Tax Evasion, Social Norms and Conditional Cooperation

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

2005

vorgelegt von
Christian Traxler

Referent: Prof. Dr. Andreas Haufler
Korreferent: Prof. Dr. Klaus Schmidt
Datum der mündlichen Prüfung: 26. Januar 2006
Promotionsabschlussberatung: 08. Februar 2006

Acknowledgements

First of all, I would like to gratefully acknowledge the supervision of Professor Andreas Haufler, who provided me with a perfect mix of instructive criticism and constant encouragement. This thesis has gained substantially from his invaluable comments and suggestions.

During the past three years, I have also received a lot of support from my colleagues at the Munich Graduate School of Economics respectively the Economics Department at the University of Munich. In particular I would like to thank Stefan Brandauer, Marcus Drometer, Florian Herold, Simone Kohnz, Andi Leukert, Katri Mikkonen, Tobi Seidel and Hans Zenger for helpful advices and many fruitful lunch and coffee break discussions.

Major parts of the work to chapter 3 and 4 of this thesis have been done during my stay at CREED, University of Amsterdam. I would like to acknowledge the hospitality and the wonderful atmosphere at CREED. I am especially grateful to Professor Frans van Winden, who supervised me during the visit, as well as to Eva van den Broek, Astrid Hopfensitz, Ernesto Reuben and Aljaž Ule for great feedback at the '*broodjes*'-seminar-series. Many thanks are also due to Ruslan Gurtovoy, Fabrice Le Lec, Friederike Mengel, Tom Truyts and the other participants at the Trento Summer School in 2004 for stimulating discussions. I am especially grateful to Friederike for detailed suggestions and brilliant comments.

While preparing this thesis, I have benefited enormously from collaborations with Ernesto, Frans, Gerlinde Fellner, Rupert Sausgruber as well as Mathias Spichtig and Sven Stöwhase. It was – and continues to be – a pleasure to work with you!

Financial support from the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged.

Finally, I want to thank my parents for their support during all these years. I am forever indebted to you for your understanding, patience and encouragement!

Christian Traxler
Munich, September 2005

Contents

[1]This chapter is based on joint work with Sven Stöwhase, Fraunhofer Institute for Applied Information Technology FIT, Bonn.

[2]This chapter is based on joint work with Mathias Spichtig, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam.

List of Figures

Preface

In February 2003, Simon Gächter gave a stimulating lecture on the 'Behavioral Economics of Trust and Voluntary Cooperation' in course of the CESifo lecture series at the University of Munich. Based on a variety of experimental economic studies, he discussed and analyzed behavioral regularities in public good games. In particular, he provided convincing evidence on the coexistence of different types of subjects: There seems to exist a 'large minority' of individuals who free-ride as predicted by standard economic theory, assuming rational and selfish agents. In addition to the free-riders, however, there also exists a larger group of individuals who can be classified as *conditional cooperators*, i.e. '*people who are willing to contribute the more to a public good, the more others contribute*' (Fischbacher, Gächter and Fehr, 2001, p.397). In the concluding discussion of his lecture, Simon Gächter pointed out the importance of these findings for many real-life situations – such as team production problems, the abuse of welfare payments, criminal behavior and tax evasion – and invited for future research to link the experimental evidence with the (theoretical) study of these policy relevant issues. In the present thesis we take up this invitation and elaborate on *tax evasion, social norms and conditional cooperation.*

In chapter 1, we follow a conventional approach, studying tax evasion of firms and the provision of optimal auditing incentives in a federal economy. In chapter 2–4, we consider – next to economic incentives – the impact of social norms on voluntary tax compliance respectively cooperation. While the first three chapters provide a policy oriented analysis, chapter 4 addresses the evolutionary origin of conditional cooperation. Though all four parts can be read independently, we will now briefly discuss the relation of this thesis to other behavioral and public economic literature and provide an outlook on each chapter.

Customs vs. Conventions, Social Norms vs. Social Preferences

In the last decades, economists became more and more interested in the study of social norms. Unfortunately, however, the term 'social norm' is now used for several very different concepts. Some economists use norms as an analog to *conventions*, where conventions can be defined as '*equilibrium behavior in games played repeatedly by many different individuals in society, where the behaviors are widely known to be customary*' (Young, 1998a, p.823). Driving on the right side of the road, for example, represents such a convention. Next to the question, how particular conventions emerge (Young, 1989b), it is of course crucial to understand why some conventions are possible at all. If we donate to charity every year at Christmas, standard economic theory, assuming rational, self-interested agents, fails to explain such an equilibrium behavior. While there are now plenty approaches which explain non-selfish behavior, one strand of literature directly links to conventional, respectively customary behavior.

The literature on *social customs* – initiated by the seminal work of Akerlof (1980) – studies the interplay of monetary and non-monetary incentives. Non-monetary incentives thereby arise from social or personal, external or internal sanctions which agents incur if they deviate from a social custom. If your child asks you to contribute to the primary schools annual Christmas charity project, you may do so, because you want to avoid being considered as greedy and cold-hearted by your social environment (e.g. your family or other parents), because you are afraid of be excluded from the next parents meeting, or simply because you do not want to expose your child to any form of punishment from classmates or teachers. While the first argument is based on an internal, personal sanction, related to emotional reactions (Elster, 1989a, 1989b), the second and third argument derives from external social sanction, associated with stigmatization and social disapproval (Ullmann-Margalit, 1977). Hence, if we deviate from the social custom – i.e. what is considered to be '*normal*' – we incur a form of punishment, which induces compliance with the custom.

In this thesis we build upon the social customs approach. We thereby follow the more recent literature (e.g. Lindbeck et al., 1999), which terms social customs in the sense of Akerlof (1980) as *social norms* (Elster, 1989b). A social norm is a pattern of behavior, which is enforced by external or internal sanctions. Note that a social norm (like charity giving) may also represent a convention in the sense of Young. Yet, not all conventions are social norms: Once a society has coordinated to drive on the right side of a street, it may become a convention. This equilibrium outcome, however, is then supported by standard rational agents. As there are no social sanctions at play, which enforce the convention, we do not call it a social norm.

In chapter 2 and 3, and to some extent also in chapter 4, we focus on the role of *internalized* social norms (Elster, 1989a). A norm is said to be internalized, if deviant behavior is accompanied '*by internal sanctions, including shame, guilt and loss of self-esteem, as opposed to purely external sanctions [...]*' (Gintis, 2003, p.407). Jon Elster, illustrates the point considering an anti-littering norm: '*I dont throw away litter in the park, even when there is nobody around to observe me.*' Even if there is no scope for external sanctions, we may follow a social norm. '*Shame or anticipation of it is a sufficient internal sanction*' to enforce norm compliance (Elster, 1989b, p. 104-105).[1] As the subject in Elsters example, tax evaders are, in a certain sense, also 'alone in the park': Given that individual evasion decisions take place in privacy, direct forms of social sanctions are mainly limited to the household level. Neglecting within-household interaction, we focus on the role of internal sanctions.

Following the tradition of the social customs literature (Akerlof, 1980; Naylor, 1989), we study the interplay of economic incentives and norm-guided behavior, where the power of the social norm, represented by the strength of norm enforcing sanctions, is endogenously determined: The more people adhere to a norm, the stronger it becomes. If everybody deviates from it, the norm has eroded. In an environment where all parents support the schools charity fund, not donating represents a more severe wrong doing – associated with stronger sanctions – as compared to an environment where hardly anybody contributes to the project. In this case, the same society could coordinate on different equilibria (conventions): self-enforcing situations where many agents comply with the norm (contribute to charity) or a self-enforcing situation where most agents deviate from it, and act according to their pecuniary self-interest (Elster, 1989a, 1989b). Hence, taking the strength of a social norm as endogenous opens the scope for *conditional cooperative behavior*.

We shall remark that patterns of conditional cooperative behavior can also be explained by theories of fairness and inequity aversion (Rabin, 1993; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) or theories of reciprocity and intentions (Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2005). However, we do not adopt one of these *social preference* approaches, summarized by Fehr and Falk (2002). While we are convinced that all these concepts provide important and relevant insights for

[1] Note that in recent years, a significant body of literature has emerged, which demonstrates the importance of emotions in economic decision making. Compare e.g. Elster (1998), van Winden (2001).

the explanation of human behavior, we think that one should be careful with their application on specific policy problems. In particular, we doubt that these models are appropriate to study cooperative behavior in a large population context, where direct interaction and therefore direct forms of reciprocity are limited – as in the case of tax evasion: As we have noted above, individual evasion mostly takes place in privacy. Since public information about the evasion decision of friends or co-workers is in general not available, direct (reciprocal) responses to evasion are of limited importance. Of course, in the context of tax evasion there is one significant direct interaction, namely that of the taxpayers with the government respectively with tax authorities. However, the two interacting players appear overly heterogenous, which renders the immediate application of the social preference models inappropriate, as these models are to a large extent based on payoff-comparisons between players.[2]

Nevertheless, one could relate a social norm for tax compliance to fairness norms. Tax compliance may be considered as the fair behavior, if most other taxpayers follow this fairness concept, while cheating on taxes may become the norm, if others do so as well. Such a pattern, however, is covered by our approach. In principle, one could interpret our concept of an internalized tax compliance norm as a 'boiled down' version of a more complex fairness rule, like the specific form of self-centered inequity aversion introduced by Fehr and Schmidt (1999) or Bolton and Ockenfels (2000). However, these approaches are based on an interpersonal payoff comparison, which is of limited plausibility for the scenario under consideration. In any case, the results derived in this thesis would also hold in these alternative frameworks. Having considered all that, we are convinced that our approach, focussing on the role of internalized norms, constitutes a reasonable and valid way to study tax evasion decisions.

On the Role of Tax Evasion and Auditing

An implicit assumption in a large part of the public economics literature is that taxes are fully enforced; consumers, respectively firms, are supposed to honestly report all their taxable activities. Although for some taxation problems, this assumption represents a useful simplification, it is evidently wrong. In the US, for example, the Internal Revenue Service (IRS) estimated the 'tax gap' – the difference between actual and hy-

[2] We shall remark, however, that theories of intentions may provide an interesting tool to study interactions between taxpayers and tax authority, as these models could explain how stricter enforcement can induce higher levels of evasion (Sheffrin and Triest, 1992; Feld and Frey, 2002). In terms of intention based approaches, such counterintuitive results can be related to the signaling of distrust associated with stricter auditing (compare Falk and Kosfeld, 2004).

pothetical tax revenues without evasion – to amount to a total of approximately $350 billion (as opposed to $1767 billion paid 'voluntarily') in the year 2001 (Sawicky, 2005). For the same year, the evasion of value-added taxes in Germany was estimated at € 15 billion, which represents a VAT gap of about 10% (Gebauer and Parsche, 2003). In Switzerland, Frey and Feld (2002) assess the income tax gap at more than 17%. This figures clearly indicate the dimension of tax evasion and highlight the relevance of incorporating imperfect tax collection systems into standard public economic analysis.

The public side of the economy faces these losses with extensive investments into tax enforcement. The IRS, for example, is endowed with an annual budget of $10 billion. This budget, however, is comparably slim, if we take into account that in 2001 the IRS raised more than $50 billion taxes from direct enforcement measures (Sawicky, 2005). Hence, the provision of incentives supporting tax compliance, e.g. related to the implementation of auditing policies, constitutes a serious undertaking. This also holds true for Germany, where authorities raised approximately € 1.5 billion from tax investigations in 2001 (Bundesministerium der Finanzen, 2003). In contrast to the US, however, the total public spendings on tax enforcement are difficult to assess, since in Germany the *Länder* are in charge of tax collection and tax enforcement. Moreover, there exists an – at least for an economist – seemingly incredible institutional arrangement: While the *Länder* bear the full costs of tax enforcement, the revenues are redistributed between different regions according to the *Länderfinanzausgleich*, the German equalizing transfer scheme. Hence, there is a clear distortion of the regions' incentives to provide an optimal auditing policy. As a consequence, one should expect an inefficiently low level of tax enforcement. In chapter 1, however, we will demonstrate that this is not necessarily the case in the context of a federal economy, where regions engage in fiscal competition.

Chapter 1: Tax Evasion and Auditing in a Federal Economy[3]

We consider the case where each region provides a local tax enforcement policy, associated with a particular auditing frequency. If regions use their local auditing policy as a strategic tool to attract mobile capital, fiscal competition will lead to an 'underprovision' of auditing. Analogous to the standard tax competition result (Zodrow and Mieszkowski, 1986), unconfined interregional competition will result in too little tax enforcement efforts and thereby in inefficiently high levels of tax evasion. In this case,

[3]This chapter is based on joint work with Sven Stöwhase, Fraunhofer Institute for Applied Information Technology FIT, Bonn.

the introduction of a German-styled revenue sharing system causes a clear distortion, as the mechanism drives a wedge between each region's marginal revenue benefit and costs. In addition, however, the revenue sharing scheme also has a beneficial effect: it provides an incentive for each region to partly internalize the positive fiscal externalities from auditing. If this latter effect dominates the distortion, revenue sharing will induce the provision of stricter tax enforcement as compared to the case without any interregional transfers.

In a next step, we propose an alternative revenue sharing system, which redistributes not only the revenues but also the costs of tax enforcement. Such a mechanism would provide a correction incentive to account for the fiscal externality of auditing, without introducing any distortion. Hence, this mechanism typically induces stricter auditing associated with less tax evasion. Yet, there are some limitations which render the implementation of such a mechanism difficult: In the context of informational asymmetries between different layers of government, this alternative mechanism could provide regions with an incentive to overreport tax enforcement costs.

Our study contributes to different strands of the public economics literature. First of all, chapter 1 relates to the literature on the tax evasion of firms (Marrelli, 1984; Wang and Conant, 1988; Yaniv, 1988). In particular, we extend the approach from Cremer and Gahvari (1993) and incorporate it into a standard tax competition setting. Thereby, our model provides a more general framework for studying the interaction of tax evasion and tax competition than Cremer and Gahvari (2000), who focus on the special case of tax competition between two regions. Secondly, the analysis of chapter 1 also contributes to the literature on fiscal federalism. We expand the analysis of optimal interregional redistribution (especially Köthenbürger, 2002), considering tax evasion and tax enforcement incentives. In addition, we highlight the role of decentralized tax collection in the context of fiscal competition.

One can conclude from this chapter that, in order to understand the level of tax evasion, it is not only relevant to study the incentives to comply with the tax laws. It is equally important to analyze different layers of incentives embedded in the institutional settings which shape the public sector's provision of tax enforcement measures and thereby the level of tax compliance.

Individual Income Tax Evasion and Tax Morale

While the behavior of firms, studied in the first chapter, is typically assumed to be unaffected by social norms, chapter 2 elaborates on the role of social norms for tax evasion decisions by consumers. The standard framework for the analysis of individual tax evasion goes back to the seminal contribution of Allingham and Sandmo (1972). They model consumers' income tax evasion as choice under risk, similar to Becker's (1968) analysis of crime. Taxpayers decide on how much of their income to conceal, taking into account the tax rate on the one and the enforcement policy on the other hand. Tax evasion resembles a 'risky gamble', since a tax evader stands a chance to avoid an audit and thereby succeed with the evasion. The individual decision problem becomes equivalent to an optimal portfolio choice problem: concealing income is analogous to an investment into a risky asset, whereas declaring income honestly resembles an investment into an asset with a safe return. Although this portfolio choice approach was highly influential for the theoretical research on tax evasion in the past decades (see e.g. Cowell, 1990; Andreoni et al., 1998), it fails to explain several important empirical facts. Most of all, the approach does not explain the level of tax compliance observed in western economies. Considering the rather low levels of auditing and relatively moderate penalties which characterize modern fiscal systems, we should – according to Allingham and Sandmo (1972) – expect far more people concealing taxes, than we actually do (Bernasconi, 1998; Graetz and Wilde, 1995; Skinner and Slemrod, 1985). As has been noted by Wilson (1993, p.3), '*[w]hat most needs explanation is not why some people are criminals but why most are not.*' Hence, the open question is simply, why do people 'voluntarily' pay taxes?

A huge body of empirical literature has demonstrated the crucial impact of 'soft issues' on individual tax compliance, which are typically neglected in traditional public economic analysis. Perceptions about the fairness of a tax system (Seidl and Traub, 2001), the treatment by tax authorities (Sheffrin and Triest, 1992; Feld and Frey, 2002), or political participation rights (Pommerehne and Weck-Hannemann, 1996) shape attitudes towards tax compliance (Lewis, 1986; Reckers et al., 1994) and the perceived civic duty to comply with tax laws (Scholz and Pinney, 1995; Orviska and Hudson, 2002). Taken together, these (and many other) factors determine what is often called *tax morale* (e.g. Torgler, 2005) – the individuals' intrinsic motivation to refrain from tax evasion (Frey, 1997; Frey and Feld, 2002). Thereby, a central aspect which determines the agents' tax morale are their beliefs about other citizens' tax compliance (e.g. Spicer and Hero, 1985; Porcano, 1988; Rothstein, 2000; Torger, 2005).

Chapter 2: Social Norms and Conditional Cooperative Taxpayers

In chapter 2 we account for these findings and provide a formal analysis of tax morale in the framework of Allingham and Sandmo (1972). Tax morale is modelled as an internalized social norm for tax compliance. The power of the social norm, and so the strength of the norm enforcing sanctions, are negatively related to the number of tax evaders in the society: The more people deviate from the norm, the weaker it becomes and the easier it is for an individual to legitimate the own norm violation. Within this framework, agents act as *conditional cooperative taxpayers*: They condition their compliance on the honesty of other citizens. We demonstrate that our model accounts for several shortcomings of the standard framework: In particular, our analysis predicts that higher taxes trigger more norm violations and typically higher levels of evasion. This is in sharp contrast to the conventional approach, according to which there exists a negative relation between the tax rate and the evasion level (Yitzhaki, 1974).

We extend the basic model and embed the individual evasion decision in a broader social context. We study a society consisting of different subgroups, where some agents are in the position of a 'moral role model': They have a decisive impact on the evasion level in society, as their behavior is crucial in the determination of the others' tax morale. Hence, a high level of norm compliance among societies' leaders – e.g. high profile individuals such as politicians – provides a significant contribution to the informal institutions supporting tax compliance. Finally, we also discuss the implications of this point for the optimal tax and enforcement policy.

Chapter 2 links two different strands of the literature: The classical tax evasion literature (Allingham and Sandmo, 1972; Kolm, 1973; Yitzhaki, 1974) on the one hand, and the formal literature on social norms on the other hand (e.g. Lindbeck, 1999). Our analysis also contributes to the recent discussion on the role of belief management as a policy tool (Rothstein, 2000; Gächter, 2005). Moreover, the study of tax evasion within a multi group context provides a theoretical complement to a recent contribution by Gächter and Renner (2005). While they demonstrate the strong impact of the 'leaders' behavior on voluntary cooperation of 'followers' in a public good experiment, our model shows how the evasion of the societies' role models influences the tax compliance of the 'followers' within the social structure of a large society.

The central conclusions one can draw from chapter 2 is that social interaction matters in the study of tax evasion. Cheating on taxes is more than a risky gamble: it is a *social* decision, embedded in a *social* environment. Considering this dimension improves the predictive power of the model and gives rise to interesting policy implications.

Chapter 3: Social Norms, Voting and the Provision of Public Goods

In chapter 3 we simplify the model from the previous chapter: First, we consider risk neutral agents and second we study a discrete instead of a continuous decision problem. The new framework then allows for a broader analysis of 'voluntary' tax compliance respectively compliance with the law in the context of a large society public good problem. Similar to the case of tax evasion, we consider non-deterrent law enforcement, i.e. the expected punishment from a law violation is too low to induce a rational, selfish agent to adhere to the legal norm. (Since standard theory predicts law violations, we call compliance 'voluntary', despite the legal obligation to do so.) In the context of legally regulated situations there typically exist clear societal expectations about how to behave 'adequately' (Cooter, 1998). In other words, the formal law reflects and expresses an informal rule, a social norm. Hence, public law enforcement together with social norm enforcement may evoke law obedience.

Within this framework, we approach several, quite different issues: First, we study individual compliance behavior. According to different levels in norm sensitivity, agents can be characterized as free-riders, unconditional cooperators and conditional cooperators. Hence, the model captures the type heterogeneity discussed at the beginning of this preface. Second, we consider the implications of these behavioral patterns for the public good provision problem. We provide a detailed discussion of the multiplicity of equilibria, which typically arise in the context of social norms. Interpreting public good contributions as taxes and free-riding as tax evasion, we study the relationship between tax level and collected revenues. We discuss the quite unusual shape of Laffer curves, associated with the multiplicity of equilibria, as well as the resulting policy implications. Third, we study the endogenous choice of law enforcement, respectively auditing policy and study the interaction between the formal institution – the strength of the law enforcement – and the informal institution – the social norm. Finally, we study the endogenous choice of a tax policy, according to a majority voting procedure. The voting outcome in the context of tax evaders is found to be always inefficient, resulting in a suboptimally low level of taxation and an underprovision of public goods.

This result contrasts sharply with the advantages of direct democratic institutes in the context of tax evasion, as discussed e.g. by Pommerehne and Weck-Hannemann (1996).

Akin to the structure of this third chapter, the literature relations are quite heterogenous. Our framework allows a structured discussion of various issues related to law enforcement in the context of social norms, which were so far tackled only in a verbal way (e.g. Posner, 1997, 2000; Kahan, 1997, 2005). The analysis contributes to the literature on law enforcement and social stigmatization (Rasmusen, 1996; Arbak, 2005) and provides scope for a Beckerian analysis of optimal deterrence (as e.g. in Kolm, 1973). However, in contrast to the previous literature, we thereby incorporate formal as well as informal law enforcement institutions. The voting analysis in chapter 3 is closely related to the approach of Lindbeck et al. (1999), who study majority voting in the context of social norms and unemployment benefits. The main difference to their approach is that we consider heterogeneity in the incentives related to the social norm, while they study heterogenous incentives related to different income levels. In addition, we focus on the welfare analysis of the voting outcome, omitted in Lindbeck et al. (1999). Thereby we find that in the context of free-riding, the voting outcome determined by the median voter will result in an inefficient equilibrium. Depending on whether cooperators or free-riders are in a majority, voting will result in an under- respectively overprovision of public goods. In this vein, we contribute to the standard literature on collective choice over tax respectively redistribution policies, which neglects the case of free-riding respectively tax evasion (Romer, 1975; Roberts, 1977; Meltzer and Richard, 1981).

Chapter 4: Social Norms and the Evolution of Conditional Cooperation[4]

In the final chapter of this thesis, we study the evolutionary origin of conditional cooperators. In particular we ask, under which circumstances the preferences considered in the previous chapter are evolutionary stable. We address this question using an indirect evolutionary approach (Güth and Yaari, 1992; Güth, 1995). Agents learn about the success ('fitness') of different types of players, characterized by heterogenous degrees of norm sensitivity. According to differences in success, individuals adapt their own level of norm internalization. In this way, the learning and adaptation process shapes the evolution of preferences; cooperative behavior evolves indirectly. Hence, this section

[4] This chapter is based on joint work with Mathias Spichtig, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam.

endogenoulsy derives the preference distribution – in particular, the distribution of the norm sensitivity within a society – which was considered as exogenously given in chapters 2 and 3.

Taking up an idea from biology (e.g. Via and Lande, 1987), we consider agents to adapt to a 'heterogenous environment', i.e. a social situation where they face two different types of environments: Occasions where the norm enforcement is powerful and free-riders suffer from harsh sanctions, as well as scenarios where norm violations are widespread, the social norm has hardly any impact and free-riding constitutes the fitness dominant strategy. We link the multiplicity of equilibria, which arises in our basic model, to the heterogeneity of the environment. We demonstrate that in such environments, conditional cooperators, i.e. agents with an intermediate level of norm sensitivity, dominate free-riders as well as unconditional cooperative players. The intuition for this result is quite straightforward: Conditional cooperators react flexibly to different environments: They cooperate, if the norm is strong and free-ride otherwise. Thereby, they avoid severe sanctions in the first and reap the high free-rider payoff in the second case. This way, they dominate the two unconditional strategies.

The main contribution of chapter 4 is the introduction of a 'multi-habitat' concept of adaptation to a heterogenous environment, which is novel in the evolutionary economic literature. Our analysis highlights a simple and plausible explanation of the evolutionary forces which shape conditional cooperative behavior. Moreover, using an indirect evolutionary approach we endogenously derive a preference structure which so far has been taken as exogenously given (e.g. Akerlof, 1980; Naylor, 1989; Lindbeck, 1999).

Chapter 1

Tax Evasion and Auditing in a Federal Economy*

1.1 Introduction

Many contributions to the taxation literature take tax collection as given or costlessly executed, assuming that tax authorities have full information about individuals or firms' tax liabilities. Of course, this is not very realistic. In fact, we observe significant levels of tax evasion in almost all developed countries. According to estimations of Schneider and Enste (2000), the informal sector accounts for approximately 10 to 30 percent of total GDP in OECD countries and for more than 50 percent in some less developed countries. In order to fight evasion, tax authorities have to spend resources on auditing. From the perspective of revenue maximization, tax authorities should raise auditing up to the point where the marginal gains (less tax evasion, higher revenues from taxation and detected evasion) equal the marginal costs from an increase in tax collection efforts.[1] However, there is some evidence that tax authorities spend too little on tax enforcement.

In Belgium, there were repeatedly public discussions arguing that the Flemish region is too lax in its tax enforcement policy (Cremer and Gahvari, 2000). Recent evidence supports this claim (Crabbé et al., 2004). For Germany, Lenk et al. (1998)

*This chapter is based on joint work with Sven Stöwhase, Fraunhofer Institute for Applied Information Technology FIT, Bonn.

[1]This is the point made in the early literature on tax evasion, see for example Kolm (1973). For the case of welfare maximizing governments Slemrod and Yitzhaki (1987) argue for a lower level of auditing.

argue that some regions put too little effort into tax investigation.[2] Anecdotal evidence supports their view: In 1998, German tax authorities inspected hundreds of banks which were under suspicion to support income tax evasion by transferring non declared income of their customers to bank accounts abroad. Although the operation was highly successful in detecting tax evasion, authorities could only examine a small fraction of all cases in detail: Due to a low number of tax investigators, most evaders escaped without any sanctions and a considerable amount of revenue was lost. If we take this evidence as indicating too low auditing efforts, we have to ask for the reasons that lead to this inefficiency.

One possible reason is tax competition at the regional level. Even if regional governments can not impose their own taxes, there may be scope for fiscal competition if these governments are - at least partially - responsible for tax collection. This is the case for several federal countries, like Australia, Canada, Germany and the US.[3] In the presence of tax evasion, tax revenues are not determined exclusively by the statutory tax rate, but also by the enforcement policy. By choosing a certain auditing level, each region can determine its effective tax rate. Accordingly, the audit policy becomes an alternative strategic tool for tax competition and regions might compete via their effective rather than their statutory tax rate.[4] As has been shown by Cremer and Gahvari (2000), countries will then end up with less than optimal audit rates, even if (statutory) tax rates are harmonized. This may be an important aspect in the context of a EU wide tax harmonization if tax collection remains in commission of the member states.

Another incentive for reducing auditing efforts may derive from fiscal equalization. As shown by Bordignon et al. (2001), tax collection incentives may be distorted if higher revenue reduces the amount of transfers received. Empirical support for this argument is provided by Knight and Li (1999) and Baretti et al. (2002). They find a negative impact of fiscal equalization on revenues collected by Chinese respectively German regions. For the German case, the latter authors argue that disincentives associated with fiscal equalization account for 15% lower tax revenues.

---

[2] See also a report by the Arbeitnehmerkammer Bremen (2001).

[3] While in many countries - e.g. Austria, Belgium, China, Denmark, Spain and the UK - auditing is carried out by regional governments (in some cases together with the central state), only in few countries, like France, tax auditing is completely centralized. Compare Bordignon et al. (2001, p.719), Knight and Li (1999).

[4] Following this reasoning, Crabbé et al. (2004) explain the relatively low effective taxation in the Flemish part of Belgium.

However, there is also scope for fiscal equalization to increase efficiency. As Köthenbürger (2002) demonstrates, interregional redistribution may induce regions to internalize the fiscal externalities associated to their policy. Hence, fiscal equalization could induce higher tax rates – respectively higher audit frequencies – as compared to the case of unrestricted fiscal competition. Taken these two opposing effects together, the net effect of fiscal equalization is ambiguous then and does crucially depend on the design of the equalization scheme.

As has been outlined above, the choice of audit rates may be affected by fiscal competition as well as by the design of the fiscal equalization scheme. While these aspects have been discussed separately in the literature, we combine them in a single framework. This allows us to study the joint incentives from interregional redistribution and competition for the decentralized choice of the audit policy.

We introduce a modified version of the standard model of tax evasion by the firm (e.g., Cremer and Gahvari, 1993) and incorporate it into a tax competition setting. In each region of a perfectly symmetric federal economy a representative firm uses a fixed factor and mobile capital to produce a consumption good. Firms have to pay a tax on capital and decide on how much of the taxes to evade. In making these decisions, each firm takes its evasion costs, the tax rate and the auditing probability into account. Local tax authorities choose their audit rates in order to maximize net revenues, trading off higher auditing costs with revenue increases. The statutory tax rate is determined at the federal level and taken as exogenously given by each region.

In this model, we compare the regions' choice of audit rates for three different cases. The first scenario describes a situation without any fiscal equalization where the choice of the audit rate is only affected by fiscal competition. We show that this will result in inefficiently low levels of auditing, which resembles the classical tax competition result (Zodrow and Mieszkowski, 1986) as well as the findings of Cremer and Gahvari (2000). In scenario two we introduce a system of gross revenue sharing (GRS). The GRS reflects the main properties of the German interstate transfer system (*Länderfinanzausgleich*) and introduces an explicit asymmetry: While tax revenues get shared, auditing costs are fully borne by each region.[5] Finally, scenario three describes an alternative fiscal equalization scheme, net revenue sharing (NRS), under which not only the revenues from taxation, but also the regions auditing costs get shared. For the case of symmetric regions, we show that a system of GRS in general leads to even lower spending on tax

[5] We pick up this example, since the mechanism is particularly illustrative for our analysis. Moreover, the findings from Baretti et al. (2002) allude to the quantitative relevance of the associated disincentives.

enforcement than in the case of unconfined competition. However, NRS would then increase audit rates in comparison to both, the benchmark case and the case of GRS.

The chapter is organized as follows: Section 1.2 presents the basic model. Section 1.3 introduces a benchmark scenario where regions choose their audit policies in the absence of any fiscal equalization scheme. In section 1.4, we analyze the decentralized choice under GRS and NRS. Before concluding with some policy implications we discuss our results in section 1.5. All proofs appear in the Appendix.

1.2 Basic Model

Consider an economy with n regions, each inhabited by a single representative household. In a perfectly competitive industry firms produce one homogenous private good (numeraire). The production process in each region i uses perfectly mobile capital k_i and a fixed, immobile factor. The technology is represented by a standard neoclassical production function $f(k_i)$, where the fixed factor is suppressed. We assume a perfectly symmetric economy where all regions use the same technology and are endowed with the same amount of the fixed factor. Firms have to pay a unit tax on capital at a rate t. This *statutory* tax rate is equal for the whole economy. However, each firm can try to evade taxes by concealing a share e_i of the capital employed. To conceal inputs requires the use of resources by the firm. Following the literature, we assume that the costs of evasion are convex in e_i and linear in the tax base:[6] $g(e_i)k_i$ with $g' > 0$ and $g'' > 0$. With a probability p_i the evasion gets detected, and the firm has to pay the statutory taxes plus a fine that is proportional to the taxes evaded (Yitzhaki, 1974). For all regions, the penalty rate is $s - 1$, with $s > 1$. With probability $1 - p_i$ the firm gets away with the evasion and pays only taxes on the declared amount of capital. Expected profits π_i^e are defined as

$$\pi_i^e = f(k_i) - rk_i - g(e_i)k_i - p_i(tk_i + (s-1)e_i tk_i) - (1-p)(1-e_i)tk_i$$

where r is the factor price for capital. We can simplify this expression to

$$\pi_i^e = f(k_i) - (r + g(e_i) + t_i^e)k_i \tag{1}$$

[6] This assumption makes the firms' evasion decision independent of the amount of capital employed (compare equation 2). Our main results derived below, do also hold for more general assumptions.

with $t_i^e \equiv t(1 - e_i + e_i p_i s)$, the expected or *effective tax rate* in region i. Note that revenues from detected evasion (including penalties) are also included in the definition of the effective tax rate.

Taking the policy variables as given[7], a risk neutral firm chooses k_i and e_i to maximize its expected profit. The firms optimal choice is then given by the following system of first order conditions

$$\begin{aligned} g'(e_i) &= (1 - p_i s)t, &(2)\\ f'(k_i) &= r + g(e_i) + t_i^e. &(3) \end{aligned}$$

For the rest of the chapter, we will assume that there is an interior solution with evasion in equilibrium (i.e., $p_i s < 1$). From (2) and (3) one can easily derive two basic results (see the Appendix). First, firms will conceal more if the statutory tax rate increases or the detection probability decreases – a standard result for models of firm tax evasion (Cremer and Gahvari, 1993). Second, an increase in the audit rate will raise (per unit) capital costs and hence decrease capital demand. This triggers an effect which is analogous to the impact of a tax increase on mobile capital in tax competition models.

Finally, we describe the capital market. In each region i a representative household is endowed with capital $\bar{k}_i$ and one unit of the immobile factor. The total capital supply to the economy $\bar{k}$ is fixed and market clearing requires

$$\sum_{i=1}^{n} \bar{k}_i = \bar{k} = \sum_{i=1}^{n} k_i \tag{4}$$

Individuals invest their capital in a large number of firms distributed over the whole economy. By holding a fully diversified portfolio, they avoid the potential risks associated with tax evasion by firms. In the capital market equilibrium the arbitrage condition

$$f'(k_i) - g(e_i) - t_i^e = r \tag{5}$$

has to be fulfilled for all regions i.

[7]For an analysis of commitment problems, see Reinganum and Wilde (1986).

1.3 Auditing Policy without Fiscal Equalization

Let us now turn to the regional planners' policy choice. Throughout the whole chapter we assume perfectly symmetric regions[8]. As a benchmark scenario we consider a federal economy without any interregional redistribution. Following the literature on tax evasion, the objective of the regional governments (or tax authorities) is revenue maximization[9].

The tax and the penalty rate are (exogenously) determined at the federal level. Hence, the only policy variable controlled by the regional government is the audit rate, which determines the capital allocation and thereby the regions' revenues from taxes and penalties. Each region bears the full costs associated with auditing. Assuming that these costs are linear in the tax base (i.e. the level of firms' capital inputs), we define the total detection costs of a region as $c(p_i)k_i$, where $c(p_i)$ denotes the auditing costs per unit of capital (as a function of p_i), with $c' > 0$, $c'' > 0$ and $c(0) = 0$.[10,11] In the absence of an interregional redistribution mechanism the net revenue of region i is given by

$$k_i(t_i^e - c(p_i)), \tag{6}$$

with the effective tax rate t_i^e as defined above.[12] While the statutory tax rate in the economy is 'harmonized' and hence there is no scope for standard tax competition between regions,[13] the detection policy acts as a strategic substitute for the regional policymaker: By reducing its auditing rate, a region can lower the effective tax rate. This will reduce capital costs and attract mobile capital from other regions. Taking the capital market responses into account and considering the policy of the other regions as fixed, the first order condition becomes

$$k_i \frac{\partial t_i^e}{\partial p_i} + t_i^e \frac{\partial k_i}{\partial p_i} = MC_i, \tag{7}$$

[8] The results derived in section 3 do also hold for the case of asymmetric regions.

[9] This can be justified either by a 'Leviathan' government or by a welfare maximizing government in the case of consumers which receive significantly higher marginal utilities from public than from private consumption (see e.g. Kanbur and Keen, 1993).

[10] We suppose that this cost function is exogenously given in the short run and that it can not be influenced by the regional government.

[11] One could consider a detection technology characterized by marginal auditing costs which are increasing or decreasing in the tax base. However, non-linearity would not affect our results in a qualitative way.

[12] At first sight it is not clear why evasion should be possible in the presence of a planner who perfectly knows the size of the tax base. However, if there are many producers and the tax authority does not know the exact distribution of the capital among the firms, there is scope for tax evasion.

[13] In this point we follow the institutional arrangements in Germany.

with

$$MC_i = c'(p_i)k_i + c(p_i)\frac{\partial k_i}{\partial p_i}.$$

(The term $\partial t_i^e/\partial p_i$ is derived in the Appendix. There we also show that t_i^e is concave in p_i.) Condition (7) implies a system of reaction functions which determine the Cournot-Nash equilibrium of the uncoordinated auditing choice.[14]

In the optimum, marginal benefits from auditing, depicted by the terms on the LHS of (7), have to be equal to MC_i which denotes the 'extended' marginal costs of auditing.[15] The marginal benefit on the LHS consists of two effects: The marginal increase in the effective tax rate (weighted with the tax base) and the marginal capital outflow which follows from an increase in p_i (weighted with the effective tax rate). These capital outflows clearly lower the marginal benefit of the region.

Let us now compare the decentralized choice to the centralized solution of the problem. Suppose that a central planner would choose a detection policy for each region in order to maximize the sum of all revenues,

$$\max_{p_1,\ldots,p_n} \sum_{i=1}^{n} k_i(t_i^e - c(p_i)).$$

The n first order conditions are given by

$$k_i\left(\frac{\partial t_i^e}{\partial p_i} - c'(p_i)\right) + \frac{\partial k_i}{\partial p_i}(t_i^e - c(p_i)) + \sum_{j\neq i}\frac{\partial k_j}{\partial p_i}(t_j^e - c(p_j)) = 0 \tag{8}$$

While the first two terms also appear in condition (7), the third term in (8) represents the revenue spillovers created by a region's detection policy.[16] For the case of symmetric regions, the planner will chose one unique auditing level $p_i = p_j\ \forall i, j$ and hence $t_i^e = t_j^e\ \forall i, j$. Since capital flows must be balanced, the second and third term in condition (8) cancel out and we get

$$\partial t_i^e/\partial p_i = c'(p_i) \tag{9}$$

Comparing (7) with (9) we can derive

[14]One can easily show that the regions' audit rates are strategic complements for all scenarios considered.

[15]Of course, a revenue maximizing planner neglects any further costs associated from auditing (e.g. compliance costs).

[16]Note that condition (8) corresponds to the case where one would provide each region with a corrective subsidy in order to internalize its fiscal externalities. This replicates the result of Wildasin (1989) for the case of fiscal competition in audit rates.

Proposition 1 *In the absence of fiscal equalization, the uncoordinated choice of auditing policies in an open economy with mobile capital will lead to audit rates which are below the auditing level a central planner would choose.*

The intuition for this result is straightforward: For an open region the capital outflows reduce the marginal benefit of an increase in detection efforts. However, the capital outflows from one region will enlarge the 'foreign' tax base and, ceteris paribus, raise net revenues in the rest of the economy. Since regional decision makers ignore these well known fiscal externalities, they set lower detection probabilities than a central planner. Hence, the uncoordinated policy choice leads to audit rates which are 'inefficiently low' from the perspective of total revenue maximization. This further means that firms will choose a level of evasion, which is 'inefficiently high' – again from a revenue maximizing perspective.

Proposition 1 describes our benchmark result. In the next section, we will compare this result with the regions' uncoordinated policy choice in the presence of different fiscal equalization schemes.

1.4 Auditing Policy with Fiscal Equalization

In this section we introduce fiscal equalization into the model. We analyse two different schemes of equalizing transfers: Gross revenue sharing (GRS) and net revenue sharing (NRS).

1.4.1 Gross Revenue Sharing

Consider the following simple redistribution mechanism: Each of the n regions contributes a share $0 < \alpha < 1$ of its gross revenues $t_i^e k_i$ to the redistribution system and receives a share $1/n$ of the total revenues distributed, $\alpha \sum t_j^e k_j$. This fiscal equalization scheme captures a central feature of the current German interstate transfer system (Länderfinanzausgleich): While revenues obtained under a given detection policy have to be shared with all other regions, the costs for maintaining a certain audit rate have to be fully borne by the region itself. The net revenue of a region is then given by

$$(1-\alpha)t_i^e k_i - c(p_i)k_i + \frac{\alpha}{n}\sum_{j=1}^{n} k_j t_j^e \tag{10}$$

We assume $t_i^e \gg c(p_i)$ and redistribution is not 'too extreme', such that $(1-\alpha)t_i^e > c(p_i)$ holds in equilibrium. The first order condition for the uncoordinated choice of a revenue maximizing regional government is

$$(1-\alpha)\left(k_i \frac{\partial t_i^e}{\partial p_i} + t_i^e \frac{\partial k_i}{\partial p_i}\right) + \frac{\alpha}{n}\left(\sum_{j=1}^{n} t_j^e \frac{\partial k_j}{\partial p_i} + k_i \frac{\partial t_i^e}{\partial p_i}\right) = MC_i \tag{11}$$

with MC_i as defined above. The first term on the LHS depicts the distortion introduced by the fiscal equalization scheme, whereas the second term shows that a part of the fiscal spillovers will be internalized via the redistribution mechanism. For a proper analysis, we have to distinguish between an economy with many *small* or few *large* open regions.

Small open regions

In the case of many small regions ($n \to \infty$), the impact of a single region's detection policy on total tax revenues in the economy becomes negligible. Hence, the second term on the LHS of (11) vanishes and the first order condition becomes

$$(1-\alpha)\left(k_i \frac{\partial t_i^e}{\partial p_i} + t_i^e \frac{\partial k_i}{\partial p_i}\right) = MC_i \tag{12}$$

If we compare (12) with condition (7), the result in the scenario without any inter-regional redistribution, one can easily see that GRS distorts the region's choice of detection efforts:[17] While the marginal costs MC_i are unaffected, the redistribution system clearly reduces the marginal gains from an increase in the audit rate. This is the case, since fiscal equalization leads to an implicit taxation of a region's gross revenue. Hence, regions will unambiguously choose a lower audit frequency than in the absence of fiscal equalization.

[17]In general, the equilibrium audit rates will differ between the compared scenarios. However, in a symmetric equilibrium between symmetric regions this will not affect the capital allocation. This is the case, since total capital supply is exogenous and for a given scenario equilibrium audit rates will be the same in all regions. Hence, for any equilibrium there holds $k_i = \bar{k}/n$, which allows us to compare first order conditions across scenarios.

Proposition 2 *Under a system of gross revenue sharing, a small open region maximizing its net revenue will choose a detection probability which is even lower than in the case without fiscal equalization.*

This result is equivalent to Proposition 3 in Köthenbürger (2002). As in his analysis, the inefficiency due to the distorted incentives adds to the inefficiency due to fiscal competition: Detection efforts will be reduced further below the 'efficient' level a central planner would set and the amount of taxes evaded will be higher than in our reference case without any interregional redistribution system. Moreover, it is straightforward to show that the audit rate will decrease as α increases.

Large open regions

Since the decision maker of a large region takes into account the impact of the local detection policy on the total revenues distributed, we can rearrange condition (11) and get

$$\left(1-\alpha\left(1-\frac{1}{n}\right)\right)\left(k_i\frac{\partial t_i^e}{\partial p_i}+t_i^e\frac{\partial k_i}{\partial p_i}\right)+\frac{\alpha}{n}\sum_{j\neq i}^{n} t_j^e\frac{\partial k_j}{\partial p_i}=MC_i \tag{13}$$

As in the case of small open regions, GRS introduces a distortion. However, since large regions consider the strategic effect of their policies – the impact of their audit rate on total revenues redistributed – the distortion becomes smaller: $1-\alpha+\alpha/n > 1-\alpha$. Stated differently, for large regions the implicit taxation of gross revenues drops below the rate α observed for small regions. Moreover, there is a further incentive embedded in the GRS: As reflected in the second term on the LHS of condition (13), a part of the fiscal externalities[18] gets internalized. Again, this is due to the fact that large regions – in contrast to small ones – take into account the strategic effect of their policies.

