider a game which features three players: Player A is a decision maker, player B is an active stakeholder, and player C is a third party, or passive stakeholder. In fact, this third party is not really a player, in the sense that he does not move during the game; however, his payoff depends on the outcome of the game played by the other two players.

[38] Fairness, however, is analysed in a standard game theory framework, without use of the tools of psychological game theory.

[39] See, among others : Fehr, Kirchsteiger and Riedl (1993), Fehr, Gächter and Kirchsteiger (1997), Abbink, Irlenbusch and Renner (2000, 2002), Falk, Fehr and Fischbacher (2000), Dufwenberg, Gneezy, Güthe and van Damme (2001)

Player A -the decision maker- determines the outcome of the game. To be more precise, let there be two distinct sets of direct (monetary) payoffs called the high and the low set, $\Pi = \{H, L\}$, with $H=(a_H, b_H, c_H)$ and $L=(a_L, b_L, c_L)$, where a, b and c are the payoffs of players A, B and C respectively. Player A must choose a payoff set at the end of the game. Payoffs are such that $a_L<a_H$, $c_L<c_H$, $b_L>b_H$, $a_H-a_L<b_L-b_H$. Hence, H gives a relatively high payoff for players A and C, whereas L is more favourable than H to player B. The last condition ensures that there are mutual gains from collusion (the definition of which is given shortly).

Before player A makes his choice, player B has the option of making a transfer payment to him; let k be the value of this payment, $k \geq 0$. This is the only action that player B can take in the game. When B makes a transfer to A, she is implicitly attempting to influence A's choice in her favour, that is, to convince him to choose the second payoff set.

Player C does not move, but obviously his payoff is directly affected by the final decision taken by player A.

I will use the term *collusion* to describe the situation in which players A and B cooperate in order to increase both their payoffs at the expense of player C. Thus, a collusive outcome involves player B offering a transfer $k \in (a_H - a_L, b_L - b_H)$ and player A choosing the low payoff set. The non collusive outcome occurs when player A chooses the high payoff set. [40]

The timing of events is as follows:

1. Beliefs are formed
2. Player B decides on the amount of the transfer, k, and makes the transfer
3. Player A chooses a payoff set

[40] Player A could also choose L even though no transfer has been offered. This is a strictly dominated strategy and should never be observed; in any case it cannot be described as collusion.

I now introduce beliefs. Let p denote the probability that A will choose H (the third party's preferred outcome), and let p' denote the third party's estimate of p. Finally, q is A's estimate of p', in other words q is the 2nd order belief of player A. In brief, $p=Pr(H)$, $p'=E^C(p)$, $q=E^A(p')$. Although player B cannot directly observe q, she has her own estimate of p', which should in expectation equal q since the two players possess the same amount of information.

The utility functions of players B and C comprise simply their direct payoffs, minus the transfer in the case of player B.[41] For player A utility is shown in Figure 1: If he chooses L, his utility is $(a_L+k-\gamma q)$. This includes the direct payoff and the transfer, but also a psychological cost (guilt) from betraying the third party's expectations. This cost is given by (γq), where the parameter γ is a measure of the intensity of guilt aversion. On the other hand, his final utility if he chooses H is (a_H+k); that is, it includes his payoff and the transfer payment (if any) from player B.

Thus, following the standard framework of psychological game theory, the decision maker's final utility does not only consist of his payoff, but also of a psychological component which depends on beliefs. It is beliefs *per se* that affect a player's utility in this context, and not only to the extent that they provide information about players' strategies. To be more precise, A cares about what he thinks others expected of him; so we are dealing with *beliefs about beliefs*.

It has already been stressed that the model does not condition guilt on the beliefs of any of the players that play the game, but on the beliefs of a third party who has some stake in the game. To make this point clearer, let us consider the situation outlined above: although A plays the game with B, it is not B's beliefs about A that come into the latter's utility, rather it is the beliefs of the third party C –in fact, it is A's estimate of C's beliefs. If player A chooses the payoff set L, he induces a reduction in the utility of player C.

[41] Strictly speaking, we do not need to specify player C's utility function because he does not take any action in the game.

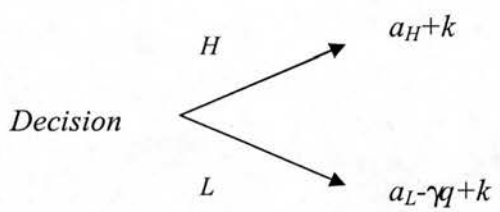
Knowing this, A experiences a psychological cost, which is proportional to his valuation of C's expectations. In other words, player A is guilt averse.

In order to be able to compare A's utilities from his two alternative strategies and say something about his choice, we must first delineate player B's strategy, in other words we must pin down the level of the transfer, k. I do so, first in a one-shot version, and then for a repeated game.

3.2. Solution

When the game is only played once, player A's total utilities for any given transfer k are shown on the following graph.

Figure 1: Player A's utility



Given the timing of events, simple backwards induction reveals that no transfer is made in equilibrium, since player B correctly anticipates that player A will choose H, which is his dominant strategy irrespective of the amount of the transfer. This is a standard commitment problem: since the transfer is made in advance, player A has no incentive to choose B's preferred payoff set once he has received the transfer. Thus, collusion cannot be sustained in equilibrium.

In this example, the presence of beliefs in A's utility does not affect the equilibrium. Indeed, the psychological factor -guilt aversion- pushes player A even stronger towards

choosing H, since L comes at a psychological cost. However, beliefs can make a difference when the game is indefinitely (even if finitely) repeated.

4. The repeated game

I now allow the game to be played repeatedly between the same players A and B, and with the same third party C holding beliefs about A's decision. Utilities are defined as in the one-shot game, and in each repetition the timing is as outlined in the previous section. Suppose that every time the game is played, the probability that it will be repeated once more equals r,[42] where r is constant across periods. In this dynamic context, players must work out their inter-temporal strategies. These can be the same as in the one-shot case (leading to a zero transfer and a choice of H), or they can differ. Players may be able to establish the collusion equilibrium $[k \in (a_H - a_L, b_L - b_H),\ \Pi = L]$ by adopting some conditional collusive strategy.

In particular, consider the possibility that player B adopts the following trigger strategy (called τ): She makes a transfer in the first period, and then keeps making a transfer as long as player A has colluded in the previous period. But if in any period A does not collude, then B switches permanently to the zero transfer strategy. Although there are infinite inter-temporal strategies that B can adopt,[43] I shall confine myself to strategy τ as a focal one. In particular, I will assume that B adopts this trigger strategy and check the conditions under which a collusion equilibrium is sustainable.

Let beliefs (q) be constant throughout the horizon of the game; this assumption is relaxed in the next section. Given that B plays τ each period, and given that none of the parameters change over time, he will make the same transfer k every period if there is collusion. Moreover, player A will either collude in every period or he will not collude in

[42] Hence, in what follows, r serves as an implicit discount factor.
[43] Indeed, the folk theorem is applicable in this type of game and implies that an infinite number of strategies can be supported in equilibrium.

any period. Call the respective strategies S^C and S^{NC}. The associated total expected payoffs are the following:

$$EU(S^{C)}) = (a_L + k - \gamma q)/(1-r)$$
$$EU(S^{NC}) = k + a_H /(1-r)$$

If player A honours player B's trust and colludes, then each period his utility will equal the low direct payoff a_L plus the transfer minus the psychological cost. If he chooses not to collude, then his utility will be high in the first period, since it will comprise both the high direct payoff and the transfer. But in all subsequent periods he will receive only the high direct payoff α_H.

Comparing the two payoffs, it follows that player A will collude as long as
$r > [(a_H - a_L) + \gamma q]/k$ *(1)*

This expression conditions the sustainability of the collusion relationship on the intensity of the remorse (the parameter γ multiplied by q), the direct monetary loss that player A suffers if he chooses to collude $(a_H\text{-}a_L)$, the expected duration of the relationship, and the level of the transfer. For player A to collude, the transfer and the expected horizon of the game must be sufficiently high with respect to the difference in monetary payoffs and to the psychological cost of disappointing the third party.

<u>*Lemma*</u>

If collusion is the equilibrium of the game, player B makes the following transfer k^* every period: $k^* = (a_H - a_L + \gamma q)/r$

This follows from *(1)*. A transfer of k^* compensates player A for the loss in utility caused by the lower direct payoff and the guilt that he suffers if he chooses L. The discount factor r enters the expression and raises the level of the transfer. This accounts for the fact that the transfer is made in advance and so, in any given period T, the benefit from

collusion for player A lies in the future (the future transfer payments), whereas the cost is borne starting at time *T*.

Proposition 1

The above game features two possible equilibria:

(i) The collusion equilibrium, with $\Pi=L$ *in every period,* $p=q=0$*, and* $k^*=(a_H-a_L)/r$

(ii) The no collusion equilibrium, with $\Pi=H$ *in every period,* $p=q=1$*, and* $k^*=0$

If $(a_H - a_L + \gamma)/r \leq b_L - b_H$*, the equilibrium is necessarily (i)*

If $(a_H - a_L)/r \geq b_L - b_H$*, the equilibrium is necessarily (ii)*

Otherwise, both equilibria are possible.

Proof: $(a_H - a_L + \gamma)/r \leq b_L - b_H$ means that B can make a transfer high enough to sustain collusion even if A were to believe that the the third party was certain of a high payoff *(q=1)*; knowing this, the third party is convinced that A and B will collude, *q* is actually zero and, absent the psychological factor, A confirms the expectations and chooses the low set. The collusion equilibrium prevails. On the other hand, $(a_H - a_L)/r \geq b_L - b_H$ means that B cannot make a transfer high enough to sustain collusion, even if beliefs are zero. Therefore everyone expects the no collusion equilibrium *(p=q=1)*, and these expectations are confirmed.

