b
UNIVERSITÄT
BERN

Faculty of Economics and
Social Sciences

Department of Economics

Endogenous Institutions and the Dynamics of Corruption

Esther Bruegger

05-04

February 2005

DISCUSSION PAPERS

Gesellschaftsstrasse 49
CH-3012 Bern, Switzerland
http://www.vwi.unibe.ch

Endogenous Institutions and the Dynamics of Corruption

Esther Bruegger*

Feb 2005

Abstract

While empirical studies which analyze large cross section country data find that corruption lowers investment and thereby economic growth, this result cannot be established for certain subsamples of countries. We argue that one reason for these mixed findings may be that a country's corruption and growth rates are tightly linked as variables of a dynamic process which can have several equilibria or have different sets of equilibria. In order to understand the circumstances in which a country converges towards a certain equilibrium, we model the individual decisions to invest and corrupt as an evolutionary game. In this model the quality of government institutions is an endogenous variable, depending on the corruption rate, the population income, and the type of institutions; the quality of institutions itself then determines the future incentives to corrupt. The comprehension of these feedback effects allows us to study the role of the type of institutions for the dynamics of corruption. We present the equilibria for different types of institutions and discuss the resulting dynamics. The results suggest that cross country studies may significantly underestimate the impact of corruption on growth for certain countries. Depending on how the quality of institutions depends on corruption and income, corruption can either lower growth, suppress it entirely, or be positively correlated with growth in some special situations.

Keywords: Corruption, Institutions, Feedback Effects, Evolutionary Game.

JEL-Classification: C73, D73

*Address: UCLA, Department of Economics, 8283 Bunche Hall, Box 951477, Los Angeles, CA 90095-1477. Email: bruegger@ucla.edu

1 Introduction

Due to the lack of time series data with sufficiently many year points that measure the prevalence of corruption in a society,[1] cross section data analysis has been the main method in economic research for studying the causes and consequences of corruption. Among the most important findings of this literature is the observation that corruption lowers economic growth (Mauro, 1995).[2] The reasons are reduced investment incentives[3] and lowered institutional quality, which both influence economic growth negatively.[4]

However, these results are not robust for certain groups of countries: Rock and Bonnett (2004) find that "corruption tends to increase growth in the large East Asian newly industrialized economies". Similarly, for a selection of Southeast Asian nations, Lim and Stern (2002) point out that "relatively high levels of corruption have been associated with decades of very rapid GDP growth" (p. 46). These controversial findings do not come as a surprise: First, the impact of corruption on growth has not been very pronounced in many studies (see e.g. Li et al., 2000), and second, there is abundant anecdotic evidence that nations going through a successful developing process experience a surge in corruption during their economic take-off (Wedeman, 2004).[5]

How can we reconcile these mixed results? One possible explanation is that a country's corruption and growth rates are endogenous variables of a dynamic process

---

[1] The indicator for the prevalence of corruption most widely used in empirical research is the Corruption Perception Index (CPI) by Transparency International (TI), an international non-governmental organization devoted to combating corruption. The first year the CPI was published is 1995.

[2] Other conclusions from cross-country studies include: Corruption increases poverty and income inequality (Gupta et al., 2002; Hindriks et al., 1999; Johnston, 1989), it augments the public deficit and public debt, increases mismanagement of public services and the respective transaction costs (Tanzi, 2002; Tanzi and Davoodi, 2000), it reduces the effectiveness of social spending, the formation of human capital (Gupta et al., 2002), expenditure on education and health (Tanzi, 2002) and tax revenue (Tanzi and Davoodi, 2000).

[3] In a corrupt environment private investment, both domestic and foreign, yields lower returns because of additional costs and a climate of heightened uncertainty (Mauro, 1995). This result has been supported by the work of Brunetti et al. (1998), Chong and Calderon (2000), Mauro (1995), Mauro (1997), Mauro (1998), Knack and Keefer (1995), Wei (2000), Tanzi and Davoodi (2000) and others. Furthermore, Ades and Di Tella (1997) present a formal model.

[4] North (1990), Murphy et al. (1993), Murphy et al. (1991) and Husted (1999) present models of how institutions affect growth, and for empirical evidence see Tanzi and Davoodi (2000), Dollar and Kraay (2003), Kaufmann et al. (2000), Buscaglia (2001), Keefer and Knack (1997); among others.

[5] See Hofstaedter (1973) for the example of the United States.

with several equilibria such that variables of different countries can converge to distinct equilibria. The impact of corruption on growth may be different on two paths not converging to the same equilibrium.

In the following we explain why we think that a dynamic consideration is essential for understanding the consequences of corruption. Based on this, we then present a model that captures the dynamic aspects of corruption. We use the results to criticize the cross-country approach and to make suggestions for future empirical work.

Why is a dynamic setup appropriate to study corruption? Because only a dynamic setup allows for the comprehension of feedback effects between the individual decisions to corrupt and population variables like growth or corruption rate. It is clear that the individual decisions determine the corruption and the growth rate: The more individuals decide to engage in corrupt activities, the higher is the corruption rate (Chakrabarti, 2001), and the less individuals decide to invest in production, the lower is the growth rate. However, the corruption and the growth rate themselves have an impact on the individual incentives to corrupt. In both cases, the quality of government institutions can be suggested as a link. Institutions that are weakened by corruption are likely to impose lower costs on corrupt behavior than strong institutions. From this we expect the corruption rate to have a positive impact on the individual incentive to corrupt.[6] Similarly, we expect the growth rate to affect the individual decisions to corrupt. The value of property rights increases in presence of growth. Since people are willing to spend more to protect property rights the more they are worth (Eggertsson, 1990), growth increases the quality of institutions.[7] So we expect growth to reduce the incentives to corrupt. Note that a comprehension of the feedback effects described requires the quality of institutions to be an endogenous variable depending on the population variables.

There is a variety of dynamic frameworks that could be employed. We choose the one of an evolutionary game. In our view, this approach has three advantages. First, feedback effects as described above can be included by extending the framework to a frequency dependent evolutionary game (Bruegger, 2005). Second, a large number of agents choose between different behavioral strategies, which explicitly represents the

[6]This reasoning is not new, Paldam (2002) and Rose-Ackerman (1999) have argued that a high corruption rate may be self-sustainable. Andvig and Moene (1990) come to the same conclusion using a slightly different argument.

[7]Chong and Calderon (2000) find empirical evidence that institutional quality itself is a consequence of economic growth.

individual decisions. Third, the setup of an evolutionary game naturally captures the situation of scarce information that prevails for illegal activities such as corruption. We assume an imitation rule as a strategy selection rule since imitation is a plausible way of mastering a situation of scarce information. In comparison, a general equilibrium model of corruption would also generate several equilibria, but not identify off-equilibrium behavior and path dependence of current population states. A growth model with a representative agent would capture the dynamic aspect of corruption, but not allow for the explicit specification of the individual incentives.

Let us briefly describe our model. The underlying game of the evolutionary game we consider consists of three strategies: agents choose between a private sector activity, being a fair government employee, and being a corrupt government employee. The private sector activity is an innovative activity, i.e. the investing strategy. It's return is a fixed surplus when playing against a fair government employee or another private entrepreneur. In case the private entrepreneur plays against a corrupt government employee, the corrupt government employee siphons the surplus and the private entrepreneur is left without return. A fair government employee earns the government wage no matter whom he plays against. The corrupt government employee also earns the government wage if playing against another government employee. However, if he plays against a private entrepreneur, he extorts the surplus from private sector activity but bears the costs of the corrupt act - that is, the probability of getting caught red-handed and losing all his income in this case. These payoffs imply that the more corruption there is (the higher the frequency of corrupt government employees), the lower the incentive for private investment.

How do population variables affect the individual decisions? As mentioned above, the quality of institutions must depend on population variables like the corruption or the growth rate. One way to establish this is to define institutional quality as the detection probability of corruption. Note that a high detection probability results in high costs of corruption. The costs of corruption enter the corrupt strategy's payoff and are therefore relevant for the decision to imitate it. The first type of institutions we study is characterized by a general functional form of the detection probability: We only assume that the detection probability depends negatively on the corruption rate. The second type of institutions we consider is an example for a more specific case: We assume that the detection probability depends positively on the population income. This is suggested by the feedback effect between growth and institutional quality.

For the general specification, the corruption dynamics features several evolutionary equilibria: A clean equilibrium, a corrupt equilibrium, and a hybrid equilibrium. Depending on the initial strategy state and other exogenous variables, a population will converge to one of these equilibrium. A clean equilibrium where corruption is extinguished can only exist if detection probability is high in absence of corruption. As a consequence, populations with notoriously inefficient government institutions are not expected to free themselves from corruption. However, if the quality of institutions increases with more private activity, a population can converge to a clean equilibrium. In a corrupt equilibrium private activity is driven out of the game and only corrupt activity prevails. Such an equilibrium always exists if the detection probability is low in presence of a high corruption rate. We conclude that if corruption in the government affects the institution's efficiency negatively, the possibility of a population being trapped in a corrupt equilibrium will prevail. A hybrid equilibrium exists for a detection probability that is not too high, but also only lightly affected by corruption. In such an equilibrium, the entire government is corrupt but private activity is not suppressed.

As countries may feature institutions which qualities may change in different manners for a given corruption rate or may start off with different corruption rates, there is no reason to assume that they converge to the same equilibrium. Therefore, cross country studies may actually blur the consequences of corruption. If countries converging to different equilibria are pooled in one sample, the consequences of corruption might be underestimated for countries converging towards the corrupt equilibrium (and overestimated for those converging towards the clean equilibrium). The reason is that a negative impact of corruption on growth is measured while corruption actually not only lowers growth but prevents an increase. Only with a deeper understanding of how institutional quality depends on growth and corruption will increase the explanatory power of empirical analysis.

2 The Basic Model

In this section we present the model of corruption as an evolutionary game but without endogenous quality of institutions yet. We refer to this version of the model as the basic corruption game. We present its stage game and discuss our choice of an imitation rule which generally leads to replicator dynamics. In addition, we introduce the notion of an evolutionary equilibrium as an equilibrium concept.

An evolutionary game describes strategic interaction over time. It is defined by the populations of players, a state space of strategies, a stage game, and an adaptation rule which determines the dynamic adjustment process.

2.1 The Population

We model corruption in a one-population game. We assume the population to consist of a continuum of infinitely-lived players. This assumption has several well-discussed implications (see Friedman, 1998, for a complete list). First, the state space of strategies is continuous. In adherence to continuous time this allows us to specify the dynamics of a game as a system of ordinary differential equations. Second, for an infinite number of players the law of large numbers can be invoked. This allows us to ignore random fluctuations and differing perceptions of the current state among players. Third, an infinite number of players motivates the myopia assumption inherent to our dynamic adjustment process specified below. Players' influence on population are so small that players do not attempt to influence other players' future actions.

2.2 The Strategies

In order to model corruption, we need three strategies: strategy 1, strategy 2, and strategy 3. Hence, the pure-strategy set of any player is $S = \{1, 2, 3\}$. To simplify interpretations, we presume that individuals only play pure strategies. It is convenient for upcoming calculations to introduce the following notation. If an individual plays strategy i, $i \in S$, we denote his strategy choice σ_i as a vector in $\mathbb{R}^3$:

$$\sigma_i \in \left\{ \sigma_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \ \sigma_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \ \sigma_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

The fraction of the population playing strategy i at time t is denoted by $x_i(t) \in [0, 1]$. The fractions of the population playing the three strategies (also called strategy frequencies) are the variables in our model which changes we intend to describe over time. The *strategy state* of the game,

$$x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix},$$

specifies the frequency of each of the three strategies in t. We will drop the time index t if there is no risk of misunderstanding.