If we compare this case with the benchmark scenario, the introduction of GRS shows an ambiguous effect on the choice of audit rates. While the implicit taxation of gross revenues tends to lower auditing efforts, the internalization of the spillovers will act in the other direction. Comparing condition (13) and (7), we can find a threshold

[18]Note that the total spillover would be $\sum_{j\neq i}\frac{\partial k_j}{\partial p_i}(t_j^e-c(p_j))$.

level such that, for $n < \hat{n}$, with

$$\hat{n} = 1 + \frac{\sum\limits_{j \neq i}^{n} t_j^e \frac{\partial k_j}{\partial p_i}}{k_i \frac{\partial t_i^e}{\partial p_i} + t_i^e \frac{\partial k_i}{\partial p_i}}, \tag{14}$$

the second effect dominates. We sum up these results in

Proposition 3 *Suppose large open regions and revenue maximizing governments under gross revenue sharing. There exists a threshold $1 < \hat{n} < \infty$, such that for $n < \hat{n}$ ($n > \hat{n}$) the detection probabilities chosen under a system of gross revenue sharing are higher (lower) than in the case without fiscal equalization.*

While GRS has an unambiguously negative effect in the case of small regions (see Proposition 2), it could increase audit rates for large open regions.[19] This result is to some extent surprising. At first glance, a GRS scheme seems to lead to a clear inefficiency: The asymmetric treatment of auditing costs on the one hand, and tax revenues on the other, unambiguously distorts a region's choice of the auditing effort. However, in the presence of fiscal competition, the redistribution system has a further effect: It works as a corrective subsidy and induces large regions to internalize a part of the fiscal externalities. While the distortion from the implicit taxation tends to lower audit rates, the corrective subsidy works in the opposite direction. If there are $n < \hat{n}$ jurisdictions, the second effect dominates the first one and GRS makes the decentralized choice more 'efficient' (in terms of revenue maximization), compared to the benchmark scenario. The intuition for this result is clear-cut: If the number of regions diminishes, the implicit taxation of gross revenues becomes smaller and the degree of internalization gets higher - irrespective of α (see condition (13)). Following this line of reasoning it is straightforward to show that an increase in α further amplifies the efficiency enhancing (reducing) effect of GRS, as long as the mechanism induces a region to raise (lower) its auditing effort for the case of $n < \hat{n}$ ($n > \hat{n}$). One can also show that an increase in n, the number of regions, would result in lower audit rates. This is the case, since more regions will always increase the distortion of the redistribution and decrease the degree of internalization.

[19] Another issue we do not take up here, is the fact that large regions face a lower capital elasticity than small regions because of a different impact of their detection policies on the interest rate. Compare e.g. Bucovetsky (1991).

1.4.2 Net Revenue Equalization

Let us now introduce an alternative system of interregional redistribution. Instead of gross revenue sharing, we consider a mechanism which is based on net revenue sharing. Each region contributes a share $0 < \alpha < 1$ of its net revenues – tax revenues net of auditing costs – and receives a share $1/n$ of the total revenues distributed, $\alpha \sum k_j(t_j^e - c(p_j))$. With this mechanism, the revenue of a region becomes

$$(1-\alpha)k_i(t_i^e - c(p_i)) + \frac{\alpha}{n}\sum_{j=1}^{n} k_j(t_j^e - c(p_j)). \tag{15}$$

The first order condition is given by

$$(1-\alpha)\left(k_i\frac{\partial t_i^e}{\partial p_i} + t_i^e\frac{\partial k_i}{\partial p_i} - MC_i\right) + \frac{\alpha}{n}\left(\sum_{j=1}^{n}\frac{\partial k_j}{\partial p_i}(t_j^e - c(p_j)) + k_i\left(\frac{\partial t_i^e}{\partial p_i} - c'(p_i)\right)\right) = 0 \tag{16}$$

As before, we discuss this condition separately for the cases of small and large open regions.

Small open regions

As under GRS the second term on the LHS of condition (16) vanishes for the case of many small regions. We can rewrite (16) as

$$k_i\frac{\partial t_i^e}{\partial p_i} + t_i^e\frac{\partial k_i}{\partial p_i} = MC_i, \tag{17}$$

which is identical to condition (7) in the benchmark scenario. While GRS distorts the auditing policy of small regions (Proposition 2), the distortion disappears under NRS. We can state

Proposition 4 *Under a system of net revenue sharing a small open region maximizing its net revenue will choose the same audit rate as in the benchmark scenario without fiscal equalization.*

The intuition for this result is clear: In contrast to the scenario of GRS, there is no asymmetric treatment of revenues and auditing costs under NRS. Therefore, revenue maximizing regions would unambiguously choose a higher audit rate after a change from GRS to NRS. The only inefficiency remaining arises from the competition for the mobile tax base. As in the benchmark scenario, auditing efforts will be 'inefficiently' low and the evasion level of the firms will be 'inefficiently' high from the perspective of total revenue maximization.

An interesting point to note is that Proposition 4 holds true for any α. Since inter-regional redistribution does not introduce any distortion, there is no equity-efficiency trade off in the choice of α.[20]

Large open Regions

As before, the government of a large open region considers the impact of its auditing policy on the tax base in the rest of the economy. We can rearrange condition (16) and get

$$k_i \frac{\partial t_i^e}{\partial p_i} + t_i^e \frac{\partial k_i}{\partial p_i} + \beta \sum_{j \neq i}^{n} \frac{\partial k_j}{\partial p_i} (t_j^e - c(p_j)) = MC_i \tag{18}$$

with

$$\beta \equiv \frac{\alpha}{n(1-\alpha)+\alpha}$$

As in the case of small open regions discussed above, NRS does not distort the decentralized choice of detection probabilities. The inefficiency due to the fiscal competition between regions still remains. However, since the decision maker of a large open region incorporates the strategic effect of the auditing policy, a fraction β of the fiscal externality gets internalized. Compared to our benchmark scenario, the NRS will lead to higher audit rates in the case of large regions. If we contrast the impact of NRS with that of GRS, the former system introduces no distortion but – in most cases – a higher corrective subsidy.[21]

Proposition 5 ***(i)*** *Under a system of net revenue sharing, a large open region maximizing its net revenue will set an audit rate which is higher than the rate chosen in the benchmark scenario without any fiscal equalization.* ***(ii)*** *Moreover, if $n > \hat{n}$, the audit rate is also higher than the rate chosen under gross revenue sharing.*

[20] This property (as well as Proposition 4 per se), only holds true since in our model we have excluded any income effect from redistribution.

[21] Although $\beta > \alpha/n$ the comparison with the case of GRS - condition (13) - is not trivial, since the redistribution volume is smaller under NRS.

Proposition 5 is the main result of our analysis: For large open regions, a NRS scheme partly internalizes fiscal externalities without introducing any distortion. Given that the sufficient condition $n > \hat{n}$ holds, the decentralized choice of detection policies under NRS will result in higher auditing efforts as compared to both, the case without any equalization transfers and the case with a GRS. Incorporating auditing costs into the redistribution mechanism will only slightly reduce the amount redistributed (for a given α) but increase detection probabilities. Firms would evade less and tax revenues would be higher. Interestingly, for $n < \hat{n}$ a GRS mechanism could in principle induce higher audit rates than a NRS mechanism. This result stems from the fact, that NRS slightly reduces the redistributive volume and that for small n the distortion embedded in the GRS vanishes. However, in the Appendix we show that $n > \hat{n}$ is only a sufficient condition, and Proposition 5 (ii) will in general also hold for $n < \hat{n}$.

From (18) one can further show, that an increase in the number of regions would lead to a lower degree of internalization and therefore to a decline in audit rates. In contrast, raising the level of redistribution would result in a stronger internalization of externalities, without causing any distortion. Hence, as discussed above in the context of Proposition 4, there is no equity-efficiency trade-off.

1.5 Discussion

Asymmetric Information

In the last section we concluded that a switch from GRS to NRS will typically enhance tax collection efforts. However, our modeling approach neglects a severe disadvantage associated with NRS which renders this transfer mechanism relatively unattractive. Under NRS, regions have a clear incentive to overstate their auditing costs, if there is asymmetric information between different layers of the government[22]. Hence, as Bordignon et al. (2001) point out, information asymmetries introduce a new source of inefficiency for the choice of tax enforcement efforts. As long as there are no institutional rules which prevent regions from abusing their informational advantage, the feasibility of NRS is unclear.[23]

[22]Regions could simply declare other administrative costs as expenditures on auditing.

[23]To the best of our knowledge, the concept of NRS is so far only a theoretical one as it is not implemented in any country.

Nevertheless, if the (gross) revenue of a region is observable to the federal government, the central authority could use this information as a proxy for the underlying audit rate and the corresponding costs. This could imply an upper bound for the declared costs. While it is not clear, whether in this case a NRS would increase enforcement efforts compared to GRS, the former mechanism would clearly reduce the level of fiscal equalization.

Asymmetric Regions

In deriving our main results, we have restricted our analysis to a perfectly symmetric economy: Evasion as well as auditing technologies are the same for all regions, jurisdictions are of the same size and production technologies are identical. The equilibrium is therefore characterized by symmetric capital allocations and equal (net) revenues. Hence, there is no need for equalization transfers. While in such a scenario the analysis of interregional redistribution seems rather artificial, the incentives embedded in different revenue sharing mechanism are equally at work in a scenario of heterogeneous regions. In the case of different production technologies, however, asymmetric capital allocations would render the comparison of GRS and NRS less clear cut. Since these size-related effects are well studied in the literature[24] and work here in a very similar way,[25] incorporating them would only blur the results of the study without gaining further insights.

Of course, one could allow for further levels of heterogeneity, e.g. with respect to auditing and evasion costs. However, this may better fit into a simulation based rather than an analytical approach. Calibrating the model and comparing the results with existing empirical evidence – especially Baretti et al. (2002) – would be an interesting task for further research.

Welfare Maximization

One could easily provide a standard welfare analysis within this framework, by explicitly modelling consumers, who receive income from their endowments of capital and the fixed factor and who derive utility from private consumption as well as from a regional public good. A welfare maximizing regional planner would then choose an audit rate that maximizes the utility of the representative consumer.

[24] Compare e.g. Bucovetsky (1991).

[25] Small regions with a lower level of capital employed tend to use 'more aggressive' strategies.

The most important difference to the case of revenue maximization is the appearance of a further externality. A welfare maximizing planner will also consider the pecuniary effects of its detection policy: A higher audit rate will reduce the capital income as well as the fixed factor income of the regional consumer. The latter effect arises, since a capital outflow will lower the marginal productivity of the fixed factor. However, the loss of capital in one region will lead to an inflow of capital and therefore to higher fixed factor incomes in the rest of the economy. On the other hand, all capital owners, domestic as well as foreign, will receive lower income from capital since an increase in audit rates will decrease the interest rate. Since a regional planner does not take into account the impact on factor incomes in the rest of the economy, there is a pecuniary externality - which can be either positive or negative. Hence, in the case of unconfined fiscal competition, there are now two externalities - a fiscal and a pecuniary one - which render the decentralized choice of audit rates inefficient.

How does the introduction of gross or net revenue equalization affect efficiency in this case? As discussed above, revenue equalization provides a mechanism which makes the decentralized planner (partially) internalize the fiscal externalities. This also holds for welfare maximizing planners. The pecuniary externality, however, will not get internalized via an interregional redistribution mechanism. With this additional externality, different revenue sharing mechanism will not induce a first best solution. In general, however, the choice of decentralized auditing policies will be more efficient under NRS than under GRS.

Central Government Policy

Another limitation of our model is the exogenous choice of the tax policy. Following the institutional framework in Germany, we have assumed that there is a harmonized tax rate, chosen by the federal government. The central government might counterbalance low detection probabilities by raising the statutory tax rate. However, in our framework this policy is not necessarily feasible. A comparative static analysis for the different scenarios shows the ambiguous result

$$\frac{dp_i}{dt} \gtreqless 0$$

Hence, an increase in the statutory tax rate might further reduce auditing efforts.

The other policy instrument available to the central government is the penalty rate. Since punishment is costless and enforces a lower level of evasion, a central planner could set $s \to \infty$ and there would be no evasion at all (see Kolm, 1973). However, strong penalties are probably not feasible: If a firm would go bankrupt because of a very severe punishment, the penalty may not be credible. Keeping this restriction in mind and assuming that s is fixed at some credible rate (with $p_i s < 1$) appears plausible.

1.6 Conclusion

In this chapter, we have analyzed the decentralized choice of audit rates for the case of symmetric regions. As a benchmark result we show that fiscal competition will lead to detection probabilities which are inefficiently low from the perspective of revenue maximization. In such a framework, a fiscal equalization scheme has an efficiency enhancing potential, since interregional redistribution provides a mechanism to internalize fiscal externalities (see Köthenbürger, 2002). We first consider a system of GRS, which makes regions bear the full auditing costs while tax revenues get shared. This asymmetric treatment of costs and revenues introduces a further distortion, which tends to lower audit rates. As an alternative to GRS, we introduce a system of NRS. Under this mechanism both costs and revenues are shared and therefore NRS does not create any distortion. We show that NRS generally induces higher detection probabilities for large regions. Results for small regions are equal to those in the benchmark case of unconfined fiscal competition.

The policy implications of these results are straightforward. A federal government which, on the one hand, equalizes tax revenues between its regions but, on the other, imposes the costs of tax collection upon these regions, will face a higher degree of tax evasion. This may well be the case for Germany, since the current fiscal equalization system corresponds to a GRS scheme. Switching to NRS should lead to more auditing, less tax evasion and to higher net revenues, while redistribution, the primary objective of fiscal equalization, would hardly be affected. However, if there is asymmetric information between central and local governments regarding a region's auditing costs, a system of NRS may not be a feasible instrument since jurisdictions could easily overstate their enforcement costs.

An alternative which could induce the first best solution would be the centralized choice of tax enforcement policies – as practiced in centralized countries such as France and as recently proposed by the German Ministry of Finance. Nevertheless, following the literature on fiscal federalism (Oates, 1972), there may exist several disadvantages of a centralized policy our model does not account for. Hence, we can not argue in favor of a centralization. Furthermore, if we consider the scenario of tax harmonization within the EU, centralized tax collection appears infeasible. Similarly, the harmonization of enforcement policies would be a difficult (if not impossible) task, since – in contrast to taxation – there are hardly any transparent and contractible indicators for the level of tax enforcement. In this case, audit policies become alternative strategic tools for fiscal competition. To limit competitive forces, it is therefore important to study different mechanisms, which help to induce efficiency in decentralized tax collection. Clearly, further research is needed in order to design a mechanism which is also feasible in the context of information asymmetries.

Appendix

Comparative Statics (Section 1.2)

Using the implicit function theorem on (2) we get

$$\frac{\partial e_i}{\partial p_i} = -\frac{st}{g''} < 0, \tag{A.1}$$

$$\frac{\partial e_i}{\partial t} = \frac{1 - p_i s}{g''} > 0 \tag{A.2}$$

Applying the implicit function theorem on equation (3) and making use of (2) we get

$$\frac{\partial k_i}{\partial p_i} = \frac{ste_i + \frac{\partial r}{\partial p_i}}{f''} < 0. \tag{A.3}$$

Effective Tax Rate (Section 1.3)

The effective tax rate is given by $t_i^e \equiv t(1 - e_i + e_i p_i s)$. We can easily derive

$$\frac{\partial t_i^e}{\partial p_i} = t\left(\frac{st}{g''}(1 - p_i s) + e_i s\right) > 0 \tag{A.4}$$

where we made use of (A.1) and $p_i s < 1$. From (A.4) and substituting (A.1) we get

$$\frac{\partial^2 t_i^e}{\partial p_i^2} = -\frac{(st)^2}{g''}\left(2 - \frac{tg'''(1 - p_i s)}{(g'')^2}\right) < 0 \tag{A.5}$$

where we assume that the first order effect dominates.

Proof of Proposition 1. We can rewrite condition (7) as

$$\frac{\partial t_i^e}{\partial p_i} + \Psi\left(t_i^e - c(p_i)\right) - c'(p_i) = 0 \tag{A.6}$$

with

$$\Psi = \frac{1}{k_i}\frac{\partial k_i}{\partial p_i}$$

and condition (9) as

$$\frac{\partial t_i^e}{\partial p_i} - c'(p_i) = 0. \tag{A.7}$$

We know from (A.3) that $\Psi < 0$. Therefore the second term in (A6) is negative and the LHS of (A.6) is smaller then the LHS of (A.7) for any p_i. ■

Proof of Proposition 2. In order to compare conditions (7), respectively (A.6), and (12), we rearrange the latter and get

$$(1-\alpha)\frac{\partial t_i^e}{\partial p_i} + \Psi\left((1-\alpha)\, t_i^e - c(p_i)\right) - c'(p_i) = 0 \tag{A.8}$$

Note that in general the equilibrium audit rates will differ between the compared scenarios. However, in a symmetric equilibrium between perfectly symmetric regions, different levels of p_i do not affect the capital allocation, since the capital supply in the economy is exogenous and audit rates will be the same in all regions. This implies that (for any p_i) Ψ is the same for the different scenarios compared.
The comparison of (A.6) and (A.8) shows that for $\alpha > 0$ the LHS of (A.8) is smaller than the LHS of (A.6) for any p_i. ■

Proof of Proposition 3. Comparing (13) with (7) we get the threshold defined in (14). Note that in the case of symmetric regions there must hold

$$\sum_{j\neq i} t_j^e \frac{\partial k_j}{\partial p_i} + t_i^e \frac{\partial k_i}{\partial p_i} = 0 \tag{A.9}$$

in any symmetric equilibrium. Applying (A.9) on (14) we get

$$\hat{n} \equiv \frac{k_i \frac{\partial t_i^e}{\partial p_i}}{k_i \frac{\partial t_i^e}{\partial p_i} + t_i^e \frac{\partial k_i}{\partial p_i}}. \tag{A.10}$$

Since the denominator of $\hat{n}$ is positive it follows from (A.3) that $\hat{n} > 1$. For the limit $\left| t_i^e \frac{\partial k_i}{\partial p_i} \right| \to k_i \frac{\partial t_i^e}{\partial p_i}$ we get $\hat{n} \to \infty$.
Using (A.9) we can rewrite condition (13) in the following way

$$\left(1-\alpha\left(1-\frac{1}{n}\right)\right)\frac{\partial t_i^e}{\partial p_i} + \Psi\left((1-\alpha)\, t_i^e - c(p_i)\right) - c'(p_i) = 0 \tag{A.11}$$

One can easily show that for $n < \hat{n}_i$ ($n > \hat{n}_i$) the LHS of condition (A.11) is higher (lower) than the LHS in (A.6) for any p_i. ■

Proof of Proposition 4. Since the capital allocation is the same for the different scenarios under consideration (see the Proof of Proposition 2), condition (17) and (7) are identical. ■

Proof of Proposition 5. **(i)** Analogous to (A.9) there also has to hold

$$\sum_{j\neq i}\frac{\partial k_j}{\partial p_i}\left(t_j^e-c(p_j)\right)+\frac{\partial k_i}{\partial p_i}\left(t_i^e-c(p_i)\right)=0. \tag{A.12}$$

Using (A.12) we can rearrange (18) and get

$$\frac{\partial t_i^e}{\partial p_i}+\gamma\Psi\left(t_i^e-c(p_i)\right)-c'(p_i)=0 \tag{A.13}$$

with

$$\gamma=\frac{1-\alpha}{1-\alpha+\frac{\alpha}{n}}.$$

Since $\Psi<0$ and $0<\gamma<1$ (for $\alpha>0$), the LHS in (A.13) is larger than the LHS of (A.6) for any p_i.

(ii) The LHS of (A.13) is larger than the LHS of (A.11) for any p_i, if

$$\alpha\left(\left(1-\frac{1}{n}\right)k_i\frac{\partial t_i^e}{\partial p_i}+t_i^e\frac{\partial k_i}{\partial p_i}\right)+\frac{\alpha}{(1-\alpha)\,n}k_i\left(\frac{\partial t_i^e}{\partial p_i}-c'(p_i)\right)>0 \tag{A.14}$$

Since the first order conditions imply $\frac{\partial t_i^e}{\partial p_i}>c'(p_i)$, the second term in (A.14) is positive. Therefore it is sufficient for condition (A.14) to hold, if the expression in the round brackets in the first term is positive. Using (A.10), one can show that this is equivalent to $n>\hat{n}$. ■

Chapter 2

Social Norms and Conditional Cooperative Taxpayers

2.1 Introduction

Income tax evasion is a serious problem in most OECD countries and even more so in less developed economies (Andreoni et al., 1998). In order to come up with appropriate policy measures to fight evasion, a deep understanding of taxpayers behavior is essential. The conventional economic approach to tax evasion (Allingham and Sandmo, 1972) only provides an unsatisfactory fundament for this understanding, as it conflicts with empirical findings in several major points. Based on Gary Beckers (1968) approach to the economics of crime, tax evasion is modeled as a risky gamble, where taxpayers trade off the risk and cost of detection with the chance of getting away with the evasion.[1] The evasion decision is thereby equivalent to an optimal portfolio choice problem, as individuals decide on how much of their income to declare – invest in a safe asset – and how much to conceal – invest in a risky asset. As most fiscal systems are characterized by rather low audit rates and penalties (Skinner and Slemrod, 1985), a rational taxpayers with a reasonable degree of risk aversion should invest into the risky asset and conceal some income (Bernasconi, 1998). In contrast to this prediction, however, we observe that many households fully comply with tax laws. The main flaw of the theory is therefore its failure to explain the level of tax compliance documented in many countries.[2] Consequently, research in the past two decades has focused on the

[1] Cowell (1990) provides a comprehensive discussion of the basic model framework and several extensions. Compare also Andreoni et al. (1998) and Sandmo (2004).

[2] Compare e.g. Alm et al. (1992), Graetz and Wilde (1995) as well as Andreoni et al. (1998).

question why people pay taxes (Alm et al., 1992; Slemrod, 1992).

Recent attempts to solve the compliance puzzle rest upon prospect theory and other deviations from standard expected utility theory.[3] However, there is evidence which cast doubt on the relevance of these approaches for the case of tax evasion. Baldry (1986), for example, shows that individuals who are prepared to take a certain amount of risk in a conventional lottery, stick to the safe asset once the lottery is framed as a tax evasion game.[4,5] Hence, cheating on the government seems to be more than just a gamble. As expressed by Agnar Sandmo (2004, p.11), '*people refrain from tax evasion* [...] *not only from their estimates of the expected penalty, but for reasons that have to do with social and morale considerations.*' This statement is supported by ample empirical evidence, which demonstrates the importance of 'tax morale' – taxpayers attitudes, moral values and perceptions about civic duty – for the decision to conceal income.[6] Several studies emphasize that the perceived evasion of other taxpayers plays a crucial role in the determination of tax morale (e.g. Porcano, 1988; Spicer and Hero, 1985). Survey data typically show that those individuals who believe that most other taxpayers are honest, consider evasion as a more serious wrongdoing than those, who overestimate the level of evasion (Körner and Strotmann, 2005; Torger, 2005). This evidence can be also linked to the observed interdependency of evasion decisions: If evasion is believed to be widespread this increases the propensity to cheat on taxes (Feld and Tyran, 2002; Geeroms and Wilmots, 1985).

In this chapter we account for these findings and provide an analytical framework which incorporates the concept of tax morale into the standard approach by Allingham and Sandmo (1972). We model tax morale as an internalized social norm for tax compliance (Elster, 1989). This norm works as a soft constraint on action, which render evasion 'costly': If a taxpayer deviates from the norm and conceals income, conduct is not in line with the individuals self-image as a 'good' member of society, who complies with societal norms and expectations. The strength of the social norm, and thereby the self-imposed sanctions associated with a norm violation, are determined by the individual specific degree of norm internalization as well as the share of evaders

[3] Bernasconi (1998), Bernasconi and Zanardi (2004), Dhami and Ali al-Nowaihi (2005), Eide (2002).

[4] According to prospect theory, framing should only matter if the 'reference point' of the prospect, e.g. related to the initial endowment, would change.

[5] Compare also Reckers et al. (1994). Further experimental evidence is discussed by Gërxhani and Schram (2005).

[6] See e.g. Alm et al. (1992, 1995), Feld and Frey (2002), Frey and Feld (2002), Güth and Sausgruber (2004), Güth et al. (2005), Lewis (1986), Orviska and Hudson (2002), Reckers et al. (1994), Scholz and Pinney (1995). Compare also Chapter 6 in Cowell (1990) and Torgler (2002) as well as vast survey evidence summarized by Torgler (2005).

in the society. The more people deviate from the norm, the weaker the compliance norm becomes. Taxpayers act *conditionally cooperative*,[7] as they condition their tax compliance on the honesty of other members of the society. This interdependence of evasion behavior will typically result in a multiplicity of equilibria.

We consider a large, heterogenous population. While some agents only maximize the expected return from the evasion gamble, others are also guided by the social norm. The power of the social norm relative to the monetary incentives is determined by an agent's norm sensitivity. For a given level of taxation and enforcement we can then derive a threshold for this norm sensitivity, which separates the population into those who adhere to the norm and those who conceal some of their income. As the tax rate increases, the threshold will rise and more taxpayers will start to evade taxes. A similar effect arises in the model of Myles and Naylor (1996).[8] In their approach, the social norm also constrains some individuals to a corner solution where they refrain from evasion. However, they consider a fixed cost of evasion, which implies that those agents who evade do so in the way predicted by Allingham and Sandmo (1972). In contrast, we assume that the utility loss of a norm violation is increasing in the own evasion level. This assumption is supported by survey evidence, which indicates that people differentiate in their moral judgements between small and large amounts of tax evasion (Aitken and Bonneville, 1980; Lewis, 1986). Hence, the evaders in our approach trade-off the monetary benefits from evasion with the non-pecuniary costs associated with a norm violation.[9] The optimal amount an individual conceals then depends – next to monetary incentives and risk preferences – also on the agent's norm sensitivity and the behavior of the other taxpayers; Norm guided individuals will not choose an optimal portfolio but 'overinvest into the safe asset'. This result is closely related to the findings in Gordon (1989), who also considers variable costs from evasion, associated to 'private stigma concerns'. In an extension he incorporates 'social stigma' costs which depend on the behavior of other agents. While Gordon's and our framework strongly overlap, our model is more compact and allows a clearer analysis. Moreover, we provide a detailed examination of tax and deterrence policies in the context of interrelated evasion decisions, which is neglected in Gordon (1989).

[7] Compare Keser and van Winden (2000) and Fischbacher et al. (2001) as well as Gächter (2005) for a comprehensive survey.

[8] Further behavioral economic approaches modelling tax evasion are provided by Bordignon (1993), Erard and Feinstein (1994), Falkinger (1995) and Schnellenbach (2002). Compare also Chapter 6 in Cowell (1990).

[9] There is clear empirical evidence which suggest that people trade off between monetary and 'emotional' incentives. Compare e.g. Bosman and van Winden (2002).

The main contribution of this chapter is the analysis of the relationship between a society's social structure and the level of tax evasion, which is completely novel in the tax evasion literature. In particular, we consider a society which consists of different subgroups, and show, that the impact of a group specific tax increase may cause a spillover on the level of evasion in other groups of the society. This is the case, since a higher tax will trigger an increase in the level of evasion within the targeted group. If this group works as a 'moral reference group' for other peers, more frequent cheating among the reference group can have an intense impact on the tax morale and thereby on the level of compliance in the rest of the society. Therefore, a change in the incentives to evade for one small group could have a large impact on the overall level of evasion. Following this line of reasoning, our model also highlights the crucial role of taxpayers' beliefs, and the 'management' of these beliefs. Hence, next to conventional enforcement policies (deterrence), appropriate information and belief management also constitute a measure to support tax compliance.

Finally, our model solves another weakness of the portfolio choice approach. For the plausible assumption of decreasing absolute risk aversion, the standard theory implies that *higher* taxes will lead to a *lower* level of evasion. If the penalty imposed on detected evaders is proportional to the concealed income, a tax increase does not trigger any substitution effect, as both the gains from evasion as well as the costs (associated with a higher expected fine) rise (Yitzhaki, 1974). The negative income effect then makes taxpayers less willing to bear risks. Therefore, they will conceal less income. We demonstrate that this counterintuitive result, which is also at odds with most empirical studies (e.g. Clotfelter, 1983), vanishes in our model, if individuals are strongly affected by the social norm. Moral taxpayers hold a biased portfolio with too little evasion, where the marginal expected return from evasion is positive. Any tax increase then further raises the marginal return from cheating and thereby provides an incentive to extend the level of evasion. We can find a threshold, where this substitution effect outbalances the income effect and the counterintuitive result turns around – evasion would increase with a higher tax rate. This result, however, is not new in the literature, as it has already been discussed by Gordon (1989).

The remainder of the chapter is structured as follows. In the next section we briefly present the Allingham and Sandmo (1972) model as a benchmark for our analysis. In section 2.3 we introduce a tax compliance norm and discuss optimal evasion within this framework. Section 2.4 provides an equilibrium concept for a heterogenous population of taxpayers. We analyze the equilibrium impact of the policy variables and study the choice of a tax and deterrence policy within this framework. In section 2.5, we extend our model to a society consisting of different subgroups, and demonstrate how group specific policy measures can cause spillovers on other subgroups of the society. The paper concludes with a discussion of the policy implications of this study.

2.2 The Allingham-Sandmo Model

Allingham and Sandmo (1972) (hereafter labeled AS) consider an individual with an exogenous gross income y. Each taxpayer i decides on how much of this income to declare and how much to conceal. Income concealed is labeled $e \in [0, y]$. The declared income, $y - e$, gets taxed with a proportional income tax at rate τ. With a fixed probability p the evasion gets detected. In this case, the tax evader has to pay the full taxes and a penalty, which is proportional to the taxes evaded (Yitzhaki, 1974). With probability $1 - p$ the evasion remains undetected and the evader only pays taxes on the declared income. The corresponding levels of net income x for state a – getting detected – respectively state b – escaping undetected – are given by

$$\begin{aligned} x^a &= x^a(e) = y - (y-e)\,\tau - \tau e - \tau e s = (1-\tau)\,y - \tau e s \\ x^b &= x^b(e) = y - (y-e)\,\tau = (1-\tau) y + \tau e \end{aligned}$$

where $s > 0$ denotes the penalty rate. The preferences of a risk averse taxpayer are represented by the following von Neumann-Morgenstern utility function

$$E\left[u(x(e))\right] = p\, u(x^a(e)) + (1-p)\ u(x^b(e)) \tag{1}$$

with $u'(x) > 0$ and $u''(x) < 0$. Agents choose e so as to maximize expected utility. The first- and second-order condition to this problem is given by

$$E\left[u\right]' \equiv -ps\tau\, u'(x^a) + (1-p)\,\tau\, u'(x^b) = 0 \tag{2}$$

respectively

$$E\left[u\right]'' \equiv p(s\tau)^2\, u''(x^a) + (1-p)\,\tau^2\, u''(x^b) < 0, \tag{3}$$

Condition (2) characterizes e^{AS}, the optimal level of income concealed.[10] One can easily show that the level of evasion decreases as tax enforcement becomes stricter:

$$\frac{de^{AS}}{dp} < 0 \tag{4}$$

$$\frac{de^{AS}}{ds} < 0. \tag{5}$$

Implicitly differentiating $E\left[u\right]'$ with respect to τ yields

$$\frac{de^{AS}}{d\tau} = \frac{(1-p)\,\tau\, u'(x^b)\left[(y-e)\left(\rho(x^b)-\rho(x^a)\right) - e\,(s+1)\,\rho(x^b)\right]}{-E\left[u\right]''} \tag{6}$$

where we made use of (2) and the Arrow-Pratt measure of absolute risk aversion, $\rho\,(x) = -u''(x)/u'(x)$. For the case of non-increasing absolute risk aversion, i.e. $\rho' \leq 0$ and hence $\rho(x_i^a) \geq \rho(x_i^b)$, the sign in the squared brackets in (6) is negative. Hence, we get the paradoxical result $de^{AS}/d\tau < 0$. According to the AS-framework, taxpayers with constant or decreasing absolute risk aversion conceal *less* income, if the tax rate rises. This counterintuitive finding is driven by the structure of the fine. Here we follow Yitzhaki (1974), who assumes that the penalty is assessed on the level of taxes evaded, rather than income concealed, as in the original AS paper. An increase in taxes will therefore raise both, the marginal gain from undetected evasion, i.e. taxes saved, as well as the marginal costs associated with higher fines in the case of detection. In the optimum, these two effects exactly offset each other. There is no substitution effect and the impact of a tax increase on the optimal evasion level is solely driven by the income effect. In the case of decreasing absolute risk aversion, a rise in taxes will reduce evasion, as a lower income makes taxpayers less willing to bear risks.

We will take this finding as a benchmark and show that it might vanish if taxpayers are not only guided by monetary incentives but also by 'moral concerns' (defined below). In order to guarantee the comparability of results, we maintain the assumption of non-increasing absolute risk aversion throughout the whole chapter.

[10] We focus on interior solutions with $e^{AS} \in [0, y]$.

2.3 A Social Norm for Tax Compliance

As we have discussed in the introduction, tax evasion seems to be more than a risky gamble. We here focus on the impact of tax morals on evasion behavior. Tax morals is modelled as a social norm for tax compliance – or equivalently, a norm against evasion. Declaring all income correctly is considered to be the 'morally right' behavior, while cheating on the taxes represents a violation of a social norm. Each member of the society has internalized these moral connotations to a certain degree. If a taxpayer conceals income, behavior is in conflict with morals. Conduct is not in line with the individuals' self-image as a 'good' member of society, who complies with societal norms and expectations. Tax evasion is then accompanied by 'emotional costs' which stem from self-imposed sanctions such as feelings of guilt or remorse. The impact of these sanctions varies with the endogenous strength of the norm. Following the literature on social norms (e.g. Akerlof, 1980), we assume that a norm is perceived as stronger, the more people adhere to it. If tax evasion becomes more common, the social norm is less strong and individuals' costs to deviate from the norm decline: It becomes easier for taxpayers to justify their wrong-doing to themselves, the more other people violate the societies' code of conduct.[11]

We incorporate such a social norm for tax compliance, considering a simple, additive preference structure expressed by the utility function

$$v\,(e_i, \theta_i) = E\,[u(x(e_i))] - \theta_i\, e_i\, c(n). \tag{7}$$

$E\,[u(x(e_i))]$ represents the expected (monetary) utility as defined in (1) and $\theta_i\, e_i\, c(n)$ represent the moral costs of tax evasion.[12] Moral costs are linearly increasing in e_i, the amount of income concealed by agent i.[13] The (marginal) moral costs of evasion are determined by $\theta_i \geq 0$, the individual specific degree of norm internalization, and the

[11]In addition to internal, self-imposed sanctions, norm deviation are typically associated with external, social sanctions, such as social disapproval, stigmatization or ostracism. We focus here on a purely private decision, assuming that any consequence of evasion (e.g. legal consequences in the case of detection) are private information. Since the scope for social sanctioning is then limited to within household interaction, we do not consider any external norm enforcement mechanism. (Note, however, that these within-household interactions seem to play an important role. As many empirical studies highlight, single households evade significantly more taxes than members of non-single household do. Compare e.g. Andreoni et al. 1998 an references therein.)

[12]As self-imposed sanctions occur irrespectively whether evasion gets detected or not, there is no risk associated with this payoff.

[13]Empirical support for this assumption is provided by Aitken and Bonneville (1980) and Lewis (1986). Note that non-linearity would not change our results qualitatively.

continuous function $c(n)$, which captures the strength of the norm for a given share of tax evaders n in the society.[14] We assume that $c(n) > 0$ and $c'(n) \leq 0$ for $n \in [0,1]$. Hence, the more taxpayer deviate from the norm, the lower are the moral costs of evasion.

Gordon (1989) considers a very similar preference structure: He considers 'private psychic costs', which are homogenous for all agents, as well as social costs, which are linearly decreasing in the share of evaders. This separation into different cost components unnecessarily complicates his analysis. While the form of $v(.)$ introduced in (7) is still very specific, it is simpler and – using the function $c(n)$ – more general than the preference structure studied by Gordon (1989).

2.3.1 Optimal Evasion Decision

Taking the policy variables as well as the number of evaders n as given, an agent maximizes $v(e_i, \theta_i)$ with respect to e_i. The first order condition[15] for an interior solution is

$$-ps\tau\, u'(x_i^a) + (1-p)\,\tau\, u'(x_i^b) = \theta_i\, c(n). \tag{8}$$

As the left hand side of (8) is equal to $E[u]'$ from (2), we can express condition (8) as $E[u]' = \theta_i\, c(n)$. Norm guided taxpayers will choose a level of evasion such that $E[u]'$, the marginal expected monetary utility, equals $\theta_i\, c(n)$, the marginal moral costs from an increase in the level of income concealed. Homo oeconomicus does not care about norms or morals, i.e. $\theta_i = 0$. Such an agent chooses an optimal portfolio, increasing evasion up to the point where the marginal expected utility is equal to zero. This yields the level of evasion predicted by the AS framework, e^{AS}. In contrast, taxpayers with high levels of θ_i may be in a corner solution and refrain from evasion. Let us define the marginal expected utility for the first unit of evasion,

$$z \equiv E[u(x(0))]' = (1 - p(1+s))\,\tau\, u'((1-\tau)\, y), \tag{9}$$

[14] Given that the total level of evasion is hardly observable and difficult to assess for a single individual, it appears plausible to assume that agents draw on the share of evaders rather than e.g. the mean level of evasion in society as a measure for the level of norm deviations.

[15] The second order condition is equivalent to (3).

with $z > 0$.[16] From (3) and (8) it then follows that any taxpayer with $\theta_i c(n) > z$ does not conceal any income. From this follows the definition

$$\hat{\theta}(n) \equiv \frac{z}{c(n)}, \tag{10}$$

which allows us to characterize the optimal individual evasion behavior $\hat{e}_i$ for a given level of n:

$$\hat{e}_i = \begin{cases} 0 & \text{for } \theta_i > \hat{\theta}(n) \\ e_i^* & \text{for } \theta_i \leq \hat{\theta}(n) \end{cases}. \tag{11}$$

Individuals with $\theta_i > \hat{\theta}(n)$ will stick to the compliance norm. On the other hand, those with $\theta_i \leq \hat{\theta}(n)$ will choose an interior solution e_i^* according to condition (8). A graphical representation of the optimal evasion level is provided in figure 2.1.[17]

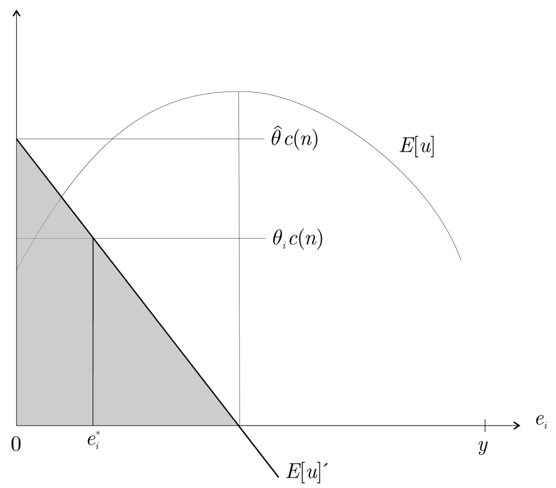

Figure 2.1: Optimal Evasion with a Social Norm

While an agent with $\theta_i = \hat{\theta}(n)$ will adhere to the tax compliance norm, taxpayers with $\theta_i = 0$ will choose e^{AS}, which maximizes $E[u]$. Individuals with $0 \leq \theta_i \leq \hat{\theta}(n)$ will choose an intermediate level of evasion, $e_i^* \in [0, e^{AS}]$. Note that e_i^* depends – next to the policy variables p, s, τ, income y and norm sensitivity θ_i – also on the n, the share of evaders in a society. Hence, evasion decisions are interdependent. Before we

[16]For $z < 0$ the would hold $(1 - p\,(1 + s)) < 0$. In this case, evasion would not be a 'fair gamble', in the sense that concealing income yields a negative expected return. The enforcement policy would be deterrent, as no rational agent would conceal any income.

[17]Despite the fact that $E\,[u]'$ is typically non-linearly decreasing in e_i, we have used a linear form for the sake of graphical simplicity.

characterize an equilibrium which accounts for this interdependence, we provide a brief comparative static analysis.

2.3.2 Partial Equilibrium Analysis

Applying the implicit function theorem on (8) we get[18]

$$\frac{\partial e_i^*}{\partial p} = \frac{s\tau\, u'(x_i^a) + \tau\, u'(x_i^b)}{E\,[u]''} < 0 \tag{12}$$

$$\frac{\partial e_i^*}{\partial s} = \frac{p\tau\,(u'(x_i^a) - s\tau e_i^*\, u''(x^a))}{E\,[u]''} < 0 \tag{13}$$

$$\frac{\partial e_i^*}{\partial \theta_i} = \frac{c(n)}{E\,[u]''} < 0 \tag{14}$$

$$\frac{\partial e_i^*}{\partial n} = \frac{\theta_i\, c'(n)}{E\,[u]''} \geq 0. \tag{15}$$

As in the AS framework, stricter enforcement will reduce evasion. Raising the penalty or the audit rate will induce agents to declare more income. Next to the formal tax enforcement institution, also the social norm has a deterrent effect: A boost in the norm sensitivity θ_i will result in a lower level of evasion. In terms of figure 2.1, the $\theta_i c(n)$ line would shift upwards and e_i^* would decline. Furthermore, (15) shows that taxpayers condition their evasion on the behavior of others. As more people start to evade taxes, agents with $0 < \theta_i \leq \hat{\theta}(n)$ react by concealing more income. The more taxpayers deviate from the compliance norm, the weaker the social norm, the lower are the moral cost of concealing income. In this sense, cheating has a positive externality on other evaders – it partially legitimates their wrong-doing, and thereby provides an incentive to evade more. From (10) we can derive

$$\frac{\partial \hat{\theta}(n)}{\partial n} = -\hat{\theta}(n)\frac{c'(n)}{c(n)} \geq 0 \tag{16}$$

since $c'(n) \leq 0$. With a higher level of n, the compliance threshold $\hat{\theta}(n)$ rises. Some agents with $\theta_i \geq \hat{\theta}(n)$, who used to refrain from evasion, will start to conceal income after an increase in n. These taxpayers condition their compliance with the tax law on the behavior of others. They act as *conditional cooperative taxpayers.* We will

[18]Note that in (12)-(14) we hold the share of evaders n constant.

elaborate on the implications of this behavior in the following section.

