TOPICS IN BEHAVIORAL ECONOMICS:
COOPERATION IN SOCIAL DILEMMAS AND INTERTEMPORAL DECISION-MAKING

INAUGURALDISSERTATION
ZUR
ERLANGUNG DES DOKTORGRADES
DER
WIRTSCHAFTS- UND SOZIALWISSENSCHAFTLICHEN FAKULTÄT
DER
UNIVERSITÄT ZU KÖLN

2015

vorgelegt
von
Karen Charlotte Heuermann, M.Sc.
aus
Bonn

Referentin: Prof. Dr. Bettina Rockenbach

Korreferent: Ph.D. Mattia Nardotto

Korreferentin: Jun.-Prof. Dr. Marina Schröder

Datum der Promotion: 19.01.2015

Acknowledgements

First and foremost, many special thanks go to my supervisor, Prof. Bettina Rockenbach for the support and advice during my PhD. I highly enjoyed all our fruitful discussions and the collaboration in our joint projects.

I also deeply thank my second supervisor Mattia Nardotto for his continuous support and help. I have strongly benefitted from his expertise and advice. I would not have made it without him.

A special thanks goes to Jun.-Prof. Dr. Marina Schröder for reviewing my thesis.

I would like to thank the Cologne Graduate School in Management, Economics and Social Sciences for the financial support during the last three years. I also want to thank all my colleagues at the CGS for always being helpful and fun to be around. A special thank goes to Jane, for being the best office buddy in the world.

Finally, I am heartily thankful to my family; my parents Beate and Arnulf, my sisters Linda and Eva and Bastian for their support, patience and confidence, for their encouragement and constant belief in my academic potential.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Traditional (neoclassical) economic theory predicts that individuals are solely interested in the maximization of their own monetary gains and their behavior is exclusively driven by the pursuit of their own narrow self-interest. However, vast evidence of experimental and empirical studies shows that behavior often deviates from this assumption. The narrow self-interest hypothesis started to be questioned in the 1980's when experimental economists began to study bilateral bargaining games and interactions in small groups in controlled laboratory settings (i.e. Roth et al., 1981; Güth et al., 1982). In recent years, numerous models of decisions assume that utility is not exclusively influenced by monetary gains but also take concerns for other players into account (i.e. Rabin, 1993; Levine, 1998; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000; Charness & Rabin, 2002; Dufwenberg & Kirchsteiger, 2004; Sobel, 2005; Fehr & Schmidt, 2006).

The core of behavioral economics is the refinement of the assumption of perfectly self-interested, rationally utility-maximizing agents and the extension of traditional economic models by integrating insights from psychology. The concept of homogenously purely self-interested agents is changed by assuming heterogeneity in interpersonal preferences among players and allowing for social preferences for some agents. Social – or other-regarding – preferences indicate that people do not solely care about themselves, but are also concerned about the payoff of others. Therefore, behavioral economics develop more realistic theories and make better predictions about behavior and, hence, increases the predictive power of economic models (Rabin, 2002; Camerer & Loewenstein, 2004).

Broadly speaking, there are three fields of "behavioral anomalies", i.e. deviations from standard theory, in a decision-making process (for a survey, see DellaVigna, 2009).[1] First, other factors than the maximization of one's own absolute payoff might influence utility (non-standard preferences). Second, subjective perceptions of the prospects of certain events might be systematically biased (non-standard beliefs). And third, economic agents might have cognitive limitations in the decision-making process and therefore use simple heuristics, or framing effects and emotions might influence their decisions (non-standard decision-making).

[1] Alternative surveys of studies in behavioral economics are provided i.e. by Earl (1990), Rabin (1998), Kahnemann (2003), and Camerer & Loewenstein (2004).

This thesis focuses on the first class of behavioral phenomena – non-standard preferences – and its application to two different fields of economics. More specifically, we scrutinize potential influences of non-standard preferences on contribution behavior in a social dilemma (chapter 2), cooperation in intertemporal decision-making (chapter 3) and the effect of responsibility on intertemporal choices (chapter 4).

Social preferences can be divided into three different types (Fehr & Schmidt, 2006). The first type suggests that an individual cares more about the benefit of another person than his own and is captured by models of altruism (Levine, 1998). Another sort of social preferences assumes a feeling of covetousness with respect to another person's advantages, which is expressed by models of envy (Bolton, 1991). The third type of social preferences assumes that a person does not only take his own absolute payoff into account but also his relative payoff in comparison to a given reference group. These distributive preferences are captured by models of inequality aversion (Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000).[2] Inequality aversion – or fairness concerns – is particularly important for our analysis of contribution behavior in a public goods game (chapter 2). Moreover, there are many theories that postulate that subjects have a concern for the perceived intentions behind economic actions, which are modeled by theories of reciprocity (see Rabin, 1993; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006). When games are repeated, reciprocity leads to the behavioral pattern of conditional cooperation. In line with strong evidence that reciprocity plays a crucial role for the evolution of contribution behavior in public goods games (for a survey see Gächter, 2006), we find that it influences the willingness of pre-commitment to (higher) public goods contributions (Chapter 2).

Second, we focus on intertemporal decision-making in (different aspects of) a social context in chapter 3 and 4. In both chapters, we analyze whether the interaction with another person has implications for intertemporal consumption decisions. By suggesting an experimental design for the joint elicitation of time and social preferences (Chapter 3) as well as for the integration of responsibility in an intertemporal choice task (Chapter 4), both chapters contribute to the understanding of behavioral regularities that might govern the choices in an intertemporal context with social interaction.

Chapter 3 investigates how individual time preferences interact with social preferences. Behavior is analyzed in an intertemporal trust game. Evidence has shown that trusting another person is fundamentally different from the assessment of stochastic risk generating the same monetary outcomes since people dislike being betrayed by another

[2] Other models also contain social efficiency concerns (see i.e. Charness & Rabin, 2002).

person as compared to a setting where chance (nature) determines the outcome (Bohnet & Zeckhauser, 2004; Bohnet et al., 2008). Betrayal aversion is in line with the theoretical and empirical evidence that people care about others' intentions. Our results confirm the existence of betrayal aversion in an intertemporal decision problem. Moreover, our main result shows that the social context has implications for intertemporal consumption decisions: Social preferences seem to outweigh individual problems of self-control when both preferences are elicited jointly (Chapter 3).

Finally, Chapter 4 analyzes the effect of responsibility for another person who cannot actively take part in the decision process but silently bears the consequences of the decision maker's choices on intertemporal decision-making. In our context, the decision maker does not know the time preferences of the dependent. Our results suggest that the decision maker cares for the preferences of the second person and integrates them into his utility function (Chapter 4).

All three research projects that are part of this thesis study the importance of the described behavioral regularities for consumption decisions in various economic settings. The starting point for each research project is the assumption that economic decision-making in the respective environments might be altered by the existence of non-standard preferences. In each chapter of this thesis, the main research method to test for these assumptions and be able to observe economic decision-making in the different contexts is the implementation of a laboratory experiment.[3] However, the studies differ in the nature of the behavioral anomalies that are investigated, the exact experimental setting that is applied to study the anomaly, and its application. Below, all studies are summarized by a brief overview of their research questions and main results.

In the second chapter "Institution Formation in Public Goods Games with a Decentralized Self-Commitment Mechanism", I explore behavioral impacts of self-commitment possibilities for the provision of a public good. The introduction of sanctioning schemes is a widely used solution to enforce cooperation in social dilemma situations. However, enforcement problems remain a core difficulty for the employment of punishment devices in a public goods context. The aim of this chapter is to analyze the endogenous formation of a sanction institution by explicitly addressing potential enforcement problems. To do so, a self-sanctioning punishment scheme is introduced that relies on voluntary ex-ante deposit lodgments that give players the possibility to (pre-)commit to contribution. The

[3] The methodological foundation of behavioral economics is experimental economics (Loewenstein, 1999). Next to laboratory experiments also field experiments are an established method to study behavioral pattern (Harrison & List, 2004; Levitt & List, 2009).

workhorse is the three-stage non-cooperative game proposed by Kosfeld et al. (2009) who examine whether players would voluntary join an institution that exerts a punishment among its members. I modify the sanction mechanism of this game, such that the punishment is not executed by a centralized institution ex-post but players regulate their punishment ex-ante in a decentralized way. The experimental results show that the vast majority of players voluntarily pre-commit to contributions and the possibility of institution formation positively influences cooperation and group welfare. Even though the pre-commitment device is independent of group contributions, players repudiate free-riding behavior on the institution. Therefore, the results stress the importance of fairness considerations for the establishment and steadiness of institutions, which is in line with findings from the common pool resource literature (Baland & Platteau, 1996; McKean, 2000) and also in the context of public choice research (Buchanan & Congleton, 1998).

Next to addressing possible enforcement problems with respect to public goods provision, this chapter tackles another problems recently discussed in the enforcement literature. Players can use the mechanism as a pre-commitment device to overcome self-control problems. Recent studies analyze whether next to the tension between the individual rationality of self-interest and collective wellbeing, a social dilemma might also represent a *self-control dilemma,* which is defined as an interpersonal conflict between "better judgment" and "temptation" (Thaler & Shefrin, 1981; Schelling, 1984; Loewenstein, 1996, 2000). The provision of public goods may impose an internal conflict on the individual who has to trade off the objective of maintaining the better judgment to act pro-socially (cooperation) and the temptation to act selfishly (Kocher et al., 2012; Martinsson et al., 2012). A suggested solution to self-control problems is to offer people to bind their behavior (Strotz, 1956; Thaler & Shefrin, 1981; Schelling, 1992; Wertenbroch, 1998).[4] So far, empirical studies on self-control strategies have found that people, when giving the opportunity, seem to use pre-commitment devices strategically in the anticipation of self-regulation failures (Wertenbroch, 1998; Trope & Fishbach, 2000; Ariely & Wertenbroch, 2002). Our experimental results are in line with these findings by suggesting that players bind their behavior ex-ante and use the deposit lodgments as a strategic device to overcome self-control problems. The underlying assumption, however, is that the social dilemma displays an individual self-control dilemma. Another possibility is that subjects simply prefer to introduce a sanction institution compared to a sanction-free environment in order to enforce discipline among the other players, which is also in line with previous findings in the literature that incorporates endogenous

[4] Binding behavior is characterized by a strategic attempt to resist future temptations by voluntarily imposing (costly) constraints on future choices (Ariely & Wertenbroch, 2002).

institutional choice into public good games (Gürerk et al., 2006; Tyran & Feld, 2006; Kosfeld et al., 2009).

There are recent studies that analyze the correlation between public goods contribution and self-control problems and those studies report experimental evidence that impatient individuals contribute less to the public good than patient ones (Curry et al., 2008; Fehr & Leibbrandt, 2011, Kocher et al., 2012) and that self-control is positively associated with cooperation (Martinsson et al., 2012). However, these studies measure the individual degree of self-control independently from the contribution decision in the public good game, where the choice tasks to elicit time preferences consist of lotteries that embrace payoffs that differ very much from those in the cooperation game. Most importantly, the typical choice task to elicit time preferences does not involve a second person. Rather, the decision maker's choices only affect his own payoff and therefore his decisions are made in a socially detached environment. However, if social dilemmas represent a self-control conflict, any choice involves *both* a temporal as well as a social dimension. Other research fields show that behavior changes when individual preferences interact with social preferences, such as a shift towards more conservative risk behavior when decisions are made in a social context than when made individually (Song, 2008; Charness & Jackson, 2009; Reynolds et al., 2009) or less liability towards overconfidence (Sniezek & Henry, 1989; Plous, 1995).

In the third chapter "Can´t Wait to be Nice: Self-Control and Cooperation in an Intertemporal Trust Game" we address this problem by analyzing the interlinkage of social and time preferences. The central research question is whether there is an interaction between intertemporal preferences and cooperative behavior, as suggested by the literature. We propose a new experimental design to study the interplay of intertemporal and social preferences based on joint elicitation. We modify a standard intertemporal choice task by replacing the lotteries with a regular trust game. We compare behavior in such intertemporal trust games to an intertemporal lottery task that does not have a social component. Our experimental results do not confirm that present-biased individuals cooperate less than time-consistent ones. Rather, when both dimensions are elicited jointly, we find that individuals that display problems of self-control when only deciding for themselves, behave similar to time-consistent individuals in the intertemporal trust game. Hence, our results suggest that the social component outweighs the temporal aspect when both preferences interact. Moreover, we find some descriptive evidence that opens the question of how responsibility might influence intertemporal choices.

Chapter 4 "Responsibility Effects in Intertemporal Decision Making" addresses this question by studying the influence of responsibility on intertemporal decision-making. We offer a novel experimental setting, where we introduce responsibility into a standard intertemporal decision problem. The decision maker is a principal who determines both his own payoff as well as the payoff of a passive dependent who is not taking part in the decision problem but (silently) observes the principal's behavior. We then compare behavior with the case where the same decisions only determine the payoff of the decision maker. Our results show that the decisions made with responsibility reflect more impatience. This finding is in contrast to predictions from the psychological literature but is consistent with a model, where the principal aggregates utility by a weighted utility function that is based on the principal's estimation of the dependent's time preferences if the principal is overconfident about his own degree of patience.

Summing up, all three chapters confirm the importance of social preferences in human decision-making. Social preferences have an influence on collective action problems (public goods provision), as well as intertemporal decision-making.

In the following chapters, each of the three studies is presented in more detail. Finally, the last chapter summarizes the main results of all three chapters and briefly discusses their implications.

Chapter 2

Institution Formation in a Public Goods Game with a Decentralized Self-Commitment Mechanism

*Karen Heuermann**

2.1 Introduction

Social dilemma situations, i.e. the provision of public goods or common pool resources depict a very prominent example for market failures. In such dilemmas, individual interest collides with societal concerns and thus no socially optimal outcome emerges. A widely used solution is the introduction of a sanction mechanism that has the authority to punish individually deviant behavior. The literature that incorporates endogenous institutional choice into public good games agrees that subjects prefer to introduce a sanction institution compared to a sanction-free environment in order to enforce discipline among players (Tyran & Feld, 2006; Gürerk et al., 2006; Kosfeld et al., 2009).[5] More importantly, the option to form an institution positively affects efficiency (Walker et al., 2000; Tyran & Feld, 2006; Kosfeld et al., 2009; Sutter et al, 2010; Dal Bó et al. 2010).