The set of feasible strategy states is called the *strategy state space*. Since strategy frequencies must add up to one, the strategy state space for an evolutionary game is a simplex. We define the simplex of dimension $n-1$ as

$$\Sigma_{n-1} = \left\{ x(t) \in \mathbb{R}^n \;\middle|\; x_i(t) \geq 0 \text{ and } \sum_{i=1}^{n} x_i(t) = 1 \text{ for } i = 1, ..., n \right\}.$$

The strategy state space of our game is Σ_2, since we can omit one dimension. We redefine the strategy state as $x = (x_1, x_2)$, from which we can always calculate $x_3 = 1 - x_1 - x_2$.

Let us now turn to the interpretations of the three strategies. Strategy 1 represents the choice of holding down a job as government employee while not exploiting the power of the position. That is, a player choosing Strategy 1 acts as a *fair government employee*. Strategy 2 also comprises to serve in public service, but in contrast to Strategy 1, the player now abuses the power of the public role for private benefits. Strategy 2 can therefore be referred to as the strategy of a *corrupt government employee*. The third option is to pursue a productive private sector activity. We refer to Strategy 3 as the strategy of a *private entrepreneur*. Sometimes it is convenient to speak of public servants or government employees generally; in this case we refer to the total of players with Strategy 1 and 2. The share of public servants is abbreviated with $x_G = x_1 + x_2$.

In every period, players are matched pairwise to play the stage game. In our application, such a pairwise encounter of two players is interpreted as one economic interaction. This means that each play of the stage game is considered as one economic act. The strategy choice of a player represents his decision which sector to direct his manpower to. The greater the share of individuals working in the public sector, x_G, the greater the share of economic activity that is processed entirely within the public sector (games played among government employees) or with the help of the public sector (games between private entrepreneurs and public servants).[8]

Note that our choice of a one-population model has three implications for the interpretation of the basic corruption game. First, the size of government, measured as the share of agents employing strategy 1 or 2, is endogenous. That means, agents find ways and means to work for the government as long as it pays for them. Second, government employees play against themselves. These interactions can be interpreted

[8] The share of economic activity taking place within the government (administration) is $\Pr(x_G \geq y)^2 = x_G^2$ and the share of economic activity happening within the private sector is $\Pr(y \geq x_G)^2 = (1 - x_G)^2 = x_3{}^2$.

as organizational and administrative tasks within a government which we believe to be a realistic feature in a model including the public sector.[9] Third, the private sector cannot elude the interaction with the public sector, no matter whether it is beneficial or damaging to its business.[10] In the next subsection we describe how the three strategies are presumed to interact.

2.3 The Stage Game

The stage game characterizes the strategic interaction of two players at any point in time. The stage game of the basic corruption game is a normal form game defined by payoff matrix A. In every period players are drawn randomly and pairwise to play the stage game and receive the average payoff $f(\sigma_i, x) = \sigma_i^T Ax$.[11]

The assumptions for A are as follows. A fair government employee receives the wage w at any point in time, independently of his opponent. The interpretation is that he is paid w no matter whether he is busy mainly delivering services to the private sector or doing work within the administration. The corrupt government employee, too, receives the wage w at any point in time, but additionally seizes a corruption income when playing against a private entrepreneur. The private entrepreneur generates a surplus s when being paired with another private entrepreneur or a fair government employee, and loses the fruits of his work when encountering a corrupt government employee. The corruption income of the corrupt public servant can be specified as $s - c$, where c depicts the individual costs of corruption. These assumptions lead to

$$A = \begin{pmatrix} w & w & w \\ w & w & w+s-c \\ s & 0 & s \end{pmatrix}. \tag{1}$$

It is a simplifying assumption that the payoff of a private entrepreneur does not vary

[9]The rate at which the economic activities taking place within the administration increase with a marginal change of x_G is inherent in the game structure: $\frac{\partial \Pr(x_G \geq x)^2}{\partial x_G} = 2x_G$.

[10]Note that this feature is similar to a type of world Niskanen proposes: sponsors (in our case these are the private entrepreneurs as we will see in the next subsection) are passive in accepting the output proposal of bureaucracy without any careful monitoring or evaluation of alternatives (Niskanen, 1996).

[11]As common in the literature, we do not differentiate between the average (expected) payoff against the population and the realized payoff of a specific stage game played. There are several reasons for that (Friedman, 1998): First, in large populations such as ours the expected payoff is a sufficient statistic. Second, payoffs are often not generated by random pairwise encounters, but by general interactions such as markets, and are therefore not stochastic.

between playing against a fair public servant and playing against a private entrepreneur. From our view this simplification is justified by the following interpretation: an entrepreneur makes the surplus s from his business activity while losing it with probability x_2, because he cannot circumvent interaction with the government.[12] Changing the payoff for Strategy 3 when playing against Strategy 1 (this is element a_{31} of matrix A), amounts to making a statement about the efficiency of public service. The reason is that x_1 would then affect the average payoff of Strategy 3. If $a_{31} = s$ though, then it is only x_2 that influences the private entrepreneur's average payoff. Although we have an idea in which direction we could change a_{31}, we do not want to include such an effect in our model. It is our aim to analyze the impact of corruption and we do not want to blur the results with other effects.

Note that the game can easily be scaled such that all agents engage in a production alternative beside the three strategy games which is not subject to corruption and does not enhance growth.[13] Even if this production output is subject to taxes, it does not change the equilibria of our model, wherefore we normalize its payoff to zero.

2.4 The Imitation Rule and the Adjustment Dynamics

Now let us describe in the following how the players select their strategies.

Our analysis is based upon the hypothesis, that strategy selection by imitation is a realistic assumption when describing illegal behavior. Players imitate the strategy of other, more successful players, where success refers to greater expected payoff. They base their decision which strategy to imitate on sporadically and imperfectly observed expected payoffs and behavior. Information is scarce because knowledgeable players hide the information for which they can be prosecuted. To derive adjustment dynamics we follow Weibull (1995, p. 155). Note that time is continuous in our model.

Assume that all agents review their strategy choice at any point in time. Let us denote the probability with which an agent playing strategy i (called i-players in the following) switches to strategy j, by $\varphi_i^j(x)$. The share of i-players that will imitate another strategy is

$$\sum_{j \in S \setminus i} x_i \varphi_i^j(x) \quad = \quad x_i - x_i \varphi_i^i(x) \, .$$

[12] Imagine that the entrepreneur needs to collect permits or certificates from the administration, has to pay taxes to the government, or is subjected to controls by law.

[13] Such a production alternative corresponds to the "production of subsistence crop" in Murphy et al. (1993).

The share of players imitating i that have played a different strategy before is

$$\sum_{j\in S\setminus i} x_j\varphi_j^i(x) \quad = \quad \sum_{j\in S} x_j\varphi_j^i(x) - x_i\varphi_i^i(x)\,.$$

This leads to a net effect in the share of i-players of

$$\dot{x}_i \quad = \quad \sum_{j\in S} x_j\varphi_j^i(x) - x_i\,. \tag{2}$$

We now further specify this general adjustment dynamics by making assumptions on $\varphi_j^i(x)$: At time t, every player samples an opponent with a probability equal for all opponents. Every player knows his own average payoff $f(\sigma_j, x)$ which depends on his strategy σ_j, he observes the opponent's strategy and, with some noise, the opponent's average payoff. Therewith, a j-player sampling an i-player observes $f(\sigma_i, x) - \varepsilon$ with probability x_i, where ε is a random variable with a continuously differentiable cumulative distribution function $\Phi : \mathbb{R} \to [0, 1]$.

We assume the following *imitation rule*: A j-player imitates strategy i when his own average payoff (known without noise) is smaller than the observed average payoff of strategy i. That is, he imitates strategy i if $f(\sigma_j, x) < f(\sigma_i, x) - \varepsilon$. The probability that a j-player imitates i is $\Pr[\varepsilon < f(\sigma_i, x) - f(\sigma_j, x)] = \Phi[f(\sigma_i - \sigma_j, x)]$. So the adjustment dynamics takes the form

$$\dot{x}_i \quad = \quad x_i \sum_{j\in S} x_j \Big(\Phi[f(\sigma_i - \sigma_j, x)] - \Phi[f(\sigma_j - \sigma_i, x)] \Big)\,.$$

Finally we must specify the cumulative distribution function Φ. We assume a uniformly distributed error term ε over the interval of possible expected payoff differences. So the function Φ is linear, $\Phi(y) = \alpha + \beta y$, and the adjustment dynamics derived from the imitation rule simplifies to

$$\dot{x}_i \quad = \quad 2\beta x_i \left(f(\sigma_i, x) - f(x, x) \right)\,.$$

Except from a time rescaling, our dynamics is thus equal to the replicator dynamics (see Taylor and Jonker, 1978; Schuster and Sigmund, 1983),

$$\dot{x}_i \quad = \quad x_i \left(f(\sigma_i, x) - f(x, x) \right) \quad \forall i \in S\,. \tag{3}$$

Since we do not focus on rates of convergence in this paper, we can continue using equation (3). We alternatively will refer to it as the replicator dynamics or the imitation dynamics. Note that in our model, by equation (3), strategy selection from myopic imitation leads to a deterministic, continuous-time, continuous-state dynamics. Furthermore, note that equation (3) is a system of ordinary differential equations. For a simplified notation, we define the system's right hand side as the (Lipschitz continuous) function $F : \Sigma \to \Sigma$, and can now write the imitation dynamics as $\dot{x} = F(x)$.[14]

2.5 The Equilibrium Concept

Since we abandon the assumption of a constant stage game in Section 3 for the analysis with endogenous quality of institution, a dynamic equilibrium concept needs to be employed. Therefore we choose the *evolutionary equilibrium* EE[15] which, unlike the widely used concept Evolutionary Stable Strategy that refers to a constant stage game (Friedman, 1998), assures stability of F in an equilibrium.

Definition 1 *A strategy state $x \in \Sigma_{n-1}$ is an evolutionary equilibrium of an evolutionary game if x is an attractor*[16] *of the dynamical system $\dot{x} = F(x)$ defining the game's adjustment dynamics.*

What is the interpretation of our equilibrium concept's definition for evolutionary games? By solving equation (3), we can describe which strategies will be imitated more or less frequently for every strategy state in the state space. Thus, we can calculate which strategy frequencies a population exhibits over time after having been in a certain strategy state. Such a "solution path" for a given initial strategy state is called a solution trajectory. An evolutionary equilibrium is a subset of state space Σ_{n-1} which a solution trajectory does not leave once reached. Additionally, if a solution trajectory of the dynamics starts sufficiently close to the evolutionary equilibrium, it remains close and converges asymptotically to the evolutionary equilibrium over time. The open set of points in Σ_{n-1} converging to a given EE are called its basin of attraction.

[14]The replicator dynamics are simplex invariant:

$$\begin{aligned}\sum_{i\in S}\dot{x}_i &= \sum_{i\in S} x_i\left(f(\sigma_i,x)-f(x,x)\right)=\sum_{i\in S} x_i f(\sigma_i,x)-f(x,x)\sum_{i\in S} x_i\\ &= f(x,x)-f(x,x)=0\,.\end{aligned}$$

[15]The term evolutionary equilibrium was introduced by Hirshleifer (1982).

[16]An attractor is defined as an asymptotically stable non-wandering set (in our case the only possible non-wandering sets are critical points and points on limit cycles or graphics). For definitions of critical points (also called equilibrium points or fixed points), limit cycles, graphics (also called separatrix cycles), asymptotic stability, and non-wandering sets see Perko (2000) or any textbook on dynamic systems.