Let us now turn to the impact of a tax increase. From (8) we obtain

$$\frac{\partial e_i^*}{\partial \tau} = \frac{1}{-E\left[u\right]''}\left\{\frac{\theta_i\, c(n)}{\tau} + ps\tau\, u''(x_i^a)\,(y+se_i^*) - (1-p)\,\tau\, u''(x_i^b)\,(y-e_i^*)\right\}. \tag{17}$$

Note that for $\theta_i = 0$ the first term in the curly brackets is zero, and the right hand side is equivalent to the effect derived in the AS framework, depicted in (6). Making use of the measure of absolute risk-aversion, we can derive from (17) a threshold $\tilde{\theta}(n)$,[19]

$$\tilde{\theta}(n) = \frac{(1-p)\,\tau^2 u'(x_i^b)\left\{\rho(x_i^a)\,(y+se_i^*) - \rho(x_i^b)\,(y-e_i^*)\right\}}{c(n)\,(1+\tau\rho(x_i^a)\,(y+se_i^*))}, \tag{18}$$

such that

$$\frac{\partial e_i^*}{\partial \tau}\begin{cases} \geq 0 & \text{for } \theta_i \geq \tilde{\theta}(n) \\ < 0 & \text{for } \theta_i < \tilde{\theta}(n) \end{cases} \tag{19}$$

While for low levels of θ_i the counterintuitive effect from the AS model carries over to our framework, the result turns around for agents with a sufficient strong tax morale: As $\theta_i > \tilde{\theta}(n)$ we get the more plausible result that tax evasion increases with a higher tax rate. Moreover, from (18) and $c'(n) \leq 0$ follows that $\tilde{\theta}(n)$ is increasing in the levels of norm deviations. The lower the share of tax evaders in the society, the lower is $\tilde{\theta}(n)$ and thereby $\frac{\partial e_i^*}{\partial \tau} \geq 0$ holds for a broader range of θ-values.[20]

The intuition for this finding, which is similar to Proposition 1 in Gordon (1989), is straightforward. As we have discussed above, a tax increase will raise the marginal benefits from evasion, as well as the marginal costs (associated with higher expected fines). For the optimal evasion level of the AS model these two effects offset each other. Hence, there is no substitution effect and the negative income effect associated with a tax rise triggers a reduction in the evasion level. In our context, however, we find a positive substitution effect for all agents with $\theta_i > 0$ (depicted in the first term in the curly brackets in (17)). These agents will choose an evasion level such that $E\left[u\right]' > 0$. Marginal expected benefits from evasion are above marginal expected costs.[21] In terms of optimal portfolio choice, moral taxpayers over-invest into the safe asset – they conceal too little and declare too much income. An increase in the tax rate then raises the wedge

[19] Compare the Appendix.

[20] Note, however, that non-increasing absolute risk aversion implies $\tilde{\theta}(n) > 0$, since for $\rho'(x) \leq 0 \Rightarrow \rho(x_i^a) \geq \rho(x_i^b)$. Hence, there is always a range $[0, \tilde{\theta}(n)]$ where the result from AS also holds in our framework.

[21] Stated more formally, there holds $(1-p)\,\tau\, u'(x_i^b) > ps\tau\, u'(x_i^a)$.

between marginal expected benefits and cost even further. This will raise marginal expected utility $E[u]'$ for all agents with $\theta_i > 0$ (respectively $e_i^* < e^{AS}$). As the moral costs of evasion are unaffected by a tax change,[22] the substitution effect provides an incentive to increase evasion. While the (negative) income effect is still present, the (positive) substitution effect dominates for $\theta_i > \tilde{\theta}(n)$. For these agents, tax evasion increases as taxes rise. Figure 2.2 provides a graphical representation of the partial equilibrium effects associated with an increase in the tax rate.

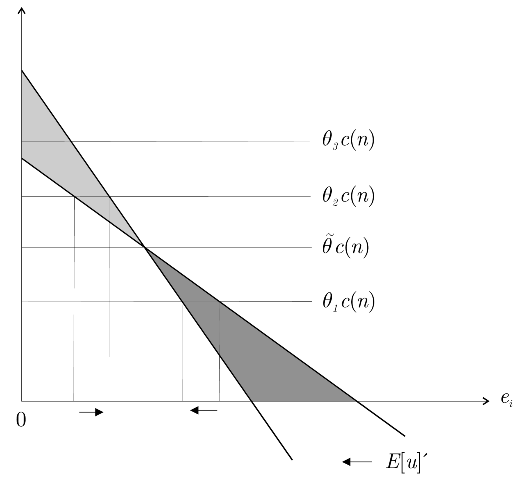

Figure 2.2: Partial Equilibrium Effect of a Tax Increase

From the definition of z above, we can easily derive $\partial z/\partial\tau > 0$. From the AS model we know that $E[u]'$ evaluated at e_i^{AS} decreases with an increase of the tax rate. Hence, the $E[u]'$ curve turns clockwise as we increase τ, with the turning-point somewhere between e_i^{AS} and $e_i = 0$.[23] The intersection of the marginal expected utility curve before and after the change in the tax rate defines the threshold $\tilde{\theta}(n)$. In the example from figure 2.2, individual 1 with $\theta_1 < \tilde{\theta}(n)$ will reduce evasion as the tax rate raises. Individual 2 however, will conceal more income. Moreover, taxpayer 3, who has been paying taxes honestly before the tax increase, will switch to an interior solution after the policy change. From this example one can also see, that a change in one policy variable typically has an impact on the share of evaders. In the following we study this effect in an equilibrium framework.

[22]If moral costs depend on the level of taxes evaded rather than on the amount of income concealed, a higher tax rate would also raise the moral evasion costs. As long as the increase in marginal expected utility dominates the increase in the marginal costs, we would still observe a rise in the evasion level for some individuals.

[23]This also implies $0 < \tilde{\theta}(n) < \hat{\theta}$.

2.4 Social Equilibrium

We consider a continuous population with unit mass. The norm parameter θ is distributed according to a continuously differentiable, cumulative distribution function $F(\theta)$, which has full support on the interval $[0, \bar{\theta}]$. The corresponding density function is $f(\theta)$ and the inverse of the distribution function is denoted $F^{-1}(n)$. As we know from (11), people choose to evade income if $\theta_i \leq \hat{\theta}(n)$. The equilibrium population share of evaders n^* is then given by the fixed-point equation

$$n^* = F\left(\frac{z}{c(n^*)}\right) \tag{20}$$

The right hand side of (20) is a continuous function in n, mapping the compact interval $[0, 1]$ into itself. Assuming that $\bar{\theta}c(1) \geq z$ – and thereby $\hat{\theta}(1) \leq \bar{\theta}$ – there always exists at least one stable equilibrium $n^* \in (0, 1]$,[24] where stability is defined by

$$\left.\frac{\partial F^{-1}(n)}{\partial n}\right|_{n^*} \geq \left.\frac{\partial \hat{\theta}(n)}{\partial n}\right|_{n^*} \tag{21}$$

For a given set of policy variables (τ, p, s) such a stable equilibrium characterizes a self supporting share of evaders. For n^*, the strength of the norm is such that a population share of $1 - n^*$ will declare their income honestly, while the remaining n^* will choose an interior solution e_i^*. While there is at least one solution to (20), the system is typically characterized by a multiplicity of equilibria. If evasion has become prevalent, the compliance norm is weak and society might find itself in a stable equilibrium with widespread cheating. For the same policy (τ, p, s) and distribution $F(\theta)$, however, the society could in principle coordinate on a different equilibrium, where most agents adhere to the norm. The social norm would be stronger and the level of tax evasion in the society would be smaller. Such a scenario is depicted in figure 2.3.

In the example of figure 2.3 we have assumed that θ is uniformly distributed. The shape of $\hat{\theta}(n)$ is defined by the function $c(n)$. In this example, there are two stable equilibria - a ‘good’ one, where only a small fraction n_l^* deviates from the norm and a ‘bad’ equilibrium with widespread evasion, n_h^*. Between these two stable equilibria, there is a third, instable one, with n_m^*. Starting from any share $n < n_m^*$, the system converges to the good equilibrium (respectively to the bad equilibrium, for any $n > n_m^*$).

[24] An equilibrium with $n^* = 0$ is not supported. As we allow for a $\theta_i = 0$, this type of agents always chooses to evade $e^{AS} > 0$ for any level of n.

In the following, we will focus our analysis on stable equilibria.

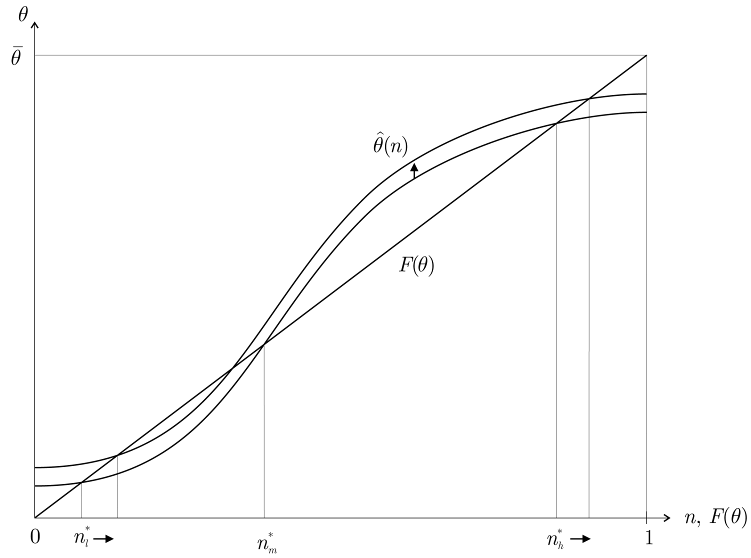

Figure 2.3: Social Equilibrium – Impact of a Tax Increase

2.4.1 Equilibrium Effect of Deterrence Policies

Let us take a look at the impact of the tax enforcement variables s and p on the equilibrium share of evaders as well as on the equilibrium level of evasion. From (20) we derive in the Appendix that in a stable equilibrium there holds

$$\frac{dn^*}{dp} < 0, \tag{22}$$

$$\frac{dn^*}{ds} < 0. \tag{23}$$

An increase in either p or s will shift the $\hat{\theta}(n)$-curve downwards and the population share of evaders drops. Using (22) and (23), we get the equilibrium impact of a stricter enforcement policy:

$$\frac{de_i^*}{dp} = \frac{\partial e_i^*}{\partial p} + \frac{\partial e_i^*}{\partial n}\frac{dn^*}{dp} < 0 \tag{24}$$

$$\frac{de_i^*}{ds} = \frac{\partial e_i^*}{\partial s} + \frac{\partial e_i^*}{\partial n}\frac{dn^*}{ds} < 0 \tag{25}$$

In addition to the first-order effects $\partial e_i^*/\partial p$ and $\partial e_i^*/\partial s$ derived above in (12) and (13), there is now a second order effect: As auditing or penalty rates become higher, the resulting drop in the share of evaders makes the social norm for tax compliance stronger – the moral costs of tax evasion increase. This will trigger a further reduction in the level of evasion. Hence, in our framework, the equilibrium impact of tax enforcement is stronger than suggested by the partial equilibrium analysis, respectively the AS model.

2.4.2 Equilibrium Effect of a Tax Increase

We now turn to the other policy variable, the tax rate. As already addressed above, individuals with θ_i above $\hat{\theta}(n)$ may switch from a corner solution to an interior solution after a tax increase. These new evaders will weaken the social norm, moral costs of evasion decline and some honest taxpayers will start concealing income. The equilibrium effect is captured by an upward shift of the $\hat{\theta}(n)$-curve in figure 2.3, which results in an extension of the share of evaders: n_l^* respectively n_h^* increase. As we show in the Appendix, we we can derive from (20)

$$\frac{dn^*}{d\tau} > 0. \tag{26}$$

With a higher tax rate, the marginal expected utility from concealing the first unit of income increases ($\frac{\partial z}{\partial \tau} > 0$). Accordingly, the evasion threshold $\hat{\theta}(n)$ rises for any n. The equilibrium impact on the optimal level of evasion is then given by

$$\frac{de_i^*}{d\tau} = \frac{\partial e_i^*}{\partial \tau} + \frac{\partial e_i^*}{\partial n}\frac{dn^*}{d\tau} \tag{27}$$

with $\partial e_i^*/\partial n$ and $\partial e_i^*/\partial \tau$ from (15) respectively (17). As we have shown above in (19), the sign of the first order effect is ambiguous and depends on the norm parameter θ_i. In contrast, the second order effect is unambiguously positive for all $\theta_i > 0$. We can derive a new threshold $\tilde{\theta}'(n^*)$,[25]

$$\tilde{\theta}'(n^*) = \frac{(1-p)\,\tau^2\,u'(x_i^b)\left\{\rho(x_i^a)\,(y+se_i^*) - \rho(x_i^b)\,(y-e_i^*)\right\}}{c(n^*)\left(1+\tau\rho(x_i^a)\,(y+se_i^*) - \tau\frac{c'(n^*)}{c(n^*)}\frac{dn^*}{d\tau}\right)}, \tag{28}$$

such that

$$\frac{de_i^*}{d\tau}\begin{cases} \geq 0 & \text{for } \theta_i \geq \tilde{\theta}'(n^*) \\ < 0 & \text{for } \theta_i < \tilde{\theta}'(n^*) \end{cases} \tag{29}$$

[25] Compare the Appendix.

In equilibrium, there are two effects which tend to raise evasion. First, there is a positive substitution effect, discussed in the partial equilibrium analysis above. Second, a tax increase is accompanied by an increase in the equilibrium share of norm breaking individuals. This second order effect lowers the moral cost of evasion and thereby provides a further incentive to conceal more income. For taxpayers with $\theta_i \geq \tilde{\theta}'(n^*)$ these two effects dominate the negative income effect – they will react with more evasion on an increase in the tax rate. Moreover, it follows from this discussion that $\tilde{\theta}'(n^*) < \tilde{\theta}(n^*)$.[26] Hence, compared to the partial equilibrium analysis, $de_i^*/d\tau \geq 0$ holds for a broader range of θ-values.

2.4.3 Policy Choice with Social Norms

We now discuss the choice of the policy variables within this framework. The total level of income concealed in a society, $\bar{e}$, is represented by the grey shaded area in figure 2.1. Formally this is

$$\bar{e} = \int_0^{\bar{\theta}} e(\theta, \tau, p)\, dF(\theta), \tag{30}$$

where $e(\theta, \tau, p)$ denotes the optimal evasion level $\hat{e}_i$ as defined in (11). From (30) we easily get

$$\frac{d\bar{e}}{d\tau} = \int_0^{\tilde{\theta}'(n^*)} \frac{de(\theta, \tau, p)}{d\tau}\, dF(\theta) + \int_{\tilde{\theta}'(n^*)}^{\hat{\theta}(n^*)} \frac{de(\theta, \tau, p)}{d\tau}\, dF(\theta) \tag{31}$$

From (29) we know that first term on the right hand side is negative, since agents with $\theta_i < \tilde{\theta}'(n^*)$ conceal less income as the tax rate rises. The second term, however, is positive, since taxpayers with $\theta_i > \tilde{\theta}'(n^*)$ evade more.[27] If, for a given tax τ and equilibrium state n^*, the population mass in the latter group is sufficiently high – that is, if there is a sufficient number of taxpayers who are significantly affected by the social norm – the second term in (31) dominates and $d\bar{e}/d\tau > 0$. Note, however, that the magnitude of the two integral terms in (31), as well as the integral limit $\tilde{\theta}'(n^*)$, depends on the tax rate respectively the equilibrium state. Hence, there may also exist equilibria with $d\bar{e}/d\tau < 0$. Let us assume that this is indeed the case for low levels of τ and that $\frac{d^2\bar{e}}{d\tau^2} \geq 0$ holds. A tax policy which minimizes the level of income concealed

[26] As the equilibrium (second order) effect is always positive, $\frac{de_i^*}{d\tau} > 0$ does also hold in a range where the substitution effect is dominated by the income effect.

[27] Figure 2.2 illustrates the partial equilibrium impact on $\bar{e}$. The light-grey shaded area captures the increase in evasion, the dark shaded area the reduction in concealed income.

is then characterized by

$$\frac{d\bar{e}}{d\tau} = 0. \tag{32}$$

The tax rate would be chosen such that the two effects depicted in (31) exactly offset each other.[28]

Consider now revenue maximization as an alternative policy objective. The problem of a revenue maximizing planner becomes

$$\max_{\tau} \quad R = \tau\left(y - \bar{e}\left(1 - p\left(1+s\right)\right)\right).$$

The first order condition for an interior solution is

$$y - \bar{e}\left(1 - p\left(1+s\right)\right) - \tau\left(1 - p\left(1+s\right)\right)\frac{d\bar{e}}{d\tau} = 0,$$

which can be expressed as

$$\frac{d\bar{e}}{d\tau}\frac{\tau}{\bar{e}} = \frac{R}{\tau y - R} \tag{33}$$

The left hand side of this condition represents the elasticity of (total) income concealed with respect to a tax increase. The right hand side gives the actual revenues relative to the tax gap – the difference between tax revenues in the case without any evasion, τy, and the actual revenues, R. A revenue maximizing planner increases the tax rate up to the point where the ratio of actual to missing revenues equals the evasion elasticity. As $\tau y \geq R > 0$, the right hand side is positive. For a revenue maximizing tax rate there has to hold $\frac{d\bar{e}}{d\tau} > 0$. As we assume $\bar{e}$ to be convex in the tax rate, it immediately follows from (32) that such a tax is higher than the tax which minimizes total incomes concealed. Note that this result is in sharp contrast with the policy choice suggested by the AS framework, where the revenue maximizing tax is equivalent to the evasion minimizing tax characterized by $\tau \to 1$.[29]

Let us now turn to the enforcement policy. Taking the penalty rate s as exogenously given,[30] we consider an authority which maximizes revenues R net of tax enforcement

[28] In case $d\bar{e}/d\tau > 0$ for all $\tau \in [0, 1]$ and $\bar{e}$ is convex in τ, the total level of incomes concealed would be minimized at $\tau = 0$.

[29] This is of course driven by the fact that in the AS model evasion is decreasing in the tax. Compare equation (6).

[30] As already noted by Kolm (1973), one could induce full compliance by setting $s \to \infty$. However, since there are several good reasons why most modern societies do not use capital punishment, we simply assume that s is fixed at a certain level. The upper bound for s might be constrained due to the credibility of penalties in the case of limited liability.

costs $d(p)$, with $d' > 0$ and $d'' > 0$ for $p \in (0,1)$. The problem is then given by

$$\max_{p} \quad \tau\left(y - \bar{e}\left(1 - p\left(1+s\right)\right)\right) - d(p)$$

and an interior solution is characterized by

$$\tau\bar{e}\left(1+s\right) - d'(p) - \tau\left(1 - p\left(1+s\right)\right)\frac{d\bar{e}}{dp} = 0, \tag{34}$$

with

$$\frac{d\bar{e}}{dp} = \int_{0}^{\hat{\theta}(n)} \frac{de(\theta,\tau,p)}{dp}\, dF(\theta) < 0. \tag{35}$$

(The second order condition holds as long as $\frac{d^2\bar{e}}{dp^2} \geq 0$.) Rearranging (34) and substituting for R we get

$$\frac{d\bar{e}}{dp}\frac{p}{\bar{e}} = \frac{p\left(\tau\bar{e}\left(1+s\right) - d'(p)\right)}{\tau y - R}, \tag{36}$$

where the left hand side captures the elasticity of concealed incomes with respect to the detection probability. The denominator on the right hand side is the revenue gap, as above in (33) and the numerator reflects the direct impact of a higher detection rate on net revenues – the first and second term in (34) – weighted by p.[31] As it follows from (24) that $\frac{d\bar{e}}{dp} < 0$, the left hand side is negative. Hence, for a revenue maximizing enforcement policy the numerator on the right hand side must be negative, i.e. the marginal costs from auditing must be larger than the direct marginal benefits.

As we have discussed in section 2.4.2, the second order effect of tax enforcement (associated with a strengthening of the social norm) renders the impact of auditing stronger as compared to the AS framework. The elasticity of evasion in our framework is therefore higher, which implies that the revenue maximizing audit rate is above the level predicted by AS.

2.5 Social Structure and Inter-Group Spillovers

In order to discuss the potential impact of a society's social structure on tax evasion, we present a simple extension of the basic model. To the best of the author's knowledge, this approach is novel in the literature on tax evasion. Consider an economy with

[31] Note that in (33), the nominator can also be interpreted as the direct revenue gains from a tax increase, weighted by the tax.

two classes of individuals: Those who receive (exclusively) capital incomes and those who earn labor income. The income level within each group is denoted by y_j where the index j indicates the income source, $j \in \{K, L\}$. Allowing the tax policy to discriminate between different income sources, the utility of an agent i from group j is given by

$$v_j\left(e_{ij}, \theta_{ij}\right) = p\, u(x_j^a(e_{ij})) + (1-p)\; u(x_j^b(e_{ij})) - \theta_{ij}\, e_{ij}\left(\alpha_j c(n_j) + (1-\alpha_j)\, c(n_k)\right) \quad (37)$$

with

$$\begin{aligned} x_{ij}^a &= x_j^a(e_{ij}) = (1-\tau_j)\, y_j - \tau_j e_{ij} s, \\ x_{ij}^b &= x_j^b(e_{ij}) = (1-\tau_j) y_j + \tau_j e_{ij}, \end{aligned}$$

where $n_j \in [0,1]$ respectively $n_k \in [0,1]$, denotes the share of evaders among agents with income source j respectively $k \neq j$. In (37) we assume that the moral costs of evasion depend on the norm compliance behavior in both groups. The parameter $\alpha_j \in [0,1]$ thereby captures the sensitivity of the norm strength in group j with respect to norm deviations in the own respectively the other group. If $\alpha_j = 1$, the norm is completely group specific, as its strength is solely determined by the behavior of the own peers. If $0 < \alpha_j < 1$, however, the norm strength is co-determined by the actions taken by agents from both income groups. If $\alpha_j \to 0$, members of group k take over the position of moral role models for group j. Norm deviations within this reference group then have a crucial impact on the tax morale of taxpayers from group j.[32]

Taking the policy variables as well as n_j and n_k as given,[33] individuals choose e_{ij} maximizing utility given by (37). Analogously to (8), the first order condition for this problem is

$$-ps\tau_j\, u'(x_{ij}^a) + (1-p)\, \tau_j\, u'(x_{ij}^b) = \theta_{ij}\ \left(\alpha_j c(n_j) + (1-\alpha_j)\, c(n_k)\right) \quad (38)$$

and the threshold for an interior solution becomes

$$\hat{\theta}_j(n_j, n_k) = \frac{z_j}{\alpha_j c(n_j) + (1-\alpha_j)\, c(n_k)} \quad (39)$$

[32] If we assume that moral costs are shaped by the structure of direct social interactions, one could also interpret the parameter α as the degree of population segregation or *viscosity (* compare Myerson et al., 1991). If α_j is high, the interaction frequency with types from the other group is low (and vice versa).

[33] We assume that groups are large such that self-perceived insignificance holds.

with

$$z_j = (1 - p\,(1+s))\,\tau_j\,u'\,((1-\tau_j)\,y_j) > 0. \tag{40}$$

For $\theta_{ij} \leq \hat{\theta}_j(n_j, n_k)$, condition (38) characterizes the optimal evasion level e^*_{ij}. All agents with θ_{ij} above this threshold are in a corner solution and pay all their taxes honestly.

2.5.1 Social Equilibrium and Policy Spillovers

How can we describe a social equilibrium for a society consisting of two groups? Let each income group be represented by a continuum $[0,1]$. Assuming that there is no difference in the degree of norm internalization between the two income groups, and $\theta_{ij} \in [0, \bar{\theta}]$ is distributed according to $F(\theta)$ for both types.[34] A stable social equilibrium is then given by the pair $\left(n^*_j, n^*_k\right)$ with

$$n^*_j = F(\hat{\theta}_j(n^*_j, n^*_{k\neq j}) \quad \text{for } j \in \{K, L\} \tag{41}$$

and stability is characterized by

$$\left.\frac{\partial F^{-1}}{\partial n_j}\right|_{n_j = n^*_j} \geq \left.\frac{d\hat{\theta}_j(n_j, n_k)}{dn_j}\right|_{n_j = n^*_j, n_k = n^*_k} \tag{42}$$

One can show that such an equilibrium exists as long as $\hat{\theta}_j(1,1) < \bar{\theta}$ holds.[35]

What is the impact of a change in the tax policy in this extended framework? Consider an increase in the tax rate for income group j. As before, we can derive that

$$\frac{dn^*_j}{d\tau_j} \geq 0 \tag{43}$$

holds in any stable equilibrium.[36] A tax increase still induces more people to cheat on taxes. However, a higher level of n^*_j now has a negative impact on the perceived norm strength in the other group, and thereby triggers more norm deviations within that

[34] Allowing for between-group heterogeneity – e.g. with respect to risk-preferences, income or the distribution of θ – would not change our results in any qualitative way.

[35] Note that $\hat{\theta}_j(0, n_k) > 0$ holds as $c(0)$ is finite.

[36] All the results of this section are derived in the Appendix.

group k. In particular, we get from (41)

$$\frac{dn_k^*}{dn_j^*} \geq 0. \tag{44}$$

The impact of a higher tax rate τ_j on the optimal evasion decision of a taxpayer with income from source j is then given by

$$\frac{de_{ij}^*}{d\tau_j} = \frac{\partial e_i^*}{\partial \tau_j} + \left(\frac{\partial e_{ij}^*}{\partial n_j} + \frac{\partial e_{ij}^*}{\partial n_k}\frac{dn_k^*}{dn_j^*}\right)\frac{dn_j^*}{d\tau_j} \tag{45}$$

where one can derive the partial equilibrium effects $\partial e_i^*/\partial \tau_j$ and $\partial e_{ij}^*/\partial n_j > 0$ analogously to the effects presented in the previous section.[37] Hence, the impact of an increase in the tax on income of type j on the evasion level of agents of type j is similar as before. The power of the social norm will decline, as the change in the group specific tax policy raises the share of evaders in both income groups. Since $\partial e_{ij}^*/\partial n_k > 0$, this provides a similar second order incentive to increase e_{ij}^*, as discussed in the previous section.[38] In addition, however, the tax policy creates an externality on tax compliance in the other income group:

$$\frac{de_{ik}^*}{d\tau_j} = \left(\frac{\partial e_{ik}^*}{\partial n_j} + \frac{\partial e_{ik}^*}{\partial n_k}\frac{dn_k^*}{dn_j^*}\right)\frac{dn_j^*}{d\tau_j} \geq 0. \tag{46}$$

Although the change in the tax rate does not alter the monetary incentives of these taxpayers, they react with an increase in evasion. As more taxpayers from the other group start cheating, tax morale and therewith compliance in the own group drops as well. Again, this result is triggered by conditional cooperative taxpayers. While in the basic framework this effect occurred within one group (the homogenous society), it now works between different peers. As a consequence, any group specific policy change typically causes a spillover on other communities within the society. Only in the case of groups with fully segregated norm perceptions (i.e. $\alpha_j = \alpha_k = 1$), tax policies will not cause any externalities. In this case, the tax compliance norm is a *group specific* rather than a *social* norm.

This point also reveals that the strength of the externality of a tax τ_j on evasion e_{ik}^* is determined by the social relevance of group j for group k. The lower α_k, the more weight is attributed to the behavior of j-agents in determining tax morals within

[37] Note that all derivatives in (45) as well as in (46) below have to be evaluated at $n_j = n_j^*$ and $n_k = n_k^*$.

[38] We could again derive a threshold for θ_{ij} where $de_{ij}^*/d\tau_j > 0$ holds.

group k, the stronger is the spillover. One could easily extend this argument to a society consisting of many different communities. If there exists one (small) group with high social respectively moral prestige, which serves as a *moral reference group* for other peers, norm violations within this small subgroup have a strong erosive impact on the social norm in the economy. In this case, the behavior of a few members of society can have a huge impact on the overall level of tax evasion. Moreover, we could characterize parameter values, where small changes in the behavior within such a reference group eliminate the existence of 'good' equilibria (as discussed in section 2.3 in the context of figure 2.3) for other peers. Even a small tax evasion scandal of a high-profile member of society (e.g. a politician or a manager) could shift a bulk of society from an equilibrium state with a large extent of tax compliance to a state with widespread evasion.[39]

2.5.2 Revenue Maximizing Taxation

Let us briefly discuss the implications of the spillovers discussed above for the choice of a revenue maximizing tax rate. Considering the case of two income groups introduced before, the planners problem becomes

$$\max_{\tau_L,\tau_K} \quad m_L R_L + R_K$$

where revenues are defined as

$$R_j = \tau_j \left(y_j - \bar{e}_j \left(1 - p\left(1+s\right)\right)\right) \tag{47}$$

and $m_L \geq 0$ depicts the share of taxpayers with labor income relative to capital income. The first-order conditions for this problem are given by

$$m_L \left(\frac{R_L}{\tau_L} - \tau_L \left(1 - p\left(1+s\right)\right) \frac{d\bar{e}_L}{d\tau_L}\right) = \tau_K \left(1 - p\left(1+s\right)\right) \frac{d\bar{e}_K}{d\tau_L}$$

and

$$\frac{R_K}{\tau_K} - \tau_K \left(1 - p\left(1+s\right)\right) \frac{d\bar{e}_K}{d\tau_K} = m_L \, \tau_L \left(1 - p\left(1+s\right)\right) \frac{d\bar{e}_L}{d\tau_K}.$$

[39]We shall further note, that different subgroups of society do not necessarily have to be in a symmetric type of equilibrium. Social structure may well supports equilibria, where some groups predominantly cheat while the members of other communities act according to the norm. Compare the discussion in Ostrom (1998) respectively Ichino and Maggi (2000), who study shirking in team-production problems embedded in different local communities.

(Second Order Conditions are assumed to hold.) Using (47) and rearranging the two first-order conditions, we can characterize the revenue maximizing taxes by

$$\tau_L \left(1 - p\,(1+s)\right) = \frac{m_L R_L\, \varepsilon_{KK} - R_K\, \varepsilon_{KL}}{m_L \bar{e}_L \left(\varepsilon_{KK}\varepsilon_{LL} - \varepsilon_{LK}\varepsilon_{KL}\right)} \tag{48}$$

and

$$\tau_K \left(1 - p\,(1+s)\right) = \frac{R_K\, \varepsilon_{LL} - m_L R_L\, \varepsilon_{LK}}{\bar{e}_K \left(\varepsilon_{KK}\varepsilon_{LL} - \varepsilon_{LK}\varepsilon_{KL}\right)} \tag{49}$$

where $\varepsilon_{jj} \geq 0$ and $\varepsilon_{jk} \geq 0$ denote the elasticity of the evasion of group j with respect to a tax change for group j respectively k (e.g $\varepsilon_{jk} = \frac{d\bar{e}_j}{d\tau_k}\frac{\tau_k}{\bar{e}_j}$). The second term in the numerator of (48) respectively (49) depicts the negative spillover the tax has on the revenues from the other income group. As both denominator are positive,[40] a revenue maximizing tax τ_j is smaller, the higher its externality is. From the discussion above follows that this would constitute an argument for lower taxes for the moral reference group. A higher tax rate on the members of such a group can trigger a significant increase in the level of evasion among other subgroups of society. Therefore, the between-group spillover provides a new effect for optimal taxation problems. Assuming that the role as reference group stems from higher social prestige associated with higher income, the effect is of particular interest in the context of redistributive policy targets. A thorough analysis of the addressed tradeoff is left for further research.

2.5.3 Tax Enforcement

Next to the impact of group specific tax rates, it is also interesting to study a discriminatory auditing policy. The results we would obtain, follow immediately from the analysis discussed so far. Stricter tax enforcement among one income group has a deterrent spillover on the other group. As a first-order effect, more frequent auditing reduces the share of evaders in the group targeted by the policy. From (44) then follows a second-order effect such that the compliance rate in the other group also rises. Hence, overall norm adherence increases, the compliance norm will become stronger and agents of both income groups will conceal less income. The implications of the spillover for a revenue maximizing auditing strategy are straightforward. The tax enforcement authority can raise higher revenues, if more resources on auditing are spent on the moral reference group. Enforcing a high level of compliance among this group has a strong spillover, and thereby yields a high marginal return from auditing. If we

[40]For both income groups, the elasticity of evasion with respect to the own tax rates should be higher than the cross-tax elasticity for a reasonable set of parameters.

again assume that the social reference group is characterized by a higher income, this provides an additional argument to implement stricter auditing for this group, since higher incomes are cet.par. associated with higher levels of evasion[41] and hence more evaded taxes to detect. Summarizing this discussion, our model suggest that a revenue maximizing enforcement policy should focus on those groups of society, who act as public role models and determine the extent of tax morale within the society.

2.6 Conclusion

Given the rather low level of deterrence applied in most economies, the standard approach to tax evasion predicts an amount of cheating which is way above the level of evasion in OECD countries (Andreoni et al., 1998). According to Allingham and Sandmo's (1972) optimal portfolio model, the present level of compliance could only be explained by extremely high degrees of risk aversion (Bernasconi, 1998; Skinner and Slemrod, 1985). Even if this would be the case, the model can not explain why we observe quite different levels of evasion for countries with very similar enforcement policies. The model presented in this paper provides an explanation to these puzzles. We studied the evasion behavior of 'moral taxpayers', where we model tax morale as an (internalized) social norm for tax compliance. As the strength of the social norm behavior depends on the level of norm adherence within the society, evasion decisions become interdependent. Evaders are conditionally cooperative – they will conceal more income the more other taxpayers deviate from the compliance norm. This behavioral pattern results in multiple equilibria: An economy with a given tax and enforcement policy, could either end up in a state with a strong social norm, where most taxpayers pay all their taxes honestly, or a state with a weak social norm, where individuals predominantly cheat on taxes. Hence, the model offers an explanation why 'Palermo is not Milan, and Stockholm is not Moscow' (compare Rothstein, 2000).

One counterintuitive prediction implicit in the standard approach is that – for decreasing absolute risk-aversion – evasion should decrease if taxes increase (Yitzhaki, 1974). For norm-guided taxpayers this result may turn around. Such individuals hold a biased portfolio, as they overinvest into the safe asset: They declare more income as compared to an immoral gambler, who maximizes the expected return of evasion. Any tax increase then triggers an incentive to raise evasion. In addition to this substitution effect, there is a further effect, related to the decline in the power of the social norm.

[41] Remember that we assume decreasing absolute risk aversion.

As the tax rate rises, some individuals will switch from norm compliance to an interior evasion level – which constitutes a norm violation. As more people cheat on taxes, it becomes easier for each taxpayer to justify the own norm-deviation to themselves, which provides a further incentive to conceal more income. For the tax enforcement measures, the audit and the penalty rate, this mechanism tends to reinforce the power of deterrence policies. A higher auditing rate, for example, makes evasion more costly in a pecuniary as well as in a non-pecuniary way, since both the expected monetary costs and the moral costs increase. Hence, our model suggest that deterrence is more effective than predicted by Allingham and Sandmo (1972). Note, however, that this result relies on the assumption that the agents' norm sensitivity is unaffected by any policy change. Empirical evidence suggest, however, that a switch towards a deterrence policy which is perceived as unreasonable harsh might lead to a negative shock of the individuals' moral obligation to pay taxes. '*Brute detection might backfire*' (Sheffrin and Triest 1992), if too strict enforcement crowds out the 'intrinsic motivation' to comply with tax authorities (Scholz and Lubell 1998, Feld and Frey 2002). While this effect points at the upper limit of tax enforcement, our finding suggest that a reasonable, i.e. not too harsh, deterrence policy can have a strong positive impact on the citizens' tax moral.

Finally, we have discussed a simple framework to understand tax evasion within a structured society. In contrast to the individualistic approach of the standard theory, the decision makers in our model are embedded in a social structure. They belong to a certain group and adjust their perception of the social norm and hence, their tax morale, according to the behavior among their own peers as well as actions taken in other subgroups of society. In such a framework, any group specific policy change creates an externality on the rest of the society. If, for example, a higher capital taxation causes more capital owners to conceal income, labor income receivers will also evade more, even though monetary incentives stay unchanged for them. The magnitude of these spillovers depends on the links between different peers. If the conduct of the members from one group strongly shape what is considered to be the social norm, this group takes over the position of a moral role model. The behavior within this group has an intense impact on the strength of the norm for all other subgroups. A tax evasion scandal within such a reference group can then have a huge negative impact on overall tax compliance in an economy. In the light of this analysis, public cases of tax offenses by politicians, managers or other high-profile members of society, can cause tremendous harm to an economies informal institutions supporting tax compliance. This is also supported by recent empirical research, which demonstrates the strong impact a group leaders' decision has on the beliefs and consequently the behavior of followers (Gächter and Renner, 2005). Implicity, our model also stresses the importance of *belief management* (Rothstein, 2000; Gächter, 2005). If, for example, a government announces the switch to a stricter tax enforcement *due to a recent rise in evasion activities*, this may convey a different signal to taxpayers – and might trigger a different behavior –, as if the same policy steps would be presented as *preventive measures to maintain the high level of compliance.*

Appendix

Ad Section 2.3:

From (19) we can easily derive

$$\tilde{\theta}(n) = \frac{-\tau^2}{c(n)} \left\{ ps\, u''(x_i^a)\,(y + se_i^*) - (1-p)\; u''(x_i^b)\,(y - e_i^*) \right\}. \tag{A.1}$$

Using $\rho\,(x)$, the Arrow-Pratt measure of absolute risk aversion, rearranging and substituting for (8), we get

$$\tilde{\theta}(n) = \frac{(1-p)\,\tau^2 u'(x_i^b)\left[\rho(x_i^a)\,(y + se_i^*) - \rho(x_i^b)\,(y - e_i^*)\right]}{c(n)\left[1 + \tau\rho(x_i^a)\,(y + se_i^*)\right]} > 0. \tag{A.2}$$

Non-increasing absolute risk aversion is sufficient for $\tilde{\theta}(n) > 0$ to hold, since for $\rho'(x) \leq 0 \Rightarrow \rho(x_i^a) \geq \rho(x_i^b)$.

Ad Section 2.4:

Equilibrium Impact of the Policy Variables

Applying the implicit function theorem on (20) yields

$$\frac{dn^*}{dp} = \frac{-\,(1+s)\,\tau\, u'\left((1-\tau)\,y\right)\frac{1}{c(n)}}{\frac{\partial F^{-1}}{\partial n} - \frac{\partial\hat{\theta}(n)}{\partial n}} < 0. \tag{A.3}$$

where we used (9) to derive $\frac{\partial z}{\partial p}$. As we know from (21), the denominator must be positive in a stable equilibrium. This determines the negative sign of (A.3).

Following the same steps as before, we get

$$\frac{dn^*}{ds} = \frac{-p\tau\, u'\left((1-\tau)\,y\right)\frac{1}{c(n)}}{\frac{\partial F^{-1}}{\partial n} - \frac{\partial\hat{\theta}(n)}{\partial n}} < 0 \tag{A.4}$$

and

$$\frac{dn^*}{d\tau} = \frac{\frac{\partial z}{\partial\tau}\frac{1}{c(n)}}{\frac{\partial F^{-1}}{\partial n} - \frac{\partial\hat{\theta}(n)}{\partial n}} > 0 \tag{A.5}$$

with
$$\frac{\partial z}{\partial \tau} = (1 - p\,(1+s))\left[u'\left((1-\tau)\,y\right) - \tau y\, u''\left((1-\tau)\,y\right)\right] > 0. \tag{A.6}$$

The Threshold $\tilde{\theta}'(n^*)$

Substituting for (15) we can express (27) as

$$\frac{de_i^*}{d\tau} = \frac{1}{-E\,[u]''}\left\{\frac{\theta_i\, c(n^*)}{\tau} + ps\tau\, u''(x_i^a)\,(y + se_i^*) - (1-p)\,\tau\, u''(x_i^b)\,(y - e_i^*) - \theta_i\, c'(n^*)\frac{dn^*}{d\tau}\right\}. \tag{A.7}$$

The sign of the expression is determined by the term in the curly brackets. Using $\rho\,(x)$ and (8), we get the threshold

$$\tilde{\theta}'(n^*) = \frac{(1-p)\,\tau^2\, u'(x_i^b)\left\{\rho(x_i^a)\,(y + se_i^*) - \rho(x_i^b)\,(y - e_i^*)\right\}}{c(n^*)\left(1 + \tau\rho(x_i^a)\,(y + se_i^*) - \tau\frac{c'(n^*)}{c(n^*)}\frac{dn^*}{d\tau}\right)}. \tag{A.8}$$

For $\theta_i > \tilde{\theta}'(n^*)$ the term in the curly brackets is positive and hence $de_i^*/d\tau > 0$. Finally, we compare this threshold with the partial equilibrium threshold, $\tilde{\theta}(n^*)$ from (A.2). As the numerator of (A.2) and (A.8) are the same, but the denominator of (A.8) is bigger, it immediately follows that $\tilde{\theta}(n^*) > \tilde{\theta}'(n^*)$. Moreover, from (A.5) also follows that the denominator of (A.8) is strictly positive. Therefore $\tilde{\theta}'(n^*) > 0$.

Ad Section 2.5:

From the first order condition (38) we obtain

$$\frac{\partial e_{ij}^*}{\partial n_j} = \frac{\theta_{ij}\,\alpha_j c'(n_j)}{E\,[u_j]''} \geq 0 \tag{A.9}$$
$$\frac{\partial e_{ij}^*}{\partial n_k} = \frac{\theta_{ij}\,(1-\alpha_j)\,c'(n_k)}{E\,[u_j]''} \geq 0. \tag{A.10}$$

Note that $\frac{\partial e_{ij}^*}{\partial n_k} = 0$ if $\alpha_j = 1$.

Implicitly differentiating (41) yields

$$\frac{dn_j^*}{dn_k^*} = \frac{-z_j\,(1-\alpha_j)\,c'(n_k^*)\left(\alpha_j c(n_j^*) + (1-\alpha_j)\,c(n_k^*)\right)^2}{F'(n_j^*)\left(\alpha_j c(n_j^*) + (1-\alpha_j)\,c(n_k^*)\right)^2 + z_j\alpha_j c'(n_j^*)} \geq 0 \tag{A.11}$$

as the denominator must be positive in any stable equilibrium.

If $\alpha_j = 1$, we get $\frac{dn_j^*}{dn_k^*} = 0$.

From (39) and (41) we can derive

$$\frac{dn_j^*}{d\tau_j} = \frac{\frac{\partial z_j}{\partial \tau_j}\left(\alpha_j c(n_j^*) + (1-\alpha_j)\, c(n_k^*)\right)}{F'(n_j^*)\left(\alpha_j c(n_j^*) + (1-\alpha_j)\, c(n_k^*)\right)^2 - z_j\left(\alpha_j c'(n_j^*) + (1-\alpha_j)\, c'(n_k^*)\frac{dn_k^*}{dn_j^*}\right)} > 0. \tag{A.12}$$

From (42) we know that in any stable equilibrium the denominator must be positive. As $\frac{\partial z_j}{\partial \tau_j} > 0$, we get $\frac{dn_j^*}{d\tau_j} > 0$.

Chapter 3

Social Norms, Voting and the Provision of Public Goods

Introductory Remark: *This part of the thesis is the only chapter which is not in the format of a conventional journal article. Here we will discuss several different results within one unifying framework.*

3.1 Introduction

Modern societies have established legal norms and public law enforcement systems which support cooperative behavior in many public good problems. Polluting the environment, evading taxes, free-riding on public transports or simply littering a public place constitutes a violation of law in almost all modern jurisdictions. Free-riders might get detected and punished by a law enforcing authority. In practice, however, we often observe low detection probabilities together with mild sanctions.[1] Hence, law is typically 'non-deterrent', in the sense that free-riding is still the dominant strategy for rational agents with a reasonable degree of risk-aversion.[2] Nevertheless, in many societies broad majorities obey the law and collective action problems are successfully solved (Tyler, 1990). One explanation for this observation is based on the role of social norms (Ullmann-Margalit, 1977, Elster, 1989a, 1989b). In legally regulated situ-

[1] Compare e.g. Andreoni et al. (1998) for the case of tax evasion and Cohen (2000) on environmental regulations.

[2] For a survey on the economic analysis of law, see Polinski and Shavell (2000).

ations, there are typically clear societal expectations about how to behave adequately (Cooter, 1998). If there exists a social norm for cooperation, a formal (legal) and an informal (social) norm concur in supporting compliance with the law.[3] In this case, public deterrence policies together with social norm enforcement may evoke pro-social behavior.