The above equilibria satisfy the definition of a Psychological Nash Equilibrium (PNE), namely players' strategies are optimal *and* beliefs correspond to the actual play. There is a multiplicity (at least within certain parameter values), meaning that we could end up in any of the two equilibria.[44]

In this repeated game, the psychological element in the behaviour of the decision maker has the effect of making it less likely that collusion will emerge in equilibrium. To see

[44] Huang and Wu (1994, p.395) view the multiple equilibria that can emerge in their psychological corruption game as "alternative social norms", which "are supported by the equilibrium beliefs and psychological emotions".

this, note that, absent the psychological parameter (i.e. when $\gamma=0$), the necessary and sufficient condition for collusion is $(a_H - a_L)/r \leq b_L - b_H$ *(2)*

This is the solution to the standard, non-psychological version of the game. Comparing *(2)* with the conditions given in Proposition 1 reveals that, for certain parameter values, collusion is always the outcome when $\gamma=0$ but not necessarily when $\gamma>0$.

5. Time-varying beliefs

5.1. General

So far it has been assumed that beliefs are stationary, and that they correspond to the equilibrium outcome which is reached in the first period and which does not change over the course the game. This assumption is now relaxed, and beliefs are allowed to vary over time. The impact of this modification on the results will be clear in the analysis which follows, but a number of things must be noted before we look at the specifics of the model.

First, allowing beliefs to be updated means that we are no longer dealing with a psychological game in the strict sense.[45] Perhaps more importantly, the notion of a PNE is no longer applicable as an equilibrium concept, because the model posits a specific behavioural assumption regarding the evolution of beliefs (in the form of an updating process), and this assumption implies that beliefs will not necessarily be correct in equilibrium. Although the literature mentions the importance of studying dynamic psychological games and allowing for updating of beliefs, no generally accepted framework or concept of equilibrium has been developed to incorporate these features into standard psychological games. In light of this fact, the model will consider a state in which players A and B are in equilibrium in their collusion game, but the third party C holds out-of-equilibrium beliefs about the outcome of the game. This point is discussed further towards the end the section.

[45] See Battigalli and Dufwenberg (2005, p.7)

One way to motivate the above ideas would be to consider the possibility of asymmetric information in the model. It was assumed that all the parameters (payoffs and psychological factors) are common knowledge; this assumption appears unrealistic in many settings, particularly in the context of psychological games.[46] If player C knows all the parameters of the model as well as the strategies that A and B follow, he must be able to deduce the solution, beliefs will jump to a corner and we will have a solution similar to the one in the previous section. If, instead, information is incomplete, it seems appropriate to not constrain beliefs to be correct in equilibrium, but to allow them to take interior values. For this, it is sufficient to think that C may not know some of the parameters of the model (for instance the psychological parameter γ); or that he simply does not know that B follows the trigger strategy τ, which would mean that from C's point of view anything could be an equilibrium of the repeated game.

5.2. The model with updated beliefs

Beliefs are initially at some exogenous level q_0 and they can change over time $(q=q_t)$; player C updates his beliefs about player A's choice every period, taking into account A's choice in the preceding period, and players A and B update their own beliefs in exactly the same way since they have rational expectations about the updating process. Before imposing a specific process, I start with a general case where the only assumption is that expectations fall every time player A chooses L, and that they rise every time he chooses H. This means:

$$\begin{cases} q_{t+1} > q_t, & if\ \Pi_t = H \\ q_{t+1} < q_t, & if\ \Pi_t = L \end{cases} \qquad (3)$$

Let us examine the strategy that player B will follow in order to establish collusion. Proceeding as in section 4, we can determine the amount of the transfer which in every

[46] "Unless one models interaction within a family or amongst friends, it is probably not realistic to assume that players know each other's psychological propensities" (Battigalli and Dufwenberg 2005, p.31). Also, Rabin (1993) motivates the study of psychological games with incomplete information, where a player's knowledge of other players' motives may be limited.

period must make player A indifferent between always colluding and never colluding (from that period onwards). This is given by:

$k^*_t = (a_H - a_L + \gamma q_0)/r$ *(4)*

Proposition 2

The unique optimal sequence of transfer payments made by player B in the collusion equilibrium is k^*_t as given in (4).

Proof: In any period T throughout the game, no transfer $k_T > k^*_T$ will be offered; since k^*_T is high enough to make player A prefer the collusive outcome subject to k^*_t being offered in all subsequent periods, a higher transfer unnecessarily reduces B's payoff. On the other hand, no $k_T < k^*_T$ is offered in period T either. Such a transfer is lower than what is needed to compensate player A, so the latter will not collude in period T with $k_T < k^*_T$, unless he expects to be compensated for the loss in T in some future period(s). But such an expectation is not rational and any such commitment by player B would not be credible, since we just saw that k_t will never exceed k^*_t. Thus, k_T must equal exactly k^*_T.

Corollary

The sequence of optimal transfer payments in the collusion equilibrium is strictly decreasing over time: $k^*_t > k^*_{t+1}$

This follows immediately from *(3)* and *(4)*.

We must now examine the conditions under which the collusion equilibrium will actually prevail. In any period, making a transfer in exchange for the choice of L is profitable for player B if $k_t < b_L - b_H$, i.e. if her cost (transfer) is lower than her benefit. Moreover, from Corollary 1, if this action is profitable in the first period, it will be profitable in all subsequent periods; the intuition is that expectations fall over time and the required compensation diminishes as guilt becomes weaker. This leads to the fact that an equilibrium which features collusion every period is definitely profitable for player B if

$(a_H - a_L + \gamma q_0)/r \leq b_L - b_H$ *(5)*

Hence, if *(5)* holds, the outcome is always collusion. This condition is interesting because it ties the emergence of collusion on initial beliefs.[47]

That is not the whole story, however. Condition *(5)* is sufficient for collusion to be profitable for player B, but it is not necessary. It may be the case that player B wants to sustain collusion even if it is unprofitable for her in the first period. The reason is that, as beliefs fall, k^*_t will fall over time so that the total expected return from always making the transfer may be positive. In other words, player B might make a "sacrifice" today in the expectation that she will make up for it tomorrow, since the effect of this sacrifice will be to drive expectations down. In order to be able to say something about the outcome of the game in that case, we must assume a specific dynamic process for beliefs and assess the benefits and costs of setting $k_t=k^*_t$ as in *(4)* throughout the horizon of the game.

Let beliefs be updated according to the following process:

$$\begin{cases} q_{t+1} = q_t + \rho(1-q_t), & if\ \Pi_t = H \\ q_{t+1} = q_t - \rho q_t, & if\ \Pi_t = L \end{cases} \qquad (6)$$

This process specifies that in any given period beliefs will move either upwards, if there was no collusion in the previous period, or downwards if there was collusion in the previous period.[48] The parameter $\rho \in [0,1]$ is a measure of the speed of beliefs adjustment; the greater this is, the faster q_t will rise or fall.[49] The process is convex, so that q_t cannot reach 1 or 0 unless $\rho = 0$, in which case beliefs jump to a corner after the first period. In the opposite case where the parameter ρ is zero, beliefs do not move over time and will remain at their initial value of q_0. This process captures the main features of the adjustment of beliefs that I have in mind, although one could think of alternative specifications.

---

[47] This is discussed further in section 5.3.

[48] The first part of the process which refers to the case where there is no collusion does not enter into any of the calculations and does not affect the results, because in the no-collusion equilibrium the level of beliefs does not matter and the payoff is constant at a_H.

[49] In fact, the speeds of upward and downward adjustment should not be constrained to be equal; this would account, for example, for the possibility that confidence might take a longer time to build than to destroy. The reason why here we only have one parameter ρ for the speed of adjustment is that the upward path of beliefs is irrelevant to the outcome of the game.

Solving the difference equations in *(6)* gives the following expressions for beliefs at any point in time as a function of initial beliefs:

$$\begin{cases} q_t = q_0(1-\rho)^t + 1-(1-\rho)^t, & \textit{in the no collusion equilibrium} \\ q_t = q_0(1-\rho)^t, & \textit{in the collusion equilibrium} \end{cases} \qquad (7)$$

From *(4)* and *(7)*, we obtain the level of the transfer in any period t as:

$$k_t^* = [a_H - a_L + \gamma q_0(1-\rho)^t]/r \qquad (8)$$

<u>*Proposition 3*</u>

When beliefs follow the dynamic process given in (6), the specified psychological game features two possible equilibria:

(i) If $\frac{(a_H - a_L)}{r} + \frac{\gamma q_0(1-r)}{r[1-r(1-\rho)]} \leq b_L - b_H$, *(9)*

the outcome is collusion in every period with $k_t = k^*_t$ *as given in (8).*

(ii) If (9) does not hold, collusion never occurs.

Proof: The collusion equilibrium prevails as long as the expected total value of the amount transferred throughout the game does not exceed the expected value of the total benefit of collusion to player B: $\sum_{t=0}^{\infty} k_t^* r^t \leq \sum_{t=0}^{\infty} (b_L - b_H) r^t$, which translates to *(9)*. In such a case, player B will always transfer k^*_t, because to her this is the least costly way of "buying" collusion. Hence, collusion is sustained indefinitely. On the contrary, if *(9)* does not hold, player B never makes a transfer and player A always chooses *H*.

5.3. Comparative statics and discussion

As it has already been mentioned, Proposition 3 defines equilibria of the repeated collusion game between players A and B. However, these are not equilibria of the extended game which includes the third party C, in the sense that C's beliefs are not necessarily accurate as required by the definition of a PNE. To be more precise, a PNE in

this case would require q_t to always be 0 in the collusion equilibrium and 1 in the no collusion equilibrium. Instead, the behavioural assumption here is that beliefs start at some arbitrary level and then move according to the observed history of play. This leads to condition *(9)* in Proposition 3, but also to the time-varying transfer given by *(4)*.

It must also be noted that, as in section 4, the inclusion of beliefs in player A's utility function has the effect of making collusion less likely than in the standard, non-psychological equivalent of the game. This is captured by the second term in the left-hand side of *(9)*, which is the only difference between *(9)* and *(2)*, the condition for the standard game.

Condition *(9)* allows us to examine the effect of the various parameters of the model on the emergence of collusion in equilibrium:

(i) Collusion is more likely the higher is the potential gain for player B (i.e. the higher is b_L-b_H), because this inceases her gains from collusion along with the upper limit on the transfer she is prepared to make. On the other hand, collusion is less likely the higher is the direct (monetary) cost of collusion to player A, given by a_H-a_L.