3 Endogenous Quality of Institutions

In Section 3.1 we include the feedback between corruption and the individual incentive to corrupt by modelling quality of institutions as an endogenous variable. In Section 3.2 we present our main result and discuss them in Section 3.3. Finally, we discuss the corruption dynamics when the feedback effect between growth and institutional quality is included in the corruption game.

3.1 The Strategy State Dependent Payoff Matrix

We change the stage game in two ways: First, we model the government wage as an endogenous variable of tax revenue. Second, we define the costs of corruption as a function of the strategy state. In order to do so, we write each component of the payoff matrix (1) as a function of the strategy state x. It is obvious that these extensions depart from standard evolutionary game theory. We now actually analyze a frequency dependent evolutionary game (Bruegger, 2005).

Let us assume that government wage is a function of tax revenue (and population income therefore), and that the government budget is financed through proportional taxes. Furthermore, we assume that the government's budget is balanced at any point in time and that all legally earned payoffs are subject to taxes. The tax rate is denoted by τ.

According to that, net income equals $(1-\tau)w(x)$ for the government employees and $(1-\tau)s(x)$ for the private entrepreneurs. The costs of corruption are specified below. Hence, we rewrite the payoff matrix of the corruption game as

$$A(x) = \begin{pmatrix} (1-\tau)w(x) & (1-\tau)w(x) & (1-\tau)w(x) \\ (1-\tau)w(x) & (1-\tau)w(x) & (1-\tau)w(x)+s(x)-c(x) \\ (1-\tau)s(x) & 0 & (1-\tau)s(x) \end{pmatrix} . \tag{4}$$

The function $c(x)$ depicts the expected costs of corrupting a private entrepreneur and absorbing his surplus $s(x)$ from private economic activity. The probability at which a corrupt activity is detected is denoted by $p(x)$, $p(x) \subseteq [0,1]$. If corrupt actions are detected with a high (low) probability we say that institutional quality is high (low). We assume that if a corrupt government employee is caught red-handed, he is punished by having drawn off his net income, i.e.

$$c(x) \quad = \quad p(x)\left((1-\tau)w(x)+s(x)\right) .$$

Note that a corrupt government employee encountering a private entrepreneur obtains the payoff $w(x) + s(x)$ if $p(x) = 0$ and obtains a zero payoff if $p(x) = 1$. We leave detection probability $p(x)$ unspecified for the moment and describe the corruption game's dynamics for a general function $p(x)$. Later we will discuss the game with a specification of the function $p(x)$. For now we make the following assumption.

Assumption 1 *An increase in the share of corrupt government employees decreases the probability that a corrupt act is detected, i.e.,*

$$\frac{\partial p}{\partial x_2}(x) < 0\,.$$

Further an increase in the share of corrupt government employees decreases the detection probability by more than an increase in the share of fair government employees, i.e.,

$$\frac{\partial p}{\partial x_2}(x) < \frac{\partial p}{\partial x_1}(x)\,.$$

Note that $\frac{\partial p}{\partial x_1}(x)$ can be positive or negative. Assumption 1 includes the feedback between institutional quality and corruption in our model.

Though it is possible to analyze the game with $w(x)$, $s(x)$, and $c(x)$ being functions of x, we choose to normalize $s(x) = 1$. By that normalization we abandon the possibility of a frequency dependent surplus from private activity. Though we are aware of interesting specifications of $s(x)$, we abstain from including such an effect. The reason is that we model the costs of corruption for the private sector in the payoff matrix directly, and do not want to lessen the explanatory power of the model by including further effects. Note that by setting $s(x) = 1$, government activities are of no direct utility for the private sector. But of course we can choose a $p(x)$ such that the share of fair government employees influences the costs of corruption.

These assumptions allow us to calculate the government wage explicitly. Tax revenue $r(x)$ is given by

$$\begin{aligned} r(x) &= \tau x^T \begin{pmatrix} w(x) & w(x) & w(x) \\ w(x) & w(x) & w(x) \\ 1 & 0 & 1 \end{pmatrix} x \\ &= \tau\left((x_1 + x_2)w(x) + (1 - x_1 - x_2)(1 - x_2)\right)\,. \end{aligned}$$

As mentioned above, the government is not able to tax the transfers from the private entrepreneur to the corrupt government employee. Thus, only legal payoffs are enlisted in the matrix used for tax revenue calculation. Now we can explicitly compute $w(x)$ using the budget constraint of the government:

$$\begin{aligned} w(x) &= \frac{r(x)}{x_1 + x_2}, \\ &= \frac{\tau\left((x_1 + x_2)w(x) + (1 - x_1 - x_2)(1 - x_2)\right)}{x_1 + x_2}. \end{aligned}$$

By solving for government salary we receive

$$w(x) = \frac{\tau}{1-\tau}\frac{(1 - x_1 - x_2)(1 - x_2)}{x_1 + x_2}.$$

Plugging the explicit expression for $w(x)$ into the definition of $r(x)$ yields

$$r(x) = \frac{\tau}{1-\tau}(1 - x_1 - x_2)(1 - x_2).$$

This allows us to evaluate how the government wage and the tax revenue depend on x_1 and x_2.

$$\frac{\partial w(x)}{\partial x_2} < \frac{\partial w(x)}{\partial x_1} < 0 \quad \text{and} \quad \frac{\partial r(x)}{\partial x_2} < \frac{\partial r(x)}{\partial x_1} < 0.$$

It is not surprising that the derivations of the government wage with respect to x_1 and x_2 are negative. Firstly, if either x_1 or x_2 increases, the frequency of private entrepreneurs decreases. However, this fraction of the population is solely responsible for the contribution of value to $r(x)$, since government employees only pay tax on their wages which are a fraction of $r(x)$; hence the government employees cannot contribute to tax revenue positively. Secondly, the higher the number of government employees is, the lower is government salary for a given $r(x)$. The reason is our assumption of a balanced government budget: if x_1 or x_2 increases, $r(x)$ has to be split among more employees, so each gets a smaller wage. We also observe that the derivation of government wage with respect to x_2 is smaller than the one with respect to x_1. The rationale for this is the following. The more corrupt government employees there are, the higher is the share of games played among corrupt employees and private entrepreneurs. This implies that more of the entrepreneurs' surpluses flow outside the taxation system because they become illegal income from corruption. This reduces $r(x)$ and therewith $w(x)$.

3.2 The Main Result - The Dynamics of Corruption

In the last subsection we specified the functions in the payoff matrix of the corruption game, except for $p(x)$ which we want to treat generally. We can rewrite the payoff matrix (4) as

$$A(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 - c(x) \\ (1-\tau)(1-w(x)) & -(1-\tau)w(x) & (1-\tau)(1-w(x)) \end{pmatrix} \tag{5}$$

without changing its dynamics (Bruegger, 2005). The replicator dynamics for frequency dependent evolutionary game yields a system of three differential equations for the corruption game. We drop the third equation for $\dot{x}_3$ and substitute x_3 by $1 - x_1 - x_2$. This leaves us with the planar system

$$\begin{aligned} \dot{x}_1 &= x_1(1 - x_1 - x_2)\left[(1-\tau)(w(x) - 1) - x_2(\tau - c(x))\right] \\ \dot{x}_2 &= x_2(1 - x_1 - x_2)\left[(1-\tau)w(x) + (1 - x_2)(\tau - c(x))\right] . \end{aligned} \tag{6}$$

We describe the dynamics of corruption in the following proposition.

Proposition 1 *Under Assumption 1, the corruption game can only have the following three critical points as evolutionary equilibria (EE):*

- $(x_1, x_2) = (\tau, 0)$ *is an EE if* $\frac{1}{2-\tau} < p(\tau, 0)$,
- $(x_1, x_2) = (0, 1)$ *is an EE if* $p(0, 1) < \tau$,
- $(x_1, x_2) = (0, \bar{x}_2)$ *exists as a critical point if there exists an* $\bar{x}_2$ *satisfying* $p(0, \bar{x}_2) = \frac{\tau}{(1-\bar{x}_2)^2\tau + \bar{x}_2}$ *and is an EE if* $-\frac{\partial c}{\partial x_2}(0, \bar{x}_2) < \frac{\tau}{\bar{x}_2^2}$.

There always exists at least one evolutionary equilibrium.

According to Proposition 1, the corruption game can have seven different combinations of evolutionary equilibria. We show these in Figure 1: The bottom left vertex of the simplex represents the strategy state $(1, 0)$ (only fair government employees), the bottom right vertex the strategy state $(0, 1)$ (only corrupt government employees), and the top vertex the strategy state $(0, 0)$ (only private entrepreneurs).

The proof of Proposition 1 is given in Appendix A on the pages 26-39.

3.3 Interpreting the Main Result

Our first observation is, that the dynamics of corruption converge. We find three strategy states that can be evolutionary equilibria under our assumptions. We discuss them in turn.

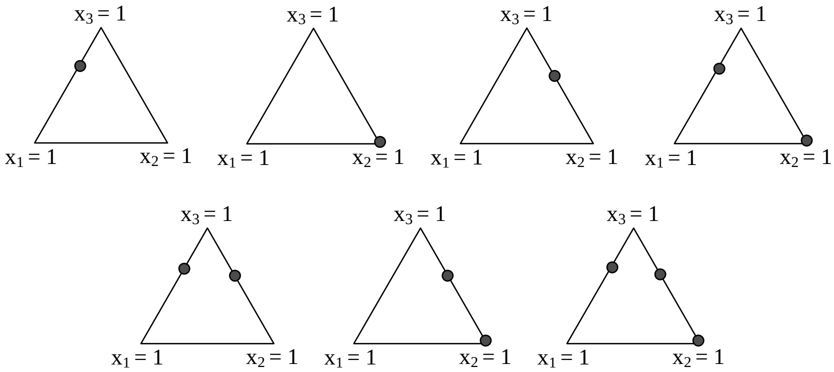

Figure 1: The possible combinations of evolutionary equilibria.

3.3.1 The Clean Equilibrium

The critical point $(\tau, 0)$ always exists and is an evolutionary equilibrium if the detection probability of corrupt behavior is sufficiently high in $(\tau, 0)$. We refer to this EE as the *clean equilibrium* because all corrupt activity is crowded out by private activity and fair government service. The higher the tax rate, the higher the share of government employees in this equilibrium. The reason for this result is the way we defined government wage: Since we distribute all tax revenue equally among the government employees, their share will rise until the government salary equals the net return from private activity.

We know that a clean equilibrium exists, if the function $p(x)$, the society's detection probability, takes a high values for $x_2 \to 0$. We conclude that for a society with a judiciary functioning efficiently at a low corruption rate, a clean equilibrium exists. This is a slightly optimistic finding: As long as institutions detect corruption well for low corruption rates, at least some initial strategy states converge towards the clean equilibrium.

Note that a population may have a clean equilibrium for some tax rates, but not for others. Let us briefly discuss the condition for the clean equilibrium, $\frac{1}{2-\tau} < p(\tau, 0)$. The left hand side of the inequality is an increasing function in τ, $p(\tau, 0)$ as a function of τ, can actually lie above, below or intersect with it. If $p(\tau, 0)$ lies above (below), the clean equilibrium exists for all tax rates (does not exist for any tax rate). If the two function intersect and $p(\tau, 0)$ is the steeper (flatter) function at the intersection point, then the clean equilibrium only exists for tax rates above (below) a certain threshold. Of course we can have several threshold and therefore intervals of tax rates in which the clean equilibrium exists.