Kosfeld et al. (2009) introduce an "endogenous institution formation game" (p. 1338), where they examine whether players would voluntary join an institution that exerts a punishment mechanism among its members. To do so, the institution formation is modeled by a three-stage non-cooperative game. In the first stage, players can vote on participating in a costly organization that punishes each member for contributing less than his full endowment.[6] In the second stage, those who voted for participation learn the number of participants and then decide whether the organization shall be implemented, where a unanimous agreement is required. The third stage comprises the contribution decision. Since the authors focus on the formation process of the institution, they abstract from any enforcement problems by letting a

* Financial support from the Deutsche Forschungsgemeinschaft (through the Gottfried Wilhelm Leipnitz Price of the DFG, awarded to Axel Ockenfels) is gratefully acknowledged.

[5] An exception is the study of Bothelo et al. (2005) who find that an environment without punishment is widely preferred. However, ex-post, it can be argued that in their particular experimental setting an environment with punishment institution was less efficient than the VCM and this is a possible explanation for their result.

[6] In contrast to the studies of Walker et al. (2000), Gürerk et al. (2006), Tyran & Feld (2006), Kroll et al. (2007) and Sutter et al. (2010), only members of the institution agree to be sanctioned, which increases the incentive for non-members to free ride. This is a "second-order free-rider problem" (cf. Oliver, 1980). Every player profits from the implementation of the institution but each player is best off when *the others* are members. Kosfeld et al. (2009) call this the "dilemma of endogenous institution formation" (p. 1336).

central authority impose the sanctions. They argue, however, that their setting captures characteristics of international agreements, such as the Kyoto Protocol or the European Union Stability and Growth Pact. But especially when global public goods are considered, enforcement problems due to the absence of strong (centralized) institutions play a crucial role. Insights from experimental economics therefore stress the importance of decentralized punishment schemes, i.e. punishment executed by the players themselves, in order to antagonize free riding and maximize social welfare (Baland & Platteau, 1996; Ostrom, 1990, 1999; Fehr & Gächter, 2000, 2002; Andreoni et al., 2003; Sefton et al., 2007).

The aim of this study is to extent the framework of Kosfeld et al. (2009) by explicitly addressing potential enforcement problems. To do so, a modification of the three-stage non-cooperative game of Kosfeld et al. (2009) is introduced, referred to as *deposit-institution formation game*. In stage 1, each player announces whether he accepts or rejects to lodge a deposit. Deposits are attached to a minimum contribution amount, such that they are not subtracted if the player contributes at least this amount but are lost in case of downward deviation. In the second stage, it is disclosed how many players are willing to lodge the deposit and then players bargain about its actual implementation (unanimous agreement). In the third stage, the public good game is played. Hence, the punishment mechanism used by Kosfeld et al. (2009) is changed; instead of letting the sanction mechanism be executed by a central authority ex-post, players are allowed to regulate their own punishment ex-ante in a decentralized way. Gerber and Wichardt (2009) theoretically analyze a two-stage mechanism, where deposits are lodged prior to the contribution stage, which are refunded in case of paying an exogenously given contribution level and lost in case of defection. The authors show that, if the deposit is sufficiently large, (full) contribution is a dominant strategy and therefore conclude that deposit lodgments might be a solution when strong sanction institutions are absent.

When analyzed theoretically, the deposit-institution formation game comprises two types of subgame perfect Nash equilibria. A *deposit equilibrium*, where people voluntarily lodge the deposit and an institution is formed along the equilibrium path, and a *status quo equilibrium*, where no one lodges the deposit. Next to the decision whether to implement an institution or not, players need to coordinate on two behalves. First, they have to coordinate on the institution size. Second, they have to solve the problem of who is part of the institution. A possible solution can be offered to the coordination problem that concern the institution size by adding strictness as equilibrium condition in every subgame. When assuming standard preferences and strictness is an equilibrium refinement, there exists a unique strict subgame perfect deposit equilibrium with respect to the size of the institution. In this equilibrium, exactly the minimum threshold of players, s^*, that makes the implementation of an institution individually profitable, lodge the deposit and contribute to the public good; the other players free ride. Hence, if $s^* > 0$, there exist an equilibrium, where players freely pre-commit to cooperation. The reason is that even though players are best off when free-riding on the institution, players are still better off when the institution emerges compared to the status quo equilibrium, where public good contributions are zero.

In order to test the theoretical predictions, an experiment was implemented. In four treatments, subjects played 20 rounds of the game described above. Since the focus is on possible enforcement problems, I vary the combination of deposit and minimum contribution amounts to manipulate the profitability of the institution. There are three main experimental results. First, the majority of players voluntarily pre-commits to (higher) contribution by lodging the deposit. Institutions are implemented in between 67.81 and 74.69 percent of the cases, even if lodging the deposit is not profitable from a payoff-maximization viewpoint. Moreover, players respond to incentives; more deposits are lodged in the two treatments, where institution formation is efficient. Second, the majority of implemented institutions consist of all players. Thus, I cannot confirm the equilibrium prediction of the standard model that expects the institution to consist of no one or no more than two players. However, it is in line with predictions based on social preferences; when analyzing the deposit-institution formation process while assuming that players suffer sufficiently from disadvantageous inequality aversion using the social preference model proposed by Fehr and Schmidt (1999), the grand institution with $s^* = n$ becomes an equilibrium. The experimental data indicate that the predominance of grand institutions is not (only) driven by miscoordination but rather results from (almost) equilibrium play. Therefore, the results stress the importance of fairness considerations for the establishment and steadiness of institutions, which is in line with findings from the common pool resource literature (Baland & Platteau, 1996; McKean, 2000) and also in the context of public choice research (Buchanan & Congleton, 1998). Finally, contribution levels are increased and stable for members when an institution is implemented. Thus, the results suggest that by using ex-ante deposit lodgments as a decentralized sanction mechanism, players can resolve the second-order free-rider problem.

The self-sanctioning scheme presented in this paper addresses three problems recently discussed in the enforcement literature. First, players can use the mechanism as a pre-commitment device to overcome self-control problems. Recent studies analyze whether next to the tension between the individual rationality of self-interest and collective wellbeing, a social dilemma might also represent a *self-control dilemma.*[7] The provision of public goods may impose an internal conflict on the individual who has to trade off the objective of acting pro-socially (cooperate) and the temptation to act selfishly (Kocher et al., 2012; Martinsson et al., 2012). Experiments find a positive correlation between patience and contribution levels (Curry et al., 2008; Fehr & Leibbrandt, 2011, Kocher et al., 2012) and that self-control is positively associated with cooperation (Martinsson et al., 2012). A suggested solution to self-control problems is to offer people to bind their behavior[8] (Strotz, 1956; Thaler & Shefrin, 1981; Schelling, 1992; Wertenbroch, 1998). So far, empirical studies on self-control strategies have found that people seem to use pre-commitment devices strategically in the anticipation of self-regulation failures (Wertenbroch, 1998; Trope & Fishbach, 2000; Ariely & Wertenbroch, 2002).

Second, the costs of sanctions are single-sided. Decentralized punishment mechanisms typically are costly and its usage therefore reduces the profits of both the punisher as well as the person that is punished. The increase in contributions to the public good thereby often does not compensate for the costs of sanctions, which in total results in lower (net) efficiency (Kroll et al., 2007; Nikiforakis & Normann, 2008; Egas & Riedl, 2008; Bothelo et al., 2005; Fehr & Gächter, 2000, 2002; Ostrom et al., 1992; Sefton et al., 2007; Nikiforakis, 2008). Only when the mere threat of the punishment establishes cooperation and the punishment does not need to be executed anymore, efficiency increases (Sefton et al., 2007; Fehr & Gächter, 2000). The self-sanctioning mechanism presented in this paper removes this problem.

Third, the self-sanctioning scheme eliminates any possibility of counter-punishment. Most studies that analyze decentralized punishment have a common property: Those who punish other subjects are immune from any actions of reprisal.[9] However, in practice, the opportunity to retaliate enacted punishment exists in almost every decentralized interaction. Hence, some of the more recent public good experiments allow for *counter-punishment* activities, i.e. subjects are informed about who has sanctioned them in the previous round and can then sanction their punishers. These studies find that the possibility of counter-punishment has the potential to entirely outweigh the increase of contributions achieved by the punishment device in the first place and therefore renders it completely ineffective

[7] A self-control dilemma is defined as an interpersonal conflict between better judgment and temptation (Thaler & Shefrin, 1981; Schelling, 1984; Loewenstein, 1996, 2000).

[8] Binding behavior is characterized by a strategic attempt to resist future temptations by voluntarily imposing (costly) constraints on future choices (Ariely & Wertenbroch, 2002).

[9] In some of the studies with partner matching, agents could punish all other group members by reducing their contribution amount or by randomly sanctioning other group members in subsequent rounds, but there could not be implemented an individually targeted sanction based on the prior sanction behavior.

(Denant-Boemont et al., 2007; Nikiforakis, 2008). This fundamentally challenges the finding that (decentralized) punishment devices offer a solution to social dilemma situations.

In section 2.2 the deposit-institution formation game is analyzed theoretically and subgame perfect Nash equilibria are deduced under the assumption of standard preferences. In section 2.3 the experiment is described and analyzed. Section 2.4 concludes.

2.2 Endogenous Institutional Arrangement: Theory

2.2.1 The General Model

In line with Kosfeld et al. (2009), a symmetric n-player public goods game with $n \geq 2$ players is considered. Every player i has an initial endowment, $w > 0$, of the private good and can contribute $0 \leq c_i \leq w$ to the public good. If player i contributes $c_i \in [0, w], i = 1, \dots, n,$ the realized amount of the public good is $z = \sum_{i=1}^{n} c_i$. Assuming (local) linearity in monetary payoffs, a player i's payoff function then becomes:

$$\Pi_i(c_i, \dots, c_n) = w - c_i + a \sum_{j=1}^{n} c_j \tag{2.1}$$

The marginal per capita return (MPCR) from the investment into the public good is captured by the parameter a, with $0 < a < 1 < na$.[10] Since the choice of participating in the institution is the choice between lodging the deposit or not, the following three-stage game is referred to as *deposit-institution formation game.* Notice that the three-stage game analyzed here has the same structure as the one in Kosfeld et al. (2009); the only modification is the different framing of the sanction mechanism. Then, the exact sequence of actions is as follows:

Participation Stage: For a given minimum contribution level, $0 < c_{min} \leq w$, players simultaneously and independently can choose to lodge a deposit d, with $0 < d \leq w$. Deposits are attached to a minimum contribution amount, such that they are lost in case contributions are below this threshold. Players who pronounce a willingness to lodge the deposit are called participants. The others are called non-participants.

Implementation Stage: Participants learn how many players want to participate. All participants then simultaneously and independently vote on the implementation of the

[10] With a < 1, $\frac{\partial U(c_1,\dots,c_n)}{\partial c_i}$ = -1 + a < 0 $\forall\, i.$ Thus, zero contribution, $c_i^* = 0$, is a strictly dominant strategy ∀i. Yet, since an > 1, $\frac{\partial \Sigma U_i(c_1,\dots,c_n)}{\partial c_i} = -1 +$ an > 0 $\forall\, i$. Thus, social optimality requires full contribution.

institution. For a successful implementation a unanimous agreement is required. If an institution is realized, participants become members and automatically lodge the deposit, whereas non-participants cannot become members and do not lodge the deposit. If no institution is implemented, no player lodges the deposit.

Contribution Stage: The final stage comprises the contribution decision. If an institution is realized, members who contribute less than the minimum contribution amount loose the deposit. Non-members did not lodge a deposit and therefore do not obey any sanctioning mechanism. If no institution emerges, no one lodged a deposit and a standard public good game is played.

Let S denote the set of members of the institution with size $|S| = s$. First, consider the case of successful institution implementation and thus $S \neq \emptyset$. Then, a player's final payoff Π_i is:

$$\Pi_i = \begin{cases} w - c_i - \hat{d}(c_i) + a \sum_{j=1}^{n} c_j & \text{if } i \in S \quad (2.2) \\ w - c_i + a \sum_{j=1}^{n} c_j & \text{if } i \notin S \quad (2.3) \end{cases}$$

with

$$\hat{d}(c_i) = \begin{cases} d & \text{if } c_i < c_{min} \\ 0 & \text{if } c_i \geq c_{min} \end{cases} \tag{2.4}$$

If, however, no institution is implemented, i.e. if $S = \emptyset$, then for all player i:

$$\Pi_i = w - c_i + a \sum_{j=1}^{n} c_j \tag{2.5}$$

Equations (2.2) and (2.4) show how the self-sanctioning mechanism of the deposit institution works: members of the institution have to lodge a deposit, which is subtracted if players contribute less than the minimum contribution amount to the public good, i.e. for whom $c_i < c_{min}$. So, members are punished for contributing too little, whereas non-members do not obey any punishment device.

2.2.2 Standard Preferences

The core question of the following analysis is whether players voluntarily implement a deposit institution and whether (full) contribution can be reached with this institutional setting. The underlying assumption for the analysis is common knowledge of rationality with perfect information. Each player therefore chooses his action in each stage by perfectly anticipating the outcomes of the game in all stages.[11] Below, I will analyze the deposit-institution formation game, and especially derive the set of subgame perfect equilibria if players' preferences are given by $U_i = \Pi_i$. Since the three-stage game analyzed here has the same structure as the one in Kosfeld et al. (2009), it can be solved in the same way.

In a subgame perfect equilibrium, each player decides on his action in every stage of the game based on backward induction. Hence, each subgame is considered separately, starting with the last stage (contribution stage). It is shown in Appendix 2.5.1 that, given that the players' utility function can be characterized by $U_i = \Pi_i$, equations (2.2) and (2.4) entail that institution members' dominant strategy in the subgame starting with the contribution stage is to contribute the minimum contribution amount once an institution is implemented if and only if $d > c_{min}(1-a)$. For non-members it is optimal to contribute nothing. If no institution is implemented, zero contribution is the dominant strategy for all players. Thus, distinguishing between the cases that an institution is formed along the equilibrium path and when it is not implemented, we have the following Nash equilibria in each subgame that starts at stage 3:

LEMMA 2.1: In stage 3, we have the following Nash equilibria in each subgame that starts in stage 3: (1) If $S \neq \emptyset$ and $d > c_{min}(1-a)$, then $c_i = c_{min}$ for all $i \in S$ and $c_i = 0$ for all $i \notin S$, (2) If $S = \emptyset$, then $c_i = 0$ for all i.