We study this idea in a model of public good provision within a large population. Individuals choose to either contribute a fixed amount to the public good or to free-ride. In the latter case, an agent might get detected by a centralized institution ('police') and has to pay the public good contribution plus a fine. We assume that these sanctions are non-deterrent, such that a rational agent would always choose to free-ride. However, next to the monetary incentives shaped by the legal system, behavior is also guided by an *internalized* social norm for cooperation (Elster, 1989a). Agents are heterogenous with respect to their norm sensitivity: While some are completely unaffected by the social norm, others suffer from self-imposed sanctions (feelings like guilt or remorse) if they deviate from the norm. The strength of these internal sanctions depends – next to the individual specific degree of norm internalization – on the strength of the social norm in the society. Following the literature,[4] we assume that the '*power of a norm to command adherence is positively associated with its level of adherence in the population.*' (Kreps, 1997, p. 359). If free-riding is widespread, the perceived obligation to follow the social norm is weak and norm-based sanctions for free-riding are low. The opposite holds for high levels of cooperation. According to this logic, the strength of the norm is then endogenously determined, raising the possibility of multiple equilibria.[5]

Within this framework we first characterize individual behavior. Depending on the degree of norm sensitivity, individuals can be characterized as free-riders, unconditional contributors or 'conditional contributors' (Keser and van Winden, 2000, Fischbacher et al., 2001). The first two types of agents – with a very low respectively high degree of norm internalization – will always free-ride respectively cooperate. Agents with an intermediate norm-sensitivity condition cooperation on the behavior of other members in society: They tend to cooperate, if norm-adherence in society is strong, but free-ride

[3]On social norms and legal institutions compare e.g. Kahan (1997), Posner (1997, 2000) and the special issue on 'Social Norms, Social Meaning and the Economic Analysis of Law' in *Journal of Legal Studies* 27 (June 1998).

[4]E.g. Akerlof (1980), Gordon (1989), Lindbeck et al. (1999), Myles and Naylor (1996), Naylor (1989).

[5]Related approaches studying multiple equilibria of criminal activities are Sah (1991), Rasmusen (1996) and Arbak (2005).

if norm violations become widespread. Here, this behavior is triggered by a cooperation norm with an endogenous strength. Hence, our model provides an alternative to the notion of a 'norm of conditional cooperation', recently introduced by Fehr and Fischbacher (2004a, 2004b).

We then determine the equilibrium level of contributions depending on the pecuniary incentives as well as on the strength of the social norm. Within this context, we discuss the role of belief management as a policy tool for equilibrium selection. Moreover, we provide an alternative explanation for the *broken windows effect* (Wilson and Kelling, 1982): The strong deterrent impact of a zero-tolerance policy, as implemented by the former NY City Mayor Rudolf Giuliani, may not only rely on its direct effect – higher (expected) sanctions associated with the strict enforcement of public order. Removing public signs of social disorder – 'fixing broken windows' – also conveys a signal about widespread compliance with norms of pro-social conduct, strengthen these norms and thereby reduce the frequency of norm violations.

While the basic model framework describes a broad class of large-scale collective action problems, we then focus on the interpretation of a tax evasion game. We show that in equilibrium, higher taxes raise the share of evasion, whereas stricter law enforcement works into the opposite direction. Taking this behavior into account, we derive the Laffer curve for this model. We find that for some levels of taxation the multiplicity of equilibria results in discontinuous jumps in collected revenues. Next, we study the endogenous choice of the tax policy. We first discuss welfare optimal taxation in the context of social norms. Following the approach taken up by Lindbeck et al. (1999) we then analyze individual voting behavior. We show that majority voting will result in a suboptimal low level of taxation and an underprovision of public goods. This result is driven by an endogenous difference between the mean marginal costs of public good provision and the costs faced by the decisive voters. In the voting equilibrium, the decisive voter is a cooperator who 'voluntarily' pays taxes and therefore bears the full costs of any tax increase. However, there are also free-riders who pay the tax only if they get detected. Hence, the mean costs of public good provision considered by the social planner are below the costs taken into account by the decisive voter. Thus, contributors will always prefer a lower level of taxation and public good provision than would be optimal.

One can generalize this result in the following way: Given that public good contributions are not fully enforced, any voting equilibrium where society consists of contributors and free-riders, will result in an inefficient level of taxation respectively public good provision. If, as in our case, the contributors form a majority, taxation will be inefficiently low resulting in an underprovision of public good. If, however, the free-riders form a majority, voting will result in an inefficiently high tax rate. This provides the main result of our analysis, which has so far been neglected in the literature on voting and public good provision. Hence, our analysis complements the conventional under- respectively overprovision result from the classical collective choice literature (Romer, 1975; Roberts, 1977; Meltzer and Richard, 1981).

The inefficiency of voting outcomes in standard textbook models is driven by *exogenous* differences in the mean and the median voters' benefits respectively costs from public good provision (compare e.g. Persson and Tabellini, 2000). In our approach the discrepancy between mean and median preferences arises from an *endogenous* mechanism: The voting outcome defines a tax, for which agents – depending on their norm sensitivity – decide on whether to evade or to contribute. The (exogenous) heterogeneity in norm sensitivity together with the endogenous level of taxation then translates into an endogenous heterogeneity in behavior, which results in different costs of taxation. This drives the wedge between the decisive voter's and the social costs of public good provision, resulting in an underprovision of public goods.

Furthermore, we also contribute to the recent discussion on the role of social norms for law enforcement. While this literature is mainly based on verbal reasoning (e.g. Posner, 1997, 2000; Kahan, 1997, 2005) or empirical studies (e.g. Galbiati and Vertova, 2005; Tyran and Feld, 2005), we provide a simple model framework which combines a conventional Beckerian approach to the economics of crime with the literature on internalized norms (Akerlof, 1980). Within this framework, we endogenize the strength of law-enforcement. We consider the revenue maximizing choice of the detection frequency together with the welfare optimal tax policy. We then study, amongst others, the effect of an exogenous shock in the norm sensitivity on the detection policy. Though such a shock always leads to a higher level of cooperation, the reaction of the law enforcing authority is ambiguous: A strengthening of the social norm could be accompanied by a stricter legal norm: In this case, formal and informal enforcement mechanism would work as complements. However, a stronger social norm could – to some extent – also substitute the legal norm. Hence, our model provides a theoretical foundation for the verbal arguments of Richard Posner (1997, p.369), according to which '*law both complements and substitutes for norms*'.

The chapter is organized as follows. In section 3.2 we introduce the basic model and discuss the individual contribution decision. Next we characterize the social equilibrium of the model. In section 3.4 we provide a welfare economic analysis, and compare it to the voting equilibrium, derived in section 3.5. Section 3.6 describes the endogenous choice of law enforcement and section 3.7 summarizes the main results. All proofs appear in the appendix.

3.2 A Model of Internalized Social Norms

A large scale society is represented by a continuum of agents with unit mass. Players simultaneously choose $x \in \{0, 1\}$, either to follow the law and contribute a fixed amount t ('tax') to the public good or to violate against the law and contribute nothing. Let us denote the first choice by $x = 1$ (contributing) and the latter by $x = 0$ (free-riding). If a player decides to free-ride, she will get punished by a public law enforcement institution with probability p and has to pay t plus a fine f.[6] With probability $1 - p$ she will get away undetected. Throughout the whole analysis we assume that $t \geq p(t + f)$ holds. In our model this implies that a rational, risk neutral agent who is only concerned about monetary payoffs (*homo economicus*) will always choose to free-ride: The law is non-deterrent.

3.2.1 Preferences

Agents preferences are defined over private consumption, a public good and a norm-based payoff. Consumption is given by an exogenous income minus the expected payments for each strategy x.[7] The utility from the public good g is given by $\delta v(g)$, with $v(.)$ being twice continuously differentiable and strictly concave, $v' > 0$, $v'' < 0$ and $\delta > 0$. Finally, the norm-based payoff is given by $-\theta_i c(n)$ in the case of free-riding and zero otherwise.[8] The individual specific parameter θ_i, with $\theta_i \geq 0$, reflects different degrees of norm-sensitivity. The function $c(n)$ reflects the 'social' strength of the norm, depending on n, the share of free-riders in the society. Assuming an additively separable preference structure and risk neutrality,[9] the utility function of agent i is

[6] The tax t and the fine f are assumed to be non-negative and $0 \leq p < 1$.

[7] In the following we will omit the constant income.

[8] In an earlier version of this chapter, we considered a negative payoff from free-riding as well as a positive norm-payoff associated with cooperation. However, since this decomposition only change our results quantitatively, we dropped this model component.

[9] Introducing risk-aversion would not change the main results of our analysis in a qualitative way.

given by

$$u_i(x,n) = -xt - (1-x)\left(p\,(t+f) + \theta_i c(n)\right) + \delta v(g). \tag{1}$$

We can express individuals' utility as

$$u_i(1,n) = -t + \delta v(g) \tag{2}$$

for $x = 1$, respectively as

$$u_i(0,n) = -p\,(t+f) - \theta_i c(n) + \delta v(g) \tag{3}$$

for $x = 0$.

This preference structure describes a society with a social norm for cooperation or equivalently, a norm against free-riding. If an agent deviates from this norm, she will incur sanctions which transform into a utility loss of $\theta_i c(n)$. While the origin of these sanctions could in principle be external (e.g. social disapproval), we here assume that they stem from an internal mechanism, based upon an *internalized* cooperation norm (Elster, 1989).[10] One important mechanism which represents such a system of 'internal', self-imposed sanctions are emotions in the context of norm guided situations.[11] A deviation from an internalized norm is typically accompanied '*by internal sanctions, including shame, guilt and loss of self-esteem, as opposed to purely external sanctions* [...]' (Gintis, 2003, p.407). Hence, emotional costs may induce an agent to abstain from breaking the norm. The parameter θ_i can be interpreted as a measure for the degree of norm-internalization:[12] Low values of θ_i describe individuals who are hardly affected by the emotions associated with a norm violation (relative to the monetary incentives), whereas the well-being of agents with a high θ_i is very sensitive towards norm-deviations: they suffer from strong feelings of guilt or remorse if they deviate from the norm.

While θ_i captures interpersonal differences in the level of norm-sensitivity, $c(n)$ depicts variations of the sanctions' magnitude for different levels of norm violations. We employ the following

[10] Compare also Becker (1996, p.225) on internalized norms. For models of social norms with external sanctions see e.g. Holländer (1990) and Rege (2004).

[11] A review on the role of emotions in economic theory is provided by Elster (1998). For empirical research, which has demonstrated the importance of emotions for economic decision making, compare Ben-Shakhar et al. (2004), Hopfensitz and van Winden (2005), Reuben and van Winden, (2005). Compare also the theoretical models in van Winden (2001), Bowles and Gintis (2003).

[12] Note that we exclude negative values of θ. While we could easily allow for 'punks', who derive benefits from a norm-violation, we would not gain any additional insights.

Assumptions: (A1) The finite-valued function $c(n)$ is continuously differentiable in n, mapping the unit interval into the $\mathbb{R}_+$. **(A2)** Moreover, the extent of the norm-enforcing sanctions is non-increasing in n, i.e. $c'(n) \leq 0$ for all $n \in [0,1]$,

where we write $c'(.)$ for $\partial c/\partial n$. This assumption is in line with the literature on social norms(Akerlof, 1980, Naylor, 1989, Lindbeck et al., 1999), assuming that a norm-deviation is (emotionally) less costly, if norm-violations are widespread. The intuition behind this assumption is straightforward: the more people in society follow a social norm, the stronger it is and the higher are the individual costs of acting against it. The more people deviate, however, the easier it becomes for an agent to justify her own norm-violation. Empirical support for this line of reasoning comes from studies on norm guided behavior in economic (e.g. Azar, 2004, Nyborg and Rege, 2003), in the law literature (e.g. Grasmick and Green, 1982, Liu, 2003) as well as in social psychology (e.g. Cialdini et al., 1990, Reno et al., 1993).

3.2.2 Individual Decision

Let us call the triplet $(t,\, p,\, f)$ a *policy*, consisting of a *tax policy* t and an *enforcement policy* $(p,\, f)$, and remember from above that we assume $t \geq p(t+f)$. Note further, that in a large community the impact of a single individual on n respectively g is negligible. Hence, taking the policy as well as n as given, a player i will choose to contribute iff $u_i(1,n) \geq u_i(0,n)$. It is easy to show that this holds iff $\theta_i \geq \hat{\theta}(n)$ with

$$\hat{\theta}(n) := \frac{t - p\,(t+f)}{c(n)}. \tag{4}$$

For agent i, who has internalized the norm to a degree $\theta_i \geq \hat{\theta}(n)$, the utility loss from the norm violation would be bigger than (or equal to) the monetary benefits from free-riding.[13,14] Hence, the player will contribute t to the public good. If the enforcement policy gets stricter or if the tax decreases, the monetary incentive to free-ride becomes weaker and $\hat{\theta}(n)$ declines. For the special case of $t = p(t+f)$ we get $\hat{\theta}(n) = 0$. In this case, law enforcement is strong enough to enforce contributions of all (risk neutral) agents with $\theta_i \geq 0$. As long as the numerator of $\hat{\theta}$ is strictly positive, the law is non-deterrent for agents with $\theta_i = 0$; i.e., expected monetary sanctions alone do not result

[13]For the case of equality we assume that the player follows the norm.

[14]We assume here that people rationally solve the trade off between monetary and emotional incentives. Empirical support for this assumption is provided by e.g. Bosman and van Winden (2001, 2002), Hopfensitz and van Winden (2005).

in cooperative behavior. However, the (centralized, external) monetary punishment together with the (decentralized, internal) norm-based sanctions may induce public good contributions.

For a given policy and distribution of θ, this model captures three different behavioral patterns: On the one hand, individuals with $\theta_i < \hat{\theta}(0)$ will always *free-ride*. These players are hardly affected by the social norm and the economic incentive to free-ride is always dominant, even if the whole society would comply with the norm. On the other hand, the action of an individual with $\hat{\theta}(0) \leq \theta_i < \hat{\theta}(1)$ depends on the level of n. Assumption A2 implies that the threshold $\hat{\theta}(n)$ is increasing in n: with a higher share of free-riders, the social norm becomes weaker and it is less costly to deviate from the norm. Hence, an agent might contribute to the public good for low levels of norm-deviations, while he would free-ride if the norm becomes weak. One can characterize such agents as *conditional contributors*.[15] (3) Finally, agents with $\theta_i \geq \hat{\theta}(1)$ act as *unconditional contributors*: the utility loss from a norm violation always dominates the monetary incentive to free-ride, even for the (hypothetical) case where everybody in society would break the norm. The model therefore explains the typical patterns of individual behavior observed in experimental economic studies of public good games (e.g. Keser and van Winden, 2000; Fischbacher et al., 2001) as well as in field studies (e.g. Croson, 2005; Frey and Meier, 2004a, 2004b). We shall emphasize that here conditional contributors appear in the context of a cooperation norm with an endogenous norm strength. Hence, conditional contributions are not necessarily triggered by a 'norm of conditional cooperation', as recently proposed by Fehr and Fischbacher (2004a, 2004b). While they consider a norm which commands agents to 'cooperate in this case' but 'free-ride under that conditions', it has been stressed in the social norms literature (e.g. Elster, 1989a, 1989b), that norms are typically simple moral rules and guidelines. Therefore, our approach seems to be more in line with the standard literature on social norms.

Digression: Congestion in the Detection Technology

Consider the case where the detection technology of the law-enforcing institution depends on the frequency of law violations in the society. It seems plausible to assume that, for a given investment into law-enforcement (e.g. a fixed number of policemen), the detection probability is decreasing with the number of illegal actions. Stated differently, there is congestion in the detection technology: If a policeman is busy sanctioning

[15] Keser and van Winden (2000), Fischbacher et al. (2001).

agent i, it is more likely that the wrongdoing of agent j stays undetect as compared to a situation where i would comply with the law and the policeman could fully focus on detecting law violations.[16]

We can easily include this idea into the model, assuming that the detection probability is given by $p = \pi(n)$ with $\pi' < 0$. The threshold level for cooperation from (4) would then be given by

$$\hat{\theta}(n) := \frac{t - \pi(n)\,(t+f)}{c(n)}.$$

It is straightforward to show, that $\hat{\theta}(n)$ still is increasing in n. As before, the emotional costs of free-riding are decreasing in n. With $\pi' < 0$, however, norm violation have an additional positive externality on other free-riders: The more people break the law, the lower are the expected monetary sanctions and the temptation to free-ride rises. Hence, the pecuniary incentives related to congestion point into the same direction as the incentives associated with the endogenous norm strength. They both trigger conditional cooperative behavior. As the congestion effect would only change our results quantitatively, we will assume $\pi' = 0$ in the remainder of this chapter.

3.3 Social Equilibrium

Let θ be distributed over the interval $[0, \infty]$ according to a cumulative distribution function $F(\theta)$, which is continuously differentiable and common knowledge. The corresponding density function $f(\theta)$ has full support on the interval $[0, \infty]$ and $f(\theta) > 0$. $F^{-1}(n)$ denotes the inverse function of $F(\theta)$.

3.3.1 Equilibrium Share of Free-Riders

For a given policy an equilibrium share of free-riders n^* is defined by $n^* = F\left(\hat{\theta}(n^*)\right)$, or equivalently

$$n^* = F\left(\frac{t - p\,(t+f)}{c(n^*)}\right). \tag{5}$$

Equation (5) is a fix-point equation in n, where $F(\hat{\theta}(n))$ maps the unit interval into itself. The solution to (5) yields a Nash equilibrium, i.e. a self-supporting share of free-riders. Given the fraction of norm-violations n^*, the strength of the norm in the

[16] Anecdotal evidence, in particular for the case of tax enforcement, supports this line of reasoning. Compare chapter 1 of this thesis. A theoretical discussion is provided by Sah (1991).

society is such that there are exactly n^* agents with $\theta_i < \hat{\theta}(n^*)$. We call such an equilibrium locally stable, if

$$\frac{\partial F^{-1}(n^*)}{\partial n} \geq \frac{\partial \hat{\theta}(n^*)}{\partial n}. \tag{6}$$

With this, we can derive

Proposition 1 ***(i)*** *For any policy (t, p, f) with $t \geq p(t+f)$, any distribution $F(\theta)$ and any $c(n)$ fulfilling A1 and A2, there exists at least one locally stable equilibrium share of free-riders, $n^* \in [0,1]$.* ***(ii)*** *If $c'(n) = 0$ for all $n \in [0,1]$, there exists a unique globally stable equilibrium share of free-riders, $n^* \in [0,1]$.*

Proof. See Appendix. ∎

Note, that the equilibrium $n^* = 0$ can only occur for the special case $t = p\,(t+f)$ and $n^* = 1$ is only supported if $t \to \infty$.[17] If we exclude these special cases, we get $n^* \in (0,1)$.

Proposition 1 guarantees the existence of at least one stable equilibrium share n^*. If the emotional costs of a norm deviation were independent of the others' behavior (i.e. $c'(n) = 0$), there would exist exactly one solution to (5). Typically, however, there will be a multiplicity of equilibria: one policy could result in several equilibrium levels of n^*. An example is depicted in figure 3.1.

In the case represented in figure 3.1 (see next page), there are three possible equilibria. A 'good' equilibrium, with a low share of norm violators n_l^*, a 'bad' equilibrium with a large fraction of free-riders n_h^*, and a third equilibrium, n_m^*, somewhere in between. As condition (6) holds, whenever the $\hat{\theta}$-curve intersects the cumulative distribution from above, there are two (locally) stable equilibria, n_l^* and n_h^*, and one unstable equilibrium, n_m^*.

While our modelling approach is a static one, the underlying dynamics of this system are straightforward. Assume that the game is played repeatedly and agents neither know the distribution $F(\theta)$ nor do they observe the level of n in the present period. If, however, they can infer from the public good level the last period's share of free-riding, society would converge towards a locally stable equilibrium: Starting from any off-equilibrium share $n < n_m^*$ $[n > n_m^*]$, society would end up in an equilibrium with n_l^* $[n_h^*]$.[18]

[17] This is driven by assuming $c(1) > 0$ together with $F^{-1}(1) = \infty$.

[18] For a more detailed discussion of the dynamics of this class of models see Lindbeck et al. (1999, 2002) and Rege (2004).

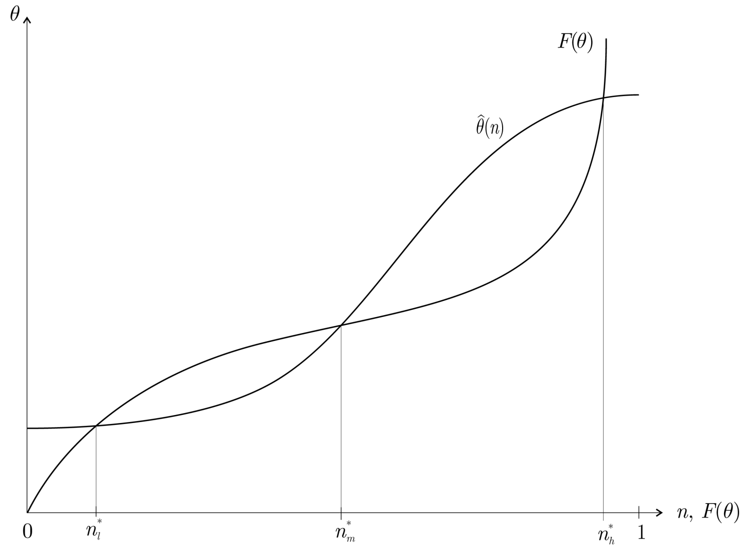

Figure 3.1: Multiple Equilibria

A comparative static analysis within this framework yields the following results:

Proposition 2 *A marginal increase in t, decrease in p or decrease in f will independently raise the monetary incentive to free-ride. As long as $p(t+f) < t < \infty$, this will lead to an increase [decrease] in the share of free-riders n^* in any stable [instable] equilibrium.*

Proof. See Appendix. ■

Proposition 2 is quite intuitive: if the monetary temptation to free-ride becomes bigger, the threshold $\hat{\theta}(n)$ rises. The $\hat{\theta}$-curve in figure 3.1 shifts upwards for any level of n, which in turn results in a higher equilibrium share of n^* for any stable equilibrium.

Digression: Beliefs and Broken Windows

As we focus on the case where agents have learned about the actual level of free-riding n, we neglect an important aspect: In our framework, 'belief management' – i.e. the manipulation of beliefs – constitutes a possible policy tool. Consider a society in an equilibrium with frequent free-riding, where all people hold correct beliefs about n: Take for example the case of an anti-littering norm in a park. A littered park provides a clear signal about the behavior of others and indicates that norm-violations are widespread. Society finds itself in a 'littering equilibrium'. If, however, every morning a cleaning squad clears the park, this could – over time – induce a downward bias in the beliefs about n. The anti-littering norm would be perceived stronger than it actually is; the threshold $\hat{\theta}(.)$ would fall and some people, who would have littered in a more dirty park, may refrain from littering. If behavior adapts this way, the initial off-equilibrium beliefs could turn into correct ones. Hence, belief manipulation constitutes a tool to induce one particular equilibrium state.

In this vein, our model also provides an alternative explanation to the *broken window effect*: It has been claimed that public signs of disorder trigger further breaches of the public order. '*If a window in a building is broken and is left unrepaired, all the rest of the windows will soon be broken.* [...] *One unrepaired broken window is a signal that no one cares, and so breaking more windows costs nothing*' (Wilson and Kelling, 1982, p.31).[19] Translated into the terminology of our model: Agents form beliefs about the law enforcement policy and signals for a high level of n are assumed to signal low levels of p (respectively f). In our framework, however, the endogenous strength of a social norm provides a different channel, explaining conditional cooperative behavior: Clear indications of social disorder signal that norms for pro-social behavior are weak – or weaker than expected. Consequently, the inhibition threshold for anti-social behavior as street-crime declines and triggers further norm violations (Kahan, 1997, 2005; Liu, 2003). Reversing this logic, policy can signal strong(er) social norms, by 'fixing broken windows'. Expressing strong community disapprobation of minor misbehavior can then induce a positive bandwagon effect (Kahan, 1997). The discrimination between competing explanations of conditional cooperation is finally an empirical task, left for future research.

3.3.2 Social Equilibrium State

Equation (5) defines a correspondence which maps a policy into – typically several – equilibrium levels of n^*. In order to avoid the technical difficulties linked to the multiplicity of equilibria, we can 'turn around' this mapping. Instead of asking which policy induces which equilibrium, we can also ask, which equilibrium n^* is supported by which (tax) policy. Rearranging (5), we get

$$t(n) = \frac{F^{-1}(n)c(n) + pf}{1 - p}, \tag{7}$$

[19]This explanation also formed the basis for the zero-tolerance policy of the former New York city mayor Rudolph Guiliani. Compare also Kahan (1997) and Posner (1997, p.367).

with $t(n) : [0,1] \rightarrow [0,\infty]$. Hence, for any $n^* \in [0,1]$, the continuous function $t(n)$ yields one tax t^* such that for $t^* = t(n^*)$ the population share n^* is an equilibrium according to (5). Stated differently, t^* is the equilibrium tax rate associated with n^*.[20]

Finally we turn to the public good g. Let the marginal rate of transforming the private good into g be constant and equal to unity. For a given policy (p, f) the public costs for law enforcement are given by d.[21] The level of the public good is equal to the total amount of contributions ('tax revenues') plus the expected payments of the punished free-riders minus enforcement costs d,

$$g = (1-n)\,t + np\,(t+f) - d. \tag{8}$$

Substituting (7) and rearranging, we get the public good level as a function of n,

$$g(n) = t(n)\,(1 - n\,(1-p)) + npf - d. \tag{9}$$

This function defines $g^* = g(n^*)$, the revenues for an equilibrium share n^* and the corresponding equilibrium tax $t(n^*)$. We now can define a social equilibrium state as follows:

Definition 1 ***(i)*** *With an enforcement policy (p, f) as well as $c(n)$, d and $F(\theta)$ exogenously given, a social equilibrium state $e = (n^*, t^*, g^*)$ is a combination of a share of free-riders $n^* \in [0,1]$, a tax $t^* = t(n^*)$ and a public good level $g^* = g(n^*)$ defined by (5), (7) and (9).* ***(ii)*** *A social equilibrium state e characterized by n^* is called stable if $t'(n^*) \geq 0$.* ***(iii)*** *The set of possible social equilibrium states in a society is characterized by $S = \{(n^*, t(n^*), g(n^*)) : 0 \leq n^* \leq 1\}$ with $S \in [0,1] \times [0,\infty] \times \mathbb{R} \cup \{+\infty\}$.*

Note that any social equilibrium state e is fully characterized by the equilibrium share of free-riders n^*, as all components of e are functions of n^*, i.e. $e = (n^*, t(n^*), g(n^*))$.[22] The characterization of a stable equilibrium state then follows immediately from Proposition 2: Since a tax increase raises n^* in any stable equilibrium, there must also hold $t' \geq 0$.

[20] For the case of multiple equilibria, $t(n)$ is non-injective, since there are different equilibrium levels $n^* \in [0,1]$ which are supported by the same tax t^*.

[21] Since the punishment policy is exogenously fixed, we can treat these costs as an exogenous parameter. When we later endogenize the choice of p, we will consider d as a function of p.

[22] By this definition we exclude corner solutions with $t^* > t(1)$ from the set of possible social equilibrium states. Corner solutions at $n^* = 0$ are ruled out by the assumption $t \geq p\,(t+f)$.

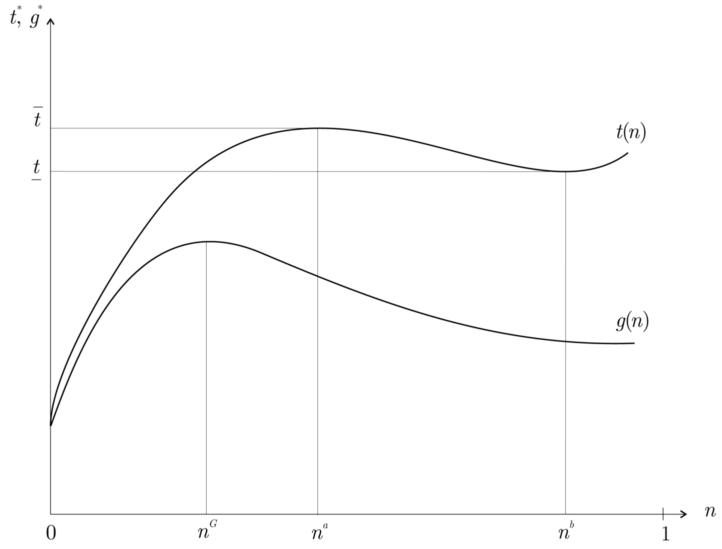

Figure 3.2: Social Equilibrium States

A graphical representation of the set S is given in figure 3.2. While the shape of $t(n)$ and $g(n)$ obviously depend on exogenous factors – in particular $F(\theta)$ and $c(n)$ – the example of figure 3.2 depicts some general properties which hold for any possible scenario within this model framework. Consider first the equilibrium state where nobody free-rides. One can show that in this state the function $t(n)$ has its global minimum.[23] From (7) we can derive

$$t'(n) = \frac{1}{1-p}\left(\frac{\partial F^{-1}}{\partial n}c(n) + F^{-1}\,c'(n)\right). \tag{10}$$

Making use of (7), we get from (9)

$$g'(n) = t'(n)\,(1 - n\,(1-p)) - F^{-1}(n)\,c(n). \tag{11}$$

As $F^{-1}(n) = 0$, there always holds $g'(0) = t'(0) > 0$. Hence, at $n^* = 0$, a tax increase triggers more norm deviations but raises at the same time more revenues. At a certain tax level, however, the first order effect of a tax increase is counterbalanced by the decline in contributions due to the increase in free-riding, and the system is at the peak of the Laffer curve. In figure 3.2, the corresponding level of free-riding is indicated by n^G. As t^* is further increased – i.e. if we move from n^G further to the right – this will result in a decrease in g^*. This range of n^* is therefore characterized

[23]Since the term $F^{-1}(n)\,c(n)$ is positive for any $n \in (0,1]$, it immediately follows from (7) that $t(n) > t(0)$ for all $n \in (0,1]$.

by $t' > 0$ together with $g' < 0$.

In the case of figure 3.2, we then reach a level of norm violations n^a, where $t'(n^a) = 0$. For $n^* > n^a$ the sign of t' becomes negative, and we get into a range of unstable equilibrium states (compare Definition 1(ii)). At a share of free-riders n^b, with $t'(n^b) = 0$, the sign of t' turns around again, reflecting a range of stable equilibrium states characterized by a high fraction of free-riders.

From figure 3.2 one can also see that in this model the multiplicity of equilibria is a local rather than a global property. Only in the range of $\underline{t} < t^* < \bar{t}$ there are several – in this case, three – possible equilibrium levels of n^* respectively g^*.[24] Given there exists a $\underline{t}$ – that is, if there is a range of t^* characterized by at least two stable equilibria – it follows that $\underline{t} > t(0)$ (since $t(n)$ has its global minimum at $n^* = 0$). Hence, there always exists a range of taxes, $t(0) \leq t^* < \underline{t}$, where a tax policy induces one unique equilibrium share of free-riders. We will come back to this point in the next section.

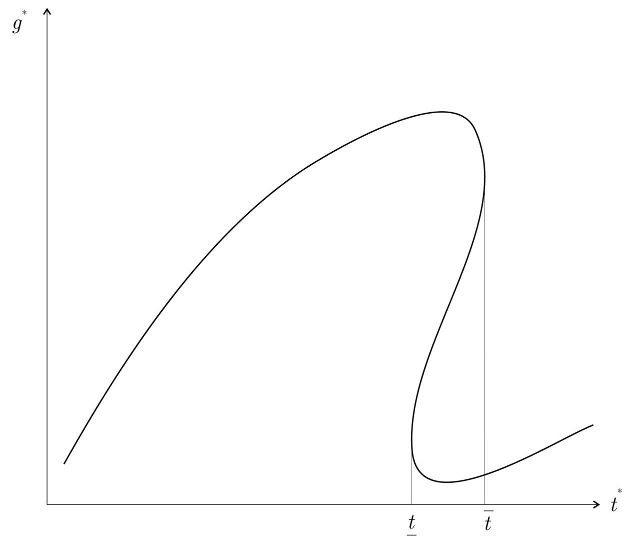

Figure 3.3: Laffer curve

We can represent the set of social equilibrium states in an alternative way, plotting all possible equilibrium tax levels t^* (horizontal axis) against the corresponding public good level g^* (vertical axis). Figure 3.3 shows that in the case of multiple equilibria the

[24]Let us consider the lowest and the highest equilibrium tax, where the $\hat{\theta}(n)$-curve (4) is tangential to $F^{-1}(n)$ (compare figure 3.1) and hence $t'(n) = 0$. (A tangential point n^* is defined by $\frac{\partial F^{-1}(n)}{\partial n} = -\hat{\theta}(n)\frac{c'(n)}{c(n)}$. Using this condition in (10), we get $t'(n) = 0$.) These tax levels are defined by $\underline{t} := \min\{t(n) : t'(n) = 0\}$ respectively $\bar{t} := \max\{t(n) : t'(n) = 0\}$, for $n \in [0, 1]$.

resulting Laffer curve of this system is folded. As we have discussed above, raising the tax from $t(0)$ would first be accompanied by an increase in g^*. On the downward-sloping side of the Laffer curve, however, there is a discontinuous jump in the revenues at $\bar{t}$. For any tax above this level, equilibria are unique and characterized by a weak social norm, widespread free-riding and low levels of public good provision.[25] Hence, at this level of taxation, a small change in the tax policy may trigger a huge increase in the evasion level. If one policy would be associated with more than three equilibria, there would exist several such levels of taxation where minor tax increases induce discontinuous changes in revenues. While standard theory considers a range with monotonically increasing tax revenues followed by monotonically decreasing tax revenues, our analysis suggest that the concept of a continuous Laffer curve may be misleading. A Laffer curve could be characterized by several upward- and downward-sloping ranges, with several peaks and several discontinuous jumps in the revenue level.

3.4 Welfare Economics

In this section we address the question of policy choice in a framework with social norms. We take the enforcement policy as exogenously given and focus on the welfare maximizing choice of a tax policy. In section 3.6 we will also endogenize the choice of the detection probability p. Applying standard welfare economics in this framework, two conceptual difficulties emerge. First: How can a social planner 'choose' between several equilibrium states? and second: How shall we incorporate norm-based payoffs into the concept of social welfare?

3.4.1 Tax Policy and Equilibrium Selection

As we have discussed in section 3.3, this model is in general characterized by multiple equilibria. In contrast to standard taxation problems, a planner can therefore not 'automatically' pick a combination of a tax and a public good level, since for one tax policy society could coordinate on several equilibria, resulting in different budget balancing public good levels. As different equilibria are associated with different levels of welfare, the question of equilibrium selection arises. How can the planner induce one particular social equilibrium state $e \in S$?

[25] The backward bending part of the Laffer curve in the range of $\underline{t} < t^* < \bar{t}$ represents the unstable states of the system.

One strategy is based on belief management, as discussed in section 3.3. Another possibility is related to the salience of certain equilibria. Consider the case, where society is in an equilibrium state e, with $t^* < \underline{t}$. From the discussion of figure 3.2 and 3.3 we know that such a tax policy induces one unique equilibrium share n^*. If the planner (marginally) increases the tax above $\underline{t}$, the coordination problem arises again. Note, however, that there is always a potential equilibrium state e', with n' being in the neighborhood of the former equilibrium share of free-riders.[26] It seems plausible, that such a close-by equilibrium state is more salient than alternative, more distant states. If agents coordinate according to this concept of salience, the social equilibrium state in the neighborhood of the previous one becomes a focal point equilibrium (Schelling, 1960). With stepwise tax policy changes, a planner can move along the Laffer curve (Figure 3.3) and society can be guided into any targeted (stable) equilibrium state.[27] In the following, we will neglect any problems of equilibrium selection and assume that the planner (respectively the law enforcing authority in section 3.6) can induce any stable equilibrium state.

Let us add a remark on the applicability of such stepwise policy adjustments. Realistically, an adaptation process from one to another equilibrium state would take quite some time. In this case, the transformation phase could be characterized by welfare levels which are even below the original (suboptimal) equilibrium state. From the perspective of dynamic welfare optimization, it could therefore indeed be optimal to stay in an inefficient equilibrium rather than engage in a very costly transformation. To stress the importance of this point, consider as an example the pattern of tax evasion in post-Soviet Russia. As Rothstein (2000) points out, many Russians do not consider tax evasion as a serious crime, since they are convinced that most citizens violate tax laws. In terms of our model, n is high and $c(n)$ is low: The economy is in the range of the Laffer curve associated with 'bad' equilibria.[28] According to our model, one measure to raise tax compliance consists in a radical tax cut, accompanied by a policy which signals that for this low level of taxation most Russians will comply with tax laws. Tax payers would change their behavior and society converges towards a low-tax-low-evasion equilibrium. Given that initially evasion would remain at a high level, the transition period to this new equilibrium state would be characterized by a period of extremely poor revenue levels. In the context of political constraints to raise

[26] Remember that $t(n)$ and $g(n)$ are continuous functions in n.

[27] If there are more than three equilibria for one tax, there may be equilibrium states which can not be implemented by stepwise tax adjustments. However, all equilibrium states in the neighborhood of $n^* = 0$ which are characterized by $t'(n) \geq 0$ can always be implemented.

[28] Additionally, a significant mass of θ-types may be concentrated around zero.

a minimal level of tax revenues, the government would be locked in a suboptimal mid-tax-high-evasion state – even if it would be desirable to engage in the transformation from a long term perspective.

3.4.2 Welfarist versus Non-Welfarist Methods of Policy Assessment

How to undertake a social welfare analysis in the context of 'social preferences' (Fehr and Falk, 2002), i.e. if agents are governed by incentives beyond the mere monetary self-interest? More specifically, how should a utilitarian welfare function look like, if individuals' utility is defined over 'social' arguments[29], such as the (payoff) inequality between different agents (Bolton and Ockenfels, 2000, Fehr and Schmidt, 1999), or – as in our case – the emotional costs of norm violations? While the thorough and detailed discussion, which is appropriate in order to address this fundamental point, is beyond the scope of this chapter, we only provide a brief discussion of two possible approaches to our specific problem.[30]

A Welfarist Approach

In standard social welfare analysis, any social evaluation of a situation is based on individual preferences. If a policy decreases the utility of an agent, this must be reflected in the social welfare measure. One possible measure of social welfare is that of classical utilitarianism, the sum of individuals' utilities. In our case, this sum also includes the disutility free-riders suffer from breaking the norm – the utility losses based upon an emotional, self-imposed sanction. Shavell (2003, p.5) discusses that *'it is of no moment from the perspective of welfare economics'*, that such sources of (dis)utility are *'different in their character from conventional springs of utility and disutility'*. According to this view, we should follow a purely *welfarist* concept, in the sense that social welfare is exclusively based upon individuals' preferences. The

[29]Arguments in a utility function are *social,* if they depend on other individuals' actions respectively payoffs.

[30]Compare Bernheim and Rangel (2005) for a detailed discussion of welfare economics in the context of non-selfish agents. A related strand of literature discusses the concept of social welfare in the presence of bounded-rational (but selfish) agents. Compare O'Donoghue and Rabin (2003), Camerer et al. (2003), Kanbur et al. (2004).

planners' taxation problem then becomes

$$\max_{t} W = -(1-n)\,t - np\,(t+f) - c(n)\int_0^{\hat{\theta}(n)} \theta\, dF(\theta) + \delta v(g). \tag{12}$$

Due to the multiplicity of equilibria, the first-order condition to this problem does in general not fully characterize the welfare optimal equilibrium state. Hence, it will be more convenient to 'turn the problem around' and solve for n^*. Using (7) and (9) we can rewrite (12) as[31]

$$\max_{n^*} W = d - g(n^*) - c(n^*)\int_0^{\hat{\theta}(n^*)} \theta\ dF(\theta) + \delta\, v\left(g(n^*)\right). \tag{13}$$

The first order condition for an interior solution is given by

$$\delta v'\left(g(n^w)\right) = 1 + \frac{\Psi(n^w)}{g'(n^w)}, \tag{14}$$

with $g' \neq 0$ and

$$\Psi(n^*) = c(n^*)\,\hat{\theta}(n^*) + c'(n^*)\int_0^{\hat{\theta}(n^*)} \theta\, dF(\theta). \tag{15}$$

is derived in the Appendix. Condition (14) defines n^w and thereby $e^w = (n^w, t^w, g^w)$, a social equilibrium state which is optimal according to this welfarist approach.[32] One can interpret (14) as a Samuelson condition for the optimal choice of a tax $t^w = t(n^w)$ which induces the optimal equilibrium share of free-riders n^w and results in the public good $g^w = g(n^w)$. On the left hand side (LHS) of (14) we find the marginal utility from the public good. The right hand side (RHS) depicts the marginal costs of the public good provision. The first term on the RHS, unity, reflects the monetary costs associated with the contributions (taxes plus fines) to the public good. The second term characterizes the non-monetary cost, the marginal change in the disutility related to the sanctions the free-riding agents incur (relative to the change in the public good): In a stable equilibrium, a rise in the tax triggers more free-riding. Consequently, more agents will suffer from feeling guilty. However, an increase in n^* lowers the costs of a norm deviation. While this second order effect, expressed in the second term of Ψ, is negative, we will focus on the case where the first order effect dominates and $\Psi \geq 0$ (in the relevant range of n^*). This restriction is also rationalized by the fact that negative

[31]Expressing the problem this way, we assume that the planner is able to establish one specific social equilibrium state $e \in S$, characterized by n^*. Compare section 4.1.

[32]In principle, there could be several solutions to (14). See the Appendix for a discussion.

values of Ψ would render equilibrium states on the downward-sloping side of the Laffer curve welfare optimal.[33] As for $g' > 0$ the second term on the RHS of condition (14) is then positive, the non-monetary costs add to the monetary costs of public good provision – the RHS in (14) is bigger than one.

A Non-Welfarist Approach

Are there any reasons to deviate from this welfarist approach? There seems to be a normative conflict in the welfare approach discussed above. In our context, a deviation from the social norm is equivalent to a violation of formal law. Given that society has agreed on establishing this legal norm, it appears inconsistent to consider the disutility from a law violation as social costs. To illustrate this point, consider the example of tax evasion: if evaders suffer from feeling guilty about the tax fraud, a planner may not incorporate the related disutility in the welfare function.[34]

Given that $c' < 0$, a free-rider would ceteris paribus – i.e., not considering the decline in the public good – prefer if others would also deviate from the norm and thereby reduce their emotional disutility. In principle, we could interpret this as anti-social preference in the sense of Harsanyi (1982), who suggests to exclude such anti-social preferences from a utilitarian social welfare function.[35]

As a consequence of these points, one could suggest to exclude the emotional disutility from the social welfare function and follow a *non-welfarist* approach: individuals' behavior is generated by one set of preferences, whereas the social planner evaluates it using different preferences. Note that non-welfarist concepts, where social welfare is evaluated according to some other criterion than individuals' utility, are common in many fields of public economics: If a planner is not only concerned about individual well-being but also about horizontal equity (Musgrave, 1990), this also represents a deviation from welfarism. Further examples are the analysis of merit

[33] Compare the discussion of the second order condition to (14) in the Appendix.

[34] Note that this argument has some critical normative implications, since it would also suggest to not consider the monetary punishment of norm-violations as social costs.