3.3.2 The Corrupt Equilibrium - The Corruption Trap

The second critical point, $(0,1)$, also exists independently of the specification of $p(x)$. Furthermore, it is an evolutionary equilibrium if the function $p(x)$ takes a value low enough in $(0,1)$. In this case we speak of a *corrupt equilibrium* or a *corruption trap* because in this equilibrium all agents choose to be corrupt government employees. A society's detection probability $p(x)$ takes low values for a high corruption rate if corruption badly affects a society's judiciary or its implementation. As soon as elements of a judiciary that are responsible for the detection of corruption can be corrupted, we expect that the society can be trapped in the corruption equilibrium for at least some initial strategy states.[17]

Of course the tax rate plays an important role for the existence of the corruption trap. The higher the tax rate is, the higher can $p(0,1)$ be such that the corruption trap still exists. So one way to escape from a path towards the corruption trap is to cut taxes. We will elaborate on this possibility further below.

3.3.3 The Hybrid Equilibrium

The last critical point, $(0, \bar{x}_2)$ does not exist for all τ and all $p(x)$. The right hand side of the condition for its existence is a function which can become greater than 1 for big τ, so this equilibrium exists for small and moderate tax rates only.

If $(0, \bar{x}_2)$ exists and if the costs of corruption are not decreasing too strongly in x_2, then it is an EE of the corruption game. We refer to it as the *hybrid equilibrium*

[17]The corrupt equilibrium may appear unlikely at first glance, because all players choose to be government employed although government wage converges to zero. We offer two arguments in favor of this evolutionary equilibrium. The first is an example: During Carlos Menem's last term of office as President of Argentina, 70% (!) of the labor force was employed by the local governments in many of the provinces (namely Chaco, Tucumán, La Rioja, and others). The majority of government employees barely worked yet collected their salary in the end of the month. At that time, Argentina certainly fulfilled the condition for the existence of the corrupt equilibrium: although the corruption rate was high, almost nobody was convicted for corrupt activities. Only 5% of Argentineans reported that they would seek judicial help in case of severe problems. The efficiency of the judiciary was too low for people to bother reporting corruption (and other crimes). The example is taken from TI's Daily Corruption News Service, http://www.transparency.org/cgi-bin/dcn-read.pl?citID=35148. The second argument is a theoretical one. The dynamics converge at an extremely low speed towards $(0,1)$, so our model does not actually suggest any observations of $(0,1)$ or its close vicinity.

because the government service is entirely corrupt but does not suppress private economic activity. If a society's detection probability $p(x)$ does not depend strongly on the corruption rate and is neither very high nor very low, then we expect the hybrid equilibrium to exist. Consequently, a society whose judiciary is not too strong but also not too badly affected by corruption may have a hybrid equilibrium and can converge to it at least for some initial strategy states.

3.3.4 Discussion

The stability conditions for the three critical points are independent of each other. Therefore the specification of $p(x)$ and the tax rate τ may imply any combination of the three evolutionary equilibria for the corruption game. By definition the EE are the attractors of (6), and since Proposition 1 denotes *all* EE, every trajectory through an initial strategy state $x_0 \in \Sigma_2$ of system (6) converges to one of the EE.[18] If two or three EE exist for a given $p(x)$ and τ respectively, then the initial state is decisive for the EE the population converges to. The sets of initial strategy states that converge to a certain evolutionary equilibrium are called the evolutionary equilibria's *basins of attraction.* We now discuss three implications of Proposition 1 for which we do not have to calculate the basins of attraction of the EE.

First, the corruption trap has the lowest population income. Reducing the tax rate can eliminate the corruption trap. The intuition for this is that the net payoff for private activity is enhanced and the government wage reduced, such that the return as a corrupt government employee is decreased. However, reducing the tax rate can also eliminate the clean equilibrium, since the institutions could be weakened depending on how institutional quality depends on the share of fair government employees. As a consequence, corruption could then not be suppressed.

Second, the model does not require that the quality of the institutions must be high at any point on a path towards the clean equilibrium. If corruption is high and therefore the quality of institutions low, a country can still converge to the clean equilibrium for some initial strategy states.

Third, a population with institutions which are not very efficient but also not vulnerable to corruption can converge to a hybrid equilibrium. Without a specific function

[18] Note that the separatrices are exceptions. The separatrices of a dynamical system are those trajectories that approach a saddle point in the limit. The corruption game can have two saddle points at most; these are $(0, \bar{x}_2)$ and $(\tau - \hat{x}_2, \hat{x}_2)$ (see Proof of Proposition 1). So we have at most two separatrices within the simplex.

for $p(x)$ it is not possible to say if the population income is higher in the hybrid than the clean equilibrium.

What does Proposition 1 imply for empirical research? Cross section data may actually contain countries that are developing along solution trajectories that converge to distinct equilibria. If this is the case, our estimation of the impact of corruption on growth will not be meaningful. For any selection of data points in the simplex containing points converging to different equilibria, an estimation of the impact of corruption on growth is far too optimistic for those countries that converge to the corrupt equilibrium: While corruption will have no long term effect for those on the trajectory to the clean equilibrium, those that converge to the corrupt one suffer big losses in income from corruption. We can now understand how a group of Asian countries can show a positive correlation between corruption and growth rate: They either converged to the hybrid equilibrium or follow a trajectory that show increasing share of private and corrupt activity before converging in one of the three equilibria. We conclude that a meaningful estimation of the impact of corruption can be obtained if the sample shows the same convergence behavior. To determine the equilibrium a country converges to is a very challenging task and we need to gain further insights about the role of institutions in a dynamic setting. The reason is that only with a specific description of how the quality of institutions depends on growth and corruption rate, we will be able to determine the basins of attractions of the different equilibria. The description of the basins of attractions would then allow us to classify countries according to their convergence behavior more reliably.

The type of institutions we have studied up to now is characterized by Assumption 1: The more corruption there is, the lower the probability that a corrupt activity is detected. In the next section, we consider another type of institutions which quality also depends on the population income. We specify the detection probability function explicitly and show how the basins of attraction can be determined. In order to describe each equilibrium's basin of attraction formally, the system would have to be solved explicitly. As it is the case for many nonlinear differential equation systems, this is not possible. Instead, we provide an example of a function $p(x)$ and describe the global behavior of the corruption game by numerical simulations.

3.4 The Feedback Effect between Income and Quality of Institutions

In our model, the legal population income is

$$l(x_1, x_2) \quad = \quad \frac{(1 - x_1 - x_2)(1 - x_2)}{1 - \tau}\,. \tag{7}$$

The higher the income of a population is, the more valuable are property rights. Consequently, populations are disposed to provide more resources to protect property rights the richer they are. Let's assume that players have the possibility to protect their property rights, for instance by establishing a judiciary that is independent of the government. We then suggest that the efficiency of the judiciary depends on the amount of financial resources available. Therefore it seems reasonable to assume that $p(x)$ depends positively on income $l(x)$.

For simplicity, we implement $p(x)$ as a linear function of $l(x)$. Since $p(x)$ is a probability it can only take values in $[0, 1]$. Consequently, we define it as

$$p(x) = (1 - \tau)l(x) = (1 - x_1 - x_2)(1 - x_2)\,.$$

Note that $\frac{\partial p}{\partial x_2} < \frac{\partial p}{\partial x_1} < 0$. An increasing share of corrupt as well as fair government employees decreases $p(x)$. The reason is that only private activities generate income where government employees are financed over taxes and do not contribute to population income.

The clean equilibrium exists for $p(\tau, 0) > \frac{1}{2-\tau}$, i.e. for $\tau < \frac{1}{2}(3 - \sqrt{5}) = 0.38$. The corrupt equilibrium exists for $p(0, 1) < \tau$, which is always satisfied because $p(0, 1) = 0$. The critical point $(0, \bar{x}_2)$ is not an attractor for $\tau < 0.25$, because condition (14) is then violated; and it does not exist for other τ because (8) does not have a solution in $[0, 1]$ for $\tau > 0.25$. So the hybrid equilibrium does not exist.

According to these results, the model predicts the following for a society in which the detection probability of corrupt activity depends positively on income: If tax burden is moderate, it either converges to the clean or the corrupt equilibrium, depending on the initial strategy state. If taxes are very high, it converges to the corrupt equilibrium.

We display the solution trajectories in Figure 2, they are calculated with the tax rates $\tau = \{0.1; 0.15; 0.2; 0.3; 0.5; 0.8\}$.

As we see in Figure 2, the basin of attraction of the corrupt equilibrium broadens with τ increasing. The reason is that a high τ reduces the incentive of private activity because of a smaller relative payoff; firstly because more taxes have to be paid and secondly because government wage is higher. This reduced incentive for private activity

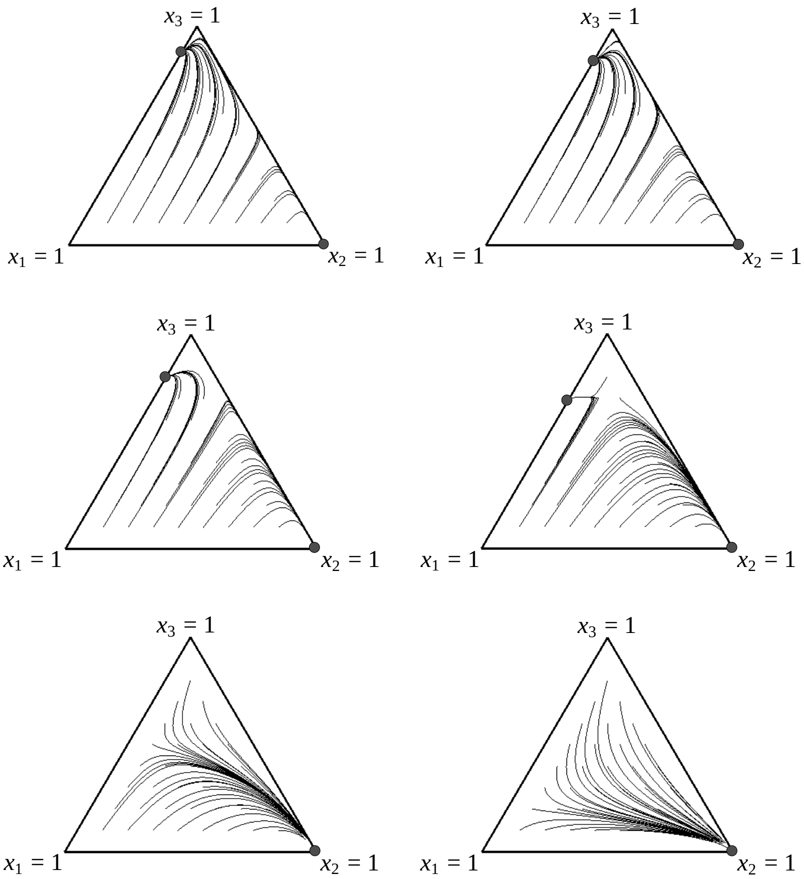

Figure 2: The dynamics of corruption with feedback effect.

is then responsible for a decrease in the population income, which itself increases the payoff of a corrupt government employee. So under a higher tax rate, initial strategy states with a lower share of corrupt government employees will converge towards the corruption trap.

How can we interpret the shape of the basins of attraction? The corruption rate equals $\frac{x_2}{x_1+x_2}$ in our model, which are straight lines from $(0, 0)$ to points of the opposite edge of the simplex. So for a given tax rate, a population with a lean government may start off with a high corruption rate and still converge to the clean equilibrium. Contrarily, a population with a large government that starts off with the same corruption rate may converge to the the corrupt equilibrium.

This example of a specific function $p(x)$ suggests that understanding the quality of institutions under different circumstances is crucial for the analysis of corruption: Only then are we able to say to which equilibrium a population converges from the current strategy state. This knowledge then allows for reliable empirical consideration.