Only if the loss of deviation (d) is larger than the gain ($c_{min}(1-a)$), deviating is not beneficial for the player. This further depends on the MPCR of the public good. If a is large, the benefit from the public good already provides the player with a higher incentive to contribute, and the required deposit amount can be smaller. Since $0 < a < 1$ and $c_{min} > 0$, condition $d > c_{min}(1-a)$ is always satisfied for $d = c_{min}$.

[11] Since the payoff functions of the players only depend on the number of participants (not who is participating) in the implementation stage and whether an institution is established (not who becomes a member) in the contribution stage, the assumption of perfect information can also be relaxed without changing the theoretical results. In this case, the notion of a subgame perfect equilibrium can be substituted with a sequential equilibrium. In the experiment described in the next section, players do not have perfect information.

Next, it is analyzed whether players are willing to join the deposit institution. Besides the free-rider problem, it might be payoff increasing for players to coordinate their contributions in the framework of the institution. Consider the implementation stage. In equilibrium, institution members earn $w - c_{min} + asc_{min}$ and every player receives w if no institution is implemented. Hence, joining an institution is beneficial if the institution size, s, is such that

$$as > 1 \tag{2.6}$$

Let s^* be the lowest number of s such that condition (2.6) is satisfied. This threshold s^* depicts the minimum number of participants needed to make the implementation of the institution individually profitable for participants. It follows that for (2.6) to hold s has to be strictly larger than 1.[12]

For convenience, any subgame perfect equilibrium, where people voluntarily lodge the deposit and hence an institution is implemented is called *deposit equilibrium.* If no one lodges the deposit, the equilibrium is referred to as *status quo equilibrium*. Then:

PROPOSITION 2.1: When assuming standard preferences, a deposit equilibrium exists if any only if $s \geq s^*$ players participate. Moreover, also a status quo equilibrium exists for all participants, s, with $1 \leq s \leq n$.

The proof is provided in Appendix 2.5.1. As is shown in Kosfeld et al. (2009), Proposition 2.1 displays that an institution might emerge endogenously. The necessary condition for a successful implementation is that leastwise s^* players participate. If this minimal threshold is met, there exists a deposit equilibrium for every $s \geq s^*$, where an institution of size s is implemented. Note that the minimal threshold s^* is not given exogenously but emerges from the players' incentives to lodge the deposit. Yet, since also a status quo equilibrium exists, the implementation of the institution is not guaranteed.[13] In addition, if $s^* < n$, all institutions of size $s \in \{s^*, s^* + 1, \dots, n\}$ are a deposit equilibrium. As long as $s \geq s^*$, there also exist mixed strategy equilibria, where the probability of participation is positive and an institution is

[12] $w - c_{min} + asc_{min} > w \Leftrightarrow -c_{min} + asc_{min} > 0 \Leftrightarrow c_{min}(as - 1) > 0$. Since $c_{min} > 0$, the necessary condition for (2.6) to hold is that $as - 1 > 0 \Leftrightarrow as > 1 \Leftrightarrow s > \frac{1}{a}$. Since a < 1, it is implied that $s > 1$. The next non-negative integer satisfying this condition is $s^* \geq 2$.

[13] The provision of the institution, so the second-order public good, in equilibrium could be seen as a step-level mechanism with crucial threshold s^*. However, the provision in this model differs from standard step-level public goods (i.e. Bagnoli & Lipman, 1989; Croson & Marks, 2000) since it is an equilibrium outcome.

implemented. Next to the decision whether to implement the institution or not, the players are also confronted to two coordination problems. First, they must coordinate how many players will lodge the deposit (the institution size). Second, players also need to coordinate on who will lodge the deposit and who stays out of the institution whenever $s < n$.

Consider stage 1 next. As discussed in Kosfeld et al. (2009), the first coordination problem of how many players will lodge the deposit can be addressed by introducing strictness as an equilibrium modification. A *strict* Nash equilibrium implies that the best response of each player is a unique strategy and hence any deviation from equilibrium play makes the player strictly worse off. Then, if the game has multiple stages, the condition for a strict subgame perfect equilibrium is that each subgame induces a strict Nash equilibrium.

PROPOSITION 2.2: When assuming standard preferences, there exist a unique strict subgame perfect Nash equilibrium with exactly s^* players being part of the institution.

For the proof see Appendix 2.5.1. Proposition 2.2 says that strictness of equilibrium yields a unique deposit equilibrium. In this equilibrium, precisely the required threshold for the minimum number of players, s^*, form the institution and lodge the deposit. This makes intuitive sense. The crucial point is that every institution where the minimum threshold s^* is met will be realized in the implementation stage because players' material payoff is still higher as in the status quo equilibrium. Yet, since the setting depicts a second-order free-rider problem, each player is best off in material terms if the institution successfully emerges and its members contribute to the public good but he free rides on the institution and contributes nothing. Every additional member above the threshold s^*, i.e. $s - 1 \geq s^*$, therefore has the incentive to opt out of the institution knowing that the institution still emerges. Only if the minimum size of the institution, s^*, is met, the free riding option becomes unattractive since an institution of size $s^* - 1$ will not be implemented. Unless $s^* = n$, there are thus two groups of participants; players who voluntarily lodge the deposit and contribute to the public good and players who free ride and contribute nothing.

Note that the unique deposit equilibrium of size s^* depends on strictness as equilibrium refinement. However, as argued above, if this refinement is relaxed and weakly dominant strategies are considered as well, institutions of size larger than s^* are also an equilibrium.[14]

[14] If, for example, players' expectations are such that an institution of size smaller than n will be rejected in the implementation stage, each player's best response is to participate in the institution.

2.3 Endogenous Institutional Arrangement: Experiment

2.3.1 Procedural Details

To reduce complexity, the deposit-institution formation game is slightly modified in the experiment. The participation and implementation stage are merged by letting participants condition their choice in the participation stage on a required minimum number of participants. If this number is met, players automatically join the institution. Everything else is the same as described above. The structure of the experiment is as follows.[15]

At the beginning of each round, groups of four players are randomly matched and each player receives an endowment of 40 tokens ($n = 4$ and $w = 40$). Each player then chooses to lodge the deposit or not, i.e. whether he wants to participate in the institution.[16] Lodging a deposit is the willingness to agree that the deposit amount will be subtracted from the final payoff in case of non-compliance. Therefore, the deposit lodgment can be used as a pre-commitment device to resist the future temptation of free-riding by voluntarily imposing a constraint (which is costly to overcome) on one´s future choices. By doing so, the subject can bind himself to cooperation, knowing that *breaking* this contract is costly (not its implementation). In a sense, it therefore can be seen as a "fictitious payment" since it is only payoff relevant in case of violation of the agreement. Consequently, the strategy space of members and non-members is held constant. The parameters are chosen such that no subject could make a negative net yield.

If the player votes for lodging the deposit, he has to condition his choice on the minimum number of participants he requires (including himself) to actually join the institution (*participation stage*). Players whose minimum number of participating players is met automatically lodge the deposit without making another active decision (*implementation stage*). Nonparticipants do not make any decision about the minimum number they require. Finally, players simultaneously can freely choose the amount they want to contribute to the public good (*contribution stage*). After each round, every participant gets a detailed feedback about his round profit.

There are four experimental treatments varying the degree of profitability of the institution. In all four treatments the MPCR of the public good is 0.65 ($a = 0.65$)[17] and the minimum contribution amount is 40 tokens ($c_{min} = 40$). Thus, when players want to

[15] Experimental instructions are provided in Appendix 2.5.2.
[16] In the experimental instructions, the term "institution" was not used. Instead, subjects were asked if they want to lodge the deposit or not, knowing the amount of the deposit and the minimum contribution amount.
[17] The parameter a=0.65 was chosen to reach an equilibrium institution size of $s^* = 2$ in IF20 and IF40. The study of Grosse et al. (2011) shows that with a threshold of 3 out of 4, players seem to form the grand institution anyways. Therefore, a lower threshold can undermine this effect.

participate in the institution, they commit to contribute their full endowment. This parameter was chosen because a pilot study, where subjects could endogenously choose a combinations of the minimum contribution amount and the deposit level, showed that the vast majority suggested the efficient amount $c_{min} = w$. This implicates that subjects prefer the entire commitment of "all or nothing" in order to use the deposit lodgment as pre-commitment device. Institution formation is therefore at its highest costs, which renders participation very unattractive and increases the incentive to free-ride on the institution. On the other hand, in case of successful implementation, it can lead to the highest level of efficiency. It is possible that subjects may use the deposit mechanism as a signal for being a conditional cooperator to coordinate their behavior. To control for this, deposit lodgments in the baseline treatment IF0 do not impose any material consequences and therefore can be used as mere communication device. The deposit amount differs among the other treatments: IF10 has a deposit of 10 tokens, IF20 and IF40 have deposits of 20 and 40 tokens, respectively. In treatments IF0 and IF10, the implementation of an institution is not profitable since $d < c_{min}(1 - a)$. IF20 and IF40 lead to a profitable institution. In each treatment, the game consisted of 20 rounds.

Since Kosfeld et al. (2009) focus on the formation process of the institution (not institutional enforcement), they merge the implementation and contribution stage in their experiment, such that – in case an organization emerges – members automatically transfer their full endowment to the public good. Moreover, treatments variation is based on varying the equilibrium institution size and the comparison to an environment where institution formation is not possible. Since in our setting, the focus is on whether the institution can be enforced, we vary the profitability of institutions rather than its size. Moreover, we do not bind players to any contribution amount but let them freely choose any contribution amount with $0 \leq c_i \leq w$. However, in line with Kosfeld et al. (2009), the institutional framework is such that members of the institution are punished for contributing less than the efficient amount. In contrast to Kosfeld et al. (2009), the institution does not impose any costs on its members.

The experiments were run at the Cologne Laboratory for Economic Research. The experimental software z-Tree (Fischbacher, 2007) was used. In total, 256 subjects recruited via the organizational software Orsee (Greiner, 2004) participated in the experiment; 64 subjects in each treatment. A stranger matching design is used to avoid any confounding reputational effects. To do so, constant groups of eight are formed, out of which groups of four players are randomly matched in each round. Therefore, there are 8 independent observations per treatment, yielding 32 independent observations in total. Each treatment consisted of two sessions. Each session lasted about 60 minutes. On average, subjects earned 10.63 Euro (about $13.50) in the experiment.

2.3.2 Hypotheses

The treatment variation is based on the profitability of the institution. Since $d < c_{min}(1-a)$ in IF0 and IF10 and thus the threat of losing the deposit is not credible, a player who participates in the institution is always strictly better off contributing zero and losing the deposit than contributing the required amount of 40 tokens. Therefore, each member of the institution would always contribute nothing in the contribution stage. However, since $d > c_{min}(1-a)$ in IF20 and IF40, the threat of losing the deposit is large enough to make members better off contributing 40 tokens, whereas non-members free-ride. Thus, we can deduce the following hypotheses for the contribution stage:

HYPOTHESIS 2.1(a): Everyone free-rides (i.e. $c_i = 0 \;\forall\, i$) in IF0 and IF10. Since $d < c_{min}(1-a)$, every member of the institution is better off contributing zero instead of c_{min} and thus every member loses the deposit in IF0 and IF10 if an institution is implemented.

HYPOTHESIS 2.1(b): Members of the institution contribute the minimum contribution amount (i.e. $c_i = 40 \;\forall\, i \in S$) and non-members free-ride (i.e. $c_i = 0 \;\forall\, i \notin S$) in IF20 and IF40. Since $d > c_{min}(1-a)$, every institution members' dominant strategy is to contribute c_{min} and thus no member loses the deposit in IF20 and IF40.

Knowing that all members of the institution will contribute 0 in the contribution stage in IF10, each player is always better off not implementing an institution and thus the only rational equilibrium in IF10 is $s^* = 0$. Using backward induction, no player should therefore initiate an institution. Note, no hypothesis can be inferred for players' behavior in the implementation stage in treatment IF0 because players are indifferent between joining the institution or not. Since $a = 0.65$, treatments IF20 and IF40 are designed such that $s^* = 2$. Therefore the following hypotheses for the implementation stage can be assumed:

HYPOTHESIS 2.2(a): No institutions are implemented (i.e. $s = s^* = 0$) in IF10.

HYPOTHESIS 2.2(b): The institution consists of two players (i.e. $s = s^* = 2$) in IF20 and IF40.

Naturally, the former hypotheses imply a hypothesis about the impact on net efficiency. Net efficiency is the percentage of potential maximal earnings reached by the group in each round. Since there are no deposits paid in IF0 and IF10 and everyone contributes 0 in equilibrium, average net efficiency is predicted to lay around 40 percent in IF0 and IF10.[18] Since in IF20 and IF40, two out of four players in each group lodge the deposit and contribute their full endowment and no deposits are lost, average net efficiency lays around 70 percent.[19] Hence, net efficiency in the profitable treatments is higher compared to the unprofitable ones:

HYPOTHESIS 2.3(a): Net efficiency in IF0 and IF10 is the same as predicted by a standard public good game and lies around 40 percent of the welfare maximum.

HYPOTHESIS 2.3(b): Net efficiency in IF20 and IF40 lays around 70 percent of the welfare maximum. Hence, net efficiency increases if the institution formation is profitable.

2.3.3 Results

The result section is structured as follows. First, contribution behavior is scrutinized under the premise of implementation rates of institutions. Then, by incorporating the credibility (and thus profitability) of institutions, it is differentiated between members' and non-members' behavior. Finally, the impact on net efficiency is analyzed.

The first result shows that contribution levels are higher and remain stable if institution formation is profitable.