(ii) The likelihood of collusion is decreasing in the psychological parameter γ, which captures the extent of guilt aversion. This is hardly surprising, given that the presence of guilt aversion has already been shown to make collusion harder to establish.

(iii) The likelihood of collusion is decreasing in q_0: Higher initial beliefs imply a greater cost of guilt -for a given speed of belief adjustment.

This point underlines the importance of initial beliefs. In a situation where the third party holds low initial expectations of the decision maker, collusion becomes more likely because the remorse that comes with it is not too strong. On the other hand, high initial expectations of the third party may prevent collusion, even though expectations would fall if collusion did occur.

(iv) Collusion is more likely the longer the horizon of the game, i.e. the larger is r. As the horizon becomes longer, the effect of guilt diminishes relative to the other factors that determine the outcome of the game,[50] making collusion less costly for the decision maker.

(v) Finally, it is interesting to note that the probability of collusion is increasing in the speed at which beliefs are updated, measured by the parameter ρ.

In other words, if the third party is quick to become "disillusioned" by the decision maker, then the latter is more likely to collude in the first place; on the contrary, a decision maker is less likely to disappoint someone whose beliefs are more resilient and who will, in a way, be more willing to "overlook" collusion. The intuition behind this result is that, if expectations decline fast following a collusive outcome, then the psychological cost of guilt also diminishes fast, and the transfer that is required to establish collusion is lower.

The last two points relate to one of the most interesting features of this model: introducing a dynamic setting and allowing beliefs to vary over time also leads to a *time-varying cost of guilt*. If the third party's expectations fall during the game, so does the psychological cost to the decision maker of betraying these expectations, and vice versa. The fall in the cost of guilt is more sustained, the longer the horizon of the game.

To sum up the main idea of the model, high perceived expectations make collusion harder to sustain because they imply a high guilt-related disutility for the decision maker; on the contrary, low expectations (i.e. the belief that collusion is very likely) can actually *lead* to collusion. This highlights the self-fulfilling property of third party beliefs. The effect of perceived third party beliefs on decision making is investigated in the economic experiment that forms Chapter 3 of this thesis.

[50] This is the case because q_t falls over time, whereas the other parameters remain constant.

6. Examples

This paper has analysed one particular game with a specific structure and three players: a decision maker, an active stakeholder, and a passive stakeholder. In principle this structure resembles a trust game between two players, which generates an externality for a third party. Crucially, it is precisely the (perceived) expectations of this third party that enter the utility function of the decision maker and grant the game its psychological dimension.

One can think of real world situations where the model is applicable. Among those, the most interesting from an economic point of view is probably corruption in public administration. Let us think of the decision maker as a bureaucrat, of the active stakeholder as (some representative of) a lobby, and of the passive stakeholder as the public.[51] The lobby can make a transfer to the bureaucrat with the purpose of eliciting some favour from him; if the transfer is high enough, the lobby can manage to sustain collusion to the detriment of the public's utility. In this context it is appropriate to talk in terms of a "bribe" instead of a transfer, and of "corruption" instead of collusion. Consistent with the widely used definition of corruption as the "misuse of public office for private gain" (Rose-Ackerman, 1999), the bureaucrat can choose an action that hurts the public but ensures a higher payoff for himself (the bribe minus the costs of corruption). There is a dual cost of corruption: the bureaucrat may have to exert some effort in order to grant the favour or in order to preserve secrecy and avoid getting caught -this cost corresponds to the difference (a_H-a_L) in the model. Alternatively this difference can be thought of as the expected cost from being caught (probability of detection multiplied by the penalty). In addition to this direct cost, the bureaucrat suffers a moral cost of corruption. This is no other than the term $(-\gamma q)$ in the model: it is the guilt from

[51] I intentionally refer to a bureaucrat and not a politician. The latter is elected and therefore the public is far from a *passive* stakeholder. Bureaucrats, on the other hand, are appointed and their positions are often permanent (although it is true that their careers can depend on political affiliations).

betraying the public's expectations.[52] Beliefs can then be interpreted as the public's perception of the extent to which corruption is widespread in public administration at a given point in time, or in a more general way as a reflection of social norms. Starting from an initial set of beliefs and observing the incidence -or not- of corruption, the public update their expectations every period. The model predicts that, as long as the exogenous parameters do not change, corruption will occur either in all periods or in none; it therefore also predicts that social norms will keep moving either towards very high levels of trust in public officials, or towards complete disillusionment where no one expects them to act honestly. If that is the case, lobbies will eventually be able to capture officials with only small amounts of bribes. The intuition for this is that an official who knows that the public considers him corrupt suffers a low moral cost when he actually behaves in a corrupt manner. In other words, the official is caught in a circle of self-fulfilling expectations.

It must be noted that the model in effect assumes the impact of public expectations on the incidence of corruption, by including them in the bureaucrat's utility function. The main interest then lies in the importance that is accorded to initial beliefs, and to the fact that corruption lowers expectations and thereby paves the way for more corruption. Hence, in a dynamic setting, corruption leads to more corruption.[53] In fact, the model predicts that lobbies are likely to act strategically and offer very high bribes at initial stages of the bribery relationship, with the purpose of marring the image of public officials and making collusion easier and cheaper in the future. This strategy is more profitable, the higher is the speed of dynamic adjustment of beliefs (the parameter ρ in the model). That is to say, corruption is more likely to find fertile ground in societies where beliefs -or social norms- are not particularly resilient.

---

[52] The basic model in Huang and Wu (1994) mentioned in the introductory section looks at a similar situation. A bureaucrat can choose to be corrupt, in which case he increases his material payoff but lowers that of voters. The main difference here is that I endogenise the benefit from corruption by introducing a briber. Moreover, the bribe varies over time, and so does the psychological cost of corruption.

[53] In a formal prinicipal-agent model, Tirole (1996) shows how an economy can become "locked" in a high corruption steady state when the principal's source of information about the agent is the agent's group reputation.

More generally, the model may be applicable to principal-agent situations where the principal has only loose control -or none- over the agent, and can therefore be seen as a more or less passive stakeholder. The agent's propensity to collude with someone else against the principal's interests should then depend on his beliefs about the principal's expectations of him. The relationship between bureaucrats and the public is cast precisely in this framework, with virtually no control of the principal (public) over the agent (bureaucrat).

One can think of further relevant examples for this game outside the field of economics. A spouse who is lured into an extra-marital affair may feel more guilt if they believe that their husband or wife (i.e. the passive stakeholder) trusts them a lot; in such a case, the third person may have to make a very large transfer (gift) in order to establish "collusion"! In a football game, or sports in general, the decision maker is often the referee, the briber may be one of the two teams and the passive stakeholder who is affected by the outcome is the other team (who cannot afford or does not want to pay a bribe) or the spectators who would like to see a fair game. We can arguably encounter the same circle of self-fulfilling expectations as in the previous example, with referees more (less) prone to bribe taking in countries or leagues where they are widely regarded as dishonest (honest).

7. Guilt aversion towards both players

A question that may arise concerning the specification of the game is the following: Why does the decision maker feel guilt *only* towards the third party? In other words, why does he play a psychological game with one of the players and a standard (neoclassical) game with the other? To motivate this assumption, one may evoke social and psychological arguments relating to the particular examples we have given for this game. In the case of corruption, it is natural to assume that a bureaucrat has some sense of mission and responsibility towards the electorate, but not towards a business lobby. Similar arguments can be applied to the other examples.

This section modifies the model of section 5 by adding guilt aversion towards the active stakeholder. In particular, I assume that if a positive transfer is made and the decision maker selects H, he suffers guilt. In this context, guilt does not arise from the choice of H per se, but from the fact that the decision maker does not reciprocate to player B's transfer.[54] The only difference that this additional psychological element makes in the game comes from the fact that, if in period t player B makes a transfer and player A selects H, the latter's payoff becomes $U^H_t=a_H+k-\xi\mu_t$, where ξ measures the strength of guilt aversion towards player B and μ_t is player A's 2nd-order belief, defined in the same way as q: it is A's valuation of B's valuation of the probability that A will select L. Note, however, that this kind of guilt can only occur once, in the period when A betrays B. This is because in any subsequent period B will switch to a zero transfer strategy. Thus, contrary to the previous case, the problem facing the decision maker is not the same in every period: If he betrays player B he will have a disutility in that particular period but will afterwards return to the high direct payoff a_H. For simplicity, let $\mu_t=\mu$: This is B's belief that A will reciprocate to a transfer. Since there is no asymmetric information between A and B, and because beliefs do not matter anymore after A does not reciprocate once, there is no need for an updating rule similar to *(3)* or *(6)*. The optimal sequence of transfers in the collusion equilibrium is determined as:

$$k^*_t = [a_H - a_L + \gamma q_t - \xi\mu(1-r)]/r \qquad (10)$$

The difference between k^*_t in *(4)* and in *(10)* is the additional term $-\xi\mu(1-r)$ which reduces the level of the transfer that is required to sustain collusion. Compared to the game of the previous sections, guilt towards player B makes it relatively cheaper for the latter to influence the decision maker, the more so the shorter is the horizon of the game. In fact, with the effect of guilt now going in two different directions, it is the relative disutilities of letting down the two players that matter for the outcome of the game. The term $(1-r)$ reflects the fact that guilt towards B can only occur once, so that only part of it enters A's

[54] This is not a crucial assumption; if player A suffers guilt whenever he chooses H irrespective of whether a transfer has been made, the results are almost identical to the ones that will be obtained here.

utility seen from any point t in time. Then, given *(6)*, the condition for collusion to be the equilibrium outcome becomes:

$$\frac{(a_H - a_L)}{r} + \frac{\gamma q_0 (1-r)}{r[1-(1-\rho)r]} - \frac{\xi\mu(1-r)}{r} \leq b_L - b_H \qquad (11)$$

In addition to the findings of section 5, we see here that collusion is more likely when player B's perceived beliefs *(μ)* are high and when A is very guit averse towards B (high ξ). The effect of the time horizon on the incidence of collusion is not clear cut any more, because it works in opposite directions through the second and third term on the left-hand side of *(11)*.