[35] A quantitative argument not to consider the emotions as a part of social welfare comes from recent findings in emotions research. While many studies show, how (anticipated) emotions affect economic behavior (e.g. Bosman and van Winden, 2001, 2002, Hopfensitz and van Winden, 2005), there is also evidence suggesting that people tend to systematically overestimate emotions when they make decisions (Loewenstein and Schkade, 1999; Gilbert et al., 2002; Wilson and Gilbert, 2003). Emotions associated with an action choice turn out to be often less strong than anticipated. In our case, this would mean that 'decision disutility' associated from guilt would be higher than the 'experienced disutility'. (Compare the discussion of a moment-based concept of utility in Kahneman, 2000).

goods (Besley, 1988) and the literature on so called 'sin taxes' (O'Donoghue and Rabin, 2003, Kanbur et al., 2004).[36]

In our specific case, we can express the problem of a non-welfarist planner, who excludes the emotional sanctions, as

$$\max_{n^*} W = d - g(n^*) + \delta\, v(g(n^*)). \tag{16}$$

The first order condition for an interior optimum is then given by

$$\delta v'\left(g(n^{nw})\right) = 1. \tag{17}$$

This condition defines n^{nw}, respectively an equilibrium state $e^{nw} = (n^{nw}, t^{nw}, g^{nw})$ which is optimal according to this non-welfarist approach.[37] As the non-welfarist approach only takes into account the monetary costs of public good provision, the non-monetary component has been dropped from the Samuelson condition (compare (14) from above). Therefore, a non-welfarist planner considers lower marginal costs of public good provision than a welfarist planner. Accordingly, the optimal social equilibrium state e^{nw} is characterized by a higher public good level, as compared to g^w, which is optimal according to the welfarist analysis: $g^{nw} > g^w$. This also implies that the tax t^{nw} as well as the share of free-riders n^{nw} are higher in the non-welfarist optimum: $t^{nw} > t^w$, $n^{nw} > n^w$.

3.5 Voting with Social Norms

How is individual voting behavior respectively the voting outcome influenced by a social norm? This is the central question we address in this section by extending the voting concept from Lindbeck et al. (1999) to our framework. Our approach differs from Lindbeck et al. (1999), who study the extent of income redistribution (respectively unemployment benefits), in two main points: We consider heterogeneity in the voting incentives related to the social norm, while they study heterogenous incentives related to different income levels. Moreover, we will focus on the welfare properties of the voting outcome – a point which is omitted in Lindbeck et al. (1999).

[36] Compare also Bernhein and Rangel (2005). For a discussion of the Pareto criterion in the context of such non-welfarist approaches compare Kaplow and Shavell (2000).

[37] If δ becomes very large, the optimum is given by the 'corner solution' characterized by $g' = 0$. Compare the discussion of the second order condition to this problem in the Appendix.

Let us first define

$$U_i(n^*) := \max\{u_i(1, n^*), u_i(0, n^*)\}. \tag{18}$$

U_i depicts the utility agent i derives in an equilibrium state $e \in S$ characterized by n^*. An agent then prefers the state e over $e' \in S$ associated with n' as long as $U_i(n^*) > U_i(n')$. We make the following definition:

Definition 2 *A social equilibrium state $\tilde{e} = (\tilde{n}, t(\tilde{n}), g(\tilde{n}))$, characterized by $\tilde{n}$, is a voting equilibrium if $\tilde{e} \in S$ is a stable social equilibrium state and there exists no other stable equilibrium state $e' \in S$ which is preferred against $\tilde{e}$ by a majority.*

In the voting mechanism underlying this definition, we assume that voters do not only compare different tax levels, but different equilibrium states. They chose between a 'package' $\tilde{e} = (\tilde{n}, t(\tilde{n}), g(\tilde{n}))$ and any (stable) alternative state e'. Hence, we follow Lindbeck et al. (1999) and neglect any problems of equilibrium selection.

3.5.1 Marginal Incentives for Voters

Let us turn to the incentives for the voters. The marginal impact of a change in the equilibrium state – that is, the joint change in n^*, t^* and g^* – on an individual who complies with the norm, is given by

$$\Delta u_i(1, n^*) \equiv \frac{\partial u_i(1, n^*)}{\partial n^*} = -t'(n^*) + \delta v'(g(n^*))\, g'(n^*). \tag{19}$$

For a free-rider, we get

$$\Delta u_i(0, n^*) \equiv \frac{\partial u_i(0, n^*)}{\partial n^*} = -pt'(n^*) - \theta_i c'(n^*) + \delta v'(g(n^*))\, g'(n^*). \tag{20}$$

The (locally) most preferred equilibrium state is then given by the state n^* which maximizes $U_i(n^*)$, characterized by $\Delta u_i(x^*, n^*) = 0$ for agent i's equilibrium strategy x^*. Depending on whether she will contribute or free-ride in state n^*, this equals

$$\delta v'(g(n^*)) = \frac{t'(n^*)}{g'(n^*)} \tag{21}$$

respectively

$$\delta v'(g(n^*)) = \frac{pt'(n^*) + \theta_i\, c'(n^*)}{g'(n^*)} \tag{22}$$

and $g' \neq 0$.[38] In the following we will assume that δ is sufficiently large such that $\Delta u_i(1,0) > 0$ holds. This is sufficient to avoid corner solutions with $n^* = 0$.[39]

Comparing (19) and (20) shows that $\Delta u_i(1,n^*) < \Delta u_i(0,n^*)$ holds for any stable equilibrium with $g' > 0$. The reason for this is straightforward: While contributors fully bear the costs of a tax increase, free-riders only pay the tax with probability p. Moreover, the disutility suffered from a norm-violation decreases in n (since $c' \leq 0$). Hence, individuals who comply with the norm face higher costs of public good provision: For $t' \geq 0$ and $g' > 0$ the RHS of (21) is bigger than the RHS of (22). This implies that cooperators will always prefer an equilibrium state with a lower tax t^* respectively a lower level of n^* than free-riders. Finally, since $v' > 0$ it follows from (21) that a cooperator always prefers a stable equilibrium state ($t' \geq 0$) in the upward-sloping range of the Laffer curve ($g' \geq 0$).

We will now discuss *local* voting equilibria in different 'regions' of S. We then compare the different local voting equilibria, and show which one is the *globally* most preferred equilibrium state. In order to provide a compact analysis, we only consider a scenario with at most three equilibria per policy, as described by figure 3.2.

3.5.2 Local Voting Equilibria

As expressed in the above definition, we only consider voting equilibria which are stable. Hence, for a society with at most three equilibrium levels of n^* for a given policy, voting equilibria are theoretically possible in the range of $n^* \in [0, n^a]$, with n^a such that $\bar{t} = t(n^a)$ and $t'(n^a) = 0$, as well as in the range of $n^* \in [n^b, 1]$, with n^b such that $\underline{t} = t(n^b)$ and $t'(n^b) = 0$ (compare figure 3.2 and 3). Let us first focus on the range with a low level of norm-violations, $n \leq n^a$.

Voting Equilibrium with a Strong Social Norm Due to the preference structure in our model, the heterogeneity in θ does not transform into heterogenous incentives for norm-adhering agents. Hence, all contributors would prefer the same state characterized by (21). If contributors are in a majority, this state also constitutes the local voting equilibrium.

[38]The second order conditions are discussed in the Appendix.

[39]Note that for very large δ, there also exists a corner solution at $g' = 0$.

Proposition 3 *Consider all social equilibrium states $e \in S$ with $n^* \in [0, n^a]$ and $n^a \leq \frac{1}{2}$. Given there exists an interior solution to (21), there always exists a unique local voting equilibrium $\tilde{e}^a = (\tilde{n}^a, t(\tilde{n}^a), g(\tilde{n}^a))$ characterized by $\Delta u_i(1, \tilde{n}^a) = 0$. For $\tilde{e}^a$ there holds $0 < \tilde{n}^a < n^a$ and $g'(\tilde{n}^a) > 0$.*

Proof. See Appendix. ■

Proposition 3 characterizes a unique local voting equilibrium for low levels of n^*, where cooperators always form a majority. The state $\tilde{e}^a$ constitutes a local optimum for all individuals who comply with the norm in the equilibrium state $\tilde{n}^a$. All those who deviate from the social norm, i.e. agents with $\theta_i > \theta(\tilde{n}^a)$, would prefer a higher tax (respectively a state with a higher level of n^*). However, since $\tilde{n}^a < \frac{1}{2}$, this latter group is in a minority: $\tilde{e}^a$ wins against any other social equilibrium state in the range of $n^* \in [0, n^a]$. Technically, the equilibrium corresponds to a median voter equilibrium.[40] In this equilibrium, however, the decisive voter with the median θ-value prefers the same equilibrium state as all other cooperative agents with $\theta_i \leq \theta(\tilde{n}^a)$.[41]

As discussed above, a norm adhering agent would never vote for a state on the downward-sloping side of the Laffer curve. Therefore $g'(\tilde{n}^a) > 0$ holds in the voting equilibrium. Moreover, as long as $\Delta u_i(1, 0) > 0$, cooperators will vote for a tax such that $\tilde{n}^a > 0$ – there is a positive share of free-riders in the voting equilibrium.

Voting Equilibrium with a Weak Social Norm Social equilibrium states in the range of $n^* \in [n^b, 1]$ are also potential voting equilibria. However, for states with a low degree of norm compliance the existence of a voting equilibrium as characterized in definition 2 is not guaranteed. A pairwise comparison of different equilibrium states, for example, could result in voting cycles, since preferences in this range of n^* are typically not single peaked. Hence, the voting outcome depends on the majority voting procedure. Since we do not gain any insights from the analysis of different possible equilibria associated with different voting rules, we will simply consider the case that there exists a local voting equilibrium $\tilde{e}^b = \left(\tilde{n}^b, t(\tilde{n}^b), g(\tilde{n}^b)\right)$ as a result of one particular voting procedure.

[40] Preferences are single peaked in the range $[0, n^a]$ for at least half of the population. For a more detailed discussion compare the Appendix.

[41] In a previous version of this chapter, where we considered a positive payoff from norm-adherence, the median voter result became more obvious, as we then considered heterogeneity among contributors. Since this only changed our results quantitatively, we dropped this model feature.

3.5.3 Global Voting Equilibrium

Given that there is scope for two local voting equilibria – one with a majority who adheres to the norm and one with widespread free-riding – the question emerges, which of these two is the global voting equilibrium? In comparing these two equilibrium states, $\tilde{e}^a$ and $\tilde{e}^b$,[42] we have to take into account that agents' decisions depend on n and θ. We have to differentiate between three possible types of individuals: (1) Those with $\theta < \hat{\theta}(\tilde{n}^a)$, who violate against the norm in both voting equilibria, (2) those with $\hat{\theta}(\tilde{n}^a) \leq \theta < \hat{\theta}(\tilde{n}^b)$, who cooperate in $\tilde{e}^a$ but free-ride in $\tilde{e}^b$, and finally (3) agents with $\theta \geq \hat{\theta}(\tilde{n}^b)$, who comply with the norm in both equilibria. Here we restrict ourself to the realistic case of $g(\tilde{n}^a) > g(\tilde{n}^b)$. Moreover we assume $t^*(\tilde{n}^a) > t^*(\tilde{n}^b)$ – which makes it harder for $\tilde{e}^a$ to be preferred.[43] When

An agent characterized by $\theta_i < \hat{\theta}(\tilde{n}^a)$ prefers $\tilde{e}^a$ over $\tilde{e}^b$ iff

$$\delta\left(v(g(\tilde{n}^a)) - v(g(\tilde{n}^b))\right) \geq p\left(t^*(\tilde{n}^a) - t^*(\tilde{n}^b)\right) + \theta_i\left(c(\tilde{n}^a) - c(\tilde{n}^b)\right), \tag{23}$$

that is, if the benefits from a higher public good level outbalance the costs from a higher (expected) tax and the stronger internal sanction ($c(\tilde{n}^a) > c(\tilde{n}^b)$ since $\tilde{n}^a < \tilde{n}^b$ and $c' \leq 0$). Remember that we here consider a scenario of mild sanctions, characterized by a low punishment probability p. Moreover, these players are those with very low levels of θ_i. Therefore, the two terms on the RHS of (23) should be relatively small.

A conditional contributor with $\hat{\theta}(\tilde{n}^a) \leq \theta_i < \hat{\theta}(\tilde{n}^b)$ prefers $\tilde{e}^a$ over $\tilde{e}^b$ iff

$$\delta\left(v(g(\tilde{n}^a)) - v(g(\tilde{n}^b))\right) \geq t^*(\tilde{n}^a) - p\left(t^*(\tilde{n}^b) + f\right) - \theta_i c(\tilde{n}^b) \tag{24}$$

Note that, in contrast of the first type of agents, the difference in (expected) tax payments between the two states is likely to be higher for these second type of individuals – they contribute $t^*(\tilde{n}^a)$ in the equilibrium where the norm is strong, while they free-ride in the other equilibrium and have expected payments $p\left(t^*(\tilde{n}^b) + f\right)$. However, these agents suffer from internal sanctions in the state $\tilde{e}^b$. Hence, this last term in (24) will reduce the costs for a switch from $\tilde{e}^b$ to $\tilde{e}^a$.

Finally, individuals with a strong norm internalization $\theta_i \geq \hat{\theta}(\tilde{n}^b)$ prefer $\tilde{e}^a$ over $\tilde{e}^b$ iff

$$\delta\left(v(g(\tilde{n}^a)) - v(g(\tilde{n}^b))\right) \geq t^*(\tilde{n}^a) - t^*(\tilde{n}^b) \tag{25}$$

[42] As we have said before, $\tilde{e}^b$ typically depends on the voting procedure. In order to show that $\tilde{e}^a$ is the global voting outcome, it must dominate any possible possible voting outcome in the range $[n^b, 1]$.

[43] Theoretically we could also get $g(\tilde{n}^a) < g(\tilde{n}^b)$ and/or $t^*(\tilde{n}^a) < t^*(\tilde{n}^b)$.

These agents comply with the norm and contribute taxes in both equilibria.

Let us denote the right hand side of condition (25) by $rhs(25)$ (and the analogous notation applies to the other two conditions). If the public good preferences are such that

$$\delta \geq \frac{\max\{rhs(23); rhs(24); rhs(25)\}}{(v(g(\tilde{n}^a)) - v(g(\tilde{n}^b)))} \tag{26}$$

holds, conditions (23)–(25) are all satisfied and $\tilde{e}^a$ would be unanimously preferred against $\tilde{e}^b$. In this case, $\tilde{e}^a$ is the global voting equilibrium. Note, however, that (26) is only a sufficient condition; even if two out of conditions (23)–(25) were violated, $\tilde{e}^a$ would still be preferred by a majority, if the one group of agents which prefers $\tilde{e}^a$ over $\tilde{e}^b$ accounts for a majority. In the following we will assume that public good preferences are sufficiently strong such that $\tilde{e}^a$ is the global voting equilibrium.

3.5.4 Social Welfare

Is the voting outcome efficient? We answer this question by comparing the voting equilibrium to the optimal equilibrium states, discussed in section 3.5. Since we did not discriminate between the two welfare approaches, we consider both, first the non-welfarist and later the welfarist concept.

Optimality in Terms of the Non-Welfarist Approach

From Proposition 3 we know that the voting equilibrium $\tilde{e}^a$ is characterized by condition (21). The welfare optimal equilibrium state according to the non-welfarist approach, e^{nw}, is given by (17). Comparing these two conditions, we get the following result:

Proposition 4 *A non-welfarist planner will always set a higher tax as compared to the tax in the voting equilibrium: $t(n^{nw}) > t(\tilde{n}^a)$. In the optimal non-welfarist equilibrium state more people free-ride and a higher level of public goods will be provided as compared to the voting equilibrium: $n^{nw} > \tilde{n}^a$, $g(n^{nw}) > g(\tilde{n}^a)$.*

Proof. See Appendix. ■

The intuition for this result is straightforward. The evaluation of the (monetary) costs of public good provision are different between the planner and the decisive voter. One the one hand, the planner considers the *mean* costs of taxation: if taxes rise, only the $1 - n$ norm-adhering agents and the np detected free-riders bear the full marginal costs. On the other hand, the decisive voter is a cooperator: She complies with the social norm and contributes t. Hence, she is confronted with the *full* marginal tax increase, t'. Since $t' > g'$ for any stable state with $n > 0$, the costs faced by the decisive voter are higher than the mean costs. A norm-adhering voter will therefore prefer a state with a lower tax and a lower level of free-riding as compared to the optimal non-welfarist equilibrium state. Since these equilibria are in the upward-sloping range of the Laffer curve, a lower tax also transforms into a lower public good level, $g(\tilde{n}^a) < g(n^{nw})$. Hence, there will be an underprovision of public goods in the voting equilibrium.

Optimality in Terms of the Welfarist Approach

Does the result of Proposition 4 carry over to the welfarist approach? While this is not clear at first sight, the comparison of the voting equilibrium with the welfare optimal equilibrium state characterized by (14) yields a clear finding: We get the same qualitative result for both welfare criteria.

Proposition 5 *A welfarist planner will always set a higher tax as compared to the tax in the voting equilibrium: $t(n^w) > t(\tilde{n}^a)$. In the optimal welfarist equilibrium state more people free-ride and a higher level of public goods will be provided as compared to the voting equilibrium: $n^w > \tilde{n}^a$, $g(n^w) > g(\tilde{n}^a)$.*

Proof. See Appendix. ∎

Although the welfarist planner also incorporates the non-monetary welfare costs reflected by Ψ, the marginal costs of public good provision for the decisive voter are still higher than the marginal welfare costs considered by the planner. More formally, $t' > g' + \Psi$ for any stable equilibrium state. For the same reason as above, the voting outcome is inefficient. There will be a suboptimal low level of taxation, free-riding and public good provision in the voting equilibrium.

Underprovision of Public Goods

Hence, majority voting will result in an equilibrium state which is inefficient from the perspective of both welfare concepts. The voting equilibrium is characterized by a suboptimally low level of taxation, a suboptimally low level of free-riding and an underprovision of public goods. This result is driven by the fact that the marginal costs faced by the decisive voter are different from the social costs of public good provision. Note that this does only indirectly rely on the impact of the social norm. The mechanism which drives a wedge between the planner's and the (decisive) voter's incentives would prevail in any voting equilibrium, where the population is divided into contributors and free-riders.

To the best of the authors knowledge, this mechanism has not been discussed and provides a quite general contribution to the literature on collective choice over tax respectively redistribution policies, which neglects the case of tax evasion (Romer, 1975; Roberts, 1977; Meltzer and Richard, 1981). The standard textbook-result on inefficient public good provision is driven by either the difference between the median voter's public good preferences and the mean preferences or the discrepancy between the median voter's income and the mean income (compare e.g. Persson and Tabellini, 2000). Both of these differences are typically exogenous. In our case, however, all agents have identical preferences for the public good, and the difference in the median voters' marginal costs and the mean marginal costs are to some extent endogenously formed: The voting outcome defines a tax, for which agents – depending on their level of θ – decide on whether to free-ride or to contribute. The (exogenous) heterogeneity in θ together with the endogenous level of taxation then translates into an (endogenous) heterogeneity in behavior, which results in different costs of taxation. This effect drives a wedge between the decisive voters' costs and the mean (social) costs of public good provision, resulting in an underprovision of public goods.

As we have mentioned above, this result only relies indirectly on the impact of the social norm. In future research we will discuss the underprovision result in a more compact framework.[44]

3.6 Public Law Enforcement

In this final section we analyze public law enforcement within this framework. More precisely, we study the endogenous choice of p, the detection respectively punishment probability. Following the tax enforcement literature, we assume that p is determined by a revenue maximizing authority, while f is exogenously fixed at a finite level. Of course, one could always enforce full cooperation by setting the fine f sufficiently high. In democratic societies, however, there is typically strong resistance against excessively

[44]For example, we assume $c(n)' = 0$. As we know from Proposition 1 (ii), there would then exist one unique equilibrium for any policy, which would in turn transform into one unique voting equilibrium. Moreover, the sign of Ψ from (15) would be unambiguously positive.

strong sanctions. With respect to pecuniary sanctions, there arise further constraints on f: Taking into account agents limited liability, high fines may not be credible.

We consider the following timing: First, the planner (respectively voters) determines a tax t. At a second stage, the law enforcement authority determines p by investing an amount $d(p)$ into the provision of auditing.[45] Finally, agents choose their actions. The public good is then determined by the resulting net revenues, defined as total revenues (including fines) minus the costs of law enforcement. Using $d = d(p)$ in (8), we get

$$g(n, p, t) = t\,(1 - n) + np\,(t + f) - d(p). \tag{27}$$

The function $d(.)$ is increasing and strictly convex for $0 \leq p < 1$ and describes how much the authority has to invest in order to implement a certain detection probability.

3.6.1 The Enforcement Subgame

Let us take the outcome of the first stage of the game – the tax t – as exogenously given and focus on the enforcement subgame. Applying backward induction, we first study the agents' behavior. Since at this last stage the policy variables are exogenously given, the individual decision is still determined by the threshold in (4) and the equilibrium share of free-riders n^* is given by (5). The law enforcing authority then chooses a detection policy p which maximizes total net revenues, taking into account the impact on n^*. We choose an indirect approach to this problem, using the technique introduced in section 3.3. Taking t as exogenous, we can use (5) in order to derive the detection probability as a function of n and t,

$$p(n, t) = \frac{t - F^{-1}(n)c(n)}{t + f}, \tag{28}$$

where we assume that $p(.) \in [0, 1)$ holds. The function $p(n, t)$ defines a detection policy p^* which is needed in order to support the equilibrium n^* (for a given t), $p^* = p(n^*, t)$.[46] Substituting (28) in (27), we can express the problem of the authority as

$$\max_{n^*} g(n^*, p(n^*, t), t) = (1 - n^*)\,t + n^*\,p(n^*, t)\ (t + f) - d(p(n^*, t)). \tag{29}$$

As in section 3.5, we study the optimization in an indirect manner: Instead of analyzing the direct choice of p, we rather allow the authority to select one equilibrium level n^*

[45] We do not consider commitment problems.

[46] Compare the analogy to $t(n^*)$, defined in section 3.3.

with the associated detection policy $p(n^*, t)$, characterized by (30). As before, we neglect any problems of equilibrium selection. The first order condition for an interior solution to the problem is given by

$$-t + (t+f)\left(p(n^+, t) + n^+ \left.\frac{\partial p(n,t)}{\partial n}\right|_{n=n^+}\right) = d'(p(n^+, t)) \left.\frac{\partial p(n,t)}{\partial n}\right|_{n=n^+} \tag{30}$$

where $d' = \frac{\partial d}{\partial p}$ and $\frac{\partial p}{\partial n}$ is derived in the Appendix.[47] Equating marginal revenues and marginal costs of a joint change in p^* and n^* (a marginal shift in the equilibrium state), condition (30) defines an equilibrium share of free riders $n^+ = n(t)$ as an implicit function of t, induced by the revenue maximizing enforcement policy $p^+ = p(n^+, t)$. Substituting n^+ and p^+ in (27), we get $g^+ = g(n^+, p^+, t)$, the maximum net revenue which can be enforced for a given tax t. With this, we can make the following definition:

Definition 3 ***(i)*** *Taking f, $c(n)$, $d(p)$ and $F(\theta)$ as exogenously given, a revenue maximizing social equilibrium state $e^+ = (t, n^+, p^+, g^+)$ is a combination of a tax t, a share of free-riders $n^+ = n(t)$, a detection probability $p^+ = p(n^+, t)$ and a public good level $g^+ = g(n^+, p^+, t)$ defined by (30), (28) and (27).* ***(ii)*** *A revenue maximizing social equilibrium state e^+ is stable, iff $\frac{\partial p(n^+,t)}{\partial n^+} \leq 0$.* ***(iii)*** *The set of revenue maximizing social equilibrium states in a society is characterized by $S^+ = \{(t, n(t), p(t), g(t)) : 0 < t \leq \infty \mid (t \geq p\,(t+f)) \wedge (p(t) \in [0,1))\}$ with $S^+ \in (0,\infty] \times [0,1] \times [0,1) \times \mathbb{R} \cup \{+\infty\}$.*

The state e^+ characterizes the outcome of the enforcement subgame for a given tax t. From Proposition 2 we know that in any stable equilibrium n^* is decreasing in p. From this follows part (ii) of the definition. Part (iii) defines the set of all possible revenue maximizing equilibrium states in the society. We will come back to this definition in the following.

Before we turn to a comparative static analysis of this equilibrium framework, we introduce a graphical representation of the system. In figure 3.4 we present an example of the functions $p(n^*, t)$ and $g(n^*, p^*, t)$, where t is fixed and n^* varies on the horizontal axis. The revenue maximizing equilibrium state is defined by the peak of the net revenue curve $g(.)$, which determines n^+, p^+ and g^+.

[47] There we also show that the problem is concave in the relevant range of n^*.

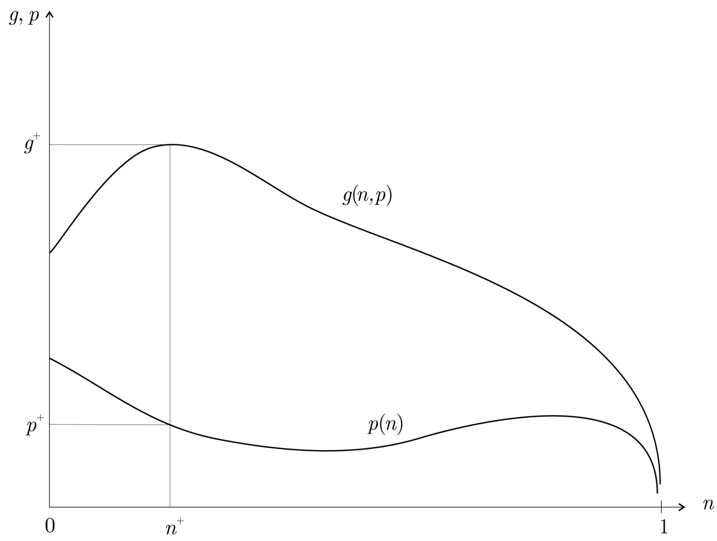

Figure 3.4: Revenue Maximizing Equilibrium

Comparative Statics – The Fine How is n^+ respectively p^+ affected by an exogenous change in the level of the fine f? In the Appendix we show that

$$\frac{dn^+}{df} < 0. \tag{31}$$

As the fine gets higher, the revenue maximizing authority chooses to enforce a lower level of free-riding. In terms of figure 3.4, the peak of the $g(.)$-curve shifts to the left. The intuition for this result is straightforward: A marginal increase in p has a stronger deterrent effect the higher the fine is. Hence it becomes 'cheaper' to enforce a higher level of cooperation and n^+ will fall as f rises.

The change in the equilibrium enforcement policy p^+ is ambiguous. As f increases, the same level of cooperation can be enforced with a lower detection probability. Hence, the $p(.)$-curve depicted in figure 3.4 shifts downward for all n^+, tending to lower p^+. As a second order effect, however, the planner will reduce n^* by increasing p^+(see above). In equilibrium, the revenue maximizing enforcement policy p^+ is partially substituted by an increase in f, only if the first order effect dominates. For this case we would get $\frac{dp^+}{df} < 0$.

Comparative Statics – The Social Norm In this framework we can also study how the formal norm enforcement changes with the strength of the informal, social norm. We consider an exogenous shock in the distribution such that agents experience an increase in their θ-level and the new distribution function first-order stochastically dominates the initial one. Let us denote the inverse function of the initial cumulative distribution as $F^{-1}(n, \phi)$ (with ϕ normalized to unity in the initial distribution). We can express the shock in the distribution of θ as an increases in F^{-1} for any $n \in [0, 1]$, i.e. $\frac{\partial F^{-1}}{\partial \phi} \geq 0$. In the Appendix we show that

$$\frac{dn^+}{d\phi} < 0. \tag{32}$$

After a positive shock in the social norm, the authority will enforce an equilibrium state e^+ with a low level of free-riding as compared to the state before the shock. The intuition for this result is similar as above. If all agents become more sensitive towards the social norm, a norm deviation is accompanied by a stronger internal sanction. Hence, the same equilibrium level of cooperation can be enforced with lower external sanctions. Law enforcement becomes 'cheaper', in the sense that the marginal costs to enforce an equilibrium level n^* fall with a stronger social norm.

Again, the overall impact on the equilibrium level of p^+ is ambiguous. The first order effect, discussed above, is clearly negative. From this perspective, formal law enforcement and informal norm enforcement are substitutes. In terms of figure 3.4, the $p(.)$-curve shifts downward after a positive norm-shock. After a shock in the norm sensitivity, however, a revenue maximizing authority will enforce a lower level of free-riding (compare above). Hence, the planer would move along the new $p(.)$-curve to the left and p^+ would increase. If this latter effect is small, the first order effect dominates and we get

$$\frac{dp^+}{d\phi} \leq 0. \tag{33}$$

In this case, a strengthening of the social norm would lead to a higher level of law compliance enforced by a less strict punishment policy. The informal institution – the social norm – would partially substitute the formal institution – the external enforcement policy. Stated differently, intrinsic incentives 'crowd out' the extrinsic incentives provided by revenue-maximizing authority. Turning around this reasoning, it also holds that society will invest more into law enforcement as the social norm erodes.

For the alternative case, where the second order effect in $dp^+/d\phi$ dominates, and the sign becomes positive, the law enforcing authority would find it optimal (in terms of revenue maximization) to provide a stricter enforcement after a positive shock on the social norm. Hence, a strengthening in the informal institution would trigger higher investments in the formal institution. Although these two mechanism work as substitutes in the enforcement of cooperation, they could be complementary in the revenue maximizing equilibrium state.

Comparative Statics – The Tax Finally, we study the impact of an (exogenous) tax increase on the revenue maximizing equilibrium state. In turns out that the sign of $\frac{dn^+}{dt}$ is ambiguous. In equilibrium, a higher tax could either lead to more or less free-riding. There are several effects, pointing into opposite directions: On the one hand, a tax increase raises the incentives to free-ride. Therefore a stricter deterrence policy is needed to enforce the same equilibrium level of cooperation. Consequently the marginal costs of law enforcement for a given level n^* rise. On the other hand, enforcing the same level of cooperation, raises more revenues as the tax is higher. Hence, marginal revenues could rise (for a given level n^*). Depending on the shape of $d(.)$ respectively the level of n and t, either the one or the other effect dominates.

If $\frac{dn^+}{dt} > 0$, the change in the equilibrium policy p^+ is ambiguous as well. The first order effect points into the opposite direction as the second order effect. However, if $\frac{dn^+}{dt} < 0$ we would get the unambiguous result $\frac{dp^+}{dt} > 0$. In this case, the authority would strongly increase p^+ as a reaction to a tax increase, and thereby implement a lower level of free-riding.

Note that the ambiguity of the effect of a tax increase on the level of free-riding is similarly reported in the standard literature on tax evasion. Empirical studies, on the one hand, often document a positive correlation between the tax rate and the level of evasion. However, there are several studies which report the opposite finding.[48] On the other hand, the standard theory of tax evasion (Allingham and Sandmo, 1972), discussed in chapter 2 of this thesis, delivers no clear prediction. Only for the case of absolute decreasing risk-aversion, Yitzhaki (1974) predicts a decrease in evasion as a response to a higher tax rate. In chapter 2 we have shown that for norm guided taxpayers, this result may turn around, such that the aggregate impact on the level of evasion becomes ambiguous. While the result derived in this section is similar, it is driven by a completely different mechanism, based on the endogeneity of the auditing rate p.

[48] Compare the survey in Andreoni et al. (1998).

3.6.2 The Complete Game

Finally, we turn to the welfare optimal choice of a tax policy in the 1st stage of the game. In order to shorten the discussion, we only consider a non-welfarist approach to the problem.[49] Note that for a given tax t, definition 3 determines one unique equilibrium state e^+. We can therefore set up the problem in the following (direct) way:

$$\max_t \; W = -\underbrace{\left\{\left(1-n^+\right)t + n^+ p^+ \left(t+f\right)\right\}}_{R} + \delta v(g^+) \tag{34}$$
$$\text{s.t. } n^+ = n(t),\; p^+ = p(n^+, t),\; g^+ = g(n^+, p^+, t)$$

By the choice of t, the planner picks the optimal equilibrium state from the set S^+. Let us define the expression in the curly brackets in (34), the gross revenues, as R. An interior solution t^+ is then characterized by the following first order condition:

$$\delta v' = \frac{\frac{dR}{dt}}{\frac{dg^+}{dt}} \tag{35}$$

where $\frac{dg^+}{dt}$ and $\frac{dR}{dt}$ are derived in the Appendix and the second order condition is assumed to hold.

Condition (35) together with definition 3 provides the subgame perfect solution to the full game, given by (t^+, n^+, p^+, g^+), i.e. an optimal tax t^+, a share of free-riders $n^+ = n(t^+)$, a revenue maximizing policy $p^+ = p(n^+, t^+)$ and the public good level $g^+ = g(n^+, p^+, t^+)$. While in section 3.3, we only considered the endogenous determination of n^*, we now have an equilibrium which characterizes a tax and a punishment policy, the resulting level of free-riding and the budget balancing public good level. The only remaining exogenous components of the model are $c(.)$, $d(.)$, the distribution $F(\theta)$ and the fine f.

Equation (35) represents the Samuelson condition for the optimal choice of t^+ (respectively g^+) in the complete game: The (sum of the) marginal benefits from the public good are set equal to the marginal rate of transformation. Here, this rate is given by the marginal gross revenues relative to the marginal increase in the net revenues – which directly transforms into the marginal increase in the public good. In contrast to section 3.4, the planner now takes into account the social costs of law en-

[49] In the Appendix we show that all results derived in this section do carry over for the welfarist approach.

forcement. Since the law enforcing authority conditions its policy p^+ upon the tax t^+, any change in taxes will alter $d(p^+)$ and thereby the *net* revenues, which determine the budget balancing public good level.[50] If in the optimal equilibrium state $\frac{dp^+}{dt} > 0$, we get $\frac{dR}{dt} > \frac{dg^+}{dt}$. In this case the marginal increase in gross revenues would be higher than the marginal increase in the public good. Therefore, the marginal costs of the public good provision would be bigger than one.

What is now the difference in the optimal policy of the (non-welfarist) planner of section 3.4, who considers the detection probability as fixed, and the planner in the present section, who takes into account the impact of the tax policy on the authorities choice of p^+? Comparing conditions (17) and (35) we can formulate the following result:

Proposition 6 *Iff* $\frac{dp^+}{dt} > 0$*, the public good level* g^+ *provided in the equilibrium state* e^+ *will be lower as compared to* g^{nw}*, the public good level characterized by (17):* $g^+ < g^{nw}$*. Iff* $\frac{dp^+}{dt} < 0$*, the opposite holds true:* $g^+ > g^{nw}$*.*

Proof. See Appendix. ■

The intuition for this result evidently follows from the discussion of the Samuelson condition above. Considering the enforcement policy as exogenously given (section 3.4), the (marginal) increase in gross revenues directly transforms into a higher level of public goods. However, if the planner takes into account the endogenous enforcement policy, a wedge is driven between (marginal) net and gross revenues increases. Depending on the direction in the equilibrium response of p^+ to a tax change, the marginal costs of public good provision will either raise or fall. Thus, public good provision will be lower or higher as compared to the optimal state discussed in section 3.4.[51]

Voting Equilibrium Let us briefly study the case where the tax is determined by majority voting, as discussed in section 3.5. Hence, instead of a planner, voters determine the tax at the first stage of the game. We can then study, whether our underprovision result derived above carries over to this extended framework. Applying the analysis from section 3.5, the equilibrium tax which is preferred by a cooperating

[50] Note that we can write $g = R - d(p)$.

[51] We can not directly compare the level of t, p respectively n between the equilibrium states e^{nw} and the optimal e^+. However, if the (exogenous) level of p in section 4 is assumed to be equal to p^+, one can show that $t^+ < t^{nw}$ and $n^+ < n^{nw}$ holds for $\frac{dp^+}{dt} > 0$.

agent is given by

$$\delta v' = \frac{1}{\frac{dg^+}{dt}}. \tag{36}$$

Assuming that the decisive voter in the global voting equilibrium complies with the norm (i.e. the logic behind Proposition 3 can be applied to this framework), the above condition – together with definition 3 – characterizes the new voting equilibrium state $\tilde{e}^+ = \left(\tilde{t}^+, \tilde{p}^+, \tilde{n}^+, \tilde{g}^+\right)$.[52] Comparing now condition (35) with (36) we can derive the following result:

Proposition 7 *Iff $\frac{dn^+}{dt} > 0$, the underprovision result from Proposition 4 still holds: According to the non-welfarist approach, majority voting in the complete game will result in a suboptimally low level of public goods: $\tilde{g}^+ < g^+$.*

Proof. See Appendix. ■

For the plausible case where $\frac{dn^+}{dt} > 0$, the mean marginal costs of public good provision considered by the planner are lower than the costs faced by the decisive voter. Therefore, our main result from section 3.5 also holds in the extended model version. Only for the special case where $\frac{dn^+}{dt} < 0$ the result would turn around and we would get an oversupply of public goods in the voting equilibrium.[53]

3.7 Conclusion

In many situation which are regulated by formal law, deviant behavior is not only associated with the possibility of a formal, legal sanction – e.g. a fine imposed by a law enforcing authority – but also with decentralized, informal sanctions: Law obedient citizens my express their disapproval or stigmatize those who do not comply with the law (Rasmusen, 1996; Arbak, 2005). However, people also adhere to legal regulations in situations when there is nobody around who could impose informal sanctions. In these situations, compliance can be explained by internalized social norms. If a norm is internalized, it is associated with self-imposed sanctions, related to emotions like guilt or remorse (Elster, 1989a, 1989b). In this chapter, we picked up this idea and

[52]For a formal derivation of this result we would have to repeat the complete analysis from section 5 for this extended equilibrium framework. As there are no reasons, why voting in the extended version of the model should deliver any qualitatively different results, the analysis is not repeated here.

[53]For this case, the mean marginal costs of taxation are very high, since a tax increase raises the share of contributors. Moreover, the raise in taxes is accompanied by a stricter enforcement policy ($\frac{dp^+}{dt} > 0$), which strongly raises the costs of taxation for free-riders.

studied how the interplay of internal sanctions together with legal norm enforcement may induce an agent to abstain from law violations.

While our framework applies to the broad class of public good problems where free-riding is associated with legal sanctions, one can interpret the approach as a tax evasion model with discrete choice. We demonstrate the interdependence in free-riding, respectively evasion behavior, which typically results in a multiplicity of equilibria. Since small changes in the tax rate can induce strong changes in aggregate free-riding respectively evasion behavior, the Laffer curve within such a system is characterized by discontinuous jumps in the level of revenues. In contrast to a conventional, 'well-behaved' tax-revenue relationship, the Laffer curve could also consist of several upward- and downward-sloping segments. Hence, instead of asking 'on which side of the Laffer curve we are', one could also ask where the nearest or where the highest 'peak' is located.

The main focus of this chapter was the analysis of voting and public good provision in the context of social norms. Our study suggests that in the context of free-riding, majority voting will result in inefficient levels of taxation and public good provision. In the standard literature, where public good contributions are considered to be fully enforceable, majority voting results in inefficient outcomes, if the median voter's preferences and the mean preferences differ. In our case, agents' preferences for the public good are ex-ante identical. However, individuals differ in their sensitivity towards the social norm, which results in different propensities to free-ride. As long as there exists a voting equilibrium where free-riders and cooperators coexist, these two types face different marginal costs of public good provision: Cooperators fully bear a (marginal) tax increase, whereas free-riders only pay taxes if they get detected. The mean marginal costs considered by a social planner will always be between these two levels. Hence, a voting equilibrium where the cooperators form a majority will be characterized by an inefficiently low level of taxation, resulting in an underprovision of public goods. The opposite holds true for a voting equilibrium where the free-riders account for a majority. In this latter case, taxes would be too high and there would be an overprovision of public goods. Though we have demonstrated this result for a very specific framework, we also have discussed that this finding generalizes to all voting equilibria where cooperators and free-riders coexist.

What can we say about the empirical relevance of this result? There is ample evidence which indicates that tax evasion is lower in jurisdictions with direct democratic institutions, as compared to indirect democratic systems (e.g. Weck-Hannemann and Pommerehne, 1989; Pommerehne and Weck-Hannemann, 1996). This result is typically explained by assuming that political participation results in a stronger 'tax morale' and thereby a higher level of compliance: If taxpayers are closely involved in the political decision process, so goes the argument, their perceived moral obligation to pay taxes is higher as compared to an indirect decision process (Pommerehne and Weck-Hannemann, 1996; Torgler, 2005). Although this argument appears convincing, our analysis suggests an alternative explanation: If we assume that politicians in a representative democracy will implement a welfare optimal fiscal policy (i.e., the median voter does not immediately determine the fiscal policy), the associated level of taxation would be higher than the level implemented by majority voting (given that cooperators form a majority). The lower level of tax evasion observed in jurisdictions with direct democratic systems would then be simply due to a lower level of taxation. However, this level of taxation – and thereby also the associated level of evasion – could be inefficiently low from a welfare perspective! Hence, it is not necessarily true that direct democratic institutions are more efficient simply because they induce a higher level of tax compliance. Even if a change from an indirect to a direct democratic system induces an increase in tax morale, it remains unclear whether overall efficiency increases, since such an institutional change shifts power towards the median voter and thereby induces a political distortion. Clearly, further research is required in order to clarify whether this objection is legitimate.

Appendix

Appendix to Section 3.3

Proof of Proposition 1. **(i)** Remember that we assume $t - p\,(t+f) \geq 0$ and that – per assumption A1 – $c(n)$ takes finite positive values. From (4) then follows that $0 \leq \hat{\theta}(n) \leq \infty$. For the inverse of the cumulative distribution functions $F^{-1}(n)$, we know $F^{-1} : [0,1] \to [0,\infty]$ with $F^{-1}(0) = 0$ and $F^{-1}(1) = \infty$. Therefore we get: $\hat{\theta}(0) \geq F^{-1}(0)$ and $\hat{\theta}(1) \leq F^{-1}(1)$. Since $F^{-1}(.)$ as well as $\hat{\theta}(.)$ are continuous functions defined over the unit interval $[0,1]$, mapping n into $[0,\infty]$, there exists at least one fix-point $n^* \in [0,1]$ where $\hat{\theta}(n^*) = F^{-1}(n^*)$ and hence $n^* = F(\hat{\theta}(n^*))$ holds. For at least for one n^* there must also hold $\frac{\partial F^{-1}}{\partial n} \geq \frac{\partial \hat{\theta}}{\partial n}$. Hence, there exists at least one stable equilibrium.
(ii) If $c'(n) = 0$ for all $n \in [0,1]$, it follows that the threshold $\hat{\theta}(n)$ is a constant $\hat{\theta}$ for any n. It follows from above that there exists exactly one fix-point $n^* \in [0,1]$ where $\hat{\theta}(n^*) = F^{-1}(n^*)$ and hence $n^* = F(\hat{\theta}(n^*))$ holds. Since we assume $f(\theta) > 0$, $\frac{\partial F^{-1}}{\partial n} > \frac{\partial \hat{\theta}}{\partial n} = 0$ must hold for any n. Hence, the equilibrium is stable. ■

Proof of Proposition 2. A locally stable equilibrium is characterized by (6). From (4) and the assumptions on $f(\theta)$ and $F(\theta)$ one can easily show that

$$\frac{\partial F^{-1}(n)}{\partial n} \gtreqless \frac{\partial \hat{\theta}(n)}{\partial n} \Leftrightarrow \frac{1}{f(\hat{\theta}(n))} \gtreqless -\hat{\theta}(n)\frac{c'(n)}{c(n)}. \tag{A.1}$$

Using the implicit function theorem on (5) we get

$$\frac{dn^*}{d\,(t - p(t+f))} = \frac{f(\hat{\theta}(n^*))}{c(n^*) + c'(n^*)\,\hat{\theta}\,f(\hat{\theta}(n^*))}. \tag{A.2}$$

From A2 we know that $c'(n) \leq 0$. Using (A.1) one can easily see that the denominator of (A.2) is non-negative [negative] for any locally stable [instable] equilibrium. Hence, for any stable [instable] equilibrium, the fraction n^* raises [falls] with an increase in $(t - p(t+f))$. ■

Appendix to Section 3.4

A. Welfarist Approach

Derivation of Ψ Using the Leibnitz Rule of integral differentiation, the derivative of the integral term in (13) is given by

$$-c'(n^*)\int_0^{\hat{\theta}(n^*)} \theta \; dF(\theta) - c(n^*)\frac{d\hat{\theta}(n^*)}{dn^*}\hat{\theta}(n^*)\, f(\hat{\theta}(n^*)). \tag{A.3}$$

Taking into account that $t^* = t(n^*)$ in (4) and simplifying, we get

$$\frac{d\hat{\theta}(n^*)}{dn^*} = \frac{1}{c(n^*)}\left[t'(n^*)(1-p) - \hat{\theta}(n^*)c'(n^*)\right].$$

Substituting $t'(.)$ from (10), and using $\frac{\partial F^{-1}}{\partial n} = \frac{1}{f(\hat{\theta}(n))}$ we can rewrite $\frac{d\hat{\theta}}{dn}$ as

$$\frac{d\hat{\theta}(n^*)}{dn^*} = \frac{1}{c(n^*)}\left[\frac{c(n^*)}{f(\hat{\theta}(n))} + F^{-1}(n^*)c'(n^*) - \hat{\theta}(n^*)c'(n^*)\right].$$

Since for any equilibrium n^* there must hold $F^{-1}(n^*) = \hat{\theta}(n^*)$, the second and third term in the squared brackets chancel out. Finally, substituting $\frac{d\hat{\theta}}{dn}$ into (A.3) and reversing signs yields Ψ as defined in (15).