RESULT 2.1: Overall, the average contribution levels are higher in treatments IF20 and IF40, where institution formation is profitable compared to treatments IF0 and IF10, where the institutions do not impose a credible threat. Moreover, contribution levels are stable in IF20 and IF40 but decreasing in IF0 and IF10.

Table 2.1 summarizes the average contributions per treatment. Average contributions are around 20 tokens in IF0, 24 tokens in IF10, 31 tokens in IF20 and 30 tokens in IF40. Average contributions are not statistically different between IF0 and IF10 (Mann-Whitney test,

18 $\frac{4*(40-0+0.65*4*0)}{4*(40-40+0.65*4*40)} = 38.46\%$ in IF0 and IF10. Note, the hypothesis about net efficiency is not different to predictions in a standard public good game without the possibility of institution formation.

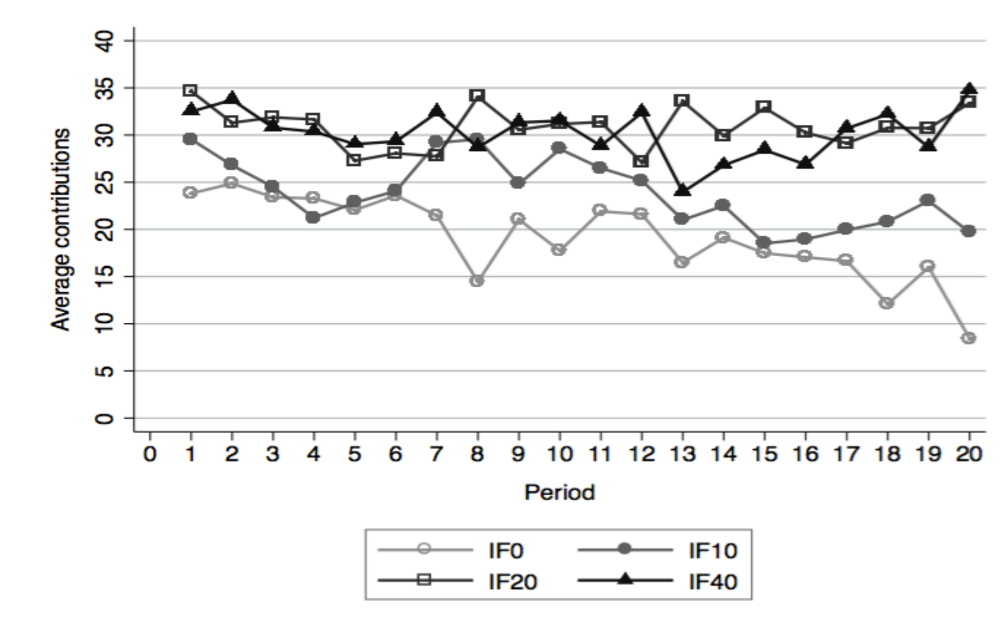
19 $\frac{[2*(40-40+0.65*2*40)+2*(40-0+0.65*2*40)]}{4*(40-40+0.65*4*40)} = 69.23\ \%$ in IF20 and IF40.

p=0.0587)[20] and also not between IF20 and IF40 (Mann-Whitney test, p= 0.8336). However, contributions are significantly higher in treatment IF20 compared to treatments IF0 and IF10 (Mann-Whitney test, IF0 vs. IF20: p=0.0008; IF10 vs. IF20, p=0.0274) and in treatment IF40 compared to IF0 (IF0 vs. IF40, p=0.0016). However, when comparing IF10 to IF40, the difference fails to be significant (IF10 vs. IF40, p=0.0587). Moreover, Figure 2.1 shows that contribution levels are constant in IF20 and IF40 but decreasing in IF0 and IF10. The decrease of average contributions over time is highly significant in IF0 and IF10 (Spearman rank order correlation; IF0: $\rho= -0.3053$, p=0.0000; IF10: $\rho= -0.1599$, p=0.0000) but there is no sigificant change in treatments IF20 and IF40 (IF20: $\rho= 0.0482$, p= 0.0848; IF40: $\rho= 0.0052$, p=0.8524).

TABLE 2.1: AVERAGE CONTRIBUTIONS

	Treatment					
	IF0	IF10	IF0 & IF10	IF20	IF40	IF20 & IF40
Average contributions						
Total	19.12	23.85	21.68	30.87	30.16	30.52
Members	29.88	32.84	31.22	37.97	39.01	38.48
Non-members	12.23	9.61	11.21	15.87	13.49	14.63
No institution	12.08	8.99	10.91	15.49	12.39	13.86

FIGURE 2.1: AVERAGE CONTRIBUTIONS PER TREATMENT



[20] The statistical tests build on group averages per independent observation. Throughout the whole paper, results that are reported are based on two-sided tests.

Next, the implementation rate of institutions is analyzed. The following result shows that in all treatments, an institution is always initiated, and implemented in between 68 and 75 percent of the cases.

RESULT 2.2: In all three treatments, in each round an institution is always initiated and the institution is implemented in 68, 75, and 72 percent of the cases in treatments IF10, IF20, and IF40, respectively.

Table 2.2 depicts the absolute and relative numbers of initiated and implemented institutions and the corresponding institution sizes per treatment. In IF10, fewer institutions are implemented than in IF20 and IF40 (217 versus 239 and 230, respectively) and this difference is significant (Mann-Whitney test, IF10 vs. IF20, p=0.0005; IF10 vs. IF40, p=0.0365) but the implementation rate does not differ between IF20 and IF40 (Mann-Whitney test, p=0.1668). The high participation rate shows that players seem to highly value the possibility of a pre-commitment device to establish cooperation. More importantly, they successfully implement institutions and thus resolve the second-order free-rider problem.

TABLE 2.2: IMPLEMENTED INSTITUTIONS

	Treatment					
	IF10		IF20		IF40	
	Nr.	%	Nr.	%	Nr.	%
Initiated Institutions	320	100.00	320	100.00	320	100.00
Implemented institutions						
Total	217	67.81	239	74.69	230	71.88
1 member	15	6.91	10	4.18	12	5.22
2 members	7	3.23	7	2.93	8	3.48
3 members	22	10.14	43	17.99	32	13.91
4 members	173	79.72	179	74.90	178	77.39

Notes: An institution is initiated if at least one player per group votes for the deposit lodgment. Percentages are computed in the following way: Initiated institutions in relation to all rounds, implemented institutions over all initiated institutions, institutions size over all implemented institutions.

In the following, the size of implemented institutions is analyzed:

RESULT 2.3: In all treatments, the majority of implemented institutions consist of all players, i.e. $s = 4$. *Institutions smaller than* s^*, *i.e.* $s < s^*$, *are very rarely observed.*

As can be seen in Table 2.2 (lower part), in all three treatments, the majority of implemented institutions consists of all four players (IF10: 79.72 percent, IF20: 74.90 percent, IF40: 77.39 percent). On average, the size of implemented institutions does not differ between treatments (Mann-Whitney test, IF10 vs. IF20, p=0.2922; IF10 vs. IF40, p=0.5992; IF20 vs. IF40, p=1.000). Moreover, institutions are almost never implemented with fewer than s^* players. Overall, only 4.18 (5.22) percent of implemented institutions consist of less than s^* players in IF20 (IF40). Hence, the theoretical threshold s^* is not a good estimation of the maximum institution size (cf. Proposition 2.2) but rather predicts its minimum size. This finding confirms the findings of Kosfeld et al. (2009). Since in the majority of the cases (67.81%) an institution is implemented in IF10, i.e. $s > 0$, Hypothesis 2.2(a) can be rejected. Moreover, in 92.11% of the cases, institutions consist of more than two players in IF20 and IF40 and therefore Hypothesis 2.2(b) can be rejected as well.

Next, the question why there are so many grand institutions can be addressed. There are two possible reasons. Either, players target to implement institutions of size s^* but miscoordinate in the participation stage. Another possibility is that players aim to implement the grand institution and reject institutions of less than four players in the implementation stage. Analyzing the required minimum number of participants may shed some light on the motivation behind subjects' behavior:

RESULT 2.4: The vast majority of players require more than two players to participate in the institution. This implies, that institutions with size smaller or equal to two have a high likelihood of failing to be implemented in the implementation stage.

Support for Result 2.4 is presented in Table 2.3. If the motivation for the high participation rate was that players miscoordinate, the required minimum number of participant should be $s_{min} = s^* = 2$ in treatments IF20 and IF40. However, only 13.74 (18.96) percent require this minimum number of participants in IF20 (IF40). The likelihood that an institution of $s < s^*$ is implemented lies below 7 (5) percent in IF20 (IF40). When the threshold s^* is met, the likelihood increases but still is very low. The result that roughly one third to one half of players require $s_{min} = 3$ might be an indicator that the predominance of grand institutions emerges because $s = 4$ depicts a focal point in the coordination problem. However, a non-trivial fraction of players reject institutions of size $s^* < s < n$. Moreover, players seem to learn the benefit of establishing the grand institution in the course of the experiment and adapt their required minimum number of participants over time. In all three treatments, the required minimum number of participants increases significantly over time (Spearman rank order correlation, IF10: ρ=0.0722, p=0.0098; IF20: ρ=0.0745, p=0.0077; IF40: ρ=0.1121, p=0.0001). Hence, the data suggest that grand institutions are not solely formed because of focal points. Rather, a possible explanation seem to be that the driving force behind subject's behavior might be additional influences, such as equality and fairness considerations.[21]

TABLE 2.3: REQUIRED MINIMUM NUMBER OF PLAYERS

	Treatment					
	IF10		IF20		IF40	
	Nr.	%	Nr.	%	Nr.	%
Required Minimum Number						
1 member	56	5.20	50	4.52	74	6.65
2 members	189	17.57	152	13.74	211	18.96
3 members	360	33.46	541	48.92	465	41.78
4 members	471	43.77	363	32.82	363	32.61

Note: Table 2.3 depicts the absolute and relative number of the minimum number required by participants over all rounds. The relative number is the frequency the required minimum number is chosen relative to all choices made by participants.

So far, it has been shown that there are significantly less institutions in treatment IF10. The next question to address is the credibility of the institution, which is the core difference between the treatments. The second part of Hypothesis 2.1(a) says that if an institution is implemented, every member is better off losing the deposit in IF0 and IF10 since the institution is not credible. In treatment IF20 and IF40, however, the threat of losing the deposit should be high enough to make members stick to the agreement (second part of Hypothesis 2.1(b)). Thus, it has to be analyzed how many members violate the agreement and thus get punished for doing so:

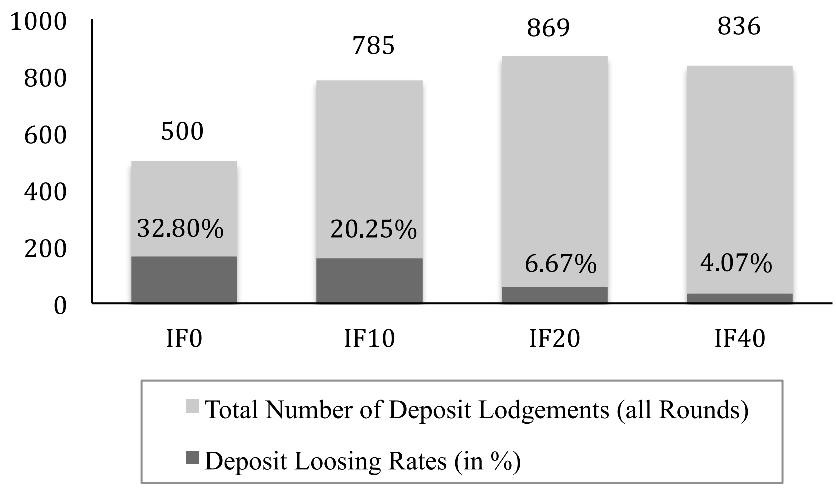
RESULT 2.5: Deposit losing rates are much higher in IF0 and IF10 than in IF20 and IF40. Moreover, over time, more people lose their deposit in IF0 and IF10 but the rate remains constant in IF20 and IF40.

Figure 2.2 illustrates the absolute number of all individually lodged deposits over all rounds and how many of these lodgments have been lost per treatment. Clearly, the second part of Hypothesis 2.1(a) cannot be rejected. Deposit losing rates are much higher in IF0 (32.80

[21] The high implementation rate of grand institutions could also be due to expectations about the required minimum number of the other players.

percent) and IF10 (20.25 percent) than in IF20 (6.67 percent) and IF40 (4.07 percent) and the difference between IF0 and IF20 and IF40, as well as between IF10 and IF20 and IF40 is strongly significant (Mann-Whitney test, IF0 vs. IF20, p=0.000; IF0 vs. IF40, p=0.000; IF10 vs. IF20, p=0.000; IF10 vs. IF40, p=0.000). Since lost deposit rates are minimal in treatments IF20 and IF40, also the second part of Hypothesis 2.1(b) cannot be rejected. Furthermore, lost deposit rates increase significantly over time in IF10 (Spearman rank order correlation, ρ= 0.1828, p=0.0000) but do not change in IF20 and IF40, respectively (IF20: ρ= -0.0307; p=0.3655; IF40: ρ= 0.0149, p=0.6673). Thus, players respond to incentives: fewer institutions are implemented in IF10 but additionally the institution is less binding.

FIGURE 2.2: DEPOSIT LOSING RATES



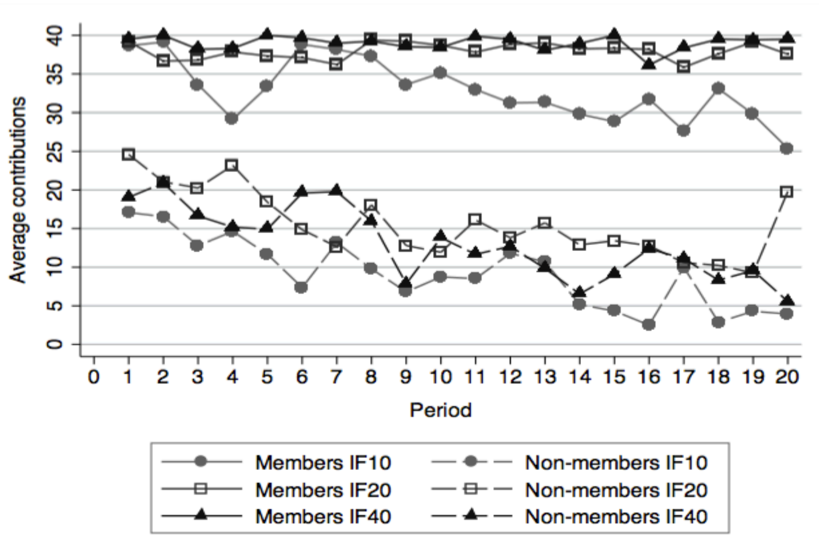
Note: The relative numbers are calculated as follows: Percentage of deposits that have been subtracted relative to all deposit lodgements over all rounds.