References

Abbink K., Irlenbusch B., Renner E. (2000): "The Moonlighting Game - An Experimental Study on Reciprocity and Retribution". *Journal of Economic Behavior and Organization* 42, 265-277

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

Battigalli P., Dufwenberg M. (2005): "Dynamic Psychological Games". Working Paper, IGIER, Bocconi University

Bacharach M., Guerra G., Zizzo D. (2007): "The Self-Fulfilling Property of Trust: An Experimental Study". *Theory and Decision* 63, 349-388

Bolton G., Ockenfels A. (2000): "ERC: A Theory of Equity, Reciprocity, and Competition". *American Economic Review* 90, 166-193

Camerer C. (1997). "Progress in Behavioral Game Theory". *Journal of Economic Perspectives* 11, 167-188

Charness G., Dufwenberg M. (2006): "Promises and Partnership". *Econometrica* 74, 1579-1601

Colman A. (2003): "Cooperation, Psychological Game Theory, and Limitations of Rationality in Social Interaction". *Behavioral and Brain Sciences* 26, 139-198

Dufwenberg M., Gächter S., Hennig-Schmidt H. (2006): "The framing of Games and the Psychology of Strategic Choice". CeDEx Discussion Paper no. 2006-20

Dufwenberg M., Gneezy U. (2000): "Measuring Beliefs in an Experimental Lost Wallet Game". *Games and Economic Behavior* 30, 163-182

Dufwenberg M., Gneezy U., Güth W., van Damme E. (2001): "An Experimental Test of Direct and Indirect Reciprocity in case of Complete and Incomplete Information". *Homo Oeconomicus* XVIII 19-30, München

Falk A., Fehr E., Fischbacher U. (2008): "Testing Theories of Fairness - Intentions Matter." *Games and Economic Behavior* 62, 287-303

Fehr E., Gächter S., Kirchsteiger G. (1997): "Reciprocity as a Contract Enforcement Device: Experimental Evidence". *Econometrica* 65, 833-860

Fehr E., Kirchsteiger G., Riedl A. (1993): "Does Fairness Prevent Market Clearing? An Experimental Investigation". *Quarterly Journal of Economics* 108, 437-459

Fehr E., Schmidt K. (1999): "A Theory of Fairness, Competition, and Cooperation". *Quarterly Journal of Economics* 114, 817-868

Geanakoplos J., Pearce D., Stacchetti E. (1989): "Psychological Games and Sequential Rationality". *Games and Economic Behavior* 1, 60-79

Huang P., Wu H.M. (1994): "More Order without More Law: A Theory of Social Norms and Organizational Cultures". *Journal of Law, Economics and Organization* 10, 390-406

Rabin M. (1993): "Incorporating Fairness into Game Theory and Economics". *American Economic Review* 83, 1281-1302

Rose-Ackerman S. (1999): "Corruption and Government: Causes, Consequences, and Reform". Cambridge University Press

Tirole J. (1996). "A Theory of Collective Reputations (with applications to the persistence of corruption and to firm quality)". *Review of Economic Studies* 63, 1-22

Varoufakis Y. (2006): "The Bonds that Impede: A Model of the Joint Evolution of Corruption and Apathy". *Indian Economic Journal* 54, 84-103

Chapter 3

Third party beliefs and collusion: An experimental study

*Third party beliefs and collusion: An experimental study**

Abstract

This experiment tests the impact of beliefs on decision making in the context of a three player collusion game. One of the players is an outsider, or "third party", in the sense that he cannot take any action in the game even though he is affected by the outcome. If the outcome is collusion between the two players, the third party stands to lose. The paper's main novelty is that it investigates the impact of the third party's beliefs on the behaviour of the two other players and on the likelihood that they increase their monetary payoffs through collusion. The statistical and econometric analysis of the experimental data reveals that 2nd order beliefs (that is, the third party's beliefs as perceived by the players) are the strongest and most significant factor which determines behaviour and the outcome of the game. This result is in line with the framework and predictions of psychological game theory. Moreover, the paper gives some insights on the relevant weight of different behavioural assumptions (fairness, reciprocity, guilt) in the cognitive process of decision making.

* I am grateful to Nick Feltovich from the University of Aberdeen, whose contribution was pivotal in every phase of this experiment. My thanks also to Yiannis Varoufakis, Ed Hopkins, Santiago Sanchez-Pages, Stephane Straub and Maia Guell for their comments. This experiment was funded by a Small Project Grant from the University of Edinburgh Development Trust.

1. Introduction

Laboratory experiments are widely used to test behavioural theories in economics and the social sciences in general. The reason is obvious: experiments allow us to observe and record the way that human subjects interact in a controlled environment, and to compare the observed behaviour to the one predicted by economic models.

This paper uses an experiment in order to give an answer to the following question: How is the behaviour of players in a game affected by their perceptions about the beliefs of an outsider? Psychological game theory suggests that players are often motivated by emotions which depend on other players' beliefs *(expectation dependent emotions)* such as guilt, embarrassment, disappointment, surprise etc. This paper goes one step further and examines whether a player can be motivated by emotions which depend on the perceived expectations of a third party, who does not directly participate in the game but who is affected by its outcome. In particular, I consider guilt and some idea of social responsibility as the psychological factors of motivation. I believe that this research topic has relevance for real world economic situations, notably in the field of public administration. This point is discussed towards the end of the paper.

I studied the behaviour of subjects divided in groups, with three subjects in each group. In a nutshell, two of the subjects were playing a game of intertemporal cooperation, in which the cooperative outcome lowered the payoff of the third player. The main points of interest that emerge from the statistical and econometric analysis of the data are the following: *(i)* cooperation is a rather rare outcome, despite the fact that in expectation it raises the payoffs of the two players, *(ii)* cooperation is strongly correlated with the perceived expectations of the third party: the more the third party sees cooperation as an unlikely event, the lower is the actual likelihood of it occurring. This result corroborates the basic hypothesis that the experiment is meant to test, and in more general terms adds to the stock of experimental evidence which supports the relevance of beliefs and the assumptions underlying psychological games.

On a more methodological note, the paper discusses the difficulty in determining the relative weight that different behavioural theories carry for the explanation of actual behaviour and attempts to shed some light on this topic. In light of the evidence, players' beliefs and the emotions (guilt) that these generate appear to have a stronger and more robust effect on behaviour than considerations of fairness.

The following section presents the theoretical model which formed the basis for this experiment, along with an overview of the relevant literature. Section 3 outlines the experimental design and methodology, and section 4 presents the statistical analysis of the data, followed by a number of regressions and an interpretation of the results. Section 5 summarises the main findings of the experiment and briefly discusses potential economic implications.

2. Theoretical framework

2.1. Related literature and the idea of social responsibility

The paper relates closely to a number of experiments which measure players' beliefs and study their impact on behaviour. This methodology, which implicitly assumes that beliefs affect behaviour in their own right and should therefore be included in utility functions, conforms to the basic structure of psychological games as developed originally in Geanakoplos et al (1989). Among the relevant experimental work, Charness and Dufwenberg (2006) and Dufwenberg et al (2006) use the term *guilt aversion*. The first of these papers introduces the term, tests experimentally and confirms the role of guilt aversion in a one shot trust game between two players. The twist here is that players are allowed to exchange messages, and the authors conjecture that communication affects behaviour through its impact on beliefs. The second paper shows that players tend to make higher contributions to a public good when they believe that others expect them to do so. In a similar spirit, Dufwenberg and Gneezy (2000) measure players' beliefs in an experimental "lost wallet" game, where a player can keep a payoff entirely for himself or choose to let a second player allocate a higher payoff between the two. The authors find

that the share given from player 2 to player 1 is positively correlated with what 2 thinks 1 expects to get, and they too advance the hypothesis that players are averse to letting others down. Bacharach et al (2007) study a two person trust game, where one player can "trust" or "withhold", and the other can "fulfil" or "violate". The basic result is qualitatively the same: a player is more likely to fulfil when she believes that the other player strongly expects her to fulfil. The authors call this phenomenon *trust responsiveness*.

In principle, the ideas of guilt aversion and of trust responsiveness can be seen as two sides of the same coin. In all four experimental papers mentioned above, a player is faced with a decision which can be either continuous or discrete, but always has some sense of social worth: honouring someone's trust, contributing a high amount to a public good, rewarding someone for not capturing the entire surplus, all these actions would by most people's standards be considered "good" or "honest" from a social and moral point of view. The main observation of interest is that people are more likely to opt for the "good" action when they believe that others expect them to do so. There seems to be some sense of social responsibility in individuals, but it is interesting to see that this is not constant; on the contrary, it depends on perceived expectations and the desire to live up to those. For this reason, social responsibility may have more to do with *emotions* such as guilt or shame, and less with considerations about *equity* or *fairness*.[55] If individuals only rely on a personal and absolute criterion of fairness when they make decisions, then their actions should be affected by (absolute and relative) monetary payoffs, and not by other's expectations. Of course, it is possible that the two motives coexist. This experiment tries to shed some light on the relevant weight of these motives, and also discusses the role of *reciprocity* between players.[56]

What distinguishes this work from the rest of the literature is mainly the fact that it extends this notion of social responsibility to a stakeholder (third party) who does not directly participate in the game. In a way, the experimental design pits the narrow self

[55] See Fehr and Schmidt (1999), Bolton and Ockenfels (2000) for formal theories of fairness
[56] See Rabin (1993) on reciprocity

interest of a decision maker against two other players, an active and a passive stakeholder. The main research question concerns the role of guilt -or social responsibility in general- towards the passive stakeholder. Two further interesting aspects of the experimental design are the asymmetry of information across players and the repeated interaction among them; especially the first of these features is uncommon in the relevant experimental literature.