Multiple Solutions As we have already stated in the main text, there might exist several solutions to (14). Numerical simulations show, however, that for reasonable high levels of δ there is a unique optimum. If there are multiple solutions, we exclude them by focusing on stable equilibrium states in the upward-sloping range of the Laffer curve. Such states are characterized by $g' > 0$ and $t' > 0$, which always holds in the broader neighborhood of $n^* = 0$. In this range of n^*, we typically get $g'' < 0$ (compare figure 3.2).

Second Order Condition The second order condition (SOC) to the problem (13) is given by

$$\psi\frac{g''}{g'} + \delta v''\,(g')^2 + \frac{d\Psi}{dn} < 0. \tag{A.4}$$

where we made use of (14) While the sign of the third term is ambiguous, the second term is negative since $v'' < 0$. Assuming that δ (and thereby the second term) is sufficiently large, and that the third term is not too strongly positive, the SOC holds if the first term is negative. We can identify several possible cases, where the sign of term 1 is negative:

(1a) $\psi < 0$, $g' > 0$, $g'' \geq 0$. (2a) $\psi > 0$, $g' > 0$, $g'' \leq 0$.
(1b) $\psi < 0$, $g' < 0$, $g'' \leq 0$. (2b) $\psi > 0$, $g' < 0$, $g'' \geq 0$.

As discussed above, we focus on solutions in a range where $g(.)$ is concave. Hence, we neglect the cases (1a) and (2b). In case (1b), $\psi < 0$ and raising taxes and thereby n^* lowers total norm-based disutility. Stated differently, in this case a tax increase has a positive non-monetary welfare effect. There could be solutions to (14) where the planner would choose an equilibrium on the downward-sloping side of the Laffer curve ($g' < 0$) in order to raise n and thereby lower norm-based welfare-costs of the free-riders. Since this does not seem to be very plausible, we exclude this scenario from our analysis. (Theoretically, there could also exist an optimum with $\psi < 0$, $g' > 0$, $g'' \leq 0$. Since such a solution is only possible for very low levels of δ, and since then the SOC is unclear, we neglect this case.) Finally, for the case (2a), $\psi > 0$. In such a situation, any increase in n (associated with a higher tax) will raise the norm-based welfare costs. The planner will operate in the upward-sloping side of the Laffer curve, with $g' > 0$ and the SOC holds for $g'' \leq 0$. (Theoretically there could also be an optimum in the range where $g' < 0$. Since then the SOC becomes unclear, we exclude this case by assumption.)

B. Non-Welfarist Approach

As in the welfarist approach, there might also here be several solutions to (17). As above, we exclude all instable equilibria and states with $g' < 0$. The SOC to (17) is given by

$$g''\left(\delta\, v' - 1\right) + \delta v''\left(g'\right)^2 < 0.$$

As an interior solution to (17) is characterized by $\delta\, v' = 1$,the SOC is fulfilled since $v'' < 0$.

Appendix to Section 3.5

Individual Voting Decision – Second Order Conditions The second order conditions for (19) and (20) are given by

$$-t'' + \delta \left(v''\, g'^2 + v'\, g'' \right) < 0$$

respectively

$$-pt'' - \theta_i c'' + \delta \left(v''\, g'^2 + v'\, g'' \right) < 0,$$

where we have omitted the functions' arguments. In a range of n where both t and g are concave, the SOC for norm-adhering agents is fulfilled. If δ is large enough, the SOC should also hold for free-riders – for low levels of θ_i even for $c'' < 0$. (For very high values of θ_i, $\theta_i c''$ could be the dominant effect. However, high θ-types will typically cooperate.)

Proof of Proposition 3. Let us denote n^G such that $g'(n^G) = 0$. From (11) follows that $t'(n^G) > 0$. Since per definition $t'(n^a) = 0$, there has to hold $n^G < n^a$. Moreover, $t'(n^G) > 0$ and $g'(n^G) = 0$ also implies that (i) $\Delta u_i(1, n^G) < 0$ and remember that we assume (ii) $\Delta u_i(1, 0) > 0$. Since $\Delta u_i(1, n^*)$ is strictly decreasing in n^* (second order condition), it immediately follows from (i) and (ii) that there exists one unique $\tilde{n}^a \in (0, n^G)$ with $\Delta u_i(1, \tilde{n}^a) = 0$. As $n^a \leq \frac{1}{2}$, we can then state: $0 < \tilde{n}^a < n^G < n^a \leq \frac{1}{2}$. Since for at least half of the population $U_i(n^*)$ is single peaked in the range $n^* \in [0, n^a]$, there can not exist any alternative state e' with $n' \in [0, n^a]$, which is preferred against $\tilde{e}^a$ by a majority. ■

Discussion of Proposition 3. Comparing (19) with (20) yields $\Delta u_i(0, n^*) > \Delta u_i(1, n^*)$ for any stable equilibrium. For all free-riders with $\theta_i < \hat{\theta}(\tilde{n}^a)$ there must hold $\Delta u_i(0, \tilde{n}^a) > \Delta u_i(1, \tilde{n}^a) = 0$ since $\frac{\partial^2 u_i(0,n)}{\partial n \partial \theta} > 0$. They would prefer an equilibrium with a higher level of taxation and free-riding. The same could hold for agents with $\hat{\theta}(\tilde{n}^a) \leq \theta_i < \hat{\theta}(n^a)$, who cooperate in the voting equilibrium. Note that $U_i(n)$ is not continuously differential for all $n \in [0, 1]$. For every agent i, there exists a level of n where $\theta_i = \hat{\theta}(n)$. At this level of n, $u_i(1, n)$ and $u_i(0, n)$ intersect, resulting in a kink in $U_i(n)$. For higher [lower] levels of n, agent i would free-ride [cooperate]. Therefore, an agent who cooperates in the voting equilibrium $\tilde{e}^a$, could turn into a free-rider for an equilibrium n' with $\tilde{n}^a < n' \leq n^a$, which theoretically could result in $U_i(n') > U_i(\tilde{n}^a)$ (since there holds $\Delta u_i(0, n^*) > \Delta u_i(1, n^*)$ for any stable equilibrium). Stated differently, for agents with $\theta_i < \hat{\theta}(n^a)$ preferences are not necessarily single peaked. However, assuming $n^a \leq \frac{1}{2}$ assures that this group is always in a minority. There can never exist an alternative e' which is preferred against $\tilde{e}^a$ by a majority.

Proof of Proposition 4. Note that the LHS of conditions (17) and (21) are identical for any n. Using (10) and (11) one can easily show that $t'(n^*) > g'(n^*)$ for any $n^* \in (0, n^a]$. Comparing the RHS of (17) and (21) it immediately follows from $g'(\tilde{n}^a) > 0$ (Proposition 3) that

$$\frac{t'(\tilde{n}^a)}{g'(\tilde{n}^a)} > 1,$$

since $\tilde{n}^a \in (0, n^a)$. As $t' > 0$ and $g' > 0$ holds for $\tilde{n}^a$ as well as for n^{nw}, $v'' < 0$ implies $t(n^{nw}) > t(\tilde{n}^a)$, $n^{nw} > \tilde{n}^a$ and $g(n^{nw}) > g(\tilde{n}^a)$. ■

Proof of Proposition 5. In order to proof Proposition 5 we simply have to show that $t'(n^*) > g'(n^*) + \Psi(n^*)$ holds in the relevant range of $n^* \in (0, n^a]$. Substituting Ψ from (15) and g' from (11), we can rewrite this condition as

$$t'(n^*) - t'(n^*)\left(1 - n^*\left(1-p\right)\right) + F^{-1}(n^*)c(n^*) > \hat{\theta}(n^*)\, c(n^*) + c'(n^*) \int_0^{\hat{\theta}(n^*)} \theta \; dF(\theta).$$

Since for any equilibrium n^* there must hold $F^{-1}(n^*) = \hat{\theta}(n^*)$, the condition simplifies to

$$n^*\, t'(n^*)\left(1-p\right) > c'(n^*) \int_0^{\hat{\theta}(n^*)} \theta \; dF(\theta)$$

which holds for any $n^* \in (0, n^a]$, since $t'(n^*) > 0$ and $c' \leq 0$. Hence, the RHS of (21) is higher than the RHS of (14). As $t' > 0$ and $g' > 0$ holds for $\tilde{n}^a$ as well as for n^w, $v'' < 0$ implies $t(n^w) > t(\tilde{n}^a)$, $n^w > \tilde{n}^a$ and $g(n^w) > g(\tilde{n}^a)$. ■

Appendix to Section 3.6

Second Order Condition to (30): The SOC to problem (30) is given by

$$SOC := 2\frac{\partial p(n,t)}{\partial n}\left(t+f\right) + \frac{\partial^2 p(n,t)}{\partial n^2}\left(n\left(t+f\right) - d'\right) - d''\left(\frac{\partial p(n,t)}{\partial n}\right)^2 < 0. \quad \text{(A.5)}$$

From (28) we can derive

$$\frac{\partial p(n,t)}{\partial n} = -\frac{\Omega}{t+f} \quad \text{(A.6)}$$

with

$$\Omega := \frac{\partial F^{-1}}{\partial n} c(.) + F^{-1}(.) c'(.) \tag{A.7}$$

From the proof of Proposition 2 we know that Ω is positive in any stable equilibrium, and therefore $\frac{\partial p}{\partial n} < 0$. Hence, the first and the third term (since $d'' > 0$) in (A.5) are negative. From the first order condition follows that the expression in the brackets in the second term must be negative. A sufficient condition for the SOC to hold is then $\frac{\partial^2 p}{\partial n^2} > 0$. This is assumed to hold for stable equilibrium states.

The Enforcement Subgame

Comparative static analysis w.r.t. the fine f From (28) we get

$$\frac{\partial p}{\partial f} = -\frac{p}{t+f} \leq 0. \tag{A.8}$$

As the fine gets stronger, one can implement the same equilibrium n^* with a lower detection probability.

Using (28) and (A.6) we can simplify the first order condition (30) and get

$$-F^{-1}(.) c(.) - n\,\Omega + \frac{d'(.)}{t+f}\,\Omega = 0. \tag{A.9}$$

Applying the implicit function theorem on (A.9) and substituting for (28), (A.6) and (A.8) we get

$$\frac{dn^+}{df} = -\frac{1}{SOC}\,\frac{\Omega}{(t+f)^2}\,(-d'(.) + d''(.)\,p) < 0$$

with $SOC < 0$ as defined above in (A.5). Note that Ω must be positive in a stable equilibrium (Proposition 2). Since $d'' < 0$, the expression in the round brackets is negative. Therefore we get $\frac{dn^+}{df} < 0$. With this result, we can derive from (28)

$$\frac{dp^+}{df} = \underbrace{\frac{\partial p(n^+,.)}{\partial f}}_{-} + \underbrace{\frac{\partial p}{\partial n}\frac{dn^+}{df}}_{+} \gtreqless 0$$

where the signs follow from (A.6) and (A.8). Hence, the sign of $\frac{dp^+}{df}$ is unclear. Only if the first order effect dominates, $\frac{dp^+}{df} < 0$.

Comparative static analysis w.r.t. the distribution of θ Let us express the inverse of the cumulative distribution function as $F^{-1}(n, \phi)$ with $\phi = 1$ for the initial distribution of θ. We study a shock in the distribution such that $\frac{\partial F^{-1}(n,\phi)}{\partial \phi} \geq 0$ for all $n \in [0, 1]$. Hence, the new distribution function first-order stochastically dominates the initial one. Substituting $F^{-1}(n) = F^{-1}(n, \phi)$ in (28) and (A.9) we can derive

$$\frac{\partial p}{\partial \phi} = -\frac{\partial F^{-1}}{\partial \phi}\frac{c}{t+f} < 0$$

and with that

$$\frac{dn^+}{d\phi} = -\frac{1}{SOC}\left\{-\frac{\partial F^{-1}}{\partial \phi}c + F^{-1}c\left[\frac{1}{\Omega}\left(\frac{\partial^2 F^{-1}}{\partial \phi \partial n}c + \frac{\partial F^{-1}}{\partial \phi}c'\right) - \frac{d''\,\Omega}{t+f}\right]\right\},$$

where we made use of condition (30) and (A.7). It is sufficient for $\frac{dn^+}{d\phi} < 0$ to hold, if the expression in the squared brackets is negative. Since $d'' > 0$ and for stable equilibria $\Omega > 0$, this holds if the only term which could be positive, $\frac{\partial^2 F^{-1}}{\partial \phi \partial n}c$, is not too large. We assume that this is fulfilled for the equilibrium state e^+.

From (28) we get

$$\frac{dp^+}{d\phi} = \underbrace{\frac{\partial p(n^+, t)}{\partial \phi}}_{-} + \underbrace{\frac{\partial p}{\partial n}\frac{dn^+}{d\phi}}_{+} \gtreqless 0$$

as $\frac{\partial p}{\partial \phi} < 0$ and $\frac{\partial p}{\partial n} < 0$ (in stable equilibria). Hence, $\frac{dp^+}{d\phi}$ is only negative, if the first order effect dominates.

Comparative static analysis w.r.t. the tax t From (28) we can derive

$$\frac{\partial p}{\partial t} = \frac{f + F^{-1}(.)c(.)}{(t+f)^2} > 0. \tag{A.10}$$

Applying the implicit function theorem on (A.9), we get

$$\frac{dn^+}{dt} = -\frac{1}{SOC}\frac{\Omega}{(t+f)^2}\left\{-d' + d''\frac{f + F^{-1}(.)c(.)}{t+f}\right\} \gtreqless 0 \tag{A.11}$$

where made use of (28), (A.6) and (A.10). The two terms in the curly brackets have different signs, hence, the sign of $\frac{dn^+}{dt}$ is ambiguous. (Note, that the result of Proposition 2, $\frac{\partial n^*}{\partial t} \geq 0$, does not apply here, since there the detection probability was exogenously given.)

The change in the equilibrium policy p^+ is given by

$$\frac{dp^+}{dt} = \underbrace{\frac{\partial p(n^+,t)}{\partial t}}_{+} + \underbrace{\frac{\partial p}{\partial n}\frac{dn^+}{dt}}_{+/-} \gtreqless 0. \tag{A.12}$$

The Complete Game: Non-Welfarist Approach

First Order Condition From the definition of R follows

$$\frac{dR}{dt} = 1 - n^+\left(1-p^+\right) - \frac{dn^+}{dt}\left(t - p^+\left(t+f\right)\right) + n^+\frac{dp^+}{dt}\left(t+f\right). \tag{A.13}$$

Since $g^+ = R - d(p^+)$, we can derive

$$\frac{dg^+}{dt} = \frac{dR}{dt} - d'\frac{dp^+}{dt}. \tag{A.14}$$

Proof of Proposition 6. From (A.14) immediately follows that $\frac{dg}{dt} < \frac{dR}{dt}$ iff $\frac{dp^+}{dt} > 0$. Therefore, the right hand side in (35) is bigger than one (as in a reasonable equilibrium there holds $\frac{dg^+}{dt} > 0$). Iff $\frac{dp^+}{dt} < 0 \Longrightarrow \frac{dg}{dt} > \frac{dR}{dt}$ and the right hand side in (35) is smaller than one. Proposition 6 then follows from $v'' < 0$. ■

Proof of Proposition 7. Substituting (A.6), (A.10) and (A.12) in (A.13), we get $\frac{dR}{dt} < 1 \Leftrightarrow$

$$-n^+\left(1-p^+\right) - \frac{dn^+}{dt}\left(t - p^+\left(t+f\right) + n^+\Omega\right) + n^+\frac{f + F^{-1}c}{t+f} < 0.$$

Substituting now p^+ from (28) in the first term, the first and the third term on the LHS cancel out. As $\Omega > 0$ in a stable equilibrium and since $t \geq p^+\left(t+f\right)$, the LHS is negative iff $\frac{dn^+}{dt} > 0$. Hence, $\frac{dn^+}{dt} > 0 \Leftrightarrow \frac{dR}{dt} < 1$. With $\frac{dR}{dt} < 1$, the RHS in (35) is smaller than the RHS of (36). From $v'' < 0$ then follows $\tilde{g}^+ < g^+$. ■

The Complete Game: Welfarist Approach

The problem of a welfarist planner is given by

$$\max_t W = -\underbrace{\left\{\left(1-n^+\right)t + n^+\,p^+\,(t+f)\right\}}_{R} - c(n^+)\int_0^{\hat{\theta}(n^*)} \theta\; dF(\theta) + \delta v(g^+)$$
$$\text{s.t. } n^+ = n(t),\; p^+ = p(n^+, t),\; g^+ = g(n^+, p^+, t).$$

The first-order condition to this problem is given by

$$\delta v' \frac{dg^+}{dt} = \frac{dR}{dt} + \Psi^+ \tag{A.15}$$

Using the Leibnitz Rule we get

$$\Psi^+ = c'(n^+)\frac{dn^+}{dt}\int_0^{\hat{\theta}} \theta dF(\theta) + c(n^+)\frac{d\hat{\theta}}{dt}\hat{\theta} f(\hat{\theta}). \tag{A.16}$$

Substituting n^+ and p^+ in (4) we can derive

$$\frac{d\hat{\theta}(n^+, p^+)}{dt} = \frac{\left((1-p) - \frac{dp^+}{dt}(t+f)\right)c(n^+) - \left(t - p^+\,(t+f)\right)c'\frac{dn^+}{dt}}{c(n^+)^2}. \tag{A.17}$$

Note that in contrast to section 3.3, the sign of $\frac{d\hat{\theta}}{dt}$ is unclear. However, there has to hold $sign\left\{\frac{d\hat{\theta}}{dt}\right\} = sign\left\{\frac{dn^+}{dt}\right\}$. From this follows, that the sign of Ψ^+ is ambiguous, since the two effects in (A.16) point into opposite directions.

Comparing now condition (A.15) with (36) we can show that the underprovision result from Proposition 5 still holds if $\frac{dn^+}{dt} \geq 0$ and $\int_0^{\hat{\theta}} \theta dF(\theta) - \hat{\theta}^2 f(\hat{\theta}) > 0$. If we use (A.13) and (A.17) we get $\frac{dR}{dt} + \Psi^+ < 1 \Leftrightarrow$

$$\left((1-p) - \frac{dp^+}{dt}\right)\left[\hat{\theta} f(\hat{\theta}) - n^+\right] - \frac{dn^+}{dt}\left(t - p^+\,(t+f) + n^+\Omega\right) <$$
$$-c'(n^+)\frac{dn^+}{dt}\left(\int_0^{\hat{\theta}} \theta dF(\theta) - \hat{\theta}^2 f(\hat{\theta})\right)$$

From the proof of Proposition 7 we know that the expression in the round brackets in the first term is zero. As we assume that the expression in the round brackets on the RHS of the inequality is positive, this condition holds for $\frac{dn^+}{dt} \geq 0$. In this case, the majority voting outcome in the complete game will result in an suboptimal low level of public goods, also from the perspective of the welfarist approach.

Chapter 4

Social Norms and the Evolution of Conditional Cooperation*

4.1 Introduction

Starting with Sonnemans et al. (1999), Keser and van Winden (2000) and Fischbacher et al. (2001), economic research on voluntary public good provisions has highlighted the role of *conditional cooperative behavior*: Agents who follow this behavioral pattern condition their cooperation on the cooperativeness of others respectively on their beliefs about others behavior. The relevance of this conditional strategy, which is well documented in the social psychology literature (e.g. Kelley and Stahelski, 1970; Dawes et al., 1977), has been also demonstrated in field experiments: Frey and Meier (2004a, 2004b) find that charitable giving follows a conditional pattern. Individuals are willing to donate more money, the more other people they expect to donate. Similarly, Croson (2005) shows that voluntary contributions in a fundraising campaign of a US public radio station could be substantially increased by informing potential donors about the high contribution level of previous contributors.[1]

One approach to explain these observations is based upon the role of social norms.[2] Social norms are rules of conduct, which are enforced by internal or external sanctions

*This chapter is based on joint work with Mathias Spichtig, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam.

[1]Further evidence on conditional cooperation is surveyed by Gächter (2005).

[2]There are further theoretical approaches which account for these observations: Theories of fairness and inequity aversion (Rabin, 1993; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), theories of reciprocity and intentions (Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2005) as well as conformity models (Bernheim, 1994) are all able to explain conditional cooperation. Compare Fehr and Falk (2002) for a comprehensive review.

(Ullmann-Margalit, 1977; Elster, 1989a, 1989b). If the magnitude of these sanctions are positively related to the degree of norm compliance, a social norm for cooperation can trigger conditional cooperation. This relationship is typically studied using the concept of internalized social norms, which is modelled as a 'preference for norm compliance'. The strength of these pro-social preference – related to the level of norm internalization respectively the norm sensitivity of an agent – is thereby considered as exogenously given.[3] We deviate from this tradition and use an indirect evolutionary approach to study the endogenous formation of norm sensitivity. In this vein, we endogenously derive the distribution of preference considered in section 2 and 3 of this thesis. For the case of evolutionary adaptation to a 'heterogenous environment', where agents are exposed to cooperative as well as to non-cooperative environments, we find those levels of norm sensitivity to be evolutionary successful, which induce conditional cooperative behavior.

In contrast to standard evolutionary game theory, which studies the direct evolution of genetically encoded strategies, *indirect evolutionary approaches*, pioneered by Güth and Yaari (1992) and Güth (1995), allow for cognition and rationality in the evolution of behavior and the underlying preferences. The main idea of an indirect approach is the following: Rational actors play a game and their preferences induce certain strategies. These strategies lead to a certain payoff which determines the 'success' of an actor. This success, related e.g. to income and social prestige, determines what is called fitness in biology. More successful players, i.e. agents with a higher social prestige, will then get more frequently imitated than others. This way, the preferences of successful players spread in the population while those of less successful actors get eliminated. Hence, the approach provides a method to endogenously study the formation of preferences.[4]

We first introduce a model of voluntary public good provision within a large scale society, which corresponds to a simplification of the framework studied in chapter 3 of this thesis. In their decision whether to cooperate or to free-ride, individuals trade off the benefits from free-riding with the costs related to norm-enforcing sanctions. The impact of these sanctions on the utility of an agent varies along two dimensions: First, we allow for preference heterogeneity with respect to the level of norm internalization, which transforms into different levels of sensitivity towards the sanctions. Second, the extent of sanctions varies with the degree of cooperation in the population. The higher the share of norm compliance, the stronger are the sanctions. As we have addressed above, this will induce conditional cooperative behavior: Agents with an

[3]Compare e.g. Akerlof (1980), Lindbeck et al. (1999), Naylor (1989), Rege (2004).

[4]See also Bester and Güth (1998) as well as Dekel et al. (2004) and the references cited therein.

intermediate level of norm sensitivity will condition their cooperation on the behavior of the other members in the society. In turn, this behavior opens the scope for multiple equilibria. The society could coordinate either on a state with a strong social norm and far-reaching cooperation or on a state with weak norm-enforcement and widespread free-riding.

The agents' evolutionary success is determined by income respectively social status. While the direct payoff from free-riding clearly dominates that of cooperation, the sanctions associated with free-riding have a negative impact on social status. This effect is driven by social exclusion or stigmatization respectively the economic consequences of these sanctions.[5] Therefore, the gain in social prestige related to the higher income of a defector can be offset by the status loss associated with the punishment of a norm-violation. Whether free-riding or cooperation is evolutionary more successful then depends crucially on the level of cooperation in the society, which determines the strength of the status-impact of the sanctions. In course of the evolution, individuals learn about the success of different agents with different levels of norm sensitivity. Accordingly they adapt their own preferences, i.e. their level of norm sensitivity, where more successful preference-types get more frequently imitated. Individual behavior, the level of cooperation within the population and the strength of sanctions then evolves indirectly, along with the endogenous change in preferences.

We first study the evolution of norm-sensitivity within a framework, where agents adapt to one particular environment associated with one equilibrium state of the public good game. If the status impact of sanctions is sufficiently strong, norm-enforcement may counterbalance the fitness advantage of free-riders for high levels of norm-compliance. In this case, there is scope for an evolutionary equilibrium with an intermediate level of norm compliance, where free-riders and cooperators coexist. However, such an equilibrium is never evolutionary stable. Typically, the evolutionary pressure would induce a decline in the norm sensitivity and cooperation would break down.

This picture dramatically changes once we study evolution within a 'heterogenous environment'. While in biology, the idea that subjects may face different 'habitats' is quite common,[6] this idea has been hardly reflected in evolutionary economics. We introduce this idea, linking the multiplicity of equilibria, which arises in our basic model,

[5]Compare Fehr and Gächter (2002) on monetary punishment and Masclet et al. (2003) as well as Falk et al. (2005) on 'informal' sanctions.

[6]Compare, among many others, Christiansen (1975), Levins (1968), Maynard Smith and Hoekstra (1980), Via and Lande (1987).

to the evolutionary process. We consider the case where the population faces sometimes an equilibrium state with strong norm-compliance and sometimes an alternative state with widespread norm violations. Agents will then interact in both, in mainly cooperative and in mainly non-cooperative environments, where the fitness-impact of sanctions is strong in the former and weak in the latter habitat. Consider for example the case of an anti-littering norm. Consider for example an individual, which crosses one public park on her daily way to work and promenades in a different park, every day after lunch. Assume that the first park is more or less clean, whereas the second one is littered. In such a heterogenous environment, we encounter three possible strategies: Individuals with a high level of norm sensitivity will unconditionally cooperate. They follow the anti-littering norm in both parks. Types with a low level of norm sensitivity, however, will free-ride in both parks: they litter, no matter what other individuals do. In contrast, individuals with an intermediate level of norm-sensitivity, however, will act as conditional cooperators: They cooperate in the 'clean habitat' but free-ride in the 'littered' environment. Thereby, they dominate both unconditional strategies. In the environment with a strong social norm – the clean park – conditional cooperators adhere to the norm and avoid social sanctions, which makes them more successful as free-riders. In the environment where the norm is weak – the littered park – they free-ride and earn a higher fitness payoff than unconditional cooperators. Hence, types with an intermediate norm-sensitivity are better adapted to a heterogenous environment as their preferences allows them to react flexibly to different social situations. Conditional cooperation dominates the unconditional strategies, which are specialized for one particular habitat. As we consider the pattern of norm-enforcement as exogenously given, we do not directly contribute to the literature on the evolution of social norms.[7] However, we provide a simple and intuitive explanation for the evolutionary origin and stability of conditional cooperation, which is novel in the literature.

While evolutionary economic approaches typically focus on discrete heterogeneity in trait values (e.g. different preference types),[8] we study the evolution of a continuously distributed trait – the norm-sensitivity. As already noted above, our indirect evolutionary analysis therefore constitutes an evolutionary underpinning of the distribution of preference types considered in chapter 2 and 3 of this thesis. We thereby use a method from quantitative genetics, originally introduced by Lande (1976). Lande's approach, takes into account for externalities in the evolutionary success of one preference type, which also emerges in our framework: As the distribution of norm sensitivity changes,

[7]Bowles and Gintis (1998), Boyd and Richerson (2002), Gintis (2003), Henrich and Boyd (2001).
[8]Compare e.g. Basin et al. (2004).

this induces different equilibrium states which renders the impact of social sanctions on status either stronger or weaker. Different distributions thereby lead to variations in the environment with implications for the success of one particular type. While the method we apply is novel in evolutionary economics, it provides a more tractable tool for our analysis than the approach based on standard replicator dynamics (Weibull, 1995). Moreover, we can show that our findings are quantitatively similar to the results one would derive according to such alternative evolutionary dynamics.

The chapter is structured as follows. We first study a simple model of social norms and cooperation in a large population. In section 4.3 we introduce an evolutionary approach from quantitative genetics. We then apply this method to our model and discuss the evolution of social norms and cooperative behavior in a homogenous respectively in a heterogeneous environment. Section 4.5 presents a simple model extension, which accounts for the endogenous formation of norm enforcing sanctions. Section 4.6 provides a critical discussion of our findings and section 4.7 concludes the chapter.

4.2 Social Norms and Cooperation

Consider a large society represented by a continuum of individuals $[0, 1]$. Each agent i chooses $x_i \in \{0, 1\}$, to contribute to the public good ($x_i = 1$, 'cooperate') or not to contribute ($x_i = 0$, 'free-ride'). The material payoff $y(x_i)$ for strategy x_i is given by

$$y(x_i) = -x_i c \tag{1}$$

where $c > 0$ depicts the material costs of the public good contribution. The action x_i also determines a payoff $z(x_i, n)$, where n depicts the share of free-riders in the society. This payoff is defined as

$$z(x_i, n) = (x_i - 1)\, s(n) \tag{2}$$

where $s(n)$ relates to the sanctions an agent incurs, if he violates the cooperation norm. The origin of these sanctions could in principle be internal, external or a mixture of both. In the context of internalized social norms (Elster, 1989), emotions represent an internal sanctioning mechanism. If an agent has internalized a cooperation norm, free-riding would be associated with emotions as guilt, remorse or the loss of self-esteem.[9] External sanctions could be monetary or non-monetary, e.g. related to social

[9]For a review on the role of emotions in economic theory compare Elster (1998). Compare also Bowles and Gintis (2003), van Winden (2001).

disapproval.[10] For the moment, we will not study the origin of these sanctions – i.e. why people engage in (costly) norm-enforcement activities – and simply assume that there is a mechanism which induces a certain punishment for free-riders. In section 4.5 we discuss a model extension which endogenously explains the origin of such norm-enforcing sanctions.

Throughout the whole chapter we employ the following assumption:

Assumption A1: The finite-valued function $s(n)$ is continuously differentiable in n. For $n \in [0, 1]$ there holds $s'(n) \leq 0$. Moreover $s(0) > 0$ and $s(1) = \varepsilon$.

Allowing the sanctions to depend on other agents' behavior captures the idea that the degree of norm compliance (co)determines the strength of norm-enforcement and thereby the strength of the social norm. Following the literature (e.g. Akerlof, 1980; Lindbeck et al. 1999), we here assume $s(n)$ to be decreasing in n: A deviant agents is supposed to suffer from weaker internal sanctions, as free-riding becomes more and more widespread: One feels less guilty violating an anti-littering norm in a dirty park as compared to a clean park. The equivalent is supposed to hold for external sanctions (Rege, 2004).[11] For the case of perfect norm compliance ($n = 0$), sanctions are strictly positive. In a society where everybody free-rides, however, the cooperation norm has completely eroded. The norm-based moral connotation of 'wrong' (free-riding) and 'right' (contributing) have vanished and sanctions are infinitesimal.

4.2.1 Preferences

Let the preferences of agent i, defined over $y(.)$, $z(.)$ and the public good payoff $v(g)$, be given by an additive separable utility function[12]

$$u_i(x_i, n) = y(x_i) + \theta_i z(x_i, n) + v(g(n)), \tag{3}$$

with the individual specific parameter $\theta_i \in [-\infty, \infty]$. The public good is defined by $g = g(n)$, $g' < 0$, and $v' > 0$. We can interpret the parameter θ_i as the degree of norm sensitivity. While a player with $\theta_i = 0$ is solely concerned about the direct pecuniary

[10] For evidence on the role of non-monetary sanctions compare Masclet et al. (2003), Rege and Telle (2003, 2004). A theoretical analysis of social sanctions is provided by Holländer (1990) and Rege (2004).

[11] For experimental evidence supporting this assumption, compare Falk et al. (2005), Masclet et al. (2003).

[12] This simple preference structure allows us to show the main points of our analysis as clear as possible. Our results do also hold for more complex preferences.

payoff from the game (and the public good), an agent with $\theta_i > 0$ also considers the norm-based payoff.[13]

Taking n as well as g as given,[14] player i will cooperate iff $u_i(1,n) > u_i(0,n)$, which holds for $\theta_i s(n) > c$. Hence, an agent will contribute to the public good, if the utility loss from the sanction dominates the costs of cooperation. This implies the threshold

$$\hat{\theta}(n) := \frac{c}{s(n)}, \tag{4}$$

which divides society into norm-adhering and norm-breaking individuals: Those with $\theta_i > \hat{\theta}(n)$ cooperate, while those with $\theta_i \leq \hat{\theta}(n)$ free-ride. The action choice of an agent is then determined by her norm sensitivity θ_i and the share of free-riders:

$$x_i = x(\theta_i, n) = \begin{cases} 0 & \text{for } \theta_i \leq \hat{\theta}(n) \\ 1 & \text{for } \theta_i > \hat{\theta}(n) \end{cases} \tag{5}$$

Note that the threshold $\hat{\theta}(n)$ is non-decreasing in n,

$$\frac{\partial \hat{\theta}(n)}{\partial n} \geq 0 \tag{6}$$

since $s'(.) \leq 0$. As more agents deviate from the norm, the sanctions associated with a norm violation become smaller. Hence, an agent who cooperates for low levels of n may turn into a free-rider for higher levels of n. Those individuals with $\theta_i \in (\hat{\theta}(0), \hat{\theta}(1)]$ condition their cooperation on the behavior of others. They act as *conditional cooperators* (Keser and van Winden, 2000; Fischbacher et al., 2001). Agents with $\theta_i \leq \hat{\theta}(0)$, however, would always free-ride, irrespectively of other subjects' behavior. Allowing for a heterogeneity in θ-types, the model therefore captures the two main patterns of behavior typically found in empirical studies (compare Gächter, 2005).

---

[13] Agents with $\theta_i < 0$ could be interpreted as 'punks'; they would derive benefits from a norm-violation. As will become clear in the following, we only include this latter group for technical convenience. Excluding non-negative values of θ would not change any of our results.

[14] In a large society the decision of a single individual has a negligible impact on the share of free-riders and on the public good level.

4.2.2 Equilibrium

Let the cumulative distribution function of the parameter θ be given by $\Phi(\theta)$. We assume that $\Phi(\theta)$ is continuously differentiable on the interval $[-\infty, \infty]$, the corresponding density function $\phi(\theta)$ has full support and $\phi(\theta) > 0$

Assumption A2: (i) The inverse function of the cumulative distribution is given by $\Phi^{-1}(n)$ for $n \in [0,1]$, with $\Phi^{-1}(0) = -\infty$ and $\Phi^{-1}(1) = \infty$. (ii) $\exists\ n \in (0,1)$: $\Phi^{-1}(n) > \hat{\theta}(n)$.

A social equilibrium state in such a society is given by a share of free-riders n^*, characterized by the fixed point equation

$$n^* = \Phi(\hat{\theta}(n^*)). \tag{7}$$

Lemma 1 *For any $s(n)$ and $\Phi(\theta)$ as characterized in A1 and A2(i) there always exists an equilibrium with $n^* = 1$. If A2(ii) holds, there always exists at least one further equilibrium with $0 < n^* < 1$.*

Proof. We can rewrite condition (7) as $\Phi^{-1}(n^*) = \hat{\theta}(n^*)$. From A2(i) we know that $\Phi^{-1}(1) = \infty$ and from $s(1) = \varepsilon$ follows $\hat{\theta}(1) \to \infty$. Hence, there always exists an equilibrium with $n^* = 1$. From A1 we know $s(0) > 0 \Rightarrow \hat{\theta}(0) > 0$ which implies $\hat{\theta}(0) > \Phi^{-1}(0)$. From this follows that there must exist at least one $n^* \in (0,1)$ where $\Phi^{-1}(n^*) = \hat{\theta}(n^*)$ holds as long as A2(ii) is fulfilled, since both $\hat{\theta}(n)$ and $\Phi^{-1}(n)$ are continuously increasing functions defined over the unit interval. ■

An equilibrium constitutes a self-supporting share of norm-violators: The threshold $\hat{\theta}(n^*)$ is such that the population share with $\theta_i \leq \hat{\theta}(n^*)$ is exactly n^*. There always exists one equilibrium where nobody contributes, $n^* = 1$. The cooperation norm has eroded, everybody free-rides and society fails to provide the public good. Given that assumption A2(ii) holds, the strength of the norm sensitivity is distributed such that there exists a level of free-riding n where the maximum level of norm sensitivity among free-riders, $\Phi^{-1}(n)$, is above the cooperation threshold $\hat{\theta}(n)$. In this case, the system is characterized by multiple equilibria. In addition to the equilibrium where the public good provision fails, there is at least one equilibrium with a positive share of contributors. A graphical representation of two possible scenarios is provided in figure 4.1. While for the example depicted in the left graph A2(ii) is fulfilled, it does not hold for the example in the right graph. In the first case, there is a multiplicity of equilibria, in the latter there is only an equilibrium with $n^* = 1$.

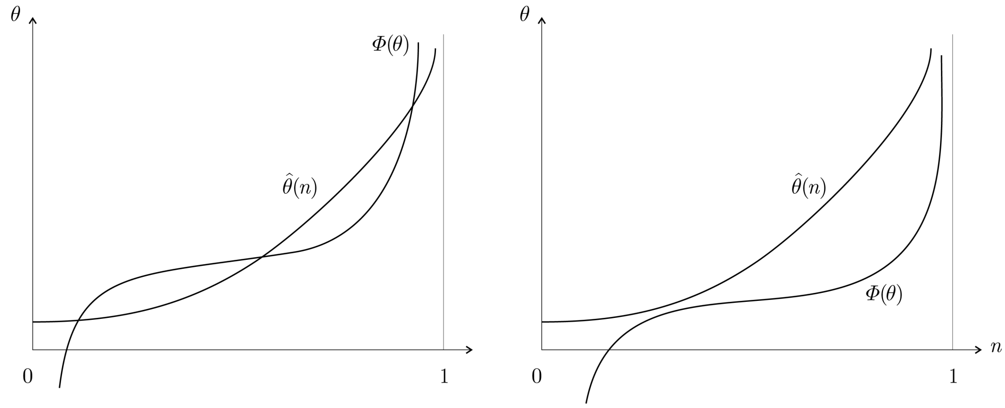

Figure 4.1: Equilibrium Share of Free-Riders

Given that the distribution $\Phi(\theta)$ is common knowledge, society immediately coordinates into one of the possible equilibria. Alternatively one could consider $\Phi(\theta)$ to be unknown, but assume that agents can induce the behavior of other members in society from the public good level. In a repeated game, agents would learn about the share of free-riders. As long as players base their next period's decision on this share – i.e. cooperate in period t iff $\theta_i > \hat{\theta}(n^{t-1})$ and free-ride otherwise – society would converge into an asymptotically stable equilibrium, characterized by

$$\frac{\partial \Phi^{-1}(n^*)}{\partial n} \geq \frac{\partial \hat{\theta}(n^*)}{\partial n}. \tag{8}$$

In the scenario depicted in the left graph in figure 4.1, there exist two instable equilibrium states – the one with an intermediate level of n^* and another one at $n^* = 1$ – and two stable equilibria: one with a high level of cooperation and another one where free-riding is widespread. In the graph on the right hand side of figure 4.1 the only equilibrium, $n^* = 1$, is also stable, since the cumulative distribution intersect the $\hat{\theta}(n)$-curve 'from below' (and therefore condition 8 holds).

In the evolutionary analysis of section 4.4, we will neglect instable states, assuming that in the long-run society always coordinates on a stable equilibrium state.

4.3 Evolutionary Quantitative Genetics

In the following we will study the evolution of the distribution $\Phi(\theta)$. For this purpose, we introduce a technique from evolutionary quantitative genetics,[15] first analyzed by Lande (1976). This approach provides a tractable method to study an evolutionary process within a continuously polymorphic population, i.e. to address the evolution of a type distribution.

Consider a large population which is heterogeneous along one trait α. The trait value is normally distributed with mean $\bar{\alpha}$ and variance σ^2, $\alpha \sim F(\alpha, \bar{\alpha}, \sigma^2)$. To simplify notation, we will denote the distribution function by $F(\alpha)$ and the corresponding density function by $f(\alpha)$. Let the fitness of an α-type for a given distribution with mean $\bar{\alpha}$ be given $w(\alpha, \bar{\alpha})$. The mean fitness of the population is then

$$\bar{w} = \int w(\alpha, \bar{\alpha})\, dF(\alpha). \tag{9}$$

Here we allow for *frequency dependent fitness.* Fitness is called frequency dependent, if the fitness of a α-type does also depend on the composition of the population.[16] In economic terms, frequency dependence is given if one group of subjects – respectively the strategy played by these α-types – creates an externality on other subjects' fitness. Consider for example a predator that hunts on a prey structured in different size classes and all predators are specialized to hunt on only one size class. If hunting has a significant effect on the prey populations, then the evolutionary success of a predator is dependent on how many other predators are specialized on the same class.[17]

Within one generation, the change in the trait mean value in response to selection is defined as

$$\Delta\bar{\alpha} = \bar{\alpha}_s - \bar{\alpha}, \tag{10}$$

[15] Compare Falconer and Mackay (1995) and Roff (1997) for an introduction to quantitative genetics. A critical review is provided by Pigliucci and Schlichting (1997).

[16] Of course, frequency dependent fitness depends also on the variance. In order to ease notation, we have suppressed this variable in $w(.)$.

[17] A trait with frequency independent fitness would be the case of different breathing technics, which lead to different capabilities in oxygen uptake. Whatever distribution of effective and ineffective breathers we assume, their cumulative impact on atmospheric oxygen is insignificant to change the amount of oxygen potentially available to other individuals. Hence, the fitness of each type is independent of the distribution of the different types.

where $\bar{\alpha}_s$, the mean trait value after selection, is defined as

$$\bar{\alpha}_s = \int \alpha \frac{w(\alpha, \bar{\alpha})}{\bar{w}} \, dF(\alpha). \tag{11}$$

The logic expressed in (11) is similar to standard replicator dynamics. While the initial frequency of a type was $f(\alpha)$, the post-selection frequency of this α-type, $\frac{w(\alpha,\bar{\alpha})}{\bar{w}} f(\alpha)$, will be higher for types with above-average fitness. Hence, in the computation of $\bar{\alpha}_s$, more successful types will get a higher weight than less successful types. If, for example, the studied trait is body size and e.g. large individuals (with high α-values) are more fit than short individuals, the average body size in the population would increase due to natural selection: $\bar{\alpha}_s > \bar{\alpha}$ and $\Delta\bar{\alpha} > 0$.