If we now compare the contributions of members and non-members, the following result can be derived.

RESULT 2.6: In all treatments, members contribute on average more than twice as much as non-members. Contributions between non-members do not differ between treatments, but average contribution levels of members are higher in the two profitable treatments than in IF10. Moreover, contribution levels of members as well as non-members decrease over time in IF10 but remain constant for members in IF20 and IF40.

Evidence is presented in Table 2.1 (lower part). The difference in contributions of members and non-members is strongly significant in all treatments (Wilcoxon signed-rank test, IF10, IF20, IF40: p=0.0117).[22] However, Table 2.1 also shows that members contribute on average less in IF10 than in IF20 and IF40 and this difference is significant as well (Mann-Whitney test, IF10 vs. IF20: p=0.0163; IF10 vs. IF40: p=0.0090). The difference between contributions of non-members, however, is not significant (Mann-Whitney test, IF10 vs IF20: p=0.2752; IF10 vs IF40: p=0.1266; IF20 vs IF40: p=0.0495). Intuitively, these findings make sense since the institution is less binding for members in IF10 than in IF20 and IF40. Since on average members and non-members contribute more than 0 in IF10, i.e. $c_i > 0 \; \forall \, i$, the first part of Hypothesis 2.1(a) can be rejected. This is not surprising, since public good experiments show that players tend to contribute more than predicted by the selfish free-riding equilibrium (for surveys, see i.e. Ledyard, 1995; Zelmer, 2003; Gächter, 2007; Chaudhuri, 2011). The first part of Hypothesis 2.1(b) cannot be rejected. Even though players contribute on average more than 0 if no institution is implemented or they are not members (i.e. if $S \neq \emptyset$, $c_i > 0 \; \forall \, i \notin S$ and if $S = \emptyset$, $c_i > 0 \; \forall \, i$), members contribute on average almost 40 tokens, i.e. $c_i = 40 \; \forall \, i \in S$. Moreover, as can be seen in Figure 2.3, both contributions of members as well as non-members decrease over time in IF10 and this decrease is highly significant (Spearman rank order correlation, members: ρ=−0.3323, p=0.000; non-members: ρ=−0.3972, p=0.000), which is an explanation for the overall decrease in contributions. In IF20 and IF40, on the other hand, contributions of members are constant (Spearman rank order correlation, IF20: ρ=0.0645; p=0.0572; IF40: ρ=0.0461; p=0.1832), whereas contributions of non-members decrease significantly as well (IF20: ρ=−0.3752, p=0.000; IF40: ρ=−0.3285, p=0.000).

FIGURE 2.3: AVERAGE CONTRIBUTIONS MEMBERS AND NON-MEMBERS



[22] This result is very strong. In all four treatments, there is not a single case, where, on average, non-members have contributed more than members in any round. Therefore, the ranking is the same in all treatments.

Finally, net efficiency is scrutinized. Net efficiency is the average realized group earnings relative to the highest potential group earnings in each round. So, I define net efficiency as $\Pi_{realized}/\Pi_{max}$, with $\Pi_{realized}$ depicting for the average realized group earnings, and Π_{max} the potential maximal group earnings. Maximal group earnings are attained when each group member contributes his whole endowment to the public good. As can be seen in Table 2.3, net efficiency levels are higher in IF20 (85.09%) and IF40 (83.84%) than in IF0 (67.88%) and IF10 (73.97%). This difference is significant (Mann-Whitney test, IF0 vs IF20: p=0.0000; IF0 vs IF40: p=0.0000; IF10 vs IF20: p=0.0000; IF10 vs IF40: p=0.004). Moreover, net efficiency levels of groups where an institution is implemented are higher than if no institution is implemented in all treatments and this difference is also strongly significant (Wilcoxon signed-rank test, p=0.0117 in all treatments). Another interesting question is whether the level of efficiency that is achieved when an institution is implemented differs between IF10 and IF20 and IF40. This is indeed the case; average net efficiency with an institution is significantly lower in IF10 (Mann-Whitney test, IF10 vs. IF20: p=0.0015; IF10 vs. IF40: p=0.0005). Finally, net efficiency levels decrease significantly over time in IF0 and IF10 (Spearman rank order correlation, IF0: ρ=‒0.3053, p=0.0000; IF10: ρ=‒0.1812, p=0.0000) and thus follow the usual pattern of public good games. Thus, Hypothesis 2.3(a) cannot be rejected. Net efficiency levels over rounds are constant in IF20 and IF40 (IF20: ρ=0.0403, p=0.1497; IF40: ρ=-0.0053, p=0.8484). Since net efficiency levels are significantly higher in IF20 and IF40 compared to the other treatments and constant over time, it can be said that Hypothesis 2.3(b) cannot be rejected as well.

TABLE 2.4: NET EFFICIENCY WITH/ WITHOUT INSTITUTIONS PER TREATMENT

	Treatment			
	IF0	IF10	IF20	IF40
Average Level of Net Efficiency				
Total	67.88	73.97	85.09	83.84
Institution	80.01	84.25	92.81	94.14
No institution	57.04	52.29	62.29	57.52

2.3.4 Possible Explanation: Social Preferences

Result 2.3 has shown that in the majority of the cases, a grand institution, where all players lodge the deposit, emerges. Moreover, the fact that many players require all four players to participate as condition for their own participation indicates that many players reject institutions of size smaller than n. Yet, as shown above, standard theory assumes that such rejections are dominated whenever $s^* \leq s < n$. Hence, the experimental results indicate that the realization of grand institutions might be influenced by preferences for equality and fairness. To address this possible explanation, it is analyzed how social preferences might influence the institution formation process, adopting the same framework as in Kosfeld et al. (2009). To do so, the inequity-aversion model proposed by Fehr and Schmidt (1999) is used.[23] For the analysis the set of all players is divided into non-members, denoted by j (with u_j, c_j, $j \notin S$), and members, denoted by k (with u_k, $c_k, k \in S$) since players who are not in the institution and those who are in the institution impose different levels of disutility from (either advantageous or disadvantageous) inequality on player i. The utility function for all $i \in S$ becomes

$$U_i = w - c_i - d + a\sum_{k \in S\setminus\{i\}} c_k + a\sum_{j \notin S} c_j + ac_i - \alpha_i \frac{1}{s-1}\sum_{i \neq k=1}^{s} \max\{u_k - u_i, 0\}$$
$$-\alpha_i \frac{1}{n-s}\sum_{j=1}^{n-s} \max\{u_j - u_i, 0\} - \beta_i \frac{1}{s-1}\sum_{i \neq k=1}^{s} \max\{u_i - u_k, 0\} - \beta_i \frac{1}{n-s}\sum_{j=1}^{n-s} \max\{u_i - u_j, 0\} \quad (2.7)$$

and the utility function for all $i \notin S$ is

$$U_i = w - c_i + a\sum_{k \in S} c_k + a\sum_{j \notin S\setminus\{i\}} c_j + ac_i - \alpha_i \frac{1}{s}\sum_{k=1}^{s} \max\{u_k - u_i, 0\} \quad (2.8)$$
$$-\alpha_i \frac{1}{n-s-1}\sum_{i \neq j=1}^{n-s} \max\{u_j - u_i, 0\} - \beta_i \frac{1}{s}\sum_{k=1}^{s} \max\{u_i - u_k, 0\} - \beta_i \frac{1}{n-s-1}\sum_{i \neq j=1}^{n-s} \max\{u_i - u_j, 0\}$$

The parameter α_i measures player i´s disadvantageous inequality aversion and β_i the utility loss caused by advantageous inequality. In line with Fehr and Schmidt (1999), it is assumed that $\beta_i \leq \alpha_i$ and $0 \leq \beta_i < 1$ for all i. Moreover, following the assumption that players suffer most from disadvantageous inequality, it is assumed that $\beta_i < 1 - a$ for all i.

Again, the contribution stage is considered first.

LEMMA 2.2: In stage 3, we have the following Nash equilibria in each subgame that starts in stage 3: If S≠ ∅ and $d > c_{min}(1 - a)$, *then (1)* $c_i = c_{min}$ *for all* $i \in S$ *and* $c_i = 0$ *for all* $i \notin S$ *if* $\beta_i < (1 - a)$ *for all* $i \notin S$, *(2)) If S= ∅ and* $\beta_i < 1 - a$ $\forall i$, *then* $c_i = 0$ *for all i.*

The proof is provided in Appendix 2.5.1. As under the assumption of standard preferences, members always contribute the minimum contribution amount if the institution is credible.

[23] Other models of social preferences could have been considered as well (i.e. Rabin, 1993; Bolton & Ockenfels, 2000; Charness & Rabin, 2002; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Cox et al., 2007).

Non-members free-ride if their disutility from advantageous inequality is rather small, i.e. if $\beta_i < \frac{1}{2}(1-a)$. If no institution is implemented, $c_i = 0\ \forall i$ is a Nash equilibrium.[24] It is interesting to analyze whether an equilibrium with $c_i = c_{min}\ \forall i$ exists without an institution if all players have social preferences with $\beta_i < 1-a$. However, no equilibrium with $c_i = c_{min}\ \forall i$ exists without an institution.[25]

Consider the implementation stage next. If $\beta_i < (1-a)$ and an institution of size s is implemented, institution member i's utility is given by $w - c_{min} + asc_{min} - \frac{\alpha_i}{n-s}(n-s)c_{min}$ and everyone's payoff is w if no institution is implemented. Players are thus better off entering into an institution if the number of members, s, satisfies

$$as - \alpha_i > 1 \tag{2.9}$$

The formula is the same as in the standard preferences case but some disutility from inequality between members and non-members is subtracted. When comparing equations (2.6) and (2.9), it is obvious that the threshold s^{**} to satisfy condition (2.9) has to be larger than s^*.[26] For the case of identical α_i's ($\alpha_i = \alpha\ \forall i$), the same analysis as implemented in the former section can be applied based on the new threshold s^{**} defined by equation (2.9). Obviously, for sufficiently strong social preferences, $s^{**} = n$ is the unique deposit equilibrium, where standard preferences forecasts various deposit equilibria with $s^* \leq n-1$. Thus, social preferences might be an explanation for the majority of grand institutions observed in the experiment.

2.4 Discussion and Conclusion

In the framework of this study, the institution formation process suggested by Kosfeld et al. (2009) is extended by the integration of a non-centralized solution based on a conditional ex-ante self-sanctioning scheme. Players were given the opportunity to lodge a deposit that is lost in case of deviation. Despite the fact that the setting comprises a second-order free-rider

[24] Let c_{-i} be any arbitrary profile for all possible contributions for all players except i. If S= $\emptyset$, a player i will contribute an amount $c_i > 0$ to the public good, if and only if $U_i(c_i, c_{-i}) > U_i(0, c_{-i})\ \forall c_i > 0, \forall c_{-i}, \forall i$, which means $w - c_i + a\sum_j c_j + ac_i - \alpha_i(c_i - c_j) - \beta_i(c_j - c_i) > w + a\sum_j c_j - \beta_i c_j$. If the other players contribute nothing, no player i has an incentive to contribute to the public good, because $U_i(c_i, 0) < U_i(0,0), \forall c_i > 0, \forall i$. If the other players contribute to the public good, still every player has an incentive to not contribute, because $U_i(c_i, c_{-i}) < U_i(0, c_{-i}), \forall c_i > c_{-i} > 0, \forall c_{-i} > c_i > 0, \forall i$.

[25] $w - c_{min} + a\sum_j c_{min} + ac_{min} > w + a\sum_j c_{min} - \beta_i c_{min} \Leftrightarrow a - 1 > -\beta_i \Leftrightarrow \beta_i \ngtr 1 - a$.

[26] The condition for standard preferences is $s > \frac{1}{a}$. Here, the condition becomes $s > \frac{1+\alpha_i}{a}$. Since $\alpha_i \geq 0$ and $0 < a < 1$, the threshold increases.

problem, the theory suggests that players principally can resolve this problem. If the institution is profitable, i.e. the deposit is large enough to depict a credible threat, there exists a set of subgame perfect Nash equilibria, where s players ($s^* \leq s \leq n$) lodge the deposit. However, also a status quo equilibrium exists and thus, success is not guaranteed.

The empirical part of the paper shows that the mechanism of deposit lodgments successfully increases contributions and can enforce cooperation among its members. A possible explanation might be that voting is used as a signal for being a conditional cooperator (Tyran & Feld, 2006), which has been shown to already promote cooperation (Brosig et al., 2003). To control for this, a baseline treatment IF0 is implemented that does not impose any payoff consequences on the players and thus can be used as mere communication (cheap talk) tool. A general observation in standard public good experiments (without the possibility of any kind of coordination) are contribution levels between 40 and 60 percent in the first round and then a steady decline to 5 to 20 percent in the last round (i.e. Falkinger et al., 2000; Ostrom, 2000). Average contributions are not higher in IF0. Hence, it does not seem that the increase in contributions is (solely) driven by the use of the sanction mechanism as a coordination device. Moreover, the results imply that players voluntarily prefer an environment with the pre-commitment mechanism to a sanction-free environment: In all treatments, in the vast majority of the cases, institutions are implemented. Third, the vast majority of implemented institutions are grand institutions. This result shows that fairness considerations are important for the formation and stability of institutions. It also is in line with the *generality principle* suggested by Buchanan & Congleton (1998), which requires nondiscrimination of political choices and equality in behavior towards all individuals. Forth, when an institution is implemented, groups reach efficiency levels close to the possible maximum level in treatment IF20 and IF40. Due to the higher deposit losing rates in IF10, efficiency levels are lower. Still, the difference between net efficiency levels when an institution emerges compared to an institution-free environment is significant in all treatments. One possible reason could be that subjects that are more cooperative also are more likely to (self-)select into more efficient institutions. However, in a comparable setting, Sutter et al. (2010) find "support against the conjecture that self-selection is an important driving force behind the positive effects of endogenous institutional choice" (p. 1561).