4 Conclusions

To our knowledge, this is the first attempt to include feedback effects from population variables such as corruption and growth rates to the individual decision to corrupt. Most importantly we show that such an analysis requires the quality of institutions to be an endogenous variable. We find that, for a general characterization of the feedback effects, a population can converge to different equilibria which leads to path dependence for the behavior of a population. Consequently, populations with distinct initial strategy frequencies may converge to different equilibria in the long run. However, not only may they follow different solution paths, but also may they feature different equilibria they can converge to. Depending on how the ability of institutions to cope with corruption under different growth and corruption rates changes, the set of equilibria actually changes.

These results have important implications for empirical research: Only if all countries in a sample converge to the same equilibrium, we expect findings with a high explanatory power. If countries converge to different equilibria, a cross country analysis is prone to underestimate the implications of corruption for those countries converging to a corruption trap. Similarly, the consequences of corruption are overestimated for those countries converging to a clean equilibrium.

In order to classify countries according to their convergence behavior, a deeper understanding of how institutions and institutional quality depend on the dynamic process they partly shape. We hope that our results stimulates economic research on this topic.

We think that the framework of an evolutionary game has served for the purposes well to make the point we intended to discuss. It has allowed us to look at aspects like feedback effects that can only be analyzed in dynamic setups. However, in order to keep dynamics tractable and still allow for individual decisions, the decision rules we have to impose on the agents are bound to strong assumptions that apply to all agents in the game. Further, the dynamics become too involved to allow for a more sophisticated comprehension of the private sector.

References

Ades, A. and R. Di Tella (1997). National champions and corruption: Some unpleasant interventionist arithmetic. *The Economic Journal 107*(443), 1023–42.

Andvig, J. C. and K. O. Moene (1990). How corruption may corrupt. *Journal of Economic Behavior and Organization 13*(1), 63–76.

Bruegger, E. (2005). Replicator dynamics with frequency dependent stage games. Working Paper.

Brunetti, A., G. Kisunko, and B. Weder (1998). Credibility of rules and economic growth: Evidence from a worldwide survey of the private sector. *World Bank Economic Review 12*(3), 353–384.

Buscaglia, E. (2001). An analysis of judicial corruption and its causes: An objective governing-based approach. *International Review of Law and Economics 21*(2), 233–249.

Chakrabarti, R. (2001). Corruption: A general equilibrium approach. Georgia Institute of Technology.

Chong, A. and C. Calderon (2000). Causality and feedback between institutional measures and economic growth. *Economics and Politics 12*(1), 69–81.

Dollar, D. and A. Kraay (2003). Institutions, trade, and growth. *Journal of Monetary Economics 50*(1), 133–162.

Eggertsson, T. (1990). *Economic Behavior and Institutions.* Cambridge University Press.

Friedman, D. (1998). On economic applications of evolutionary game theory. *Journal of Evolutionary Economics 8*(1), 15–43.

Gupta, S., H. Davoodi, and R. Alonso-Terme (2002). Does corruption affect income inequality and poverty? *Economics of Governance 3*(1), 23–45.

Hindriks, J., M. Keen, and A. Muthoo (1999). Corruption, extortion and evasion. *Journal of Public Economics 74*(3), 395–430.

Hirshleifer, J. (1982). Evolutionary models in economics and law: Cooperative versus conflict strategies. *Research in Law and Economis 4*, 1–60.

Hofstaedter, R. (1973). *The American Political Tradition and the Men Who Made It.* Random House.

Husted, B. W. (1999). Wealth, culture, and corruption. *Journal of International Business Studies 30*(2), 339–359.

Johnston, M. (1989). Corruption, inequality, and change. In P. M. Ward (Ed.), *Corruption, Development and Inequality: Soft Touch or Hard Graft?* Routledge.

Kaufmann, D., A. Kraay, and P. Zoido-Lobatón (2000). Governance matters: From measurement to action. *Finance and Development 37*(2), 10–13.

Keefer, P. and S. Knack (1997). Why don't poor countries catch up? a cross-national test of institutional explanation. *Economic Inquiry 35*(3), 590–602.

Knack, S. and P. Keefer (1995). Institutions and economic performance: Cross-country tests using alternative institutional measures. *Economics and Politics 7*(3), 207–227.

Lim, L. Y. C. and A. Stern (2002). State power and private profit: The political economy of corruption in southeast asia. *Asian Pacific Economic Literature 16*(2), 18–52.

Li, H., L. C. Xu, and H. F. Zou (2000). Corruption, income distribution, and growth. *Economics and Politics 12*(2), 155–182.

Mauro, P. (1995). Corruption and growth. *Quarterly Journal of Economics 110*(3), 681–712.

Mauro, P. (1997). The effects of corruption on growth, investment, and government expenditure: A cross-country analysis. In K. A. Elliott (Ed.), *Corruption and the Global Economy*, Chapter 4, pp. 83–107. Washington, D.C.: Institute for International Economics.

Mauro, P. (1998). Corruption and the composition of government expenditure. *Journal of Public Economics 69*(2), 263–279.

Murphy, K. M., A. Shleifer, and R. W. Vishny (1991). The allocation of talent: Implications for growth. *American Economic Review 106*(2), 503–530.

Murphy, K. M., A. Shleifer, and R. W. Vishny (1993). Why is rent-seeking so costly to growth? *American Economic Review 83*(2), 409–414.

Niskanen, W. A. (1996). *Bureaucracy and Public Economics.* Edward Elgar Pub.

North, D. C. (1990). *Institutions, Institutional Change and Economic Performance.* Cambridge University Press.

Paldam, M. (2002). The cross-country pattern of corruption: Economics, culture and the seesaw dynamics. *European Journal of Political Economy 18*(2), 215–240.

Perko, L. (2000). *Differential Equations and Dynamical Systems.* Springer.

Rock, M. T. and H. Bonnett (2004). The comparative politics of corruption: Accounting for the east asian paradox in empirical studies of corruption, growth and investment. *World Development 32*(6), 999–1017.

Rose-Ackerman, S. (1999). *Corruption and Government.* Cambridge University Press.

Schuster, P. and K. Sigmund (1983). Replicator dynamics. *Journal of Theoretical Biology 100*, 533–538.

Strogatz, S. H. (1994). *Nonlinear Dynamics and Chaos.* Westview Press, Perseus Books Group.

Tanzi, V. and H. R. Davoodi (2000). Corruption, growth, and public finances. International Monetary Fund Working Paper: WP/00/182 November 2000; 26.

Tanzi, V. (2002). Corruption around the world. *Applied Econometrics and International Development 2*(1), 110–115.

Taylor, P. D. and L. B. Jonker (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences 40*, 145–156.

Wedeman, A. (2004). Great disorder under heaven: Endemic corruption and rapid growth in contemporary china. *China Review 4*(2), 1–32.

Weibull, J. W. (1995). *Evolutionary Game Theory.* The MIT Press.

Wei, S. (2000). How taxing is corruption on international investors? *Review of Economics and Statistics 82*(1), 1–11.

A Appendix: Proofs of Propositions

A.1 Proof of Proposition 1

We accomplish this proof in two parts. First, we apply the local theory of nonlinear systems and second, we use theorems of global theory of nonlinear systems to show that we have found all attractors of (6).

A.1.1 Local Theory of Nonlinear Systems

From the conditions for the existence and uniqueness of a solution in a frequency dependent evolutionary game (Bruegger, 2005) and the differentiability (and therefore Lipschitz-continuity) of the functions $w(x)$ and $c(x)$ we know that the differential equation system (6) has a unique solution. We redefine $F : \mathbb{R} \to \mathbb{R}$ as the right hand side of system (6).

$$\begin{aligned} F(x) &= \begin{pmatrix} F_1(x) \\ F_2(x) \end{pmatrix} \\ &= \begin{pmatrix} x_1(1-x_1-x_2)\,[(1-\tau)(w(x)-1) - x_2(\tau - c(x))] \\ x_2(1-x_1-x_2)\,[(1-\tau)w(x) + (1-x_2)(\tau - c(x))] \end{pmatrix} . \end{aligned}$$

In order to find the critical points of the system, we set $F(x) = 0$:

$$\begin{aligned} x_1(1-x_1-x_2)\,[(1-\tau)(w(x)-1) - x_2(\tau - c(x))] &= 0 \\ x_2(1-x_1-x_2)\,[(1-\tau)w(x) + (1-x_2)(\tau - c(x))] &= 0\,. \end{aligned}$$

There are nine possibilities which are solution candidates for this equation system.

1) $x_1 = 0 \;\wedge\; x_2 = 0$
2) $x_1 = 0 \;\wedge\; (1-x_1-x_2) = 0$
3) $x_1 = 0 \;\wedge\; (1-\tau)w(x) + (1-x_2)(\tau - c(x)) = 0$
4) $(1-x_1-x_2) = 0 \;\wedge\; x_2 = 0$
5) $(1-x_1-x_2) = 0 \;\wedge\; (1-x_1-x_2) = 0$
6) $(1-x_1-x_2) = 0 \;\wedge\; (1-\tau)w(x) + (1-x_2)(\tau - c(x)) = 0$
7) $(1-\tau)(w(x)-1) - x_2(\tau - c(x)) = 0 \;\wedge\; x_2 = 0$

$$
\begin{aligned}
&8) \qquad (1-\tau)(w(x)-1) - x_2(\tau - c(x)) = 0 \;\;\wedge\;\; (1-x_1-x_2) = 0 \\
&9) \qquad (1-\tau)(w(x)-1) - x_2(\tau - c(x)) = 0 \\
&\qquad\quad \wedge\;\; (1-\tau)w(x) + (1-x_2)(\tau - c(x)) = 0\,.
\end{aligned}
$$

Conditions 1), 2), and 4) state that the vertices of the simplex are fixed points.

Condition 5) gives us the edge of the simplex where $x_3 = 0$ as a set of critical points, conditions 6) and 8) give two single points in this set as fixed points.

Condition 3) gives us a critical point on the edge of the simplex where $x_1 = 0$. It only exists if there is a solution $\bar{x}_2$ which satisfies

$$(1-\tau)w(0,\bar{x}_2) + (1-\bar{x}_2)(\tau - c(0,\bar{x}_2)) = 0\,. \tag{8}$$

By plugging

$$w(0,\bar{x}_2) \;=\; \frac{\tau}{1-\tau}\frac{(1-\bar{x}_2)^2}{\bar{x}_2}$$

into (8), we receive

$$c(0,\bar{x}_2) = \frac{\tau}{\bar{x}_2}\,. \tag{9}$$

We substitute the expression for $c(0,\bar{x}_2)$ given by (11) into the last equation and get

$$p(0,\bar{x}_2) \;=\; \frac{\tau}{\tau(1-\bar{x}_2)^2 + \bar{x}_2}\,.$$

So the potential critical point $(0,\bar{x}_2)$ only exists if $p(0,x_2)$ intersects with the function $g(x_2) = \frac{\tau}{\tau(1-x_2)^2+x_2}$ at least once on $]0,1[$ for a given τ. The function $g(x_2)$ is strictly increasing in the parameter τ and can take values that are greater than one. We conclude that for high τ, an $\bar{x}_2$ will not exist. The function $g(x_2)$ is decreasing in x_2 when its image is in the interval $]0,1[$.

Condition 7) implies

$$
\begin{aligned}
(1-\tau)(w(x_1,0)-1) &= 0 \\
w(x_1,0) &= 1 \\
\frac{\tau}{1-\tau}\frac{1-x_1}{x_1} &= 1 \quad \Rightarrow \quad x_1 = \tau\,,
\end{aligned}
$$

so $(\tau, 0)$ is a fixed point.