The analysis so far describes selection within one generation. In order to address the (inter-generational) evolution of the trait α, we follow Lande (1976) and introduce the following structure of reproduction: First, only selected individuals produce the next generation of offspring. Second, sexual reproduction is assumed with random partner choice. That is, two random subjects who survived selection mate, produce offsprings and die thereafter. For a reasonable selective pressure[18] this mechanism transforms the initial distribution back to a normal distribution with constant variance σ^2 but a different mean. Starting from a normal distribution with mean α, selection will first lead to a distribution, which deviates from the normal distribution. The mean of this (non-normal) distribution after selection is given by $\bar{\alpha}_s$ from (11). After random mating and reproduction, however, the distribution of α-values in the new generation is again normal with $F(\alpha, \bar{\alpha}_s, \sigma^2)$. While the variance is unaffected by this process, the mean of the distribution changes from $\bar{\alpha}$ to $\bar{\alpha}_s$. The direction of evolution is therefore determined by selection, characterized in (10) and (11). This now allows us to analyze the evolutionary process in more detail.

From (9) we can derive the change in mean fitness from a marginal change in $\bar{\alpha}$,

$$\frac{\partial \bar{w}}{\partial \bar{\alpha}} = \int w(\alpha, \bar{\alpha}) \, \frac{\partial f(\alpha)}{\partial \bar{\alpha}} \, d\alpha + \int \frac{\partial w(\alpha, \bar{\alpha})}{\partial \bar{\alpha}} \, dF(\alpha). \tag{12}$$

While the first term characterizes the direct change in the mean fitness due to a change in the composition of the population, the second term depicts the indirect, frequency dependent fitness impact. From the density of the normal distribution we can easily compute $\partial f(\alpha)/\partial\bar{\alpha}$. Substituting in (12) and rearranging yields[19]

[18]For a detailed discussion see Lande (1976).

[19]The derivation of (13) and (14) is provided in Appendix A.

$$\Delta \bar{\alpha} = \frac{\sigma^2}{\bar{w}} \int w(\alpha, \bar{\alpha}) \, \frac{\partial f(\alpha)}{\partial \bar{\alpha}} \, d\alpha \tag{13}$$

which can be also expressed as

$$\Delta \bar{\alpha} = \frac{1}{\bar{w}} \int \left[w(\alpha, \bar{\alpha}) \, (\alpha - \bar{\alpha}) \right] dF(\alpha). \tag{14}$$

The right hand side in equation (13) respectively (14) characterizes pace and direction of the evolutionary process. As $\bar{w} > 0$, the direction of the evolutionary change in the mean trait value, $\bar{\alpha}$, is determined by the sign of the integral in (14). Note, that the integral term represents only the direct change in mean fitness (the first term in equation (12)). From (14) therefore follows that the evolution of $\bar{\alpha}$ is *independent* of the frequency dependent fitness change associated with a variation in $\bar{\alpha}$. If the direct fitness change is positive, the distribution will evolve towards a higher mean $\bar{\alpha}$. An evolutionary equilibrium is reached if $\Delta \bar{\alpha} = 0$. Such an equilibrium is characterized by

$$\int \left[w(\alpha, \bar{\alpha}^e) \, (\alpha - \bar{\alpha}^e) \right] dF(\alpha) = 0, \tag{15}$$

where $\bar{\alpha}^e$ denotes the equilibrium mean trait value.

4.4 The Indirect Evolution of Conditional Cooperation

We now apply the method introduced in the previous section in order to study the evolution of the distribution $\Phi(\theta)$ and the associated co-evolution of cooperation. As we do not believe that the norm sensitivity θ is genetically determined, we interpret evolution as a cultural process, related to social transmission and learning mechanisms. Fitness then describes the success of a certain θ-types – related to social status or prestige – rather than fitness in the biological sense. In course of evolution, individuals learn about the social status of different θ-type (respectively the behavior of these types) and accordingly adapt their preferences, i.e. their θ values. Hence, the evolutionary process endogenously shapes preferences.[20] Individual behavior and thereby the level of cooperation within society evolves indirectly with the change in preferences from one generation to the next.[21]

[20] Compare e.g. Güth and Yaari (1992), Güth (1995), Bester and Güth (1998).

[21] The term generation thereby describes a population with a given distribution of preferences $\Phi(\theta)$, rather than a parent and offspring-population in the biological sense.

We assume that fitness, apart from the direct pecuniary payoff of the game $y(x_i)$, is also determined by the status consequences of the sanctions imposed on free-riders. If, for example, norm-violators get stigmatized as untrustworthy and are excluded from some economic interactions, this results in the decrease of income and social prestige. In section 4.5 we discuss a model extension which captures an endogenous mechanism inducing stigma related sanctions. For the moment, however, we take the existence of such a norm-enforcement mechanism as exogenously given. The fitness impact of norm-enforcing sanctions is assumed to depend on the share of norm violations n, characterized by $\lambda s(n)$. The fitness-impact of stigmatization declines with the share of norm-violators. [22] With this we can express the fitness for an action x_i as[23]

$$w(x_i) = y(x_i) + \lambda\, z(x_i, n). \tag{16}$$

If $\lambda = 0$, fitness is solely determined by $y(.)$. In case $\lambda > 0$, the sanctions do reduce the evolutionary success of free-riders.[24] While the fitness costs of free-riding are $\lambda s(n)$, agents consider a disutility of $\theta_i s(n)$. In the evolutionary process they learn about the status consequences of their actions and rationally adapt θ_i. This learning may take place within the family (vertical transmission of preferences), within peer-groups (horizontal transmission) or it is guided by societal institutions (oblique transmission). We assume that the outcome of the adaptation process can be described by the evolutionary model introduced above.[25]

4.4.1 Evolutionary Adaptation to a Homogenous Environment

Let θ be normally distributed according to $\theta \sim \phi(\theta, \bar{\theta}, \sigma^2)$, and the corresponding cumulative distribution is given by $\Phi(\theta, \bar{\theta}, \sigma^2)$. Substituting for $y(x_i)$, $z(x_i, n)$ and $x_i = x(\theta_i, n)$ from (1), (2) and (5), we can express individual fitness as a function of θ

[22] Alternatively one could also consider sanctions related to ostracism, e.g. in form of exclusion from the public good consumption. Note that the impact of ostracism would also depend on the share of free-riders in society: Getting excluded from public good consumption in a cooperative society, with a high level of public goods provided, represents a more severe punishment, than exclusion in a society where cooperation fails.

[23] We do not include the public good payoff into the fitness function, since this would not alter our results.

[24] Note that it is only the heterogeneity in actions − determined by different levels of θ − which results in fitness differences. Within the group of cooperators and free-riders the heterogeneity in θ_i does not result in different levels of fitness.

[25] In section 4.6 we discuss the crucial differences of this approach to an evolutionary process according to standard replicator dynamics (Weibull, 1995).

and $\bar{\theta}$,

$$w(\theta, \bar{\theta}) = \begin{cases} -c & \text{for } \theta > \hat{\theta}(n^*) \\ -\lambda\, s(n^*) & \text{for } \theta \leq \hat{\theta}(n^*) \end{cases} \tag{17}$$

where $n^* = \Phi(\hat{\theta}(n^*), \bar{\theta}, \sigma^2)$ is an equilibrium, analogous to (7), for a normal distribution with mean $\bar{\theta}$ (and σ^2 is exogenously fixed). Note that individual fitness as described by (17) is frequency dependent; as the θ-distribution changes, the share of free-riders n^* will change and thereby the fitness costs from the norm deviation. As the approach introduced in section 4.3 takes account of such payoff spillovers, the approach is applicable to this framework.

The mean fitness is defined by $\bar{w} = \int w(\theta, \bar{\theta})\, \phi(\theta)$. Using (17), we can express $\bar{w}$ as[26]

$$\bar{w} = -c + (c - \lambda\, s(n^*)) \int_{-\infty}^{\hat{\theta}(n^*)} d\Phi(\theta). \tag{18}$$

Following (14), the evolution of $\bar{\theta}$ is determined by

$$\Delta\bar{\theta} = \frac{1}{\bar{w}}\, (\lambda\, s(n^*) - c) \left(\bar{\theta} n^* - \bar{\theta}^*\right) \tag{19}$$

where $\bar{\theta}^*$ represents the mean level of θ among the n^* agents who defect in an equilibrium (compare Appendix B). As long as $0 < n^* < 1$, we can distinguish between the following three cases:

$$\begin{array}{ll} \Delta\bar{\theta} < 0 & \text{if } \lambda\, s(n^*) < c \\ \Delta\bar{\theta} = 0 & \text{if } \lambda\, s(n^*) = c \\ \Delta\bar{\theta} > 0 & \text{if } \lambda\, s(n^*) > c \end{array} \tag{20}$$

From this the following results derive:

Proposition 1 ***(i)*** *An evolutionary equilibrium where cooperators and free-riders coexist is characterized by* $\lambda\, s(n^e) = c$ *and* $n^e = \Phi(\hat{\theta}(n^e), \bar{\theta}^e, \sigma^2)$*, where* $0 < n^e < 1$ *constitutes an equilibrium share of free-riders which is supported by a normal distribution with mean* $\bar{\theta}^e$*.* ***(ii)*** *In such an equilibrium,* $w(\theta, \bar{\theta}^e)$ *is the same for all* θ *and there holds* $\lambda = \hat{\theta}(n^e)$*.* ***(iii)*** *An evolutionary equilibrium where cooperation fails,* $n^{e1} = 1$*, is characterized by* $n^{e1} = \Phi(\hat{\theta}(n^{e1}), \bar{\theta}^{e1}, \sigma^2)$*, supported by a normal distribution with mean* $\bar{\theta}^{e1}$*.*

[26]Note that the integral expression is equal to $n^* = \Phi(\hat{\theta}(n^*), \bar{\theta}, \sigma^2)$. (Compare Appendix B.)

Proof. The proof of **(i)** follows immediately from (19). From (4) we know that $c = \hat{\theta}(n^*)s(n^*)$ must hold for any equilibrium state. $\lambda\, s(n^e) = c$ then implies $\lambda = \hat{\theta}(n^e)$. Using this in (17) and substituting for (4) proofs **(ii)**. Part **(iii)** derives from $n^* = 1 \Rightarrow \bar{\theta}n^* = \bar{\theta}^*$. Hence, for $n^{e1} = 1$ the term in the last brackets in (19) is zero and $\Delta\bar{\theta} = 0$. ■

The evolutionary equilibrium described in part (i) of the proposition is characterized by a positive share of cooperators such that there is no fitness differential between free-riders and cooperators. In equilibrium, the preferences of agents with $\theta_i = \hat{\theta}(n)$, who are indifferent between defection and cooperation, coincide with the fitness function since $\hat{\theta}(n) = \lambda$. In other words, these indifferent θ-types are 'perfectly adapted'. In addition, there is also scope for an evolutionary equilibrium where cooperation breaks down; in such an equilibrium, the social norm has eroded and everybody free-rides. From Lemma 1 we know that $n^* = 1$ constitutes a possible equilibrium state for *any* distribution. Therefore, any level $\bar{\theta}$ could be the mean of the distribution in an evolutionary equilibrium with $n^{e1} = 1$. By the time the whole society free-rides, the evolutionary pressure on $\bar{\theta}$ to decline vanishes and the system reaches a rest point.[27]

Let us now turn to the existence of these different types of equilibria.

Proposition 2 ***(i)*** *Iff $\lambda\, s(0) > c$, there exists an evolutionary equilibrium with $0 < n^e < 1$.* ***(ii)*** *There always exists an evolutionary equilibrium with $n^{e1} = 1$. If $c > \lambda\, s(0)$, this is the only equilibrium.*

Proof. **(i)** Since $c > \lambda s(1) = \varepsilon$ and $s(.)$ is continuously decreasing in n, $\lambda\, s(0) > c$ assures that there exists a level of n where $\lambda\, s(n) = c$ holds. Moreover, we can always find a distribution $\phi(\theta, \bar{\theta}, \sigma^2)$, a function $s(n)$ and a level c, which supports such an equilibrium share of free-riders n^e. **(ii)** From Lemma 1 we know that $n^* = 1$ is supported by any distribution as long as A1 and A2(i) hold. Proposition 1(iii) implies that any equilibrium with $n^* = 1$ constitutes an evolutionary equilibrium n^{e1}. From A1 follows $c > \lambda\, s(0) \Rightarrow c > \lambda\, s(n)$ for all $n \in [0, 1]$. It therefore follows from $c > \lambda\, s(0)$ that there can not exist an equilibrium with $n^e < 1$, as $\nexists\, n$ with $\lambda\, s(n) = c$. ■

A graphical representation of this result is provided in figure 4.2, where we have plotted $\lambda s(n)$-curves for two different levels of λ. In the case of the higher curve,

[27] Theoretically, we could also describe an evolutionary equilibrium with $n^e = 0$. For this case, $n^* = 0 \Rightarrow \bar{\theta}n^* = \bar{\theta}^*$. Hence, the last bracket term in (19) would equal zero and $\Delta\bar{\theta} = 0$. However, an equilibrium state with $n^* = 0$ would only be supported by a distribution with $\bar{\theta} \to \infty$. Since this would violate assumption A2(i), we exclude this case from our analysis.

$\lambda\, s(0) > c$ and there exists an equilibrium with $n^e < 1$. In the case of the lower curve, there is no intersection with the c-line. The costs of cooperation are higher than the the fitness impact of sanctions, even for the state where $n^* = 0$; free-riding dominates cooperation in terms of fitness for any level of norm-violations. Starting from any $n^* < 1$, the evolutionary process induces $\bar{\theta}$ to fall and society moves towards an equilibrium with $n^{e1} = 1$.

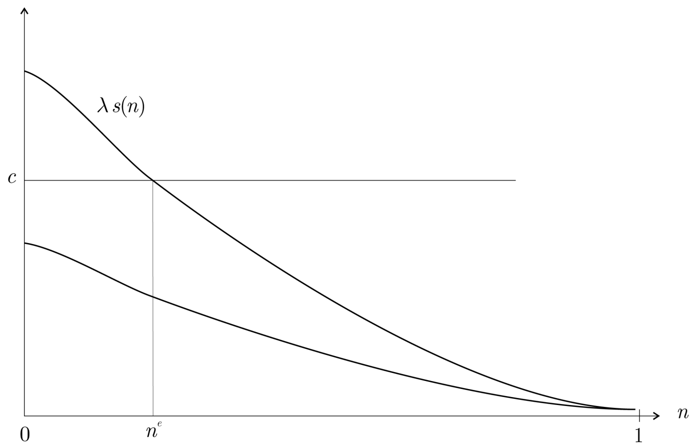

Figure 4.2: Fitness Payoffs

Finally, we address the stability of the system. An evolutionary equilibrium is locally stable if $\frac{\partial \Delta \bar{\theta}}{\partial n} > 0$. If this condition holds, small mistakes in the adaptation process would not affect the evolutionary equilibrium. Consider for example that, starting from an equilibrium distribution with $\bar{\theta}^e$, some agents would acquire a 'too' low level of θ. In this case, the share of free-riders would exceed n^e and the stability condition would imply $\Delta\bar{\theta} > 0$. An increase in the mean norm sensitivity would then provide a pressure to adapt 'back' towards the initial equilibrium. In our case, however, an evolutionary equilibrium where cooperators and free-riders coexist can never be stable.

Proposition 3 *An evolutionary equilibrium with $0 < n^e < 1$ is never stable. In contrast, an evolutionary equilibrium with $n^{e1} = 1$ is locally stable.*

Proof. See Appendix B. ■

Due to assumption A1, $s'(n) \leq 0$. Hence, any small deviation from n^e would tip the balance in fitness-payoffs between the two strategies. If the level of free-riding would slightly exceed n^e, the norm payoff would become less important and we get $c > \lambda\, s(n)$. Free-riders, i.e. types with low θ-values, earn a higher level of fitness and consequently

$\bar{\theta}$ decreases. The system moves to an equilibrium with $n^{e1} = 1$.[28] Note that the system would return to such an equilibrium after small shocks – e.g. if some agents mistakenly cooperate – as in the neighborhood of $n^{e1} = 1$ there holds $c > \lambda\, s(n^{e1})$ since $s(1) = \varepsilon$. Hence, $\bar{\theta}$ would decline and thereby trigger an increase in free-riding which would push behavior back towards the equilibrium n^{e1}.

The analysis provided so far yields an unsatisfactory result. While there may exist an evolutionary equilibrium where free-riders and cooperators coexist, such an equilibrium turns out to be instable. With the pattern of sanctions $s(n)$ considered in A1, the evolutionary adaptation induces disruptive selection respectively *disruptive evolution*: Typically, either one or the other strategy dominates in terms of fitness. The system either evolves towards an equilibrium where the norm has eroded and everybody free-rides, or $\bar{\theta} \to \infty$ and society would evolve towards full cooperation (compare footnote 27). While this is in conflict with the coexistence of free-riders and cooperators observed in real life social outcomes, we are nevertheless convinced that the crucial model assumption from A1, $s'(n) \leq 0$ – which is quite common in the literature[29] – as well as the fitness function introduced in (16) do make sense and can result in reasonable, stable, evolutionary outcomes. Here we have studied the adaptation to a homogenous environment. Agents encounter one particular situation, and evolution shapes their preferences such that they are fit for this particular environment. In reality, however, we typically face heterogeneous environments, as social interaction are repeated in different situations with quite diverse outcomes. The level of cooperation varies for different collective action problems, along time and along space. In the next section we show how we can capture such a heterogeneous environment within our model framework. In contrast to the case of a homogenous environment, the evolutionary process can result in stable equilibria where cooperators and free-riders coexist.

[28] If, on the one hand, the share of free-riding falls short of n^e, we get $\lambda\, s(n) > c$. Cooperators would be more successful than free-riders, $\bar{\theta}$ would increase and n^* would decline further. The system would evolve towards $\bar{\theta} \to \infty$. (Compare Footnote 27.)

[29] Compare e.g. Akerlof (1980), Corneo (1995), Lindbeck et al. (1999), Naylor (1989), Romer (1984).

4.4.2 Evolutionary Adaptation to a Heterogeneous Environment

Consider a society which faces one public good game which is repeated (finitely) many times within each generation. We focus on the case where there is scope for two possible equilibria,[30] n_a^* and n_b^*. Without loss of generality, we assume that $n_a^* < n_b^*$. Sometimes cooperation works rather well, sometimes it breaks down (e.g. due to exogenous shocks). Let the frequency that society coordinates on the equilibrium state n_j^* be given by π_j, for $j \in \{a, b\}$.[31] The fitness of a strategy $\left(x_i^a, x_i^b\right)$ is then given by

$$w\left(x_i^a, x_i^b\right) = \sum_{j=a,b} \pi_j \left(y(x_i^j) + \lambda\, z(x_i^j, n_j^*)\right), \tag{21}$$

where x_i^j denotes the action of agent i in equilibrium j. From $n_a^* < n_b^*$ and (6) follows $\hat{\theta}(n_a^*) < \hat{\theta}(n_b^*)$. Hence, we will observe three different strategies: On the one hand, agents with $\theta_i \leq \hat{\theta}(n_a^*)$ will free-ride in both equilibrium states. Agents with $\theta_i > \hat{\theta}(n_b^*)$ on the other hand, will cooperate in both states. A third group of subjects, those with $\hat{\theta}(n_a^*) < \theta_i \leq \hat{\theta}(n_b^*)$, behaves conditionally cooperative; such individuals would cooperate in equilibrium state a but defect in b. Making use of (1), (2), (5) and $\pi_b = 1 - \pi_a$, we can express the success of a type θ in the following way:

$$w(\theta, \bar{\theta}, \pi_a) = \begin{cases} -c & \text{for } \theta > \hat{\theta}(n_b^*) \\ -\pi_a c - (1-\pi_a)\,\lambda\, s(n_b^*) & \text{for } \hat{\theta}(n_a^*) < \theta \leq \hat{\theta}(n_b^*) \\ -\pi_a \lambda s(n_a^*) - (1-\pi_a)\,\lambda\, s(n_b^*) & \text{for } \theta \leq \hat{\theta}(n_a^*) \end{cases} \tag{22}$$

with $n_j^* = \Phi(\hat{\theta}(n_j^*), \bar{\theta}, \sigma^2)$. The crucial difference to the fitness function from (17) is the fact that intermediate θ-types obtain a fitness-payoff from two different actions. The success of the conditional cooperative strategy consists of the cooperation payoff for equilibrium state a plus the payoff from free-riding once the society has coordinated on state b.

[30] Remember that there are (at least) two possible stable equilibrium levels of free-riding, as long as assumptions A2 is fulfilled. Compare Lemma 1.

[31] The frequency π_a could be derived endogenously, according to the basin of attraction of equilibrium n_a^* relative to n_b^*. We will come back to this idea in section 4.6.

From (22) we can compute the mean fitness of the population for a given π_a,

$$\bar{w} = -c + \pi_a \left(c - \lambda s(n_a^*)\right) \int_{-\infty}^{\hat{\theta}(n_a^*)} d\Phi(\theta) + (1 - \pi_a)\left(c - \lambda\, s(n_b^*)\right) \int_{-\infty}^{\hat{\theta}(n_b^*)} d\Phi(\theta). \tag{23}$$

According to (14), the evolution of $\bar{\theta}$ is then determined by $\Delta\bar{\theta} = \frac{1}{\bar{w}}\Psi$ with[32]

$$\Psi \equiv \pi_a \left(\lambda s(n_a^*) - c\right)\left(\bar{\theta} n_a^* - \bar{\theta}_a^*\right) + (1 - \pi_a)\left(\lambda\, s(n_b^*) - c\right)\left(\bar{\theta} n_b^* - \bar{\theta}_b^*\right), \tag{24}$$

and $\bar{\theta}_j^*$ captures the mean level of θ among the free-riders in equilibrium n_j^*. The evolutionary dynamics on $\bar{\theta}$ are then given by

$$\begin{array}{ll} \Delta\bar{\theta} > 0 & \text{if } \Psi > 0 \\ \Delta\bar{\theta} = 0 & \text{if } \Psi = 0 \\ \Delta\bar{\theta} < 0 & \text{if } \Psi < 0 \end{array} \tag{25}$$

We can set up the following proposition:

Proposition 4 *Consider a heterogeneous environment with* $0 < \pi_a < 1$. ***(i)*** *An evolutionary equilibrium is characterized by* $\Psi = 0$ *with* $n_a^e = \Phi(\hat{\theta}(n_a^e), \bar{\theta}^e, \sigma^2)$ *and* $n_b^e = \Phi(\hat{\theta}(n_b^e), \bar{\theta}^e, \sigma^2) < 1$, *where* n_a^e *and* n_b^e *are supported by a normal distribution with mean* $\bar{\theta}^e$. ***(ii)*** *In equilibrium there holds* $\lambda s(n_a^e) > c > \lambda s(n_b^e)$.

Proof. Part **(i)** follows immediately from (25). Part **(ii)** derives from (24): Note that $\bar{\theta} n_j^* > \bar{\theta}_j^*$ as long as $n_j^* < 1$. Hence, the first term in (24) would be negative if $c > \lambda s(n_a^e)$. Since $n_a^e < n_b^e$, (6) implies that the second term would be negative as well. We would get $\Psi < 0$. Therefore $c > \lambda s(n_a^e)$ can not hold in an equilibrium. Iff $\lambda s(n_a^e) > c$, the first term in (24) is positive. In order to get $\Psi = 0$, the second term in (24) must be negative, which holds for $c > \lambda s(n_b^e)$. ■

As long as $n_b^e < 1$,[33] the distribution in an evolutionary equilibrium supports two equilibria such that $\lambda s(n_a^e) > c > \lambda s(n_b^e)$. In terms of fitness, cooperation dominates free-riding in equilibrium state a, since $-c > -\lambda s(n_a^*)$. For state b, however, the opposite holds: As free-riding becomes more widespread, the (fitness) costs from sanctions

[32]The derivation of $\Delta\bar{\theta}$ respectively Ψ is analogous to the one of (19). Compare Appendix B.

[33]There could also exist an equilibrium with $n_b^e = 1$ and $\lambda s(n_a^e) = c$. This case represents a heterogeneous environment with a positive share of cooperation in equilibrium a and a complete break-down of cooperation in equilibrium state b. As the properties of this case are similar to the scenarios discussed in the previous subsection, we focus on equilibria with $n_b^e < 1$.

are lower than the costs of cooperation. From this follows that conditional cooperators have more success than 'unconditional' cooperators respectively free-riders.

Corollary 1 *In an evolutionary equilibrium within a heterogeneous environment with* $n_a^e < n_b^e < 1$ *and* $0 < \pi_a < 1$*, conditional cooperators have a strictly higher fitness than both, free-riders and cooperators.*

Proof. From Proposition 4(ii) we know that $\lambda s(n_a^e) > c > \lambda s(n_b^e)$. Using this in (22) proofs the Corollary. ■

Figure 4.3 (on the next page) graphically represents such an equilibrium. The graph on the left hand side captures a system with a distribution $\Phi(\theta)$ and a function $\hat{\theta}(n)$, supporting two stable equilibrium states $n_a^* < n_b^* < 1$. The graph on the right hand side depicts the fitness difference between the strategies for the two equilibria. As in this example the advantage of cooperators compared to free-riders in equilibrium state a, $\lambda s(n_a^e) - c > 0$, is smaller than the disadvantage in b, $\lambda s(n_b^e) - c < 0$, the weight on state a – expressed by π_a – must be sufficiently high in order that $\Psi = 0$ holds.

From figure 4.3 as well as from the analysis above (compare Proposition 2) it is clear that $\lambda s(0) > c$ is a necessary condition for such an evolutionary equilibrium to exist. In addition, assumption A2(ii) must be fulfilled, such that there are (at least) two equilibrium states.

Analogously to before, the necessary conditions for local stability is $\frac{\partial \Delta \bar{\theta}}{\partial n_a} + \frac{\partial \Delta \bar{\theta}}{\partial n_b} > 0$. The formal analysis yields the following result:

Proposition 5 *It is sufficient for an evolutionary equilibrium with* $n_a^e < n_b^e < 1$ *to be stable, if* $-s'(n_a^e) < \Gamma_a$ *and* $-s'(n_b^e) > \Gamma_b$ *holds, with*

$$\Gamma_j \equiv \bar{\theta} \left(\frac{\lambda \left(\bar{\theta} n_j^e - \bar{\theta}_j^e \right)}{\lambda\, s(n_j^e) - c} + \frac{\hat{\theta}(n_j^e)^2}{s(n_j^e)} \phi(\hat{\theta}(n_j^e)) \right)^{-1}$$

Proof. See Appendix B ■

As we know from Proposition 4, there holds $c > \lambda\, s(n_b^e)$. Hence the denominator of the first term in the round brackets of Γ_b is negative and we could get $\Gamma_b < 0$. In this case $-s'(n_b^e) > \Gamma_b$ would always be fulfilled. From Proposition 4 also follows $\Gamma_a > 0$. Therefore $s'(n_a^e) = 0$ would be sufficient for $-s'(n_a^e) < \Gamma_a$ to hold.

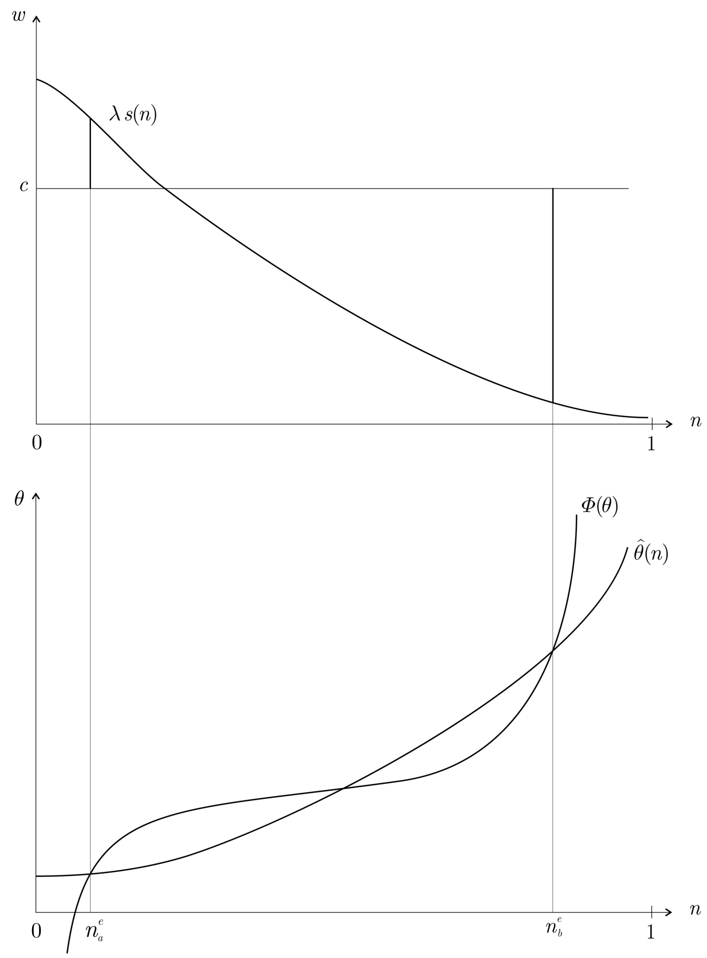

Figure 4.3: Evolutionary Equilibrium in a Heterogenous Environment

Since the stability of an evolutionary equilibrium as characterized by Proposition 4 is in general ambiguous, we conducted a series of numerical simulations. We could not find even one single parameter combination, where the stability condition did not hold. Even in the case where one of the two (sufficient) conditions in Proposition 5 were violated, the equilibrium proved to be stable. We are therefore confident, that evolutionary equilibria within a heterogeneous environment are stable for a wide range of parameters.

The intuition for this finding is straight forward: Small shocks would not change the result from Corollary 1. Conditional cooperation would still perform more successfully than the two unconditional strategies. As conditional cooperators have intermediate values of θ, preferences in the 'middle' of the θ-range are more successful and dominate more extreme – either low or high – θ-values. In contrast to the evolutionary adaptation to a homogenous environment, there is no scope for disruptive evolution. Hence, in contrast to the analysis of the previous subsection, we find (potentially) stable equilibria where cooperators and defectors coexist.

The evolutionary dominance of conditional cooperators is the main result of our analysis. Holding extreme ‘anti-social’ (low θ values) or extreme ‘pro-social’ preferences (high θ values) induces agents to play one particular strategy, irrespectively of the other agents behavior. In a stable evolutionary equilibrium within a homogenous environment, one of these two strategies will dominate the other. In a heterogeneous environment, there is scope for a third type of strategy, that of conditional cooperation. If individuals adapt to such a heterogeneous habitat, where they face either a ‘good’ or a ‘bad’ outcomes, the unconditional strategies prove less successful that the conditional strategy. Agents who free-ride in ‘bad’ equilibrium states but cooperate in a ‘good’ states, dominate the (unconditional) cooperators in the former and the free-riders in the latter environment. As compared to conventional one-game-one-equilibrium scenarios, such heterogeneous environments appear as a more realistic description of social interactions. Therefore, the evolutionary pressure to adapt to several possible outcomes provides a simple explanation for the evolutionary success of conditionally cooperative behavior.

4.5 Extension: Endogenous Social Sanctions

In this section we describe a model extension which provides an endogenous explanation for the origin of norm-enforcing sanctions. While the integration of this extension into the above analysis is left for future research, we briefly discuss some basic results which support that sanctions are decreasing in n (our assumption A1).

Assume that, next to the public good game described in section 4.2, subjects play a second game. Each agent gets randomly matched with one other member of society. One of the two, the proposer labeled P, can choose a strategy $p \in \{e, l\}$; either P enters the game and proposes an economic interaction or she leaves the game. If the proposer chooses this latter option l, both players earn a zero payoff. In case player P enters the game, the second player, responder R, chooses a strategy $r \in \{+, -\}$. When the responder decides for a cooperative move, $+$, both players earn a payoff $a > 0$. Alternatively, player P could also decide for an explorative move, $-$, such that P earns a payoff $b < 0$, and R gets $d > a$. In this case, however, the second mover suffers from the violation of the social norm for cooperation, similarly as in the public good game. He would obtaining an emotional payoff $-\theta_i$, associated to purely internal sanctions.[34] The extensive form of the game, which can be interpreted as a discrete variation of the trust game (Berg et al. 1995), is represented in figure 4.4.

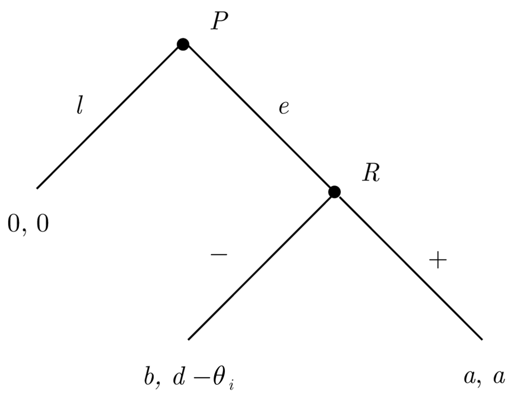

Figure 4.4: Trust Game

Given that player P can observe the type of R, we get the following subgame perfect Nash equilibrium $\{p^*, r^*\}$

$$\begin{array}{ll} \{e, +\} & \text{if } \theta_i > d - a \\ \{l, -\} & \text{if } \theta_i \leq d - a \end{array} \tag{26}$$

For $\theta_i > d - a$, the responders' norm sensitivity θ_i is strong enough to overcome the pecuniary temptation to exploit the proposer. Player P will propose an interaction, R behaves cooperatively and both players earn a payoff a. If $\theta_i < d - a$, the first mover will not enter the game since R would exploit him. Hence, if the matched responder has a norm sensitivity above the threshold $\tilde{\theta}$,

$$\tilde{\theta} \equiv d - a, \tag{27}$$

the proposer enters the game and leaves it otherwise.

Let us now turn to the more interesting case where types are non-observable. Remember that we consider the type distribution $\Phi(\theta)$ to be common knowledge. In order to study a link between trust and the public good game, we introduce the following structure: (a) Agents play the public good game myopically. (b) After the public good game, they get randomly matched with finitely many other members of society. In each match the chance of being the proposer is equal to the chance of becoming the responder. (c) With probability λ, the action of R in the public good game is observed and becomes public knowledge. It is obvious that the behavior of myopic agents in the

[34]The norm strength is assumed to be independent of responders behavior in other matches.

public good game is still characterized by $x_i = x(\theta_i, n)$ from (5). We now provide a brief analysis of the trust game.

As in the case of complete information, the strategy of a responder is given by

$$r = \begin{cases} - & \text{for } \theta_i \leq \tilde{\theta} \\ + & \text{for } \theta_i > \tilde{\theta} \end{cases} \tag{28}$$

Given that the proposer receives no signal about the responder's type – i.e., the responder's action in the public good game remains unobserved – his expected payoff (based on Bayesian beliefs) of entering the game is

$$E[e] = \Phi(\tilde{\theta})b + \left(1 - \Phi(\tilde{\theta})\right) a. \tag{29}$$

With a probability of $\Phi(\tilde{\theta})$ he would face a responder with a θ-value below $\tilde{\theta}$ who would exploit him. With probability $1 - \Phi(\tilde{\theta})$ the responder would cooperate. As long as the share of exploiting responders in the population is below a certain level,

$$\Phi(\tilde{\theta}) \leq \frac{a}{a-b}, \tag{30}$$

the proposer enters the game since $E[e] > E[l] = 0$.[35] If the responder's action x_i was observed, this provides the proposer with a signal about the type θ_i of player R. Given that player R was observed free-riding in an equilibrium state n^*, his θ-type must be below $\hat{\theta}(n^*)$. The expected payoff of trusting a free-rider is then

$$E[e|x_R = 1] = \begin{cases} b & \text{for } \hat{\theta}(n^*) \leq \tilde{\theta} \\ \frac{\Phi(\tilde{\theta})}{n^*} b + \frac{n^* - \Phi(\tilde{\theta})}{n^*} a & \text{for } \hat{\theta}(n^*) > \tilde{\theta} \end{cases} \tag{31}$$

where we have substituted for $n^* = \Phi(\hat{\theta}(n^*))$. From this immediately follows that the signal which derives from the observation that R was free-riding, is more informative, the lower n^*. In an equilibrium with a low level of free-riding, the threshold for norm-violations is rather low. Those who deviate from the cooperation norm in a society with a high level of cooperation must have a particularly low level of θ_i. If the cooperation threshold for the public good game $\hat{\theta}(n^*)$ drops below the threshold $\tilde{\theta}$, observing someone cheating is an unambiguous signal that this type will exploit a partner in the trust game. In such an equilibrium n^*, a proposer would not enter the

[35] Remember that $E[l] = 0$.

game if the matched partner was observed free-riding. If, however, norm-violations in the public good game are more widespread, $\hat{\theta}(n^*) > \tilde{\theta}$. For such equilibrium states n^* there are free-riders with $\hat{\theta}(n^*) > \theta_i > \tilde{\theta}$, who would defect in the public good game but cooperate in the trust game. Given that this latter group is big enough such that $E[e] > 0$, proposers enter the game despite their partner is a free-rider. This occurs, as long as the following condition holds:

$$\Phi(\tilde{\theta}) \leq \frac{a}{a-b} n^* \tag{32}$$

In an equilibrium state n^* where (32) holds, a player P enters the game in any case, since then condition (30) is also fulfilled. The higher n^*, the less informative is the signal about the θ-type of player R. If $n^* = 1$, the signal is completely uninformative and condition (32) boils down to the basic condition (30).

This preliminary analysis yields the following insights: 1) If n^* is low enough, a free-rider will get excluded from the game, if his norm-violation in the first game is observed. 2) The higher λ, the likelihood of getting observed, the more often a free-rider would get excluded. 3) In the case of exclusion, all observed free-riders – irrespectively of their level of θ_i – miss the chance to earn a positive payoff. 4) The higher n^*, the more likely a free-rider gets into the game and reaps a positive payoff. 5) If condition (30) is violated, the norm sensitivity in society is too low. Trust breaks down and proposers will never enter the game.

Note, that the properties of the exclusion mechanism induces sanctions on free-riders, which are roughly in line with the form $\lambda s(n)$ used in the fitness function (16). The evolutionary adaption of θ would then be driven by the agent's learning that their action in the public good game has a payoff impact in a future game. The social norm would work as a mechanism to take into account the (pecuniary) consequences neglected by myopic agents.

As the payoff structure in the Nash equilibrium of this second game does not exactly correspond to the fitness payoffs considered in section 4.4, the results of our evolutionary analysis can not immediately be applied to the 'combined game', based on this model extension. In future research we will apply the evolutionary analysis to the joint equilibrium payoffs from the public good and the trust game.

4.6 Discussion

4.6.1 The Applicability of a Quantitative Genetic Approach

In section 4.4 we have applied a method from quantitative genetics to a cultural, social learning process. According to this approach, originally studied by Lande (1976), the trait θ follows a normal distribution and the frequency of a trait changes according to the fitness-differential, $w(\theta)/\bar{w}$. If the fitness of a θ-type relative to the mean population fitness is greater than unity, the frequency of these types will increase (and shrink otherwise). The resulting (non-normal) distribution is then transformed back to a normal distribution with a new mean. This methodology provides a tractable tool to study the evolution of a (normal) distribution, which also allows us to incorporate frequency dependent fitness components. Of course, the application of this method has several limitations.[36] In our case, it implies an imperfect learning process, as the initial variance in θ is maintained during the course of evolution. Hence, by using this method we exclude the case where all agents adapt one unique θ value (e.g. $\theta = \lambda$). In an evolutionary equilibrium those agents with $\theta_i > \lambda$ ($\theta_i < \lambda$) are in some sense 'overadapted' ('underadapted'), since they consider the impact of norm-enforcement to be stronger (weaker) than the actual (fitness-) impact of these sanctions. One could justify this implication by a systematic noise embedded in the social learning process. If the errors in the adaptation process are normally distributed, the deviations from the perfect adaptation in θ would maintain a normal distribution $\Phi(\theta)$.[37] However, as our analysis – at least section 4.4.1 – is also applicable for a distribution with an infinitesimal small variance, this point is only of minor importance.

[36]If, for example, there would be an evolutionary pressure on low *and* high θ-types to grow, this would suggest an evolution towards a bimodal distribution, which is excluded by assumption in the Lande approach. However, a disruptive evolution pointing in two directions can not occur in our framework.

[37]Once an evolutionary equilibrium is reached, these errors should of course decline, such that the variance of the distribution shrinks to zero.

4.6.2 Replicator Dynamics

Would our results still hold if standard replicator dynamics are at work? Consider any initial distribution of θ and let the frequency of a type, $f(\theta)$, evolve according to

$$\dot{f}(\theta) = f(\theta)\,[w(\theta) - \bar{w}] \tag{33}$$

(compare Weibull, 1995). From the analysis in section 4.4.1 immediately follows that any distribution which supports an equilibrium share n^e with $\lambda s(n^e) = c$ also constitutes an evolutionary equilibrium according to (33). If $\lambda s(n^e) = c$ holds, there are no fitness-differences between free-riders and cooperators (compare Proposition 1) and we would get $w(\theta) = \bar{w} \Rightarrow \dot{f}(\theta) = 0$ for all θ. Similarly, the stability properties of such an equilibrium $0 < n^e < 1$ carries over. Any small deviation from n^e would either lead to a break down in cooperation or a move towards full cooperation.

The comparability of the adaptation to a heterogeneous environment analyzed in section 4.4.2, turns out to be slightly more complicated. Intuitively, it is clear that the evolutionary success of conditional cooperation also holds under the evolutionary process described by (33). Therefore, the share of agents who follow a conditional cooperative strategy would increase. In an evolutionary equilibrium there must hold $w(\theta) = \bar{w}$ for all θ-types with $f(\theta) > 0$. This can only hold if the whole population consist of conditional cooperators. Hence, the evolutionary process described by (33) will eliminate all preferences which induce unconditional strategies. In equilibrium all agents will cooperate in one equilibrium ($n_a^* = 0$) and defect in the other state ($n_b^* = 1$). While this outcome qualitatively resembles the results of our analysis, there is no scope for the coexistence of free-riders, unconditional and conditional cooperators according to the standard replicator dynamic approach.

4.6.3 Heterogeneous Environments

Alternatively to the interpretation of a heterogeneous environment provided in section 4.4, one could also reinterpret a heterogeneous scenario in terms of two identical, 'parallel' public good games within one generation. Consider for example one anti-littering norm and two completely identical public parks. On the one hand, park a is rather clean, the norm is perceived as strong and agents predominantly comply with the social norm. Park b, on the other hand, is littered, and society is in a stable equilibrium with widespread free-riding. The probabilities π_a and π_b then reflect the agents' likelihood of playing on one particular playground – i.e. in one specific park. Note, that this probability is considered to be independent of their type. In reality, different θ-types may sort in different environments. One park may be cleaner than the other, because the visitors in this park are more sensitive towards the anti-littering norm than the visitors of the other park. In the current version of the model, we abstract from this case, related to the level of viscosity within a population (Myerson et al., 1991). The main point of our analysis would also hold for a segregated society (a viscous population), as long as there are some agents who interact in different environments.[38]

Next to the incorporation of population viscosity, there is another interesting way one could endogenize the probability of facing one particular equilibrium outcome: We could relate π_j to the relative size of the basin of attraction for equilibrium state n_j^*. From the discussion in section 4.2 it is clear, that the basin of attraction for one stable fixed point n_j^* is defined by the position of the surrounding instable equilibria (fixed points). For the case of two stable equilibria, as depicted in the left graph in figure 4.3, it is the location of the instable equilibrium with an intermediate level of free-riding, which separates the distinct basins of attraction. As an increase in the mean $\bar{\theta}$ would shift the $\Phi(\theta)$-curve upwards, the level of free-riding for the instable fixed point would increase. Hence, with an increase in the mean norm sensitivity, the basin of attraction for the equilibrium with a low level of free-riding becomes larger and the one of the other equilibrium shrinks. Accordingly, the probability π_a (π_b) would be increasing (decreasing) in $\bar{\theta}$. This effect would only quantitatively alter the properties of an evolutionary equilibrium in a heterogeneous environment. Endogenous probabilities π_j, however, could add some restrictions on the existence of an evolutionary equilibrium in a heterogeneous environment.