The main differences to the experimental design of Kosfeld et al. (2009) are as follows.[27] First, since the authors focus on the formation of institutions, they abstract from any potential enforcement problems and therefore – in case an institution emerges – members

[27] The decision trees of both experiments are provided in Appendix 2.5.3.

automatically have to transfer their full endowment to the public good account. Second, players get information about the exact number of participants. Moreover, a partner matching is used. With this setting, expectation formation about others' willingness to form an institution is very easy. The analysis of elicited beliefs in Kosfeld et al. (2009) shows that the majority of players expect all other players to participate in the institution, which implies that – if expectations are right – each player's best response then is to participate in the institution as well. Hence, players can easily coordinate on the grand institutional equilibrium.[28] In this study, players only learn whether their required minimum number of participants is met not the exact number of participants. Moreover, a stranger matching is used. As argued in the theoretical part of this paper, any institution of size larger than s^* is an equilibrium if weakly dominant strategies are considered. Players might coordinate on the grand institution, if they know or expect that players require all subjects to participate in order to implement the institution (and thus all players require s_{min} = 4). However, with strictness as equilibrium refinement, this requirement is always dominated (cf. Proposition 2.2). Clearly, in the laboratory, data are more noisy. Yet, the experimental design chosen in this study reduces the possibility to coordinate on the grand institution by making expectation formation more difficult.

Clearly, a direct comparison to the results of Kosfeld et al. (2009) is not possible due to the differences in the experimental design of both studies. However, there is one point to mention. In this study, besides the fact that learning is more difficult than in a partner matching, there still is a significant increase in the grand institution. Moreover, on net, the possibility of institution formation is equally successful in both studies; cooperation rates in the two profitable treatments (IF20 and IF40) of this study are very similar (and even slightly higher) to those of the corresponding treatment (IF65) of Kosfeld et al. (2009).[29]

Hence, there are three main contributions of this study. First, the institution formation game proposed by Kosfeld et al. (2009) can also be applied to a decentralized sanction mechanism. The conclusion of the authors that "institution formation can be an important and effective solution in social dilemma situations" (p. 1353) can be supported. Second, a novel institutional setting is presented. The experimental setting contributes to the literature by combining the endogenous formation of institutions in a public good game with a pre-commitment mechanism; thus it can contribute to the understanding of endogenous

[28] If players expect (or even know) that institutions of size smaller than 4 will be rejected, each players' best response is to participate in the institution as well since the individual payoff of each player when all four players contribute to the PG is larger than the status quo equilibrium where no PG is realized.

[29] Data can only be compared to the treatments with the same MPCR (a=0.65).

institutional choice. Third, the self-sanctioning mechanism has one important advantage: The single-sidedness of punishment costs. When comparing net efficiency of IF0 (costless punishment) and IF20 and IF40 (costly one-sided punishment), it can be seen that the costs accompanied by the deposit device do not outweigh the gain in profits reached by the higher cooperation level. Even though the use of the sanction mechanism is very high, efficiency increases as well. Next, net efficiency levels reached in this study are higher compared to those found in studies using a decentralized ex-post punishment mechanisms. Typical net efficiency levels lay between 70 percent (Nikiforakis, 2008; Sefton et al., 2007) and 80 percent (Fehr & Gächter, 2000a; Kroll et al., 2007). Net efficiency levels achieved in the two profitable treatments with the sanction institution in this study lay around 93 percent. Besides the positive impact on efficiency levels compared to a costly ex-post punishment device, the self-sanctioning scheme moreover removes any possibility of revengeful action and thus eliminates the problem of counter-punishment.

Furthermore, the mechanism can help players to overcome a self-control problem that might be present in their decision whether to contribute to the public good or not. The results depict a participation rate of 100 percent, which means that in every single round, there is always at least one player who wants to lodge the deposit. Seeing the deposit lodgment as a pre-commitment device, the experimental results therefore suggest that players seem to value the possibility to bind their behavior ex-ante and use the deposit lodgments as a strategic device to overcome self-control problems in the anticipation of self-regulatory failure. However, only around 5 percent of players require a minimum number of participants of one person (themselves). This is an indicator that the high participation rate might not be caused by the motivation to castigate one's own self-control problems but rather that the mechanism is used to enforce discipline among the other players. However, the study cannot explicitly disentangle this motivation. This remains an interesting question for future research.

2.5 Appendix

2.5.1 Proof of Lemmas and Propositions

Proof of Lemma 2.1. In a subgame perfect equilibrium, it is assumed that each player rationally anticipates the equilibrium outcomes of future stages by applying backward induction. Following this line of reasoning, each player considers stage 3, the contribution stage, first.

Case (1): If i is a member of the institution, i.e. $i \in S$:

In stage 3, all members of the institution have lodged the deposit d and then have to decide about their contribution amount. They can contribute any amount c_i with $0 \leq c_i \leq w$ and face the following payoffs:

For all $i \in S$:

$$U_i = \begin{cases} w - c_i + a\left(\sum_{j \in S \setminus \{i\}} c_j + c_i\right) & \text{if } c_i \geq c_{min} \quad \text{and} \\ w - c_i - d + a\left(\sum_{j \in S \setminus \{i\}} c_j + c_i\right) & \text{if } c_i < c_{min} \end{cases}$$

Now let c_{-i} be any arbitrary profile for all possible contributions for all players except i. If $0 \leq c_i < c_{min}$, a player will always chose $c_i = 0$ because this yields the highest payoff for him. Players will not contribute more than c_{min} because the MPCR from contributing to the public good is smaller than 1, i.e. $a < 1$. Hence, if $c_i \geq c_{min}$, it is sufficient to show that a player i will contribute $c_i = c_{min}$ if and only if

$$U_i(c_{min}, c_{-i}) > U_i(0, c_{-i}) \qquad \forall\, c_{-i}\ \forall i \in S$$

$$\Leftrightarrow \quad w - c_{min} + a\sum_{j \in S \setminus \{i\}} c_j + ac_{min} > w - d + a\sum_{j \in S \setminus \{i\}} c_j$$

$$\Leftrightarrow \quad ac_{min} - c_{min} > -d$$

$$\Leftrightarrow \quad d > c_{min}(1 - a)$$

This means that only if the sanction of not complying is larger than the gain, deviating is not beneficial for the player and thus this strategy will not be chosen. Hence, if condition

$d > c_{min}(1-a)$ holds, then the strategy profile $c_i = c_{min}$ is the strictly dominant strategy for all institution members in the contribution stage.

Case (2): If i is not a member of the institution, i.e. $i \notin S$:

Since $a < 1$, for all non-members, i.e. for all $i \notin S$, the strategy profile $c_i = 0$ is the strictly dominant strategy in the contribution stage.

Therefore, we have the following Nash equilibria in each subgame that starts at stage 3: (1) If S$\neq \emptyset$ and $d > c_{min}(1-a)$, then $c_i = c_{min}$ for all $i \in S$ and $c_i = 0$ for all $i \notin S$, (2) If $S = \emptyset$, then $c_i = 0$ for all i.

Additional analysis if $\boldsymbol{d = \delta w}$. Note, that the deposit can also be a fraction of the initial endowment, i.e. d = δw. Then, the analysis looks at follows:

Case (1)': If i is a member of the institution, i.e. $i \in S$:

Players will prefer to contribute $c_i = \mathrm{c_{min}}$ over $c_i = 0$, hence any $c_i < \mathrm{c_{min}}$, if and only if

$$U_i(\mathrm{c_{min}}, c_{-i}) > U_i(0, c_{-i}) \qquad \forall\, c_{-i}\ \forall i \in S$$

$$\Leftrightarrow \quad w_i - \mathrm{c_{min}} + a\sum_{j\in S\setminus\{i\}} c_j + a\mathrm{c_{min}} > w - \delta w + a\sum_{j\in S\setminus\{i\}} c_j$$

$$\Leftrightarrow \quad a\mathrm{c_{min}} - \mathrm{c_{min}} > -\delta w$$

$$\Leftrightarrow \quad \mathrm{c_{min}}(1-a) > \delta w$$

$$\Leftrightarrow \quad \delta > \frac{\mathrm{c_{min}}}{w}(1-a) \in]0,1[$$

Since $0 \leq \mathrm{c_{min}} \leq w$ and $w > 0, 0 < \frac{\mathrm{c_{min}}}{w} < 1$ and since $a < 1$, $(1-a) > 0$. If $\mathrm{c_{min}} = w$, then $\delta > (1-a)$. If the contribution is any fraction of the initial endowment, i.e. $\mathrm{c_{min}} = kw$, with k ∈] 0,1[, then $\delta > k(1-a)$. This means that the contribution strategy depends on the size of the deposit paid, or more explicitly the fraction δ:

$$\hat{c}_i^{\,*}(\delta) = \begin{cases} \mathrm{c_{min}} & \text{if } \delta > \frac{\mathrm{c_{min}}}{w}(1-a) \quad \text{and} \\ 0 & \text{if } \delta < \frac{\mathrm{c_{min}}}{w}(1-a) \end{cases}$$

Case (2)': If i is not a member of the institution, i.e. $i \notin S$:

As before, since $a < 1$, for all $i \notin S$, the strictly dominant strategy in the contribution stage is the strategy profile $c_i = 0$.

Thus, we have the following Nash equilibria in each subgame that starts at stage 3: (1) If $S \neq \emptyset$ and $\delta > \frac{c_{\min}}{w}(1-a)$, then $\hat{c}_i = c_{\min}$ for all $i \in S$ and $\hat{c}_i = 0$ for all $i \notin S$, (2) If $S \neq \emptyset$ and $\delta < \frac{c_{\min}}{w}(1-a)$, then $\hat{c}_i = 0$ for all i, (3) If $S = \emptyset$, then $\hat{c}_i = 0$ for all i.

Proof of Proposition 2.1. It is continued with the application of backward induction. First the implementation stage is considered. It follows from the definition of s^* that if $s < s^*$, no institution will be formed and thus only action profiles without lodging a deposit are Nash equilibria. If $s \geq s^*$ there exist two Nash equilibria: A deposit and a status quo equilibrium. In the deposit equilibrium, the unanimity rule implies that not a single player rejects the implementation of the institution. The status quo equilibrium is a Nash equilibrium, in which at least one participant rejects it. Thus, in case $s \geq s^*$, an institution can be formed and the following strategy possibilities for a deposit equilibrium exist: (1) a subset of players, s, lodge the deposit and the other players do not, (2) If precisely s^* players participate, every single participant agrees on the actual lodgment of the deposit (unanimity rule). Else, each participant rejects the implementation of the institution, and (3) if an institution is implemented, institution members contribute $c_{\min}$ and non-members contribute zero.

Next, the participation stage is considered. If $s < s^*$, no institution is formed and hence no one will contribute to the public good. Thus, $U_i = w \; \forall i$ in the participation stage. Consider next the case if $s \geq s^*$. Note that in the deposit equilibrium, everyone agreed upon the implementation of the institution. For all i, the payoffs in the deposit equilibrium are as follows:

$$U_i = \begin{cases} w - c_{\min} + asc_{\min} & \text{if } i \in S \\ w + asc_{\min} & \text{if } i \notin S \end{cases}$$

A participant receives $w - c_{\min} + asc_{\min}$ if an institution emerges and receives w if he deviates and thus no institution is formed. Since we have $s \geq s^*$, the former payoff is always strictly larger than w, i.e. $w - c_{\min} + asc_{\min} > w$. Additionally, non-participants are also strictly better off by not participating, i.e. $w + asc_{\min} > w - c_{\min} + asc_{\min}$. Hence, no one has an incentive to deviate and thus a deposit equilibrium induces a Nash equilibrium in the first stage.

However, for every institution size s ($1 \leq s \leq n$) there also exist a status quo equilibrium. Consider a Nash equilibrium without an institution (a status quo equilibrium) in the implementation stage. If this is the case, each player who considers whether to participate in the institution or not, is indifferent between the two options, irrespective of the behavior of the other players and hence each action profile induces a non-strict Nash equilibrium in the participation stage.

Proof of Proposition 2.2. Proposition 2.1 states that the necessary condition for a deposit equilibrium is that at least s^* players participate, i.e. $s \geq s^*$. Now, it is shown that the case, where exactly s^* players lodge the deposit, is the unique strict subgame perfect Nash equilibrium. Consider the following cases.

Case (1): Let $s > s^*$ and an institution is implemented. Players know that one participant can opt out of the institution but the institution of size $s - 1$ will still be implemented. Since $w + a\,(s-1)c_{\min} > w - c_{\min} + asc_{\min}$, any one can increase his payoff by choosing to leave the institution and thus has a material incentive to do so.

Case (2): If $s = s^*$, deviation for a member of the institution would lead to a *reduction* in payoff from $w - c_{\min} + as^*c_{\min}$ to w. Furthermore, there is no incentive for any additional player to join since $w + as^*c_{\min} > w - c_{\min} + a\,(s^*+1)c_{\min}$.

Case (3): If $s = s^* - 1$, participation for a non-participant of the institution leads to an *increase* in payoff from w to $w - c_{\min} + as^*c_{\min}$. Hence, each non-participant has an incentive to join. Finally, any action profile with $s < s^* - 1$ can be a non-strict Nash equilibrium.