2.2. Third party beliefs and collusion: A model

The experiment is based on the theoretical model developed in Chapter 2 of this thesis, which is briefly summarised again here. The following game is played between three players: one of them (Player A) must choose between two payoff sets for all players. The two sets are called high and low, $H= (a_H, b_H, c_H)$ and $L= (a_L, b_L, c_L)$ respectively, such that $a_H>a_L$, $c_H>c_L$, $b_H<b_L$, and $a_H-a_L<b_L-b_H$. Player B has the option to make a transfer of k to Player A before the latter chooses a set, while Player C simply observes the game and cannot take any action at any time. The game is a repeated one, with a probability of breakdown equal to $(1-r)$ each period, where r is constant across periods. Hence, this game can be interpreted as a problem of intertemporal cooperation between two players (A and B) who can collude and increase both their expected total payoffs while lowering that of a third player. Collusion in this context means that A chooses the low payoff set L in response to a transfer $k \in (a_H - a_L, b_L - b_H)$.

The model follows the framework of psychological game theory and includes beliefs directly into A's utility function. The key variable is A's 2nd order beliefs, which are denoted by q_t and measure the extent to which he believes C expects him to choose the payoff set H in a given period t. Player A's period utility is a_H+k_t if he chooses H, and $a_L+k_t-\gamma q_t$ if he chooses L. The term γq_t captures the idea of guilt aversion: every time A chooses the set that gives a low payoff to C, he suffers an emotional cost which is proportional to his 2nd order belief q_t. The parameter $\gamma>0$ measures the intensity of guilt aversion. Notice that, contrary to most of the literature on psychological games, beliefs

are not constrained to be constant and they are allowed to vary over time. The implication of this feature is that beliefs are not necessarily accurate in equilibrium.[57]

Assuming a trigger strategy for Player B and a specific convex updating process for beliefs,[58] the model predicts that there are two possible equilibria: Player A will choose set L in every period if and only if the following condition is satisfied:

$$\frac{(a_H - a_L)}{r} + \frac{\gamma q_0 (1-r)}{r[1-r(1-\rho)]} \leq b_L - b_H \qquad (1)$$

where q_0 is the initial level of beliefs and ρ is some measure of the speed at which beliefs are updated every period based on the history of play. If *(1)* does not hold, collusion never occurs and the outcome is H in every period. A further result of the model is that, if the equilibrium of the game is collusion in every period, Player B will always make a transfer equal to $k^*_t = (a_H - a_L + \gamma q_t)/r$ *(2)*

2.3. Experimental hypotheses

Based on the assumptions of the above model and its predictions, the main hypotheses which this experiment aims to test are the following:

Hypothesis 1: The likelihood of collusion decreases in q_t (A's beliefs about C's beliefs)

Hypothesis 2: The minimum transfer that is necessary to sustain collusion falls over time.

This follows from the value of k^*_t in *(2)*. The intuition is that the effect of guilt falls every time collusion occurs, as expectations fall; Player B should realise this and make a progressively lower transfer.

---

[57] For a detailed discussion on this issue and on the concept of equilibrium in this model, see Chapter 2.

[58] The process is the following: $\begin{cases} q_{t+1} = q_t + \rho(1-q_t), & if\ \Pi_t = H \\ q_{t+1} = q_t - \rho q_t, & if\ \Pi_t = L \end{cases}$

In addition, the experiment attempts to shed some light on the relative impacts of guilt and fairness in this context by testing the following hypothesis:

Hypothesis 3: The likelihood of collusion is independent of the size of the externality, c_H-c_L.

Indeed, condition *(1)* does not include C's payoffs, suggesting that A's decision is not affected by *how much* he hurts C when he chooses L. If A were motivated by considerations about fairness, he should be less likely to choose L when c_L is very low relative to a_L and b_L.[59]

3. Experimental design and procedure

The experiment took place in May 2008 at the Scottish Experimental Economics Laboratory (SEEL), which is based at the University of Aberdeen. All subjects were students, recruited by means of an online database as well as ads posted on campus. No subject took part in more than one session. The experiment was programmed and conducted with the software z-tree (Fischbacher 2007).

The experimental setup was designed in order to follow the theoretical model as closely as possible. Subjects played a game in groups of three. This was based on the following payoff table:

[59] Abbink, Irlenbusch and Renner (2002) include a treatment where collusion between two players generates an externality for all other participants in the experiment. The results show that the externality has no effect on decision making. However, the structure of that experiment is very different: the externality is small and spread among all players, and it is observable to them only at the end of the experiment.

	Payoff for Player A	*Payoff for Player B*	*Payoff for Player C*
Row1	10	10	10
Row2	9	14	5

In every period, Player A had to decide whether to implement Row1 or Row2. I will also refer to Row1 and Row2 as the high and low payoff set respectively.[60] Player B had the option to make a transfer to Player A before the latter reached a decision, while Player C could not take any action. The above payoff table captures the main features of the game that I wanted to test: Row1 gives a higher payoff than Row2 to players A and C; Row2, on the other hand, is the preferred outcome of Player B by quite a margin. If B makes a transfer of at least 1 and A chooses Row2, then both these players are (weakly) better off than if the choice were Row1. On the contrary, Player C always suffers from a choice of Row2.

In order to replicate the indefinite horizon of the model while ensuring a minimum number of observations, I implemented the following structure: Each treatment was played for at least five rounds. Then, at the end of the fifth round and of every round thereafter, the game broke down with a probability of 40%.

Before the start of each session a first set of instructions was handed to the subjects and read aloud, outlining the basic characteristics of the experiment. These instructions gave information on the structure of the experiment and on payment, but did not reveal any specific information on the sequence of play. At the beginning of the session players were randomly divided into groups of three and assigned one of three roles: Player A, Player B or Player C. They were informed that they would keep their role and remain in the same group throughout the experiment –thus playing the game always with the same partners.

[60] Row2 gives a lower payoff than Row1 to Player A (who makes the choice) and also to Player C. In addition, the aggregate payoff for all three players is lower in Row2.

After the assignment of roles and before players had to start making their decisions, they were each given a further sheet of instructions according to their role (which was appearing on their screens). These sets of instructions explained the game in detail. Information was asymmetric in the following way: Players A and B were given all the information on the payoff structure, whereas Player C only knew the value of his (her) own potential payoffs. The reason why players received different sets of instructions is that I wanted to ensure a greater variation in reported beliefs by introducing asymmetric information into the design. Players C would be more likely to report corner beliefs, had they been given the full payoff structure. The theoretical model also invokes asymmetric information as a motivation for the assumption that beliefs should vary over time and need not correspond to equilibrium play.

Players were also informed about the sequence of play in each round, which was as follows:

1. Guessing stage: Measuring beliefs

Player C was asked to report his (her) estimate of the proportion (in %) of Players A across groups who would choose Row2.[61] At the same time, Player A was asked to report his (her) estimate of Player C's estimate. This estimate was used to construct our measure of A's 2nd order beliefs. Players were told that they would be rewarded at the end of the experiment depending on the accuracy of their guesses, and also that their guesses would not be made known to any other payer at any time during or after the experiment. This last feature was employed so that subjects would not have any incentive to misreport their beliefs; this might be the case if for instance Player C thought that Player A was guilt averse and therefore more likely to choose the high set if the communicated belief was very high.[62]

[61] We asked for the average across groups rather than just for the group of each particular Player C. The reason for this is that in the latter case subjects might have an incentive to "gamble" and report 0 or 100 in order to maximise their payoff. Dufwenberg and Gneezy (1996) follow a similar procedure.

[62] For a more detailed discussion on some methodological aspects of belief measurement see Bacharach et al (2004)

2. Transfer stage: Player B had the opportunity to transfer any amount between 0 and 10 to Player A, before the latter made his (her) choice.

3. Decision stage: After observing B's transfer (if any), Player A had to choose between Row1 and Row2.

4. *Profit Display:* The outcome of the game for the round was communicated to the players, along with their respective payoffs -excluding the payoffs from the guessing game which were only revealed at the end of the session.

After the end of stage 4, the game was repeated again from stage 1. From the 5th round onwards, the 40% probability of breakdown was introduced. During the course of a treatment, subjects could see the history of play on their screens.

Second treatment

In order to test Hypothesis 3, a second treatment was carried out in each session. In this, Player C's payoff in Row2 was 0 instead of 5, while A's and C's payoffs were not different from the first treatment. In other words, A's decision to collude with B would be even more detrimental to the Player C's payoff, and would lower his absolute *and relative* payoff. Everything else (group composition, roles, probability of breakdown) remained the same as in the first treatment.

Payment

Subjects were paid according to the average payoff that they earned during the entire session. The payment included the payoff from the guessing stage for Players A and C. In each period this was calculated as follows: For Player A the formula used was *[10 - abs (GuessA - GuessC)/10]*, rewarding players for an accurate guess of player C's guess in their group. For Player C the formula was *[10 - abs (GuessC - average Row2)/10]*,

rewarding players for an accurate guess of the average of Row2 choices across groups. In addition, all subjects received a show up fee of £5.

The number of periods used to calculate the final payment is a methodological question that has found different answers in the literature: Some experimenters use every round played; others use a small number (usually 1 or 2) of randomly selected rounds for payment, in order to avoid endowment effects and to make sure that subjects are highly motivated throughout the session. In this experiment we used the average payoff of all rounds played during each session, because the random selection of paying rounds presented some difficulties given the fact that the total number of rounds was not known with certainty in advance –not even to the experimenter. The average final payment (including the show up fee) across all participating subjects was £13.8.

We ran three sessions of the experiment. 15 subjects participated in the first session, 24 in the second and 18 in the third, for a total of 57 subjects or 19 groups. In the first two sessions treatment 1 was played first followed by treatment 2; the order was reversed in the third session. Given the random probability of breakdown, the number of periods played in each treatment of each session was different: in the first session the first treatment was played 7 times and the second treatment 8 times; in the second session the rounds played were 7 and 6, while in the third session both treatments only lasted for 5 rounds. This gave a total of 239 rounds played.

4. Results

4.1. Data analysis and statistics

The dependent variable that I am trying to explain is *choice*: this variable takes the value 1 if Player A has chosen Row2 (the low payoff set), and 0 otherwise. Table 1 shows the frequencies for *choice*. Table 2 shows some summary statistics for the two continuous variables: *transfer*, which is the transfer from Player B to Player A, and q_t, which is the

measure of A's 2nd order beliefs (that is, A's estimate of C's belief on how likely A is to choose Row1).