Finally, condition 9) supports a critical point if there is a solution to

$$\begin{aligned}(1-\tau)(w(x_1,x_2)-1)-x_2(\tau-c(x_1,x_2)) &= 0\\(1-\tau)w(x_1,x_2)+(1-x_2)(\tau-c(x_1,x_2)) &= 0\,.\end{aligned}$$

We rewrite the system as

$$\begin{aligned}(1-\tau)w(x_1,x_2)-(1-\tau)-x_2(\tau-c(x_1,x_2)) &= 0\\(1-\tau)w(x_1,x_2)+(\tau-c(x_1,x_2))-x_2(\tau-c(x_1,x_2)) &= 0\end{aligned}$$

and find $c(x_1,x_2)=1$ by subtracting the two equations. Plugging this result into the first of the equations of the system gives

$$\begin{aligned}(1-\tau)(w(x_1,x_2)-1)-x_2(\tau-1) &= 0\\(w(x_1,x_2)-1)+x_2 &= 0 \quad\rightarrow\quad w(x_1,x_2)=1-x_2\,.\end{aligned}$$

We now use the explicit expression for $w(x_1,x_2)$ and find

$$\begin{aligned}w(x_1,x_2)=1-x_2 \quad\rightarrow\quad \frac{\tau}{1-\tau}\frac{(1-x_1-x_2)(1-x_2)}{x_1+x_2} &= 1-x_2\\\frac{\tau}{1-\tau}(1-x_1-x_2) &= x_1+x_2\\\tau &= x_1+x_2\,.\end{aligned}$$

Let us denote $\hat{x}_2$ as the solution of $c(\tau-\hat{x}_2,\hat{x}_2)=1$, and we derive further

$$\begin{aligned}c(\tau-\hat{x}_2,\hat{x}_2) &= 1\\p(\tau-\hat{x}_2,\hat{x}_2)((1-\tau)(1-\hat{x}_2)+1) &= 1\\p(\tau-\hat{x}_2,\hat{x}_2) &= \frac{1}{(1-\tau)(1-\hat{x}_2)+1}\,.\end{aligned}$$

So, only if we assume a $p(x_1,x_2)$ such that $p(\tau-\hat{x}_2,\hat{x}_2)$ intersects with a function $h(x_2)$,

$$h(x_2) = \frac{1}{(1-\tau)(1-x_2)+1}\,,$$

on $[0,\tau]$, there exists a fixed point $(\tau-\hat{x}_2,\hat{x}_2)$. Note that $h(x_2)$ is increasing in τ and increasing in x_2. The image of the function $h(x_1)$ is in $[\frac{1}{2},1]$.

We enlist the critical points we have found in the first column of Table 1.

In a next step we want to analyze the stability of the critical points we have found above.

If a critical point x_0 is hyperbolic,[19] it is either a sink,[20] a saddle,[21] or a source[22] (Definitions e.g.by Perko, 2000, p.102). It follows from the Hartman-Grobman Theorem[23] that sinks of a differential equation system are asymptotically stable and sources and saddles are unstable. So in order to determine if a hyperbolic critical point is asymptotically stable (and thus an EE) or not we only need to calculate the eigenvalues of the Jacobian of $F(x)$ evaluated at the critical point. Therefore, we first calculate the elements of the Jacobian $DF(x_1, x_2)$.

$$\begin{aligned}
\frac{\partial F_1}{\partial x_1} &= (1-2x_1-x_2)\Big((1-\tau)(w(x)-1)-x_2(\tau-c(x))\Big) \\
&\quad + x_1(1-x_1-x_2)\left((1-\tau)\frac{\partial w(x)}{\partial x_1}+x_2\frac{\partial c(x)}{\partial x_1}\right) \\
\frac{\partial F_1}{\partial x_2} &= -x_1\Big((1-\tau)(w(x)-1)-x_2(\tau-c(x))\Big) \\
&\quad + x_1(1-x_1-x_2)\left((1-\tau)\frac{\partial w(x)}{\partial x_2}-(\tau-c(x))+x_2\frac{\partial c(x)}{\partial x_2}\right) \\
\frac{\partial F_2}{\partial x_1} &= -x_2\Big((1-\tau)w(x)+(1-x_2)(\tau-c(x))\Big) \\
&\quad + x_2(1-x_1-x_2)\left((1-\tau)\frac{\partial w(x)}{\partial x_1}-(1-x_2)\frac{\partial c(x)}{\partial x_1}\right) \\
\frac{\partial F_2}{\partial x_2} &= (1-x_1-2x_2)\Big((1-\tau)w(x)+(1-x_2)(\tau-c(x))\Big) \\
&\quad + x_2(1-x_1-x_2)\left((1-\tau)\frac{\partial w(x)}{\partial x_2}-(\tau-c(x))-(1-x_2)\frac{\partial c(x)}{\partial x_2}\right).
\end{aligned}$$

For convenience we also restate

$$w(x) = \frac{\tau}{1-\tau}\frac{(1-x_1-x_2)(1-x_2)}{x_1+x_2} \quad \text{and} \tag{10}$$

$$c(x) = p(x)\Big((1-\tau)w(x)+1\Big) \quad \text{with} \tag{11}$$

$$\frac{\partial w}{\partial x_1} = -\frac{\tau}{1-\tau}\left(\frac{1-x_2}{(x_1+x_2)^2}\right) \tag{12}$$

[19] None of the eigenvalues of $DF(x_0)$ has a zero real part.

[20] All eigenvalues of $DF(x_0)$ have negative real parts.

[21] At least one eigenvalue of $DF(x_0)$ has a positive and at least one has a negative real part.

[22] All eigenvalues of $DF(x_0)$ have positive real parts.

[23] The Hartman-Grobman Theorem states that if F is differentiable then there exists a homeomorphism that maps the trajectories in an open set around a hyperbolic critical point x_0 onto trajectories near x_0 of the linear system $\dot{x} = Ax$ with $A = DF(x_0)$. That is to say that near a hyperbolic critical point x_0 the nonlinear system $\dot{x} = F(x)$ has the same qualitative structure as the linear system $\dot{x} = Ax$ with $A = DF(x_0)$.

$$\frac{\partial w}{\partial x_2} = \frac{\tau}{1-\tau}\left(1 - \frac{1+x_1}{(x_1+x_2)^2}\right). \quad (13)$$

We treat the critical points one by one in the order of Table 1.

Critical point $(\mathbf{0}, \mathbf{0})$: It is easy to see that $\frac{\partial F_1}{\partial x_2}(0,0) = 0$ and $\frac{\partial F_2}{\partial x_1}(0,0) = 0$. The eigenvalues of $DF(0,0)$ are thus $\frac{\partial F_1}{\partial x_1}(0,0)$ and $\frac{\partial F_2}{\partial x_2}(0,0)$.[24] We find

$$\begin{aligned}
\frac{\partial F_1}{\partial x_1}(0,0) &= (1-\tau)w(0,0) \\
\frac{\partial F_2}{\partial x_2}(0,0) &= (1-\tau)w(0,0)(1-p(0,0)) + \tau - p(0,0).
\end{aligned}$$

The expression $(1-\tau)w(0,0)$ is clearly positive, because of

$$\lim_{\substack{x_1 \to 0 \\ x_2 \to 0}} w(x_1, x_2) = \infty.$$

If $p(0,0)$ such that $\frac{\partial F_2}{\partial x_2}(0,0) \neq 0$, then the critical point $(0,0)$ is hyperbolic and one positive eigenvalue is enough to know that it is unstable (see e.g. Perko, 2000, Theorem 2, p. 130). If $p(0,0)$ such that $\frac{\partial F_2}{\partial x_2}(0,0) = 0$, then the second eigenvalue is equal to zero and $(0,0)$ is not hyperbolic. It is then an unstable node, a saddle, or a saddle-node and unstable therefore (see e.g. Perko, 2000, Theorem 1, p. 151).

Critical point $(\mathbf{0}, \mathbf{1})$: The critical point $(0,1)$ does always exist and is not hyperbolic. In fact, we even have $DF(0,1) = 0$, which indicates a very complex behavior of the system near the critical point. For most other critical points it is more convenient to analyze the behavior near them in $\mathbb{R}^2$ (although we are only interested in the dynamics on the simplex).[25] For showing asymptotic stability of $(0,1)$ however, we will restrict our analysis to the simplex, which makes our efforts more comprehensible in this case.

One method to show stability for critical points that are not hyperbolic is due to Liapunov. The theorem (see e.g. Perko, 2000, p. 131, Theorem 3) that states under which conditions the existence of a Liapunov function (defined below) implies (asymptotic) stability of a critical point only applies to critical points that are interior points of the definition space of $F(x)$. The critical point $(0,1)$ is a boundary point of the simplex. Therefore we have to apply an extended theorem proved in Bruegger (2005)which

[24] The eigenvalues of a matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ are equal to a and d if either $b = 0$ or $c = 0$ or both.

[25] One reason is that the Hartman-Grobman Theorem requires open subsets containing the hyperbolic critical points. Most of our critical points are on the boundary of the simplex, however.

states that the existence of a Liapunov function guarantees asymptotic stability for a boundary point of the simplex if the simplex is invariant under $\dot{x} = F(x)$.

Theorem 1 (Bruegger, 2005) *Let E be an open subset of Σ and $x_0 \in \overline{E}$.*[26] *Suppose that $F(x) \in C^1(\overline{E})$ and $F(x_0) = 0$, where the simplex is invariant under $\dot{x} = F(x)$. Suppose further that there exists a real valued function $V \in C^1(\overline{E})$ satisfying $V(x_0) = 0$ and $V(x) > 0$, $\forall\, x \in \overline{E} \setminus x_0$. If $\dot{V}(x) < 0$ $\forall\, x \in E$, x_0 is asymptotically stable.*

We now have to show that there exists a Liapunov function for system (6) as defined in Theorem 1. We will give evidence of existence by presenting an example: We show in the following that function $V(x)$,

$$V(x) \;=\; x_1^2 + (1 - x_2)^2 \,,$$

is a Liapunov function. It is clear that $V(x) > 0\; \forall x \in \mathbb{R}^2 \setminus (0,1)$, hence $V(x) > 0\; \forall x \in \Sigma_2$. Further $V(0,1) = 0$. Let us now look at $\dot{V}(x)$.

$$\begin{aligned}
\dot{V}(x) &= 2x_1\dot{x}_1 + 2(1-x_2)(-\dot{x}_2) \\
&= 2x_1^2(1-x_1-x_2)[(1-\tau)(w(x)-1) - x_2(\tau - c(x))] \\
&\quad - 2(1-x_2)x_2(1-x_1-x_2)[(1-\tau)w(x) + (1-x_2)(\tau - c(x))] \,.
\end{aligned}$$

We analyze $\dot{V}(x)$ in E, the environment of $(0,1)$, which requires to evaluate $w(x)$ and $c(x)$ in E.

$$\begin{aligned}
\lim_{x_2 \to 1} w(x) &= \lim_{x_2 \to 1} \frac{\tau}{1-\tau} \frac{(1-x_1-x_2)(1-x_2)}{x_1 + x_2} = 0 \\
\lim_{x_2 \to 1} c(x) &= \lim_{x_2 \to 1} p(x)\Big((1-\tau)w(x) + 1\Big) = p(0,1) \,.
\end{aligned}$$

We conclude that for a sufficiently small environment of $(0,1)$ and for $p(0,1) < \tau$ we have

$$\begin{aligned}
\dot{V}(x) &= \underbrace{2x_1^2(1-x_1-x_2)}_{>0}\,[\,\underbrace{(1-\tau)(w(x)-1)}_{<0} - \underbrace{x_2(\tau - c(x))}_{>0}\,] \\
&\quad - \underbrace{2(1-x_2)x_2(1-x_1-x_2)}_{>0}\,[\,\underbrace{(1-\tau)w(x)}_{>0} + \underbrace{(1-x_2)(\tau - c(x))}_{>0}\,] \\
&\Rightarrow \; V(x) < 0 \,.
\end{aligned}$$

[26] Notation: $\overline{E}$ is the set of osculation points of E. Point x is an osculation point of E if $E \cap U_\rho(x) \neq \emptyset, \forall \rho \in \mathbb{R}_+$. The set $U_\rho(x)$ is the ρ-neighborhood of x (or the open sphere around x). It is defined as $U_\rho(x) = \{y \in \mathbb{R}^n; |x - y| < \rho\}$.