Proof of Lemma 2.2: If $\beta_i < 1 - a \ \ \forall i \notin S$ and $d > c_{min}(1-a) \ \ \forall i \in S$, then members contributing c_{min}, i.e. $c_i = c_{min} \ \forall i \in S$, and non-members contributing zero, i.e. $c_i = 0 \ \forall i \notin S$, is a Nash equilibrium. To see this, consider members first. In stage 3 all $i \in S$ have lodged the deposit d. The utility function for all $i \in S$ is (Equation (2.7)):

$$U_i = w - c_i - d + a \sum_{k \in S \setminus \{i\}} c_k + a \sum_{j \notin S} c_j + ac_i - \alpha_i \frac{1}{s-1} \sum_{i \neq k=1}^{s} \max\{u_k - u_i, 0\}$$
$$-\alpha_i \frac{1}{n-s} \sum_{j=1}^{n-s} \max\{u_j - u_i, 0\} - \beta_i \frac{1}{s-1} \sum_{i \neq k=1}^{s} \max\{u_i - u_k, 0\} - \beta_i \frac{1}{n-s} \sum_{j=1}^{n-s} \max\{u_i - u_j, 0\}$$

Given $c_j = 0$ and $c_k = c_{min}$, player i will contribute $c_i = c_{min}$ instead of $c_i \neq c_{min}$ if and only if:

$$U_i^*(c_{min}, 0, c_{min}) \geq U_i^*(c_i, 0, c_{min}) \qquad \forall c_i \neq c_{min}, \forall i \in S$$

$$\Leftrightarrow w - c_{min} + a \sum_{k \in S \setminus \{i\}} c_{min} + a\, c_{min} - \alpha_i c_{min} \geq$$
$$w - c_i - d + a \sum_{k \in S \setminus \{i\}} c_{min} + ac_i - \alpha_i (c_i - c_{min} + d) - \alpha_i (c_i + d) - \beta_i (c_{min} - c_i - d)$$

There are two possible cases: Player i can either contribute zero, i.e. $c_i = 0$ or he can chose an amount larger than zero but below the minimum contribution level, i.e. $0 < c_i < c_{min}$. Theoretically, also the case where $c_i > c_{min}$ exists. Since in the experiment, however, the minimum contribution level was chosen to be the whole endowment, the strategy space was limited and thus the third case can be ruled out here. The two possible cases are analyzed separately.

Case (1): First consider the case $c_i = 0$. If $c_i = 0$, then $d = d \ \forall \ i \in S$. Player i will contribute $c_i = c_{min}$ instead of $c_i = 0$ if and only if

$$U_i^*(c_{min}, 0, c_{min}) \geq U_i^*(0, 0, c_{min}) \qquad \forall \ i \in S$$

By contributing zero, $i \in S$ has an advantageous disutility compared to $k \in S$. But since $c_i = 0$ implies that i has to pay the deposit d, the advantageous disutility from the lower contribution and the disadvantageous disutility from paying the deposit compared to all members $k \in S$, have opposing forces; either player i is worse off than the members of the institution (if

$d > c_{\min}$ then $-\alpha_i(d - c_{\min})$) or he is better off (if $d < c_{\min}$ then $\beta_i(c_{min} - d)$). Moreover, since player i loses the deposit by contributing zero, he has some disadvantageous disutility compared to non-members ($-\alpha_i d$):

$$\Leftrightarrow w - c_{min} + a \sum_{k \in S \setminus \{i\}} c_{min} + a c_{min} - \alpha_i c_{\min} \geq w - d + a \sum_{k \in S \setminus \{i\}} c_{\min} - \alpha_i (d - c_{\min}) - \alpha_i d - \beta_i (c_{min} - d)$$

$$\Leftrightarrow a c_{min} - c_{min} - \alpha_i c_{\min} \geq -d - \alpha_i (d - c_{\min}) - \alpha_i d - \beta_i (c_{min} - d)$$

$$\Leftrightarrow d \geq c_{min}(1 - a)$$

$$\Rightarrow d > c_{min}(1 - a)$$

Thus, contributing $c_i = c_{min}\ \forall\, i \in S$ strictly dominates contributing $c_i = 0\ \forall\, i \in S$ if $c_j = 0$ and $c_k = c_{min}$ as long as the institution is credible, i.e. $d > c_{min}(1 - a)$.

Case (2): Next, consider the case $0 < c_i < c_{min}$. If $c_i < c_{min}$, then $d = d\ \forall\, i \in S$. Player i will contribute $c_i = c_{min}$ instead of $c_i < c_{min}$ if and only if

$$U_i^*(c_{min}, 0, c_{min}) \geq U_i^*(c_i, 0, c_{min}) \qquad \forall\, 0 < c_i < c_{min}, \forall\, i \in S$$

$$\Leftrightarrow w - c_{min} + a \sum_{k \in S \setminus \{i\}} c_{min} + a c_{min} - \alpha_i c_{\min} \geq w - c_i - d + a \sum_{k \in S \setminus \{i\}} c_{min} + a\, c_i$$
$$- \alpha_i (c_i + d - c_{min}) - \alpha_i (c_i + d) - \beta_i (c_{min} - c_i - d)$$

$$\Leftrightarrow c_{min}(a - 1 - 2\alpha_i + \beta_i) \geq c_i (a - 1 - 2\alpha_i + \beta_i) - d(1 + 2\alpha_i - \beta_i)$$

$$\Rightarrow c_{min}(a - 1 - 2\alpha_i + \beta_i) > c_i (a - 1 - 2\alpha_i + \beta_i) - d\,(1 + 2\alpha_i - \beta_i)$$

Since $c_i < c_{min}$ and $d\,(1 + 2\alpha_i - \beta_i) > 0$, $c_{min}(a - 1 - 2\alpha_i + \beta_i)$ is strictly larger than $c_i(a - 1 - 2\alpha_i + \beta_i) - d\,(1 + 2\alpha_i - \beta_i)$. Thus, contributing $c_i = c_{min}\ \forall\, i \in S$ strictly dominates contributing $c_i < c_{min}\ \forall\, i \in S$ if $c_j = 0$ and $c_k = c_{min}$.

This leads to the intermediate result:

$\boldsymbol{U_i^*(c_{min}, 0, c_{min}) > U_i^*(c_i, 0, c_{min})\ \forall c_i \neq c_{min}, \forall i \in S}$ **if** $\boldsymbol{d > c_{min}(1 - a)}$**.**

Consider non-members next; it has to be shown that also no $i \notin S$ has an incentive to deviate if $\beta_i < 1 - a$. The utility function for all $i \notin S$ is (Equation (2.8)):

$$U_i = w - c_i + a\sum_{k\in S} c_k + a\sum_{j\notin S\setminus\{i\}} c_j + ac_i - \alpha_i \frac{1}{s}\sum_{k=1}^{s} \max\{u_k - u_i, 0\}$$
$$- \alpha_i \frac{1}{n-s-1}\sum_{i\neq j=1}^{n-s} \max\{u_j - u_i, 0\} - \beta_i \frac{1}{s}\sum_{k=1}^{s} \max\{u_i - u_k, 0\} - \beta_i \frac{1}{n-s-1}\sum_{i\neq j=1}^{n-s} \max\{u_i - u_j, 0\}$$

Given $c_j = 0$ and $c_k = c_{min}$, player i will contribute $c_i = 0$ instead of $c_i \neq 0$ if and only if:

$$U_i{}^*(0, 0, c_{min}) \geq U_i{}^*(c_i, 0, c_{min})\ \forall c_i > 0, \forall i \notin S$$

If everyone not in the institution contributes nothing (including $i \notin S$), whereas members contribute c_{min}, then player i gets advantageous disutility compared to members of the institution because of their contribution c_{min}. Thus we have:

$$U_i{}^*(0, 0, c_{min}) = w + a\sum_k c_{min} - \beta_i c_{min}$$

Non-members can get both disadvantageous utility compared to members as well as non-members, and advantageous disutility compared to members only. Thus the utility function becomes:

$$U_i{}^*(c_i, 0, c_{min}) = w - c_i + a\sum_k c_{min} + ac_i - \alpha_i(c_i - c_{min}) - \alpha_i c_i - \beta_i(c_{min} - c_i)$$

Again, there exist two cases: Player i can either contribute more than zero but less than the minimum contribution amount, i.e. $0 < c_i < c_{min}$ or he can chose to pay the minimum contribution amount, i.e. $c_i = c_{min}$. Again, a contribution above the minimum contribution amount, i.e. $c_i > c_{min}$, is ruled out by the experimental setting.

Case (1): First, consider the case $0 < c_i < c_{min}$. Player i will contribute $c_i = 0$ instead of $0 < c_i < c_{min}$ if and only if

$$\mathrm{U_i}^*(0, 0, c_{min}) \geq \mathrm{U_i}^*(\mathrm{c_i}, 0, c_{min}) \qquad \forall\, 0 < c_i < c_{min}, \forall \mathrm{i} \notin \mathrm{S}$$

$$\Leftrightarrow \qquad w + a\sum_k c_{min} - \beta_i c_{min} \geq \; w - c_i + a\sum_k c_{min} + ac_i - \alpha_i c_i - \beta_i(c_{min} - c_i)$$

$\Leftrightarrow 0 \geq c_i(a-1) - \alpha_i c_i + \beta_i c_i$

$\Leftrightarrow \beta_i c_i \leq c_i(1-a) + \alpha_i c_i$

$\Leftrightarrow \beta_i \leq 1 - a + \alpha_i$

$\Rightarrow \beta_i < 1 - a$

This condition strictly holds because, by assumption, $\alpha_i \geq \beta_i$ and $1 - a > 0$ because $0 < a < 1$. Thus, $U_i^*(0, c_{min}, c_{min}) > U_i^*(c_i, c_{min}, c_{min})$ $\forall c_i > 0, \forall i \notin S$ if and only if $\beta_i < 1 - a$.

Case (2): If $c_i = c_{min}$. Then

$$w + a\sum_k c_{min} - \beta_i\, c_{min} \geq w - c_{min} + a\sum_k c_{min} + ac_{min} - \alpha_i c_{min}$$

$\Leftrightarrow -\beta_i c_{min} \geq ac_{min} - c_{min} \; - \alpha_i c_{min}$

$\Leftrightarrow \beta_i \leq 1 - a + \alpha_i$

$\Rightarrow \beta_i < 1 - a$

This condition strictly holds because, by assumption, $\alpha_i \geq \beta_i$ and $1 - a > 0$ because $0 < a < 1$. Thus, $U_i^*(0, c_{min}, c_{min}) > U_i^*(c_i, c_{min}, c_{min})$ $\forall c_i > 0, \forall i \notin S$ if and only if $\beta_i < 1 - a$.

Thus, we have:
$\boldsymbol{U_i^*(0, 0, c_{min}) > U_i^*(c_i, 0, c_{min}) \; \forall c_i \neq c_{min}, \; \forall \, i \notin S}$ **if** $\boldsymbol{\beta_i < 1 - a}$.

Therefore, $\boldsymbol{c_i = c_{min}}$ **for all** $\boldsymbol{i \in S}$ **and** $\boldsymbol{c_i = 0}$ **for all** $\boldsymbol{i \notin S}$ **is a strict Nash equilibrium if** $\boldsymbol{d > c_{min}(1-a) \; \forall i \in S}$ **and** $\boldsymbol{\beta_i < 1 - a \; \forall \, i \notin S}$.

2.5.2 Experimental Instructions

General Information

Welcome to this experiment and thank you very much for your participation. Please read these instructions carefully. Please switch off your mobile phone and do not communicate with any other participant anymore. If you have any questions, please raise your hand. One of the experimenters will then come to your table and answer your question personally.

You can earn money in this experiment. The amount earned depends on your own decisions and the decisions made by the other participants of this experiment. All decisions made during the experiment as well as your final earnings remain anonymous.

In this experiment the currency Experimental Currency Unit (ECU) is used. At the end of the experiment all ECU amounts that you have earned in the course of the experiment will be summed up and converted to EURO, using the following exchange rate:

200 ECU = 1 Euro.

The Euro amount will then be paid out in cash. Additionally, each participant gets an amount of 2.50 Euro, which will be paid independently of the decisions made during the experiment.

You find the exact description of the experiment on the following pages.

Contribution Decision

The experiment consists of 20 identical decision rounds. At the beginning of each round all players are divided into groups of 4 players. The composition of the group is thereby randomized, which means that each group of four will be randomly determined before each new round starts.

At the beginning of each round, each participant gets an endowment of 40 ECU. Each player decides how much of the 40 ECU he wants to keep for himself and how much he wants to invest in a group project.

After all players have made their decision, each player is informed about the total amount, which has been invested in the group project. The individual contributions however remain anonymous. The individual earnings of each participant in one round consist of two parts:

1. The amount the participant kept for himself
2. Plus 0.65 times the sum of all contributions to the group project.

Therefore, the individual earnings of each participant in one round can be calculated with the following formula:

Earnings = 40 – individual contribution + 0.65 * joint contributions to the group project

Here are some examples. If all four group members (including yourself) decide to invest 0 ECU in the group project, then your individual earnings from this round are 40 – 0 + 0.65 x (4 x 0) = 40 ECU.

If you decide to invest 40 ECU, another group member also invests 40 ECU and the other two group members invest 0 ECU in the group project, then your individual earnings from this round are 40 – 40 + 0.65 x (40 + 40 + 0 + 0) = 52 ECU. (If you would have been one of the two players who invested 0 ECU, your individual earnings from this round would have been 40 – 0 + 0.65 x (40 + 40 + 0 + 0) = 92 ECU.)

If all four group members (including yourself) decide to invest 40 ECU in the group project, then your individual earnings from this round are 40 – 40 + 0,65 x (4 x 40) = 104 ECU.

The individual earnings from each round are publicized to the corresponding participant on the screen.