Table 1: Frequencies for the dependent variable, *choice*

choice	*Frequency*	*Percent*
0	201	84.1
1	38	15.9
Total	239	100.0

Table 2: Summary statistics of the continuous independent variables

Variable	*Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
transfer	239	0.908	1.803	0	10
q_t	239	69.49	31.145	0	100

In general, despite the fact that there were substantial monetary gains to cooperation, Players A and B did not exploit them very often. In a total of 239 observations, Row1 (the high payoff set) was chosen 201 times, and Row2 (the low set) 38 times. Moreover, in only 26 cases was the choice of the low set preceded by a positive transfer. Thus, following the definition in the theoretical model, collusion in this experiment only occurred 11% of the time.

In order to explain this fact, I look in turn at the behaviour of Players A and B. Starting with Players B and the distribution of transfers, a striking 66% of them were zero. The mean transfer was 0.9, with a standard deviation of 1.8. Among nonzero transfers, the value of 2 was focal (40 cases) and led to collusion in more than half the cases. It is interesting to look at transfers in the first period, as it is natural to believe that these carry an added weight in terms of establishing collusion. It turns out that across all first periods only 39% of the transfers were zero, while the mean transfer was 1.82 –about double the average for the entire game. This hints to the possibility that Players B were generally

willing to cooperate and signalled their willingness with positive initial transfers, but later responded to A's choice of Row1 by switching to a zero transfer strategy.

Turning to Players A, they generally did not respond to positive transfers by choosing the low set; only 28% of positive transfers led to a choice of Row2. By definition, this prevented cooperation from being established or caused it to break down. On the contrary, a striking observation is that, in 12 out of 157 times (7.6%) when the transfer was zero, the low payoff set was still chosen. This is a very counterintuitive outcome; a potential explanation is that some players were signalling their willingness to cooperate over the horizon of the game.

Beliefs

In theory, expectations should fall every time there is collusion between Players A and B. I look at the response of beliefs to the incidence of collusion in order to give some insights on the way these are updated. The pattern of beliefs reveals that the third party indeed updates his(her) expectations based on the previous period's choice. Row2 was chosen 38 times altogether, 6 of which were in the last period –and therefore their impact on next period's beliefs cannot be estimated, which leaves 32 observations. A choice of Row2 led to a reduction in Player C's guess in 17 of these cases and to no change in 11 cases. Only 4 times were beliefs higher immediately after Row2 was chosen. Overall, the mean observed reduction in beliefs following a choice of Row2 was 14.25%. Hence, beliefs seem to follow the predicted pattern, falling every time Player A chooses to collude. It is also interesting to look at the response of q_t (A's 2nd order beliefs) to collusion: a choice of Row2 led to lower beliefs in 12 cases, while in a further 12 cases these remained constant.[63] The mean fall in q_t was 4.66%, which is much more modest than in the case of 1st order beliefs.

[63] It must be noted, however, that in 9 out of the 12 cases where q_t did not change its value was zero and therefore could not fall any further.

I now look at the implications of the experimental results for each of the three main hypotheses put forward in section 2.3.

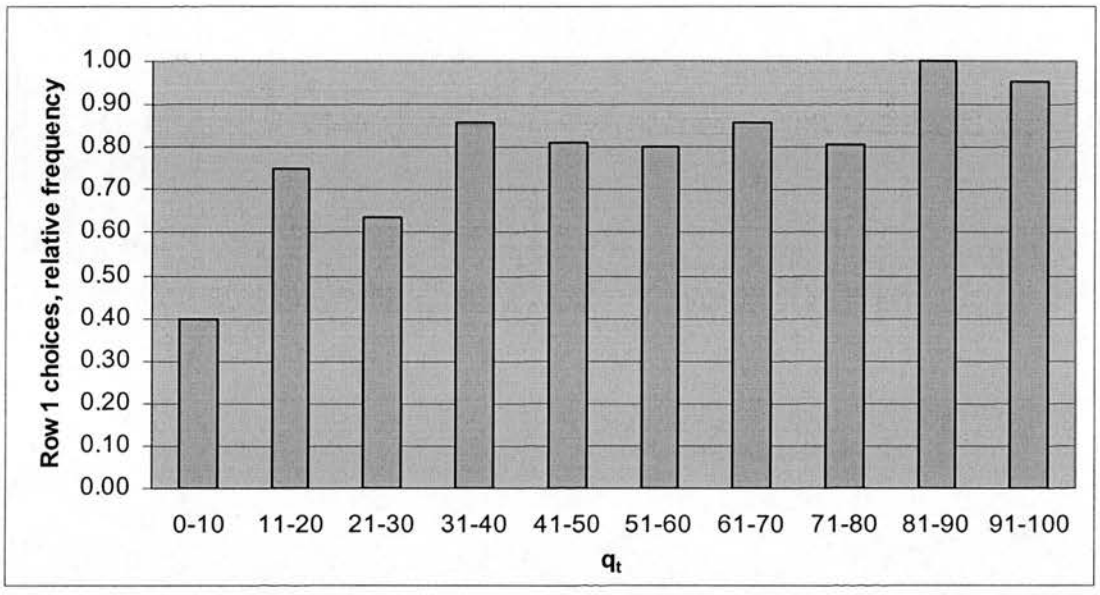
Hypothesis 1: The likelihood of collusion decreases in q_t

There is strong statistical evidence that supports Hypothesis 1. The Spearman correlation coefficient between q_t and *choice* is -0.36, and the null of no correlation is rejected.[64] Figure 1 gives an overview of the relationship between choices and beliefs, by showing the relatively frequency of a choice Row1 (no collusion) for different levels of beliefs. The choice of Row1 usually dominates, as a result of the low overall incidence of collusion, and it is generally increasing in frequency as q_t gets higher. This is particularly evident at very high levels of beliefs (above 80), where collusion almost never occurs. On the contrary, when perceived expectations are very low (between 0 and 10), the relative frequency of Row1 choices is 0.40. This means that, at very low levels of 2nd order beliefs, collusion is actually the most common outcome.

The clear connection between beliefs and the choice of payoff set supports the basic idea behind the theoretical model, which is that players are motivated by psychological factors when they make decisions. The notion of social responsibility towards a passive third party seems to be relevant, and the tools of psychological game theory –basically, the direct inclusion of beliefs in the utility function- appear to be appropriate to explain this idea. The significance of beliefs, as well as the magnitude of their effect, are further investigated in the econometric analysis which follows in section 4.2.

[64] However, the Spearman test results must be interpreted with caution because some data points are not statistically independent of each other –beliefs as well as choices may be serially correlated for a given subject.

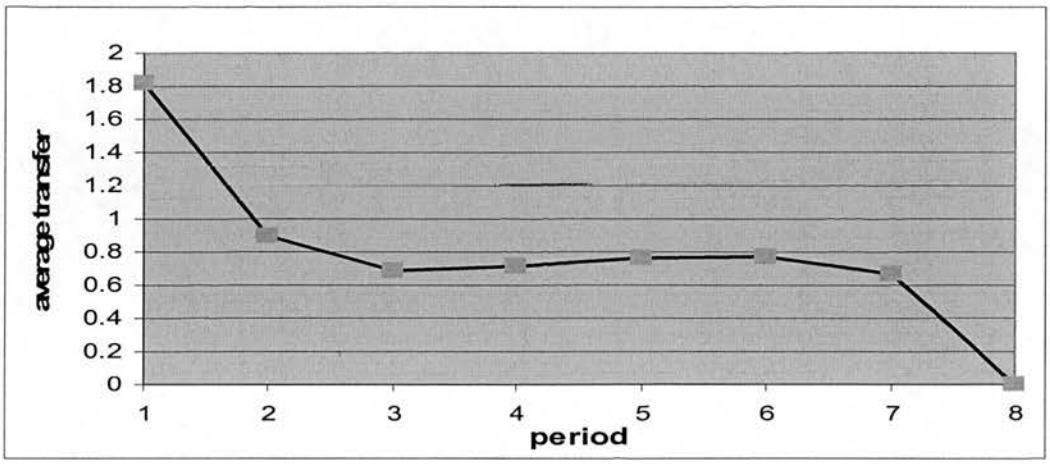
Figure 1: Relative frequency of Row1 choices at each level of beliefs



Hypothesis 2: The minimum transfer that is necessary to sustain collusion falls over time

Figure 2 shows the evolution of average transfers over time. There does appear to be a negative trend, with average transfers starting quite high and then falling rapidly over the next two periods. They then remain stable, before dropping to zero in period 8. However, this graph only depicts the average transfer across groups, and with a varying number of periods in each treatment. A more disaggregated analysis reveals that collusion -that is, a choice of the low set as a response to a positive transfer- is seldom the outcome for more than two periods in a row. There are in fact only three such occasions: collusion is sustained twice for five periods, and once for six periods. In all these cases the transfer remained constant at 2. Hence, Players B do not appear to take into account the possibility that A's guilt may be falling over time as the third party gets progressively disappointed.

Figure 2: Evolution of average transfers over time



Hypothesis 3: The likelihood of collusion is independent of the size of the externality

Put differently, hypothesis 3 states that the probability of choosing the low payoff set does not differ across treatments. Table 3 shows the cross tabulation of choice and treatment. A lower proportion of Players A chose the low payoff set in the second treatment: 15 out of 118 observations or 12.7%, compared to 23 out of 121 observations or 19% for the first treatment. These figures lend some support to the possibility that Players A were relatively more reluctant to choose the low set when they thought this was hurting the third party "too much". Nevertheless, formal statistical testing fails to detect a significant relationship. I performed a Wilcoxon signed-rank test, where each observation was the average frequency of a choice of 1 for an individual Player A over all rounds of a treatment. The 2-tailed test returns a p-value of 0.42, failing to reject the null hypothesis that the median frequencies are equal across the two treatments. Hence, the evidence is in favour of Hypothesis 3.