So if $p(0,1) < \tau$, $(0,1)$ is asymptotically stable.

Critical point $(\mathbf{0}, \bar{\mathbf{x}}_\mathbf{2})$: The critical point $(0, \bar{x}_2)$ does not necessarily exist for all τ and $p(x_1, x_2)$. Because of $\frac{\partial F_1}{\partial x_2}(0, \bar{x}_2) = 0$ the eigenvalues of the jacobian $DF(0, \bar{x}_2)$ are $\frac{\partial F_1}{\partial x_1}(0, \bar{x}_2)$ and $\frac{\partial F_2}{\partial x_2}(0, \bar{x}_2)$. By rearranging (8) to

$$(1-\tau)w(0, \bar{x}_2) - x_2(\tau - c(0, \bar{x}_2)) = c(0, \bar{x}_2) - \tau\,,$$

we can write

$$\begin{aligned} \frac{\partial F_1}{\partial x_1}(0, \bar{x}_2) &= (1-\bar{x}_2)\Big((1-\tau)w(0, \bar{x}_2) - \bar{x}_2(\tau - c(0, \bar{x}_2))\Big) \\ &= (1-\bar{x}_2)\Big(c(0, \bar{x}_2) - \tau\Big) \\ &= \tau \frac{(1-\bar{x}_2)^2}{\bar{x}_2} \ > \ 0\,, \end{aligned}$$

using (9) for the last equation. For the second eigenvalue we note that

$$(1-\tau)\frac{\partial w}{\partial \bar{x}_2}(0, \bar{x}_2) \ = \ \tau \frac{\bar{x}_2^2 - 1}{\bar{x}_2^2}$$

and can then derive

$$\begin{aligned} \frac{\partial F_2}{\partial x_2}(0, \bar{x}_2) &= \bar{x}_2(1-\bar{x}_2) \cdot \\ & \quad \left[(1-\tau)\frac{\partial w}{\partial x_2}(0, \bar{x}_2) - (\tau - c(0, \bar{x}_2)) - (1-\bar{x}_2)\frac{\partial c}{\partial x_2}(0, \bar{x}_2)\right] \\ &= \bar{x}_2(1-\bar{x}_2) \cdot \\ & \quad \left[\tau\left(\frac{\bar{x}_2^2 - 1}{\bar{x}_2^2} - 1\right) + c(0, \bar{x}_2) - (1-\bar{x}_2)\frac{\partial c}{\partial x_2}(0, \bar{x}_2)\right] \\ &= \bar{x}_2(1-\bar{x}_2)\left[-\frac{\tau}{\bar{x}_2^2} + \frac{\tau}{\bar{x}_2} - (1-\bar{x}_2)\frac{\partial c}{\partial x_2}(0, \bar{x}_2)\right] \\ &= \bar{x}_2(1-\bar{x}_2)^2\left[-\frac{\tau}{\bar{x}_2^2} - \frac{\partial c}{\partial x_2}(0, \bar{x}_2)\right]\,. \end{aligned}$$

We now have to determine the algebraic sign of the two eigenvalues. The second eigenvalue is negative if

$$\left[-\frac{\tau}{\bar{x}_2^2} - \frac{\partial c}{\partial x_2}(0, \bar{x}_2)\right] \ < \ 0\,.$$

By replacing $\frac{\partial c}{\partial x_2}(0,\bar{x}_2)$ by

$$\begin{aligned}
& \frac{\partial p}{\partial x_2}(0,\bar{x}_2)\left[(1-\tau)w(0,\bar{x}_2)+1\right]+p(0,\bar{x}_2)(1-\tau)\frac{\partial w}{\partial \bar{x}_2}(0,\bar{x}_2) \\
= & \frac{\partial p}{\partial x_2}(0,\bar{x}_2)\left[\tau\frac{(1-\bar{x}_2)^2}{\bar{x}_2}+1\right]-p(0,\bar{x}_2)\tau\frac{1-\bar{x}_2^2}{\bar{x}_2^2}
\end{aligned}$$

and rearranging terms we find the condition

$$p(0,\bar{x}_2)\tau(1-\bar{x}_2^2)-\frac{\partial p}{\partial x_2}(0,\bar{x}_2)\left[\tau\bar{x}_2(1-\bar{x}_2)^2+\bar{x}_2^2\right] \quad < \quad \tau \tag{14}$$

for a negative second eigenvalue of $DF(0,\bar{x}_2)$. We have assumed that $p(x_1,x_2)$ is decreasing in x_2, however it is possible to find a $p(x_1,x_2)$ that is consistent with (14). We conclude that the critical point $(0,\bar{x}_2)$ is then a saddle (and otherwise a source, since the first eigenvalue is positive) which can attract solution trajectories in the simplex. Condition (14) is satisfied, if $p(0,\bar{x}_2)$ is not too high and $p(0,x_2)$ does not increase strongly in x_2.

Critical point $(\mathbf{1},\mathbf{0})$: This critical point always exists but it is not hyperbolic. We will show instability by analyzing the system (6) in the vicinity of $(1,0)$. This can be done by looking at

$$-\frac{\partial F_1}{\partial x_1}(1,0)\,.$$

The intuition is as follows. Since $(1,0)$ is a fixed point, we have that $\dot{x}_1(1,0) = F_1(1,0) = 0$. We now check which sign $F_1(x)$ takes if we marginally deviate from $(0,1)$ by decreasing x_1 marginally (and increasing x_2 and x_3 marginally). If $-\frac{\partial F_1}{\partial x_1}(1,0)$ is negative (positive), we know that $\dot{x}_1$ is negative (positive) in the vicinity of $(1,0)$ since it is zero in $(1,0)$. We find

$$-\frac{\partial F_1}{\partial x_1}(1,0) \quad = \quad -(-1)\left[(1-\tau)(w(1,0)-1)\right] \;=\; -(1-\tau)\,.$$

So marginally deviating from $(1,0)$ by marginally decreasing x_1 causes the function F_1 to take a negative value. That means that a solution curve $x(t)$ of system (6) starting in the vicinity of $(1,0)$ will move away from $(1,0)$. So $(1,0)$ cannot be asymptotically stable.

Critical point $\{(\mathbf{x_1},\mathbf{1}-\mathbf{x_1})\,|\,\mathbf{x_1}\in]\mathbf{0},\mathbf{1}[\}$: This set of critical points always exists and its elements are neither hyperbolic nor isolated critical points. In order to show instability we can use the arguments made for the critical point $(1,0)$. We will check what signs $F_1(x_1, x_2)$ and $F_2(x_1, x_2)$ take in the vicinity of $(x_1, 1 - x1)$. Deviating from the edge of the simplex $(x_1, 1 - x_1)$ means that we marginally decrease x_1 and x_2 at the same time. So we are interested in the signs of

$$\begin{aligned}
&-\frac{\partial F_1}{\partial x_1}(x_1, 1 - x_1) - \frac{\partial F_1}{\partial x_2}(x_1, 1 - x_1) \quad \text{and} \\
&-\frac{\partial F_2}{\partial x_1}(x_1, 1 - x_1) - \frac{\partial F_2}{\partial x_2}(x_1, 1 - x_1) .
\end{aligned}$$

Note first that

$$w(x_1, 1 - x_1) = 0 \quad \text{and} \quad c(x_1, 1 - x_1) = p(x_1, 1 - x_1) .$$

We use the expressions of we have given for the elements of the Jacobian $DF(x)$ and find

$$\begin{aligned}
&- \frac{\partial F_1}{\partial x_1}(x_1, 1 - x_1) - \frac{\partial F_1}{\partial x_2}(x_1, 1 - x_1) \\
= \quad & -\Big\{ - x_1 \Big[(1 - \tau)(w(x_1, 1 - x_1) - 1) - (1 - x_1)(\tau - c(x_1, 1 - x_1)) \Big] \Big\} \\
& -\Big\{ - x_1 \Big[(1 - \tau)(w(x_1, 1 - x_1) - 1) - (1 - x_1)(\tau - c(x_1, 1 - x_1)) \Big] \Big\} \\
= \quad & 2x_1 \Big[- (1 - \tau) - (1 - x_1)(\tau - p(x_1, 1 - x_1)) \Big] \\
= \quad & -2x_1 \Big[1 - p(x_1, 1 - x_1) - x_1(\tau - p(x_1, 1 - x_1)) \Big] .
\end{aligned}$$

If $\tau > p(x_1, 1 - x_1)$ then $1 - p(x_1, 1 - x_1) > \tau - p(x_1, 1 - x_1) > 0$ and hence $1 - p(x_1, 1 - x_1) - x_1(\tau - p(x_1, 1 - x_1)) > 0$. If $p(x_1, 1 - x_1) > \tau$ then $0 > x_1(\tau - p(x_1, 1 - x_1))$ and again $1 - p(x_1, 1 - x_1) - x_1(\tau - p(x_1, 1 - x_1)) > 0$. If $p(x_1, 1 - x_1) = \tau$ then $1 - p(x_1, 1 - x_1) - x_1(\tau - p(x_1, 1 - x_1)) = 0$ again. So we can state

$$-\frac{\partial F_1}{\partial x_1}(x_1, 1 - x_1) - \frac{\partial F_1}{\partial x_2}(x_1, 1 - x_1) \quad < \quad 0 .$$

That means that all solution curves starting in the vicinity of $(x_1, 1 - x_1)$ will move away from $x_1 = 1$. Because this result holds for all $x_1 \in (0, 1)$ the critical points in the set $\{(x_1, 1 - x_1) \,|\, x_1 \in]0, 1[\}$ cannot be stable; it is impossible that trajectories starting in the vicinity of this edge of the simplex move towards a point in this set. It is redundant to determine the sign of $F_2(x)$ in the vicinity of $\{(x_1, 1 - x_1) \,|\, x_1 \in]0, 1[\}$.

Critical point $(\tau, \mathbf{0})$: This critical point does always exist. It is easy to see that $\frac{\partial F_2}{\partial x_1}(\tau, 0) = 0$ since $x_2 = 0$. So we know that the two eigenvalues of $DF(\tau, 0)$ are

$\frac{\partial F_1}{\partial x_1}(\tau, 0)$ and $\frac{\partial F_2}{\partial x_2}(\tau, 0)$. Note that

$$\begin{aligned} w(\tau, 0) &= 1\,, \\ \frac{\partial w}{\partial x_1}(\tau, 0) &= -\frac{1}{\tau(1-\tau)}\,, \\ c(\tau, 0) &= (2-\tau)p(\tau, 0)\,. \end{aligned}$$

The eigenvalues of $DF(\tau, 0)$ are

$$\begin{aligned} \frac{\partial F_1}{\partial x_1}(\tau, 0) &= -(1-\tau) \\ \frac{\partial F_1}{\partial x_1}(\tau, 0) &= (1-\tau)[1-(2-\tau)p(\tau, 0)]\,. \end{aligned}$$

We conclude that if $\frac{1}{2-\tau} < p(\tau, 0)$, the critical point is a sink. If $p(\tau, 0) < \frac{1}{2-\tau}$, the critical point is a saddle.