4.7 Conclusion

While the impact of heterogenous environments on evolutionary processes is well studied by biologists (e.g. Christiansen, 1975; Levins, 1968; Maynard Smith and Hoekstra, 1980; Via and Lande, 1987), this issue has been so far neglected in evolutionary economics. In this chapter we take a first step to close this gap in the literature, studying the indirect evolution of cooperation within a heterogenous environment. We consider a society with a social norm for cooperation and an established sanctioning mechanism

[38]A minimal degree of migration, i.e. population mobility between different environments, is sufficient to support this case. Compare Mengel (2005), for a recent evolutionary analysis of social norms in the context of migration.

for free-riders. The power of these norm-enforcing sanctions depends on the level of norm-compliance, thereby raising the scope for multiple equilibria. Society may coordinate on a cooperative equilibrium, where norm-compliance is high and sanctions are strong, or on a state with widespread free-riding and week norm-enforcement. These equilibria then determine the habitat-structure for the evolutionary process. Accordingly, the success of an agent – in terms of fitness – is determined by the payoff from interactions in both, cooperative and non-cooperative environments.

In our model framework, there arise three possible strategies, associated with different levels of norm-sensitivity: Agents who base their decisions mainly on the pecuniary payoff of the game free-ride in both environments, whereas individuals with a strong norm-sensitivity cooperate in both habitats. In contrast to these unconditional strategies, agents with a medium-level of norm-sensitivity follow a conditional strategy: they will cooperate in the one but free-ride in the other habitat. It is straightforward to show, that conditional cooperation dominate the unconditional strategies. In the cooperative environment, conditional cooperators follow the norm and avoid the punishment a free-riders incurs. In the habitat where the norm has eroded and sanctions do hardly play a role, conditional cooperators reap the same payoff as a free-rider, which dominates that of an (unconditional) cooperator. Hence, the preferences underlying conditional cooperation are well adapted to a heterogeneous environments. An intermediate level of norm sensitivity allows individuals to react flexibly to different environments. Thereby, they dominate unconditional strategies, which are specialized on one particular habitat.

As it seems quite realistic to consider human interaction to take place in heterogenous environments, where cooperation sometimes fails and sometimes works quite well, our approach provides a simple and plausible explanation of the evolutionary forces which shape conditional cooperative behavior. The main limitation of our approach, is the exogenously given pattern of norm-enforcement. In future research we will extend the present framework accounting for an endogenous sanctioning mechanism. This will allow us to study the robustness of our result in the broader context of the evolution of social norms.

Appendix

Appendix A (Section 4.3)

For the density of the normal distribution, $f(\alpha)$, one can easily derive

$$\frac{\partial f(\alpha)}{\partial \bar{\alpha}} = f(\alpha)\frac{(\alpha - \bar{\alpha})}{\sigma^2}. \tag{A.1}$$

Making use of this term in (12) and rearranging, we get

$$\frac{\partial \bar{w}}{\partial \bar{\alpha}} = \frac{1}{\sigma^2}\int [\alpha\, w(\alpha,\bar{\alpha})\, f(\alpha) - \bar{\alpha}\, w(\alpha,\bar{\alpha})\, f(\alpha)]\; d\alpha + \int \frac{\partial w(\alpha,\bar{\alpha})}{\partial \bar{\alpha}} dF(\alpha). \tag{A.2}$$

From (9) respectively (11) follows that the first expression in the first integral equals $\bar{\alpha}_s \bar{w}$, and the second expression is $\bar{\alpha}\bar{w}$. We arrive at

$$\frac{\partial \bar{w}}{\partial \bar{\alpha}} = \frac{\bar{w}}{\sigma^2}\,(\bar{\alpha}_s - \bar{\alpha}) + \int \frac{\partial w(\alpha,\bar{\alpha})}{\partial \bar{\alpha}} dF(\alpha). \tag{A.3}$$

Rearranging and substituting for (10) yields

$$\Delta\bar{\alpha} = \frac{\sigma^2}{\bar{w}}\left(\frac{\partial \bar{w}}{\partial \bar{\alpha}} - \int \frac{\partial w(\alpha,\bar{\alpha})}{\partial \bar{\alpha}} dF(\alpha)\right) \tag{A.4}$$

which is equivalent to (13). Making use of (A.1) we can then express (13) as

$$\Delta\bar{\alpha} = \frac{1}{\bar{w}}\int [w(\alpha,\bar{\alpha})\,(\alpha - \bar{\alpha})]\, dF(\alpha). \tag{A.5}$$

Appendix B (Section 4.4)

The mean fitness is given by

$$\bar{w} = -\lambda\, s(n^*) \int_{-\infty}^{\hat{\theta}(n^*)} d\Phi(\theta) - c \int_{\hat{\theta}(n^*)}^{\infty} d\Phi(\theta). \tag{A.6}$$

As $\Phi(\hat{\theta}(n^*)) = n^*$, we can rearrange $\bar{w}$ and get

$$\bar{w} = -\,(1 - n^*)\, c - n^*\lambda\, s(n^*). \tag{A.7}$$

From this follows (18).

As we have demonstrated in the section 4.3, only the direct fitness impact of a change in $\bar{\theta}$ is important for the evolution of this variable. The indirect effect – related to the frequency dependent fitness from $s(n)$ – is irrelevant. Hence, we follow (14) and derive

$$\Delta\bar{\theta} = \frac{\sigma^2}{\bar{w}} \left(c - \lambda\, s(n^*)\right) \int_{-\infty}^{\hat{\theta}(n^*)} \frac{\partial\phi(\theta, \bar{\theta}, \sigma^2)}{\partial\bar{\theta}} d\theta. \tag{A.8}$$

For the density of the normal distribution we get analogously to (A.1)

$$\frac{\partial\phi(\theta, \bar{\theta}, \sigma^2)}{\partial\bar{\theta}} = \frac{1}{\sigma^2} \left(\phi(\theta)\theta - \phi(\theta)\bar{\theta}\right). \tag{A.9}$$

With this, we can rewrite $\Delta\bar{\theta}$ as

$$\Delta\bar{\theta} = \frac{1}{\bar{w}} \left(\lambda\, s(n^*) - c\right) \int_{-\infty}^{\hat{\theta}(n^*)} \left(\phi(\theta)\bar{\theta} - \phi(\theta)\theta\right) d\theta, \tag{A.10}$$

where the first term in the integral is equal to $n^*\bar{\theta}$. The second expression in the integral depicts the mean level of θ for agents with $\theta_i \leq \hat{\theta}(n^*)$. Using

$$\bar{\theta}^* \equiv \int_{-\infty}^{\hat{\theta}(n^*)} \theta d\Phi(\theta) \tag{A.11}$$

we finally arrive at (19).

Proof of Proposition 3. From $\Delta\bar{\theta}$ we get

$$\begin{aligned}\frac{\partial\Delta\bar{\theta}}{\partial n} \quad = \quad & -\frac{1}{\bar{w}}\frac{\partial\bar{w}}{\partial n} \left(\lambda\, s(n^*) - c\right) \left(\bar{\theta}n^* - \bar{\theta}^*\right) \\ & +\frac{1}{\bar{w}}\lambda\, s'(n^*) \left(\bar{\theta}n^* - \bar{\theta}^*\right) + \frac{1}{\bar{w}} \left(\lambda\, s(n^*) - c\right) \left(\bar{\theta} - \frac{\partial\bar{\theta}^*}{\partial n}\right)\end{aligned} \tag{A.12}$$

From Proposition 1 we know that an equilibrium with $0 < n^e < 1$ is characterized by $\lambda\, s(n^e) = c$. Therefore, the first and the third term in (A.12) vanish in such an equilibrium. Since $s'(n) \leq 0$ and $\bar{\theta}n^* > \bar{\theta}^*$ for $n^* < 1$, there holds $\frac{\partial\Delta\bar{\theta}}{\partial n} \leq 0$. Hence, an evolutionary equilibrium with $n^e < 1$ is never locally stable.
In an equilibrium with $n^{e1} = 1$, $\bar{\theta}n^* = \bar{\theta}^*$. Therefore, the first and the second term in (A.12) vanish in such an equilibrium. Making use of the Leibnitz Rule of integral

differentiation and (4), we get

$$\frac{\partial \bar{\theta}^*}{\partial n} = \frac{\partial \hat{\theta}(n)}{\partial n} \hat{\theta}(n) \phi(\hat{\theta}(n)) = -\hat{\theta}(n)^2 \frac{s'(n)}{s(n)} \phi(\hat{\theta}(n)). \tag{A.13}$$

Using this we can rewrite the third term of (A.12) as

$$\frac{1}{\bar{w}} (\lambda\, s(n) - c) \left(\bar{\theta} + \hat{\theta}(n)^2 \frac{s'(n)}{s(n)} \phi(\hat{\theta}(n)) \right) \tag{A.14}$$

where we have substituted for (4). Note that $s(n^{e1}) = \varepsilon$ and $\hat{\theta}(n^{e1}) \to \infty$. Hence, for $n^{e1} = 1$, the term in the first brackets would equal $-c$ and the second term would be $-\infty$ (note that $s'(.) \leq 0$). From this follows $\frac{\partial \Delta \bar{\theta}}{\partial n} \geq 0$. ■

Proof of Proposition 5. Analogously to (A.12) we can derive from (23)

$$\begin{aligned} \frac{\partial \Delta \bar{\theta}}{\partial n_a} + \frac{\partial \Delta \bar{\theta}}{\partial n_b} &= \underbrace{-\frac{1}{\bar{w}^2} \left(\frac{\partial \bar{w}}{\partial n_a} + \frac{\partial \bar{w}}{\partial n_b} \right) \Psi}_{0} \\ &+ \frac{1}{\bar{w}} \pi_a \left[\underbrace{\lambda\, s'(n_a^*) \left(\bar{\theta} n_a^* - \bar{\theta}_a^* \right)}_{-} + \underbrace{(\lambda\, s(n_a^*) - c)}_{+} \underbrace{\frac{\partial \left(\bar{\theta} n_a^* - \bar{\theta}_a^* \right)}{\partial n_a^*}}_{+} \right] \\ &+ \frac{1}{\bar{w}} (1 - \pi_a) \left[\underbrace{\lambda\, s'(n_b^*) \left(\bar{\theta} n_b^* - \bar{\theta}_b^* \right)}_{-} + \underbrace{(\lambda\, s(n_b^*) - c)}_{-} \underbrace{\frac{\partial \left(\bar{\theta} n_b^* - \bar{\theta}_b^* \right)}{\partial n_b^*}}_{-} \right] \end{aligned} \tag{A.15}$$

Since in an evolutionary equilibrium $\Psi = 0$ (compare Proposition 4), the first term on the right hand side of (A.15) vanishes. As $s'(n) \leq 0$, the first term in the two squared brackets is *negative*. From the properties of the normal distribution follows that $\bar{\theta} n_j^* = \bar{\theta}_j^*$ for $n_j^* = 0$ and $n_j^* = 1$. Moreover, one can show that $\left(\bar{\theta} n_j^* - \bar{\theta}_j^* \right)$ reaches its maximum at $n_j^* = \frac{1}{2}$. Hence, we get

$$\frac{\partial \left(\bar{\theta} n_j^* - \bar{\theta}_j^* \right)}{\partial n_j^*} \gtreqless 0 \text{ for } n_j^* \lesseqgtr \frac{1}{2}. \tag{A.16}$$

For the case $n_a^* < 0.5 < n_b^* < 1$, this implies that the second term in the both squared brackets is positive, since in an evolutionary equilibrium as characterized in

Proposition 4 there holds $\lambda\, s(n_a^e) > c > \lambda\, s(n_b^e)$. If

$$-\lambda s'(n_j^e) < \frac{\lambda\, s(n_j^e) - c}{\bar{\theta} n_j^e - \bar{\theta}_j^e} \frac{\partial\left(\bar{\theta} n_j^e - \bar{\theta}_j^e\right)}{\partial n_j^e} \tag{A.17}$$

holds for $j \in \{a, b\}$ the sign of the two squared brackets is positive. Making use of (A.13) we get

$$-\lambda s'(n_j^e) < \frac{\lambda\, s(n_j^e) - c}{\bar{\theta} n_j^e - \bar{\theta}_j^e} \left(\bar{\theta} + \hat{\theta}(n_j^e)^2 \frac{s'(n_j^e)}{s(n_j^e)} \phi(\hat{\theta}(n_j^e))\right)$$

After some rearranging one arrives at the following two conditions

$$\begin{aligned} -s'(n_a^e) &< \Gamma_a \\ -s'(n_b^e) &> \Gamma_b \end{aligned}$$

with

$$\Gamma_j := \bar{\theta} \left(\frac{\lambda\left(\bar{\theta} n_j^e - \bar{\theta}_j^e\right)}{\lambda\, s(n_j^e) - c} + \frac{\hat{\theta}(n_j^e)^2}{s(n_j^e)} \phi(\hat{\theta}(n_j^e))\right)^{-1}. \tag{A.18}$$

■

Bibliography

Aitken, Sherie S. and Laura Bonneville (1980), A General Taxpayer Opinion Survey, Internal Revenue Service, Washington DC.

Akerlof, George A. (1980), 'A Theory of Social Custom, of Which Unemployment May be One Consequence', *The Quarterly Journal of Economics* 94(4), 749-775.

Allingham, Michael and Agnar Sandmo (1972), 'Income Tax Evasion: A Theoretical Analysis', *Journal of Public Economics* 1, 323-338.

Alm, James, Gary H. McClelland and William D. Schulze (1992), 'Why do people pay taxes?', *Journal of Public Economics* 48, 21-38.

Alm, James, Isabel Sanchez and Ana de Juan (1995), 'Economic and Noneconomic Factors in Tax Compliance', *Kyklos* 48, 3-18.

Andreoni, James, Brian Erard and Jonathan Feinstein (1998), 'Tax Compliance', *Journal of Economic Literature* 36, 818-860.

Arbak, Emrah (2005), 'Social status and crime', Mimeo, University of Lyon, Paper presented at the PET-Meeting 2005, Marseille.

Arbeitnehmerkammer Bremen (2001), 'Steuerliche Betriebsprüfung 1999 im Ländervergleich'.

Azar, Ofer H. (2004), 'What sustains social norms and how they evolve? The case of tipping', *Journal of Economic Behavior & Organization* 54(1), 49-64.

Baldry, Jonathan C. (1986), 'Tax Evasion is not a Gamble', *Economics Letters* 22, 333-335.

Bardsley, Peter (1994), 'Complexity, Honest Mistakes, and Tax Evasion', Working Papers 94-23, Department of Economics - La Trobe.

Baretti, Christian, Bernd Huber and Karl Lichtblau (2002), 'A tax on tax revenue. The incentive effects of equalizing transfers: Evidence from Germany', *International Tax and Public Finance* 9, 631-649.

Becker, Gary S. (1968), 'Crime and Punishment: An Economic Approach', *Journal of Political Economy* 76, 169-172.

Becker, Gary S. (1996), *Accounting for Tastes*, Harvard University Press, Cambridge, MA.

Ben-Shakhar, Gershon, Gary Bornstein, Astrid Hopfensitz and Frans van Winden (2004), 'Reciprocity and Emotions: Arousal, Self-Reports, and Expectations', Working Paper, CREED, University of Amsterdam.

Berg, Joyce, John Dickhaut and Kevin McCabe (1995), 'Trust, reciprocity and social history', *Games and Economic Behavior* 10, 122-142.

Bernasconi, Michele (1998), 'Tax evasion and orders of risk aversion', *Journal of Public Economics* 67(1), 123-134.

Bernasconi, Michele and Alberto Zanardi (2004), 'Tax Evasion, Tax Rates, and Reference Dependence' *Finanzarchiv* 60(4), 422-445.

Bernheim, Douglas B. (1994), 'A Theory of Conformity', *Journal of Political Economy* 102(5), 841-877.

Bernheim, Douglas B., and Antonio Rangel (2005), 'Behavioral Public Economics: Welfare and Policy Analysis with Non-Standard Decision-Makers', NBER Working Paper No. 11518.

Besley, Timothy (1988), 'A simple model for merit good arguments', *Journal of Public Economics* 35(3), 371-84.

Bester, Helmut and Werner Güth (1998), 'Is altruism evolutionary stable?' *Journal of Economic Behavior & Organization* 34, 193-209.

Bisin, Alberto, Giorgio Topa and Thierry Verdier (2004), 'Cultural Transmission, Socialization and the Population Dynamics of Multiple State Traits Distributions', Mimeo, NYU.

Bolton, Gary E. and Axel Ockenfels (2000), 'ERC: A Theory of Equity, Reciprocity, and Competition', *American Economic Review* 90(1), 166-193.

Border, Kim and Joel Sobel (1987), 'Samurai accountant: A theory of auditing and plunder', *Review of Economic Studies* 54, 525-540.

Bordignon, Massimo (1993), 'A fairness approach to income tax evasion', *Journal of Public Economics* 52(3), 345-362.

Bordignon, Massimo, Paolo Manasse and Guido Tabellini (2001), 'Optimal regional redistribution under asymmetric information', *American Economic Review* 91(3), 709-723.

Bosman, Ronald and Frans van Winden (2001), 'Anticipated and Experienced Emotions in an Investment Experiment', Discussion Paper 2001-058, University of Amsterdam.

Bosman, Ronald and Frans van Winden (2002), 'Emotional Hazard in a Power-to-Take Experiment', *Economic Journal* 112, 147-69

Bowles, Samuel and Herbert Gintis (2003), 'Prosocial Emotions', Working Paper, Sante Fe Institute.

Bucovestsky, Sam (1991), 'Asymmetric tax competition', *Journal of Urban Economic* 30, 167-181.

Bundesministerium der Finanzen (2003), 'Ergebnisse der Steuer- und Zollfahndung 2001', Bundesministerium der Finanzen, Berlin, Mai 2003.

Camerer, Colin F., Samuel Issacharoff, George Loewenstein, Ted O'Donoghue and Matthew Rabin (2003), 'Regulation for Conservatives: Behavioral Economics and the Case for Asymmetric Paternalism', *University of Pennsylvania Law Review* 151(3), 1211-1254 .

Chander, Parkash and Louis L. Wilde (1998), 'A general characterization of optimal income tax enforcement', *Review of Economic Studies* 65, 165-183.

Christiansen, Freddy Bugge (1975), 'Hard and soft selection in a subdivided population', *American Naturalist* 109, 11-16.

Cialdini, Robert B., Raymond. R. Reno and Carl A. Kallgren (1990), 'A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places', *Journal of Personality and Social Psychology* 58, 1015-1026.

Clotfelter, Charles T (1983), 'Tax Evasion and Tax Rates: An Analysis of Individual Returns', *The Review of Economics and Statistics* 65(3), 363-373.

Cohen, Mark A. (2000), 'Monitoring and enforcement of environmental policy', in: H. Folmer and T. Tietenberg (Eds.), *The International Yearbook of Environmental and Resource Economics 1999/2000*, Edward Elgar Publishers, Northampton, MA.

Cooter, Robert (1998), 'Expressive Law and Economics', *Journal of Legal Studies* 27, 585-608.

Corneo, Giacomo (1995), 'Social custom, management opposition, and trade union membership', *European Economic Review* 39(2), 275-292.

Cowell, Frank A. (1990), '*Cheating the Government*', MIT Press, Cambridge, MA.

Crabbé, Karen, Boudewijn Janssen and Hylke Vandenbussche (2004), 'Is there regional tax competition? Firm level evidence for Belgium', CEPR Discussion Paper No. 4721.

Cremer, Helmuth and Firouz Gahvari (1993), 'Tax evasion and optimal commodity taxation', *Journal of Public Economics* 50, 261-275.

Cremer, Helmuth and Firouz Gahvari (2000), 'Tax evasion, fiscal competition and economic integration', *European Economic Review* 44, 1633-1657.

Croson, Rachel (2005), 'Field Experiments in Charitable Contribution: The Impact of Social Influence on the Voluntary Provision of Public Goods', Working Paper, University of Pennsylvania.

Dawes, Robyn M., Jeanne McTavish and Harriet Shaklee (1977), 'Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation', *Journal of Personality and Social Psychology* 35(1), 1-11.

Dekel, Eddie, Jeffrey C. Ely and Okan Yilankaya (2004), 'Evolution of Preferences', Micro Theory Working Papers Dekel-04-08-13-01-21-07.

Dhami, Sanjit and Ali al-Nowaihi (2005), 'Why Do People Pay Taxes? Prospect Theory Versus Expected Utility Theory', Discussion Paper 05/23, University of Leicester.

Dufwenberg, Martin and Georg Kirchsteiger (1999), 'A Theory of Sequential Reciprocity', *Games and Economic Behavior* 47(2), 268-298.

Eide, Erling (2002), 'Tax evasion with rank dependent expected utility', Working Paper, Department of Private Law, University of Oslo.

Elster, Jon (1989a), 'Social Norms and Economic Theory', *Journal of Economic Perspectives* 3 (4), 99-117.

Elster, Jon (1989b), '*The Cement of Society: A study of social order*', Cambridge University Press, Cambridge, MA.

Elster, Jon (1998), 'Emotions and Economic Theory', *Journal of Economic Literature* 36(1), 47-74.

Erard, Brian and Jonathan S. Feinstein (1994), 'The role of moral sentiments and audit perceptions in tax compliance', *Public Finance/Finance Publiques* 49 (Supplement), 70-89.

Falconer, Douglas S. and Trudy F. Mackay (1995), '*Introduction to Quantitative Genetics*', 4th Edition. Addison Wesley Longman, New York.

Falk, Armin, Ernst Fehr and Urs Fischbacher (2005), 'Driving Forces Behind Informal Sanctions', *Econometrica*, forthcoming.

Falk, Armin and Urs Fischbacher (2005), 'A Theory of Reciprocity', *Games and Economic Behavior*, forthcoming.

Falk, Armin and Michael Kosfeld (2004), 'The Hidden Costs of Control', Institute for Empirical Research in Economics, Working Papers 250, University of Zurich.

Falkinger, Josef (1995), 'Tax evasion, consumption of public goods and fairness', *Journal of Economic Psychology* 16(1), 63-72.

Fehr, Ernst and Armin Falk (2002), 'Psychological Foundations of Incentives', *European Economic Review* 46, 687-724.

Fehr, Ernst and Urs Fischbacher (2004a), 'Social norms and human cooperation', *Trends in Cognitive Sciences* 8(4), 185-190.

Fehr, Ernst and Urs Fischbacher (2004b), 'Third Party Punishment and Social Norms', *Evolution and Human Behavior* 25, 63-87.

Fehr, Ernst and Simon Gächter (2000), 'Cooperation and Punishment in Public Goods Experiments', *American Economic Review* 90(4), 980-994.

Fehr, Ernst and Klaus Schmidt (1999), 'A Theory of Fairness, Competition and Cooperation', *The Quarterly Journal of Economics* 114, 817-868.

Feld, Lars P. (1997), 'Exit, Voice and Income Taxes', *European Journal of Political Economy* 13, 455-478.

Feld, Lars P. and Bruno S. Frey (2002), 'Trust Breeds Trust: How Taxpayers Are Treated', CESifo Working Paper No. 322.

Feld, Lars P. and Jean-Robert Tyran (2002), 'Tax Evasion and Voting: An Experimental Analysis', *Kyklos* 55(2), 197-222.

Fischbacher, Urs, Simon Gächter and Ernst Fehr (2001), 'Are People Conditionally Cooperative? Evidence from a Public Goods Experiment', *Economics Letters* 71, 397-404.

Frey, Bruno (1997), 'A Constitution for Knaves Crowds Out Civic Virtues', *Economic Journal* 107, 1043-53.

Frey, Bruno and Lars Feld (2002), 'Deterrence and Morale in Taxation: An Empirical Analysis', CESifo Working Paper No. 760.

Frey, Bruno S. and Felix Oberholzer-Gee (1997), 'The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-out', *American Economic Review* 87, 746-755.

Frey, Bruno S. and Stephan Meier (2004a), 'Pro-social behavior in a natural setting', *Journal of Economic Behavior & Organization* 54, 65-88.

Frey, Bruno S. and Stephan Meier (2004b), 'Social Comparisons and Pro-social Behavior: Testing 'Conditional Cooperation' in a Field Experiment', *American Economic Review* 94(5), 1717-1722.

Gächter, Simon (2005), 'Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications', paper presented at the CESifo Workshop on 'Economics and Psychology', Venice Summer Institute 2005.

Gächter, Simon and Elke Renner (2005), 'Leading by Example in the Presence of Free Rider Incentives', Mimeo, University of Nottingham.

Galbiati, Roberto and Pietro Vertova (2005), 'Law and Behaviours in Social Dilemmas: Testing the Effect of Obligations on Cooperation', Working Paper, Department of Economics, University of Siena.

Gebauer, Andrea and Rüdiger Parsche (2003), 'Evasion of value-added taxes in Europe', *DICE Report: Journal for Institutional Comparisons* 1(2), 40-44.

Geeroms, Hans and Hendrik Wilmots (1985), 'An Empirical Model of Tax Evasion and Tax Avoidance', *Public Finance/Finances Publiques* 40(2), 190-209.

Gërxhani, Klarita and Arthur Schram (2005), 'Tax Evasion and the Source of Income: An Experimental Study in Albania and the Netherlands', *Journal of Economic Psychology*, forthcoming.

Gilbert, Daniel T., Erin Driver-Linn and Timothy D. Wilson (2002), 'The trouble with Vronsky: Impact bias in the forecasting of future affective states', 114-143, in: L.F. Barrett and P. Salovey (Eds.), *The wisdom in feeling: Psychological processes in emotional intelligence*, Guilford Press, New York.

Gintis, Herbert (2003), 'The Hitchhikers Guide to Altruism: Gene-culture Coevolution, and the Internalization of Norms', *Journal of Theoretical Biology* 220, 407-418.

Glaeser, Edward L., Bruce Sacerdote and Jose A. Scheinkman (1996), 'Crime and Social Interaction', *The Quarterly Journal of Economics* 106, 507-548.

Gneezy, Uri and Aldo Rustichini (2000), 'A Fine is a Price', *Journal of Legal Studies* 29(1), 1-18.

Gordon, James P.F. (1989), 'Individual Morality and Reputation Costs as Deterrents to Tax Evasion', *European Economic Revue* 33, 797-805.

Graetz, Michael J. and Louis L. Wilde (1985), 'The economics of tax compliance: Facts and fantasy', *National Tax Journal* (38), 355-363.

Grasmick, Herold G. and Donald E. Green (1980), 'Legal punishment, social disapproval and internalization as inhibitors of illegal behavior', *Journal of Criminal Law and Criminology* 71, 325-35.

Güth, Werner and Rupert Sausgruber (2004), 'Tax Morale and Optimal Taxation', CESifo Working Paper No. 1284.

Güth, Werner, Maria Vittonia Levati and Rupert Sausgruber (2005), 'Tax Morale and (De-)Centralization: An Experimental Study', *Public Choice*, forthcoming.

Güth, Werner (1995), 'An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives,' *International Journal of Game Theory* 24, 323-344.

Güth, Werner and Menahem Yaari (1992), 'An Evolutionary Approach explain reciprocal behavior in a simple strategic game', in: U. Witt (Ed.), *Explaining Process and Change – Approaches to Evolutionary Economics*, University of Michigan Press, Ann Arbor.

Harsanyi, John C. (1982), 'Morality and the theory of rational behaviour', in: A. Sen and B. Williams (Eds.), *Utilitarianism and Beyond*, Cambridge University Press, Cambridge, MA.

Holländer, Heinz (1990), 'A social exchange approach to voluntary cooperation', *American Economic Review* 80(5), 1157-1167.

Hopfensitz, Astrid and Frans van Winden (2005), 'On the Interaction between Investment and Emotions: An Experimental Study', Mimeo, CREED, University of Amsterdam.

Kahan, Dan (1997), 'Social Influence, Social Meaning and Deterrence', *Virginia Law Review* 83(3), 349-395.

Kahan, Dan (2005), 'The Logic of Reciprocity: Trust, Collective Action and Law', in: Herbert Gintis, Samuel Bowles, Robert T. Boyd, Ernst Fehr (Eds.), *The Moral Sentiments: Origins, Evidence, and Consequences*, MIT Press, Cambridge, MA.

Kahneman, Daniel (2000), 'Experienced Utility and Objective Happiness: A Moment-Based Approach', in: D. Kahneman and A. Tversky (Eds.), *Choices, Values and Frames*, Cambridge University Press, New York.

Kanbur, Ravi and Keen, Michael (1993), 'Jeux sans frontières: Tax competition and tax coordination when countries differ in size', *American Economic Review* 83(4), 877-892.

Kanbur, Ravi, Jukka Pirttilä and Matti Tuomala (2004), 'Non-Welfarist Optimal Taxation and Behavioral Public Economics', CESifo Working Paper No. 1291.

Kaplow, Louis and Steven Shavell (2000), 'Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle', *Journal of Political Economy* 112(1), 249-251.

Kelley, Harold H. and Anthony J. Stahelski (1970), 'The Social Interaction Basis of Cooperators' and Competitors' Beliefs about Others', *Journal of Personality and Social Psychology* 16, 66-91.

Keser, Claudia and Frans van Winden (2000), 'Conditional Cooperation and Voluntary Contributions to Public Goods', *Scandinavian Journal of Economics* 102, 23-39.

Knight, John and Li, Shi (1999), 'Fiscal decentralization: Incentives, redistribution and reform in China', *Oxford Development Studies* 27, 5-32.

Kolm, Serge-Christophe (1973), 'A note on optimal tax evasion', *Journal of Public Economics* 2, 265-270.

Körner, Martin and Harald Strotmann (2005), 'Steuermoral - Das Spannungsfeld von Freiwilligkeit der Steuerzahlung und Steuerhinterziehung', Monatsbericht des BMF, Berlin, März 2005, 49-57.

Köthenbürger, Marko (2002), 'Tax competition and fiscal equalization', *International Tax and Public Finance* 9, 391-408.

Kreps, David M. (1997), 'Intrinsic Motivation and Extrinsic Incentives', *American Economic Review* 87(2), Pap&Proc., 359-64

Lande, Russell (1976), 'Natural selection and random genetic drift in phenotypic evolution', *Evolution* 30, 314-334.

Lenk, Thomas, Heidi Fuge and Friedrich Schneider (1998), 'Zurück zu mehr Föderalismus: Ein Vorschlag zur Neugestaltung des Finanzausgleichs in der BRD unter besonderer Berücksichtigung der ökonomischen Theorie der Politik', Working Paper 9805, Department of Economics, Johannes Kepler University Linz.

Levins, Richard (1968), '*Evolution in changing environments*', Princeton University Press, Princeton.

Lewis, Alan (1986), 'Fiscal Policy: The Importance of Perceptions and Attitudes', in: B. Gilad and S. Kaish (Eds.), *The Handbook of Behavioural Economics*, JAI Press, New York.

Lindbeck, Assar, Sten Nyberg and Jörgen W. Weibull (1999), 'Social Norms and Economic Incentives in the Welfare State', *The Quarterly Journal of Economics* 114, 1-35.

Lindbeck, Assar, Sten Nyberg and Jörgen W. Weibull (2002), 'Social Norms and Welfare State Dynamics', *Journal of the European Economic Association* 1, 533-542.

Liu, Ruth Xiaoru (2003), 'The Moderating Effects of Internal and Perceived External Sanction Threats on the Relationship between Deviant Peer Associations and Criminal Offending', *Western Criminology Review* 4(3), 191-202.

Loewenstein, George and David Schkade (1999), 'Wouldn't it be nice? Predicting future feelings', 85-105, in: D.Kahneman, E. Diener, and N. Schwarz (Eds.), *Well-Being: The Foundations of Hedonic Psychology*, Russell Sage Foundation Press, New York.

Masclet, David, Charles Noussair, Steven Tucker and Marie-Claire Villeval (2003), 'Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism', *American Economic Review* 93(1), 366-380.

Marrelli, Massimo (1984), 'On Indirect Tax Evasion', *Journal of Public Economics* 25, 181-196.

Matsusaka, John G. (2005), 'Direct Democracy Works', *Journal of Economic Perspectives* 19(2), 185-206.

Maynard Smith, John and Rolf Hoekstra (1980), 'Polymorphism in a varied environment: How robust are the models?' *Genetical Research* 35, 45-57.

Meltzer, Allan H. and Scott F. Richard (1981), 'A rational theory of the size of government', *Journal of Political Economy* 89, 914-927.

Mengel, Friederike (2005), 'A Model of Immigration, Integration and the Cultural Transmission of Social Norms', Mimeo, University of Alicante.

Mookherjee, Dilip and Png, Ivan (1990), 'Enforcement Costs and the Optimal Progressivity of Income Taxes', *Journal of Law, Economics and Organization* 6(2), 410-431.

Musgrave, Richard A. (1990), 'Horizontal Equity, Once More', National Tax Journal 43, 113-122.

Myerson, Roger B., Gregory B. Pollock and Jeroen M. Swinkels (1991), 'Viscous Population Equilibria', *Games and Economic Behaviour* 3, 101-109.

Myles, Gareth D. and Robin A. Naylor (1996), 'A Model of Tax Evasion with Group Conformity and Social Customs', *European Journal of Political Economy* 12, 49-66.

Naylor, Robin (1989), 'Strikes, Free Riders, and Social Customs', *The Quarterly Journal of Economics* 104(4), 771-785.

Nyborg, Karine and Mari Rege (2003), 'On social norms: the evolution of considerate smoking behavior', *Journal of Economic Behavior & Organization* 52(3), 323-340.

O'Donoghue, Ted and Matthew Rabin (2003), 'Studying optimal paternalism, illustrated by a model of sin taxes', *American Economic Review* 93(2), 186-191.

Oates, Wallace (1972), '*Fiscal federalism*', Harcourt Brace Jovanovich, New York.

Orviska, Marta and John Hudson (2002), 'Tax evasion, civic duty and the law abiding citizen', *European Journal of Political Economy* 19(1), 83-102.

Ostrom, Ellinor (1998), ‘A Behavioral Approach to the Rational Choice Theory of Collective Action’, *American Political Science Review* 92, 1-23.

Persson, Torsten and Guido Tabellini (2000), ‘*Political Economics: Explaining Economic Policy*’, MIT Press, Cambridge.

Pigliucci, Massimo and Carl D. Schlichting (1997), ‘On the limits of quantitative genetics for the study of phenotypic evolution, *Acta Biotheoretica* 45, 143-160.

Polinsky, Mitchell A. and Steven Shavell (2000), ‘The Economic Theory of Public Enforcement of Law’, *Journal of Economic Literature* 38(1), 45-76.

Pommerehne, Werner W. und Hannelore Weck-Hannemann, (1996), ‘Tax Rates, Tax Administration and Income Tax Evasion in Switzerland’, *Public Choice* 88, 161-170.

Porcano, Thomas M. (1988), ‘Correlates of Tax Evasion’, *Journal of Economic Psychology* 9, 47-67.

Posner, Richard A. (1997), ‘Social Norms and the Law: An Economic Approach’, *American Economic Review* 87(2), Pap&Proc., 365-369.

Posner, Richard A. (2000), ‘*Law and Social Norms*’, Harvard University Press, Cambridge.

Rabin, Matthew (1993), ‘Incorporating Fairness into Game Theory and Economics’, *American Economic Review* 83, 1281-1302.

Rasmusen, Eric (1996), ‘Stigma and Self-Fulfilling Expectations of Criminality’, *Journal of Law and Economics* 39, 519-541.

Reckers, Philip M., Debra L. Sanders, and Stephen J. Rork (1994), ‘The Influence of Ethical Attitudes on Taxpayer Compliance’, *National Tax Journal* 47(4), 825-836.

Rege, Mari (2004), ‘Social Norms and Private Provision of Public Goods’, *Journal of Public Economic Theory* 6(1), 65-77.

Rege, Mari and Kjetil Telle (2003), ‘Indirect Social Sanctions from Monetarily Unaffected Strangers in a Public Good Game’, Discussion Papers 359, Research Department of Statistics Norway.

Rege, Mari and Kjetil Telle (2004), ‘The Impact of Social Approval and Framing on Cooperation in Public Good’, *Journal of Public Economics* 88(7-8), 1625-1644.

Reinganum, Jennifer F and Louis L. Wilde (1986), 'Equilibrium Verification and Reporting Policies in a Model of Tax Compliance', *International Economic Review* 27(3), 739-760.

Reno, Raymond. R., Robert B. Cialdini and Carl A. Kallgren (1993), 'The Transsituational Influence of Social Norms', *Journal of Personality and Social Psychology* 64(1), 104-112.

Reuben, Ernesto and Frans van Winden (2005), 'Negative Reciprocity and the Interaction of Emotions and Fairness Norms', Mimeo, CREED, University of Amsterdam.

Roberts, Kevin W.S. (1977), 'Voting over income tax schedules', *Journal of Public Economics* 8, 329-340.

Roff, Derek A. (1997), '*Evolutionary Quantitative Genetics*', Chapman and Hall, New York.

Rothstein, Bo (2000), 'Trust, Social Dilemmas and Collective Memories', *Journal of Theoretical Politics* 12(4), 477-501.

Romer, Thomas (1975), 'Individual welfare, majority voting and the properties of a linear income tax', *Journal of Public Economics* 4, 163-185.

Romer, David (1984), 'The Theory of Social Custom: A Modification and Some Extensions', *The Quarterly Journal of Economics* 99(4), 717-27.

Sah, Raaj K. (1991), 'Social Osmosis and Patterns of Crime', *Journal of Political Economy* 99(6), 1272-95.

Sandmo, Agnar (2004), 'The theory of tax evasion: A retrospective view', Discussion Paper 31/04, Norwegian School of Economics and Business Administration.

Sawicky, Max B. (2005), 'Do-It-Yourself-Tax-Cuts: The Crisis in U.S. Tax Enforcement', E*conomic Policy Institute Briefing Paper* 160.

Schelling, Thomas (1960), '*The Strategy of Conflict*', Harvard University Press, Cambridge.

Schneider, Friedrich and Dominik Enste (2000), 'Shadow economies: Size, causes and consequences', *Journal of Economic Literature* 38(1), 77-114.

Schnellenbach, Jan (2002), 'Tax Morale, Leviathan and the Political Process: A Theoretical Approach', Paper presented at the Royal Economic Society Annual Conference 2002.

Scholz, John T. and Mark Lubell (1998), 'Trust and Taxpayers: Testing the Heuristic Approach to Collective Action', *American Journal of Political Science* 42(2), 398-417.

Scholz, John T. and Neil Pinney (1995), 'Duty, Fear, and Tax Compliance: The Heuristic Basis of Citizenship Behavior', *American Journal of Political Science* 39, 490-512.

Seidl, Christian and Stefan Traub (2001), 'Taxpayers' attitudes, behavior, and perception of fairness', *Pacific Economic Review* 6(2), 255-267.

Shavell, Steven (2003), 'Economic Analysis of Welfare Economics, Morality and the Law', NBER Working Paper 9700.

Sheffrin, Steven M. and Rober K. Triest (1992), 'Can Brute Deterrence Backfire? Perceptions and Attitudes in Taxpayer Compliance', in: J. Slemrod (1992).

Skinner, Jon and Slemrod, Joel (1985), 'An Economic Perspective on Tax Evasion', *National Tax Journal* 38, 345-353.

Slemrod, Joel (1992), (Ed.), '*Why People Pay Taxes: Tax Compliance and Enforcement*', University of Michigan Press, Ann Arbor.

Slemrod, Joel and Yitzhaki, Shlomo (1987), 'The optimal size of a tax collection agency', *Scandinavian Journal of Economics* 89(2), 183-192.

Sonnemans, Joep, Arthur Schram and Theor Offerman (1999), 'Strategic behavior in public good games: when partners drift apart', *Economic Letters* 62, 35-41.

Spicer, Michael W. and Rodney E. Hero (1985), 'Tax Evasion and Heuristics', *Journal of Public Economics* 26, 263-267.

Torgler, Benno (2002), 'Speaking to Theorists and Searching for Facts: Tax Morale and Tax Compliance in Experiments', *Journal of Economic Surveys* 16, 656-683.

Torgler, Benno (2005), '*Tax Compliance and Tax Morale: A Theoretical and Empirical Analysis*', Edward Elgar, forthcoming.

Tyler, Tom R (1990), '*Why People Obey Law*', Yale University Press, New Haven.

Tyran, Jean-Rober and Lars Feld (2005), 'Achieving Compliance when Legal Sanctions are Non-Deterrent', *Scandinavian Journal of Economics*, forthcoming.

Ullmann-Margalit, Edna (1977), '*The Emergence of Norms*', Oxford University Press, Oxford.

van Winden, Frans (2001), 'Emotional hazard exemplified by taxation-induced anger', *Kyklos* 54, 491-506.

Via, Sara and Russell Lande (1987), 'Evolution of genetic variability in a spatially heterogenous environment: Effects of genotype-environment interaction', *Genetical Research* 49, 147-156.

Vega-Redondo, Fernando (1996), '*Evolution, Games and Economic Behavior*', Oxford University Press, Oxford.

Wang, Leonard F.S. and John L. Conant (1988), 'Corporate Tax Evasion and Output Decisions of the Uncertain Monopolist', *National Tax Journal* 41, 579-581.

Weck-Hannemann, Hannelore and Werner W. Pommerehne (1989), 'Einkommensteuerhinterziehung in der Schweiz: Eine empirische Analyse', *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* 125, 515-556.

Weibull, Jörgen W. (1995), '*Evolutionary Game Theory*', MIT Press, Cambridge.

Wildasin, David (1989), 'Interjurisdictional capital mobility: Fiscal externality and a corrective subsidy', *Journal of Urban Economics* 25, 193-212.

Wilson, James Q. (1993), '*The Moral Sense*', Free Press, New York.

Wilson, James and George Kelling (1982), 'Broken Windows', *The Atlantic Monthly* 249, 29-38.

Wilson, Timothy D., and Daniel T. Gilbert (2003), 'Affective forecasting', *Advances in experimental social psychology* 35, 345-411.

Yaniv, Gideon (1988), 'Withholding and Non-withheld Tax Evasion', *Journal of Public Economics* 35, 183-204.

Yitzhaki, Shlomo (1974), 'A note on income tax evasion: A theoretical analysis', *Journal of Public Economics* 3, 201-202.

Young, Peyton H. (1998a), 'Social Norms and Economic Welfare', *European Economic Review* 42, 821-830.

Young, Peyton H. (1998b), '*Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*', Princeton University Press, Princeton, NJ.

Zodrow, George and Peter Mieszkowski (1986), ‘Pigou, Tiebout, property taxation, and the underprovision of local public goods’, *Journal of Urban Economics* 19, 356-370.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebenen Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

München, 23. September 2005

Christian Traxler

Curriculum Vitae

February 2006	Ph.D. in economics (Dr. oec. publ.) Ludwig-Maximilians-Universität, Munich
November 2004 – May 2005	Marie-Curie Fellow CREED, Universiteit van Amsterdam
October 2002 – September 2005	Ph.D. program in economics Munich Graduate School of Economics Ludwig-Maximilians-Universität, Munich
October 2001	Diploma in economics Universität Wien, Vienna
October 2000 - June 2001	Visiting student in economics Universidad Carlos III de Madrid
October 1995 - October 2001	Studies in economics Universität Wien, Vienna
October 1987 - June 1995	High School Bundesrealgymnasium Steyr
21 February 1977	born in Steyr, Austria