Table 3: Choice and Treatment

		Treatment		Total
		1	2	
choice	0	98	103	201
	1	23	15	38
Total		121	118	239

4.2. Econometric analysis

In order to give a more rigorous analysis of the parameters of interest I ran a number of probit regressions, reported in Tables 4 and 5. The dependent variable in all regressions is the choice of payoff set; the variable takes the value 1 if Player A chose the low payoff set, and 0 otherwise. The data are in the form of an unbalanced panel: There are 19 groups and therefore 19 individuals who determine the dependent variable, *choice*. Depending on the random probability of breakdown, each individual played between 10 and 15 rounds of the game. To exploit the panel nature of the data and allow for unobserved heterogeneity across individuals, random effects are used as the main estimation method. The basic specification regresses *choice* on a constant, the 2nd order belief q_t and the variable *transfer*, and the results are shown in Table 4. Both explanatory variables are significant and have the expected signs.

Table 4

Choice of payoff: Basic specification

Dependent variable: *choice*			
variable	*coefficient*	*standard error*	*p value*
constant	-0.175	0.293	0.551
q_t	-0.0175	0.004	0.000
transfer	0.120	0.058	0.038
N=239. Pseudo R^2= 0.145. Estimation method used: Random effects			

Specifications 2-4 in Table 5 add a number of controls. In *(2)* I include the variable *treatment* which takes the value 0 for the first treatment and the value 1 for the second treatment. Specification *(3)* includes the dummy variable *order* to control for order effects; this dummy takes on the value 1 if subjects played the second treatment first, and 0 otherwise. Finally, *(4)* tests for time effects by including *period* as a regressor, as well as a dummy variable *(dt5)* which has the value 1 from period 5 onwards.

Random effects require that any unobserved heterogeneity across Players A be uncorrelated with the regressors. In this case this is true for *treatment,* the time controls, and the transfer (which is made by another player). It may, however, not hold for the measure of beliefs q_t. To account for this possibility I have also estimated the basic specification using fixed effects, with the results reported in column *(5)* of Table 5. It must be noted though, that all cross-section identifiers are insignificant. Finally, column *(6)* shows the results of fitting a simple pooled probit model to the data. Both these last two alternative specifications yield almost identical results with random effects in terms of the magnitude and significance of the parameters of interest, perhaps with the exception of a slightly less significant coefficient on *transfer* in the fixed effects regression. Finally, I ran a number of logit regressions, which give very similar estimates to the ones shown in the tables (once the necessary coefficient transformations are taken into account), and are therefore not reported here.

The additional control variables are insignificant in all specifications and their inclusion does not affect the coefficients of q_t and *transfer* or their significance. The joint F-test on all additional controls returns a p-value of 0.97: these results strongly suggest that q_t and *transfer* are the only variables that affect the choice of payoff set from Player A.

Hence, the estimated final probit model is:

$$f=\Phi(-0.175 - 0.0175*q_t + 0.12*transfer), \qquad (3)$$

where f is the probability that a Player A will choose the low payoff set *(choice=1)* and $\Phi(.)$ denotes the cumulative distribution function of the standard normal distribution.

Table 5

Choice of payoff: Alternative specifications

Dependent variable: *choice*					
	(2)	**(3)**	**(4)**	**(5)**	**(6)**
const	-0.162	-0.211	-0.236	0.232	-0.069
	(0.300)	(0.297)	(0.393)	(0.546)	(0.224)
qt	-0.0174 ***	-0.0176 ***	-0.0176 ***	-0.0174 ***	-0.0174 ***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
transfer	0.119 **	0.121 **	0.123 **	0.124 *	0.110 **
	(0.058)	(0.057)	(0.059)	(0.074)	(0.053)
treatment	-0.050				
	(0.238)				
order		0.194			
		(0.337)			
period			0.033		
			(0.108)		
dt5			-0.159		
			(0.446)		
Pseudo R^2	0.146	0.147	0.146	0.248	0.188
Estimation method	Random effects	Random effects	Random effects	Fixed effects	Pooled probit

*N=239. Robust standard errors in brackets. *, **, *** denote significance at the 10%, 5% and 1% level respectively.*

Based on the regression output the following statements can be made regarding Hypotheses 1 and 3 (Hypothesis 2 cannot be tested further using these regressions):

Hypothesis 1: The likelihood of collusion decreases in q_t

The central hypothesis of the model is overwhelmingly borne out by the formal regressions. The variable q_t has the expected negative sign in all specifications and it is always highly significant. The coefficient and its standard error are practically the same regardless of the number of right hand side variables or the estimation method. Player A's choice seems to be strongly affected by what he thinks the third party C expects of him, and expectations do indeed turn out to be self fulfilling.

Turning to the transfer, it is always significant and has the expected positive sign: a higher transfer makes collusion more likely. Nevertheless, *transfer* is sometimes only significant at the 5% level, and in any case less significant than q_t. This is a very interesting result: the impact of beliefs on *choice* seems to be more robust than the impact of the monetary transfer. This highlights the prevalence of psychological factors and motivations in decision making.

In order to estimate the magnitude of the effect of beliefs on choice, I use the results of the basic specification as summarised in equation *(3)*. The probit coefficients are not as straightforward to interpret as those of a linear regression. A general way of interpreting the coefficient on q_t is the following: keeping the transfer constant at a given level, say T, the coefficient -0.0175 on q_t implies that the probability of choice being equal to 1 falls from $\Phi(-0.175 - 0.0175*q_t1 + 0.12*T)$ *to* $\Phi(-0.175 - 0.0175*q_t2 + 0.12*T)$,when q_t increases from q_t1 to q_t2. Table 6 gives the predicted probabilities f for selected values of q_t and *transfer*. 70 is the mean of q_t, while 1 and 2 are the most frequent positive transfer levels.

Table 6

Predicted probabilities of *choice=1*, for various levels of q_t and *transfer* (from basic specification)

	transfer=1	*transfer=2*
$q_t=0$	0.478	0.526
$q_t=70$	0.100	0.123
$q_t=100$	0.035	0.046

The magnitude of the impact of q_t on *choice* is very strong, perhaps even more so than expected. Keeping the transfer at 2, we see that Player A will choose to collude with Player B more than half of the time if he thinks that Player C does not expect anything; on the contrary, if A believes that C is absolutely convinced she will receive a high payoff, the predicted probability of collusion falls to just 4.6%! For the mean belief of 70, the probability of *choice=1* lies at the intermediate level of 12.3%. The same pattern holds for a transfer of 1.

The effect of a change in the level of the transfer can be examined in the same manner, by comparing the predicted probabilities across the first and second column of Table 6. For instance, if we keep q_t constant at its mean value of 70, raising the transfer from 1 to 2 increases the probability of *choice=1* from 10% to 12.3%. Reducing the transfer to 0 would lead to a predicted probability of 8%. Hence, the impact of *transfer* on *choice* is evident in the data, but at the same time it is admittedly more modest than the impact of beliefs.

Hypothesis 3: The likelihood of collusion is independent of the size of the externality

We have already seen that the statistical tests do not detect any significant difference in choices between the two treatments. This is confirmed in the probit regressions: the

variable *treatment* is always insignificant, although it has the expected negative sign in every specification. Hence, relative payoffs and the size of the externality do not seem to matter. This finding justifies the way the psychological component of A's behaviour is modelled, i.e. as a function of beliefs but independent of the size of the externality. Including beliefs in A's utility seems to be necessary in order to capture fundamental elements of his motivation; on the contrary, the fairness motive alone as embodied in the structure of payoffs cannot explain his behaviour.

4.3. Discussion: Identification of psychological factors and the effect of the transfer

A methodological question which arises here concerns the identification of the various effects that can be at play. Social responsibility towards the third party appears to be the dominant force in the relationship between Players A and C. Since C does not take any action, the theory of reciprocity is not applicable because it would require for instance that A perceives C's *actions* as "kind" and therefore reciprocates to her. Fairness is another candidate explanation for the low incidence of collusion; we have seen, however, that collusion is not less common in the second treatment where the distribution of payoffs is even more skewed. Finally, one cannot rule out the possibility that the relationship between the decision maker's choices and his beliefs can be -at least partly- explained by the presence of cognitive biases such as the false consensus effect or the self-serving bias.[65]

On the other hand, the behavioural motivations induced by the presence and the actions of Player B are not clearly identified. In general, one can think of a positive transfer as having two potential effects on A's behaviour: a "material" one owing to the possibility of increased payoffs through cooperation in the repeated game, and a "psychological" one arising from the fact that A may feel compelled to reciprocate to B's transfer.[66]

---

[65] The false consensus effect means that A may project his own beliefs on C; the self-serving bias means that A's beliefs may follow his choices because he is subconsciously trying to "justify" his actions as "good" or reasonable. In any case, both these effects imply that there may be some reverse causality, with choices causing beliefs. See also Charness and Dufwenberg (2006, p.1594).

[66] The potential psychological component of A's behaviour towards B is discussed in the theoretical model of Chapter 2 (section 7).

The analysis on Hypothesis 2 in 4.1 makes the observation that the choice of the low set is usually a one-period event, which comes as a response to a positive transfer but is not sustained. Thus, with few exceptions, it may be more appropriate to think of this choice as the result of reciprocity between players A and B, rather than as a deliberate strategy with the objective to reap the benefits of cooperation. In practice, it is not possible to disentangle the two effects here because the experiment does not measure beliefs about Player B; but the fact that sustained collusion hardly ever occurs suggests that reciprocity may be the main factor at play.[67] In any case, the focus of this experiment is not on reciprocity between the two players, but on the psychological effect of the third party's perceived expectations.

5. Concluding remarks

5.1. Summary of main findings

The experimental data provide very solid support in favour of the central hypothesis that this paper aims to test. Third party beliefs do matter and they affect the probability of obtaining cooperation in a repeated game between two players. The magnitude of this effect was found to be considerably large: ceteris paribus, raising the perceived expectations of the third party from 0 to 100 percent reduces the predicted probability of cooperation from around a half to less than 1 percent. The only other factor which turned out to be significant in the regressions was the transfer made from Player B to Player A, although this effect was quantitatively weaker than the effect of beliefs.