Critical point $(\tau - \hat{\mathbf{x}}_2, \hat{\mathbf{x}}_2)$: In order to analyze the stability of $(\tau - \hat{x}_2, \hat{x}_2)$, we evaluate the elements of $DF(\tau - \hat{x}_2, \hat{x}_2)$. Note that

$$\begin{aligned} w(\tau - \hat{x}_2, \hat{x}_2) &= 1 - \hat{x}_2 \\ \frac{\partial w}{\partial x_1}(\tau - \hat{x}_2, \hat{x}_2) &= -\frac{1-\hat{x}_2}{\tau(1-\tau)} \\ \frac{\partial w}{\partial x_2}(\tau - \hat{x}_2, \hat{x}_2) &= -1 - \frac{1-\hat{x}_2}{\tau(1-\tau)}\,. \end{aligned}$$

We find

$$\begin{aligned} \frac{\partial F_1}{\partial x_1}(\tau - \hat{x}_2, \hat{x}_2) &= (1-\tau)(\tau - \hat{x}_2)\Big(-\frac{1-\hat{x}_2}{\tau} + \hat{x}_2 \frac{\partial c}{\partial x_1}(\tau - \hat{x}_2, \hat{x}_2)\Big) \\ \frac{\partial F_1}{\partial x_2}(\tau - \hat{x}_2, \hat{x}_2) &= (1-\tau)(\tau - \hat{x}_2)\Big(-\frac{1-\hat{x}_2}{\tau} + \hat{x}_2 \frac{\partial c}{\partial x_2}(\tau - \hat{x}_2, \hat{x}_2)\Big) \\ \frac{\partial F_2}{\partial x_1}(\tau - \hat{x}_2, \hat{x}_2) &= (1-\tau)\hat{x}_2\Big(-\frac{1-\hat{x}_2}{\tau} - (1-\hat{x}_2) \frac{\partial c}{\partial x_1}(\tau - \hat{x}_2, \hat{x}_2)\Big) \\ \frac{\partial F_2}{\partial x_2}(\tau - \hat{x}_2, \hat{x}_2) &= (1-\tau)\hat{x}_2\Big(-\frac{1-\hat{x}_2}{\tau} - (1-\hat{x}_2) \frac{\partial c}{\partial x_2}(\tau - \hat{x}_2, \hat{x}_2)\Big)\,. \end{aligned}$$

Because we have assumed that $\frac{\partial p}{\partial x_2} < 0$ we have $\frac{\partial c}{\partial x_2} < 0$, and therefore we know that

$$\frac{\partial F_1}{\partial x_2}(\tau - \hat{x}_2, \hat{x}_2) < 0\,.$$

Again from Assumption 1 we derive

$$\begin{aligned} \frac{\partial F_1}{\partial x_1}(\tau - \hat{x}_2, \hat{x}_2) &> \frac{\partial F_1}{\partial x_2}(\tau - \hat{x}_2, \hat{x}_2) \\ \frac{\partial F_2}{\partial x_1}(\tau - \hat{x}_2, \hat{x}_2) &> \frac{\partial F_2}{\partial x_2}(\tau - \hat{x}_2, \hat{x}_2)\,. \end{aligned}$$

Let us abbreviate the elements of $DF(\tau - \hat{x}_2, \hat{x}_2)$ by j_i, such that

$$DF(\tau - \hat{x}_2, \hat{x}_2) = \begin{pmatrix} j_1 & j_2 \\ j_3 & j_4 \end{pmatrix}.$$

The standard formula for eigenvalues of 2×2-matrices yields for $DF(\tau - \hat{x}_2, \hat{x}_2)$

$$\begin{aligned} \lambda_1 &= \frac{1}{2}\Big(j_1 + j_4 + \sqrt{j_1^2 - 2j_1j_4 + j_4^2 + 4j_2j_3}\Big), \\ \lambda_2 &= \frac{1}{2}\Big(j_1 + j_4 - \sqrt{j_1^2 - 2j_1j_4 + j_4^2 + 4j_2j_3}\Big). \end{aligned}$$

We are only interested in the real parts of the eigenvalues, for they determine the stability of the critical point. A square root of a real discriminant Δ always either has a zero real part (if $\Delta \leq 0$) or a positive real part (if $\Delta > 0$). Therefore we have that $\lambda_1 \geq \lambda_2$. If $(\tau - \hat{x}_2, \hat{x}_2)$ is stable if and only if $\lambda_1 < 0 \wedge \lambda_2 < 0$. So the necessary condition for stability of $(\tau - \hat{x}_2, \hat{x}_2)$ is

$$\begin{aligned} \lambda_1 &< 0 \\ \frac{1}{2}\Big(j_1 + j_4 + \sqrt{j_1^2 - 2j_1j_4 + j_4^2 + 4j_2j_3}\Big) &< 0 \\ j_1 + j_4 &< -\sqrt{j_1^2 - 2j_1j_4 + j_4^2 + 4j_2j_3} \\ (j_1 + j_4)^2 &> j_1^2 - 2j_1j_4 + j_4^2 + 4j_2j_3 \\ j_1j_4 &> j_2j_3\,. \end{aligned}$$

We substitute the explicit elements of $DF(\tau - \hat{x}_2, \hat{x}_2)$ into the last equation. Simplifying the expression leaves us with

$$\begin{aligned} &\frac{1}{\tau}(1-\tau)^2(\tau - \hat{x}_2)\hat{x}_2(1-\hat{x}_2)\left(\frac{\partial c}{\partial x_2}(\tau - \hat{x}_2, \hat{x}_2) - \frac{\partial c}{\partial x_2}(\tau - \hat{x}_2, \hat{x}_2)\right) > 0 \\ &\qquad\Rightarrow \quad \frac{\partial c}{\partial x_2}(\tau - \hat{x}_2, \hat{x}_2) > \frac{\partial c}{\partial x_2}(\tau - \hat{x}_2, \hat{x}_2)\,. \end{aligned}$$

This is contradictory to Assumption 1. Thus, as long as we adhere to Assumption 1, $(\tau - \hat{x}_2, \hat{x}_2)$ cannot be stable.

Table 1: Critical points of the corruption game.

Critical point (x_1, x_2)	Conditions on τ and $p(x_1, x_2)$ for existence of critical point	Asymptotic stability
$(0,0)$	none	unstable
$(0,1)$	none	asymptotically stable if $p(0,1) < \tau$
$(0, \bar{x}_2)$	$\bar{x}_2$ such that $p(0, \bar{x}_2) = \frac{\tau}{\tau(1-\bar{x}_2)^2+\bar{x}_2}$	asymptotically stable if $\tau < -\bar{x}_2^2 \frac{\partial c}{\partial x_2}(0, \bar{x}_2)$
$(1,0)$	none	unstable
$(x_1, 1-x_1)$ with $x_1 \in]0,1[$	none	unstable
$(\tau, 0)$	none	asymptotically stable if $\frac{1}{2-\tau} < p(\tau, 0)$
$(\bar{\tau} - \hat{x}_2, \hat{x}_2)$	$\hat{x}_2$ such that $p(\tau - \hat{x}_2, \hat{x}_2) = \frac{1}{(1-\tau)(1-\hat{x}_2)+1}$	unstable

A.1.2 Global Theory of Nonlinear Systems

It is left to show that no other attracting sets exist [27] and attractors [28] than those found in the last subsection.

The Generalized Poincaré-Bendixson Theorem for Analytic Systems states that the ω-limit set [29] of any trajectory of a two-dimensional, relatively-prime, analytic system is either a critical point, a cycle, or a compound separatrix cycle. We show below that the solution trajectories of system (6) cannot be closed. This allows us to exclude limit cycles and compound separatrix cycles as ω-limits. We will then be able to conclude that only evolutionary equilibria can be attractors.

[27] A closed invariant set $A \in E$ is called an *attracting set* of a system $\dot{x} = F(x)$ if there is some neighborhood U of A such that for all $x \in U$, $\phi_t(x) \in U$ for all $t \geq 0$ and $\phi_t(x) \to A$ as $t \to \infty$.

[28] An *attractor* is an attracting set containing a dense orbit (a dense orbit is an orbit that comes arbitrarily close to each point in the attractor).

[29] A point $p \in E$ is an ω-limit point of the trajectory $\phi(t, x_0)$ of the system $\dot{x} = F(x)$ if there is a sequence $t_n \to \infty$ such that $\lim_{n\to\infty} \phi(t_n, x) = p$.

In order to show that our system does not have any closed trajectories, we apply index theory, a method that describes global behavior of the solutions to a differential equation system (see e.g. Strogatz, 1994, Chapter 6). In Proof A.1 we have calculated all critical points of system (6). They all are on the boundary of the simplex except $(\tau - \hat{x}_2, \hat{x}_2)$, which is a saddle. We now assume that there exists a closed trajectory to (6). Figure 3 shows all qualitatively different locations a closed trajectory could occupy, they are indicated by the dotted curves T_1, T_2, and T_3. The index at each of the critical points is also shown in the figure (for an explanation of how to calculate the index at critical points, see Strogatz, 1994, Chapter 6).

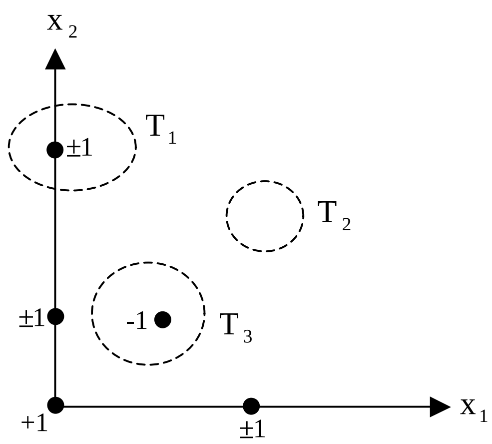

Figure 3: Locations of closed trajectories.

We can rule out closed trajectories as follows. Trajectories like T_1 are impossible because they cross the boundary of the simplex. The reason is the following. From Bruegger (2005) we know that the simplex boundary is invariant under system (6). So the boundary of the simplex contains straight-line trajectories. Since trajectories cannot cross,[30] we can exclude trajectories like T_1. Trajectories like T_2 can be excluded as well because they do not enclose any fixed points at all. And trajectories like T_3 violate the requirement that the indices inside the closed trajectories must sum up to 1 (see e.g. Strogatz, 1994, Theorem 6.8.2., p. 180). We conclude that system (6) does not have any closed trajectories. Consequently, the ω-limit set of any trajectory of our system is a critical point. It is clear from the definition of an attractor and the concept of asymptotic stability, that only an asymptotically stable critical point can

[30]This follows directly from the Fundamental Existence-Uniqueness Theorem.

be an attractor. We conclude that almost every trajectory through a point $x \in \Sigma$ approaches an EE in the limit. The sole exceptions are those trajectories that are the separatrices of the system.

From the Poincaré-Bendixson Theorem we know that if a trajectory of a planar system is confined to a closed, bounded region, then the trajectory is either attracted by a critical point or a closed trajectory. Since the simplex is invariant under the dynamics of system (6) and since we have shown that system (6) has no closed trajectories, we can conclude, that there always exists an attractor in the corruption game. This results holds for all $p(x)$ and all τ.