Varying the level of the externality did not matter significantly for the probability of cooperation between Players A and B. I interpret this as evidence that the decision maker does not care about the third party's payoff per se; his psychological motivation does not

[67] Several experiments in the literature provide evidence that players are motivated by some kind of reciprocity. See, for instance, Fehr, Gächter and Kirchsteiger (1997), Falk, Fehr and Fischbacher (2000), Abbink, Irlenbusch and Renner (2000, 2002), Dufwenberg et al (2001).

depend on what the third party *gets*, but on what she gets relative to what –the decision maker thinks- she *expects*.

Finally, the experiment did not find support for the idea that guilt diminishes over time and leads to lower transfers. In fact, collusion occurred rather rarely, and it was only sustained in a couple of cases. Hence, the results do not tell us much about the evolution of cooperation over time and its dependence on beliefs.

5.2. Implications

The findings of this experiment have implications for economic -or other- situations where outsider beliefs are potentially relevant. A few such examples are discussed in Chapter 2 of the thesis, paying particular attention to corruption in public administration. In terms of the experiment, we can think of Player A as a public official, of Player B as a lobby who can bribe the official (e.g. in order to secure a procurement contract) and of Player C as the public. Abstracting from electoral considerations (because for instance the official has a permanent job), the experimental results suggest that the official is less likely to be corrupt if she believes that the public expect her to be honest. In addition, hypothesis 3 implies that the counter motive to corruption lies not so much in the level of the externality that this would generate, as in the official's fundamental reluctance to betray the public's expectations. In other words we are dealing with a story of self-fulfilling expectations, which means that efforts to build trust in public administration could actually be the key to improving its quality.

References

Abbink K., Irlenbusch B., Renner E. (2000): "The Moonlighting Game - An Experimental Study on Reciprocity and Retribution". *Journal of Economic Behavior and Organization* 42, 265-277

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

Bacharach M., Guerra G., Zizzo D. (2007): "The Self-Fulfilling Property of Trust: An Experimental Study". *Theory and Decision* 63, 349-388

Bolton G., Ockenfels A. (2000): "ERC: A Theory of Equity, Reciprocity, and Competition". *American Economic Review* 90, 166-193

Charness G., Dufwenberg M. (2006): "Promises and Partnership". *Econometrica* 74, 1579-1601

Dufwenberg M., Gächter S., Hennig-Schmidt H. (2006): "The Framing of Games and the Psychology of Strategic Choice". CeDEx Discussion Paper no. 2006-20

Dufwenberg M., Gneezy U. (2000): "Measuring Beliefs in an Experimental Lost Wallet Game". *Games and Economic Behavior* 30, 163-182

Dufwenberg M., Gneezy U., Güth W., van Damme E. (2001): "An Experimental Test of Direct and Indirect Reciprocity in case of Complete and Incomplete Information". *Homo Oeconomicus* XVIII 19-30, München

Geanakoplos J., Pearce D., Stacchetti E. (1989): "Psychological Games and Sequential Rationality". *Games and Economic Behavior* 1, 60-79

Falk A., Fehr E., Fischbacher U. (2000): "Testing Theories of Fairness - Intentions Matter." Working paper no. 63, Institute for Empirical Research in Economics, University of Zurich

Fehr E., Gächter S., Kirchsteiger G. (1997): "Reciprocity as a Contract Enforcement Device: Experimental Evidence". *Econometrica* 65, 833-860

Fehr E., Schmidt K. (1999): "A Theory of Fairness, Competition, and Cooperation". *Quarterly Journal of Economics* 114, 817-868

Fischbacher U. (2007): "z-Tree: Zurich Toolbox for Ready-made Economic Experiments". *Experimental Economics* 10, 171-178.

Rabin M. (1993): "Incorporating Fairness into Game Theory and Economics". *American Economic Review* 83, 1281-1302

Appendix A
Written instructions

[The following set of instructions was handed out to all players at the start of the experiment, along with a consent form.]

Instructions for all players

You are about to participate in an experiment in the economics of decision-making. Please read these instructions carefully, as the amount of money you earn in the experiment will depend on how well you understand them. All money you earn will be paid to you privately in cash at the end of the experiment. If you have a question at any time, please feel free to ask the experimenter. We ask that you do not talk with the other participants during the experiment.

This experimental session consists of two different games. Each game will be played for at least 5 rounds, but the actual number of rounds is not known with certainty in advance: After the 5th round round and in every round thereafter, there is a 40% chance that the game ends, and a 60% chance that the game continues for at least one more round.

Each game will be played by you and two other players. At the beginning of the session you will be randomly divided into groups of three, and you will remain in the same group throughout the entire session. You will not be told the identity of the players matched to you, nor will they be told your identity - even after the end of the session.

Within each group, one player will be randomly chosen as Player A, one as Player B and one as Player C. You will keep your role throughout the entire session.

Once the roles have been assigned, you will receive a further sheet of instructions describing to you how the game is played and what actions you can take according to your role.

Payments: At the end of the experiment you will be rewarded on the basis of the average profit you have earned during all rounds, at an exchange rate of 4 experimental units/£. Each participant will additionally receive £5 for completing the session. Payments are made in cash.

Are there any questions?

[At the start of the first treatment, roles were randomly assigned and a message appeared on each player's screen revealing their role (e.g. "Your role is Player A"). At that point a second set of instructions was given to all participants, according to their role.]

Instructions for Player A (Game #1)

You have been randomly assigned the role of Player A in your group.
Your actions will determine the payoffs of all players: In each round, you will be asked to choose one of two alternative sets of payoffs, Row1 or Row2. These are shown in the table below:

	Payoff for Player A	*Payoff for Player B*	*Payoff for Player C*
Row1	10	10	10
Row2	9	14	5

Before you make your choice, Player B may decide to make a transfer to you. If such a transfer is made, it is added to your payoff and subtracted from Player B's payoff. The transfer cannot be taken back, irrespective of your subsequent choice of payoffs.

Please keep in mind the following:

(1) Player B can see the above table of potential payoffs. On the contrary, Player C only knows his (her) own potential payoffs, and (s)he does not know the level of the transfer. In other words, Player C only knows that (s)he will get 10 if you chose Row1 and 5 if you choose Row2.
(2) Remember that the game is repeated for a number of times, which is not known with certainty in advance
(3) Remember that you remain in the same group for the entire session.

At the beginning of each round we will be asking you a question that relates to your beliefs about that particular round. Please think carefully before you give us your guess, as you will earn an additional payoff depending on the accuracy of your guesses. Note also that your guesses will not be disclosed to any other player during or after the session.

At the end of each round, you will be shown a summary of the outcome of the game and the payoffs for the round.

Are there any questions? If yes, please raise your hand.

Instructions for Player B (Game #1)

You have been randomly assigned the role of Player B in your group.
The game is played as follows: In each round, Player A will be asked to choose between two alternative sets of payoffs, Row1 and Row2. These are shown in the table below:

	Payoff for Player A	*Payoff for Player B*	*Payoff for Player C*
Row1	10	10	10
Row2	9	14	5

Before Player A makes his (her) choice, you are given the possibility to make a transfer to him (her).
If you do make a transfer, its value is subtracted from your payoff and added to Player B's payoff. The transfer cannot be taken back, irrespective of Player A's subsequent choice of payoffs.

At the end of each round, you will receive information on the outcome of the game and the payoffs for the round.

Please keep in mind the following:

(1) Player A knows the above table of potential payoffs. On the contrary Player C only knows his (her) own potential payoffs, and (s)he does not know the level of the transfer. In other words, Player C only knows that (s)he will get 10 if Player A chooses Row1 and 5 if Player A chooses Row2.
(2) Remember that the game is repeated for a number of times, which is not known with certainty in advance
(3) Remember that you remain in the same group for the entire session.

Are there any questions? If yes, please raise your hand.

Instructions for Player C (Game #1)

You have been randomly assigned the role of Player C in your group.
The game is played as follows: In each round, Player A will be asked to choose between two alternative sets of payoffs, Row1 and Row2. Your respective payoffs are shown below:

	Payoff for Player C
Row1	10
Row2	5

Before Player A makes his (her) choice, Player B has the possibility to make a transfer to him (her). However, you will not be informed about the level of this transfer (if any).

Please keep in mind the following:

(1) In the above table, you are not shown the payoffs of Players A and B in Row1 and Row2; however, Players A and B both possess this information.
(2) Remember that the game is repeated for a number of times, which is not known with certainty in advance
(3) Remember that you remain in the same group for the entire session.

At the beginning of each round we will be asking you a question that relates to your beliefs about that particular round. Please think carefully before you give us your guess, as you will earn an additional payoff depending on the accuracy of your guesses. Note also that your guesses will not be disclosed to any other player during or after the session.

At the end of each round, you will receive information on the outcome of the game and the payoffs for the round.

Are there any questions? If yes, please raise your hand.

[At the start of the second treatment players were given new sets of instructions according to their role. Those were identical to the instructions for the first treatment, with the only exception that in the table of payoffs the value of 5 was substituted by 0 in Player C's Row2 payoff.]

Appendix B

On-screen instructions

1st stage – Measuring beliefs

Player C had to answer to following question:

"Please state your guess on the following question:
On average, what proportion of players A in all groups will choose Row2 in the current round? (0 to 100 percent)
You will be rewarded at the end of the experiment depending on the accuracy of your guess."

Player A had to answer to following question:

"We asked player C in your group to give us his(her) guess on the average proportion of players A in all groups who will choose Row2 in the current round.
What do you think his(her) guess was? (from 0 to 100 percent).
You will be rewarded at the end of the experiment depending on the accuracy of your guess."

2nd stage – Transfer

Player B was shown the following message:

"Remember that Player A will be asked to choose between Row1 and Row2.
You have the option of making a transfer to Player A before (s)he makes his(her) choice.
Your transfer to Player A (minimum=0, maximum=10): "

3rd stage – Decision

Player A was shown the transfer made from B and asked to decide between Row1 and Row2

[In the first three stages all players could see the history of play on their screens. Players A and B were also shown the full table of potential payoffs, while Player C was only shown his (her) potential payoffs]

4th stage – Profit display

All players were informed of A's decision and of their payoff for the period. Players A and C were additionally shown the following message: *"This [payoff] does not include the payoff from the guessing stage, which will be calculated and added to your total profit at the end of the game."*