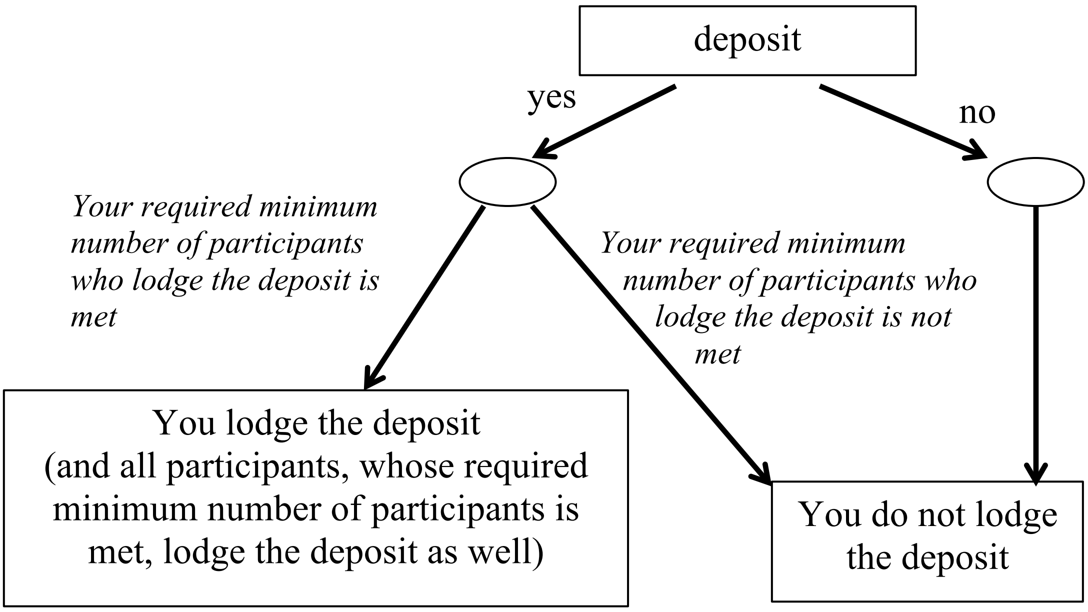
Possibility to Lodge a Deposit

Before making the above-explained contribution decision, each participant has the possibility to decide at the beginning of each round whether he wants to lodge a **deposit of 0 ECU [10**

ECU, 20 ECU, 40 ECU][30]. If a participant lodges the deposit, it will be only refunded if the participant invests **40 ECU** in the group project. The ECU amount, which is lodged as the deposit, can be provided **independently of the endowment of 40 ECU.**

A participant who wants to lodge the deposit can bind his choice on the following condition: He can decide on how many members of his group in total, **including himself**, at least have to lodge the deposit in order to make him want to lodge the deposit. Only if his required minimum number of participants is met, the participant actually lodges the deposit.

The following figure summarizes all possible courses of events of the deposit lodgments:



Thereafter, each player makes his contribution decision, i.e. each player decides how much of the 40 ECU he wants to keep for himself and how much he wants to invest in the group project (see section contribution decision, p. 2).

If you lodge the deposit but invest less than 40 ECU in the group project, then the deposit amount will be subtracted from your earnings in that round. Though, the laboratory team assures that your earnings in one round - with or without lodging the deposit – cannot be less than 0 ECU.

30 The instructions for IF0, IF10, IF20, and IF40 only differed in the deposit amounts. Putting the numbers used in IF10, IF20, and IF40 between brackets indicates this here. These were, of course, not part oft he initial instructions.

Thus, each round consists of the following two steps:

Step 1: Possibility to lodge a deposit conditioned on the minimum number of participants who lodge the deposit

Step 2: Contribution decision

The total earnings of each participant in this experiment are the sum of all earnings from all 20 rounds.

End of the Experiment

At the end of the experiment, the earnings from all rounds are summed up, converted into EURO and the amount plus 2.50 Euro is paid out to you in cash.

Please remain seated and wait until the experimenter will call your seat number to pay you the amount you have earned during this experiment.

Thank you very much for your participation!

When you have finished reading the instructions and all your questions are clarified, please press the button "forward" on your screen.

Control Questions

1. How many ECU do you get at the beginning of each round?

   ____

2. How many ECU can you maximally keep for yourself?

   ____

3. How many ECU can you maximally invest in the group project?

   ____

4. Does the composition of the groups of four players change during the experiment?

   ☐ yes ☐ no

5. Will your individual contribution to the group project be revealed to the other players?

   ☐ yes ☐ no

6. Will the sum of all contributions to the group project be revealed?

   ☐ yes ☐ no

7. Can you freely choose whether to lodge the deposit or not?

   ☐ yes ☐ no

8. If you lodged the deposit and choose to invest **less** than 40 ECU in the group project, will the deposit be refunded?

   ☐ yes ☐ no

9. If you lodged the deposit and choose to invest 40 ECU in the group project, will the deposit be refunded?

   ☐ yes ☐ no

10. When is your individual earning from one round higher? Possibility (1): If no one in the group lodges the deposit and all group members (including yourself) decide to invest 0 ECU in the group project. Possibility (2): If all group members lodge the deposit and all group members (including yourself) decide to invest 40 ECU in the group project?

    ☐ Possibility 1 ☐ Possibility 2

11. If you indicated that you lodge the deposit if at least 3 participants lodge the deposit, but there are only 2 participants **in total** willing to lodge the deposit, will you then lodge the deposit?

☐ yes ☐ no

12. If you indicated that you lodge the deposit if at least 3 participants lodge the deposit and there are 4 participants willing to lodge the deposit, will you then lodge the deposit?

☐ yes ☐ no

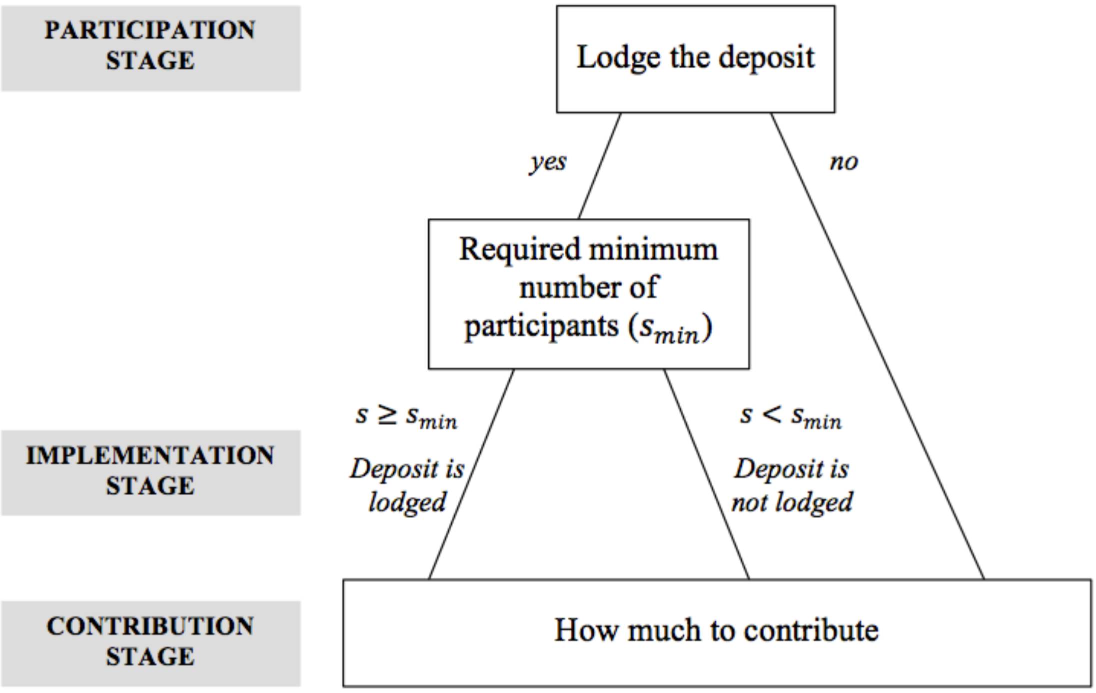
13. If you want **next to yourself at least 2 other** participants lodging the deposit, to lodge the deposit yourself, what is the required minimum number of participants you have to indicate?

☐ 1 ☐ 2 ☐ 3 ☐ 4

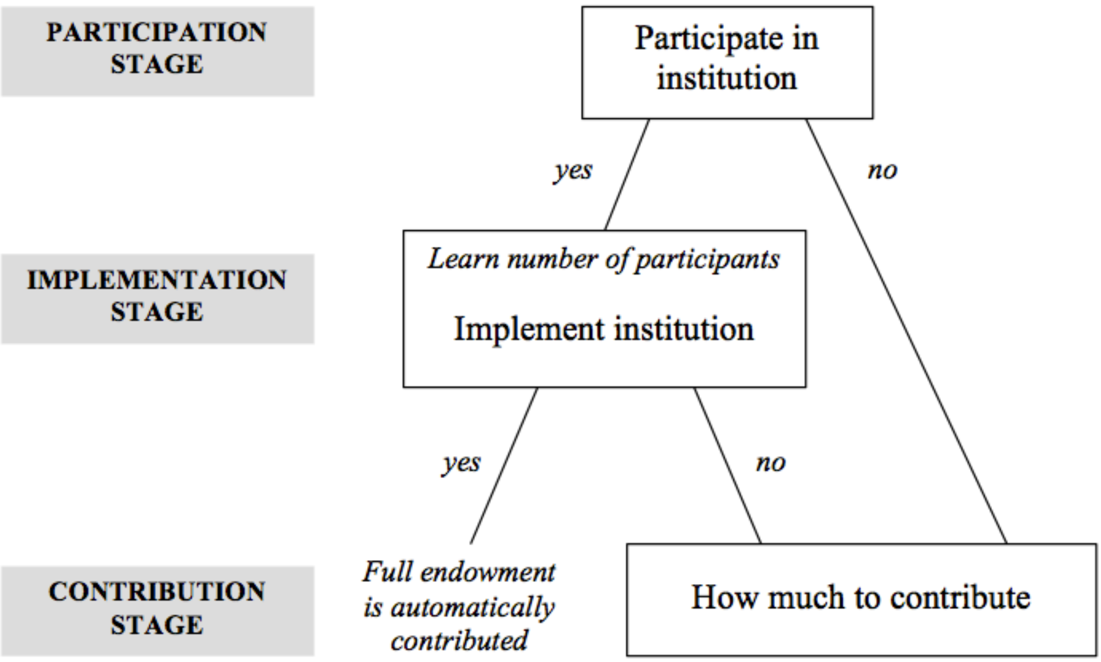
2.5.3 Decision Trees

FIGURE 2.4: OVERVIEW OF DECISIONS

(A) DECISION TREE EXPERIMENT THIS STUDY



(B) DECISION TREE EXPERIMENT KOSFELD ET AL. (2009)



Chapter 3

Can't Wait to Be Nice: Self-Control and Cooperation in an Intertemporal Trust Game

*The content of the following chapter was produced in collaboration with Mattia Nardotto and Bettina Rockenbach**

3.1 Introduction

Social dilemmas are situations, where private interests collide with societal concerns and, independent of individual intelligent pursuit of interest, no socially optimal outcome emerges. Classical examples are the provision of public goods or common pool resources. A vast literature analyzes behavioral pattern to understand potential underlying mechanisms for defective behavior in social dilemmas. Recent studies investigate whether the tension between the individual rationality of self-interest and collective wellbeing might also represent a *self-control dilemma,* where the individual has to trade off the objective of acting pro-socially (cooperative) and the temptation to act selfishly (Kocher et al., 2012; Martinsson et al., 2012).[31] Curry et al. (2008) argue that if a social dilemma evokes a self-control conflict, then individuals with stronger preferences for future over immediate rewards should be more likely to be more cooperative and that there is a positive correlation between (the individual degree of) self-control and cooperative behavior in a social dilemma. The experimental evidence reported so far suggests that a high level of self-control leads to significantly higher contributions in a public goods game if this is perceived as a self-control conflict (Kocher et al., 2012; Martinsson et al., 2012). Moreover, studies that link impatience and cooperation in a social dilemma agree that more impatient individuals contribute less to a public good than patient ones (Curry et al., 2008; Fehr & Leibbrandt, 2011) and defect more in a prisoner`s dilemma (Harris & Madden, 2002).

All these studies measure the individual degree of self-control independently from the contribution decision in the public good game, where the choice tasks to elicit time preferences consist of lotteries that embrace payoffs that differ very much from those in the cooperation game. Most importantly, the typical choice task to elicit time preferences does

* Financial support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

31 A self-control dilemma is defined as an interpersonal conflict between better judgment and temptation (Thaler & Shefrin, 1981; Schelling, 1984; Loewenstein, 1996, 2000).

not involve a second person. Rather, the decision maker's choices only affect his own payoff and therefore his decisions are made in a socially detached environment. However, if social dilemmas represent a self-control conflict, any choice involves *both* a temporal as well as a social dimension. Other research fields show that behavior changes when individual preferences interact with social preferences, such as a shift towards more conservative risk behavior when decisions are made in a social context than when made individually (Song, 2008; Reynolds et al., 2009; Charness & Jackson, 2009) or less liability towards overconfidence (Sniezek & Henry, 1989; Plous, 1995). If time and social preferences interact, there is the possibility of reinforcement or counterbalance. A person's social preferences might outweigh his individual present-bias. Or, ones' own self-control problem dominates the social element, as suggested by the literature. Eliciting both preferences separately and measure their correlation thereafter might therefore not be appropriate in all situations. Arguing that both preferences are relevant, it is not clear why the temporal component (individual degree of self-control) should predominate the social aspect and necessarily lead to less cooperation. Instead, it is possible that the choice differs profoundly for the decision maker if the intertemporal decision is made in interaction with another person. Addressing this problem is important for two reasons. Firstly, intertemporal decisions made in a social environment are ubiquitous, i.e. in a family, firm or institution. Second, if intertemporal decisions are different when being made in a social context, it is especially relevant for (the timing of) policy decisions. Moreover, it can affect the performance of a broad range of economic and social institutions (Frederick et al., 2002).

The aim of this study is to analyze whether there is an interaction between intertemporal preferences and cooperative behavior, using a novel experimental setting that addresses the problem of joint elicitation. To do so, we modify a standard time preference experiment to bring it into a social dimension. Behavior is then compared to the same choices made in a socially detached environment. This set up allows us to investigate the interplay between time preferences and cooperation in a more structured way. Especially, it offers the opportunity to analyze the correlation in behavior when both preferences interact.

Building on a standard experimental design to measure time preferences (i.e. Thaler, 1981; Tversky & Kahnemann, 1986; Benzion et al., 1989; Shelley, 1993; Green et al., 1994), we implement an experiment, where we let individuals state their preferences over intertemporal money streams of a smaller-sooner and larger-later reward. To be able to identify self-control problems, we ask subjects the same choice pair in the close and in the far future.[32] To capture the social component, we replace the usual lotteries of the time experiment by the trust game employed in Bohnet & Zeckhauser (2004). By doing so, we can explicitly disentangle stochastic risk from social risk. This is essential because both types of risk play an important role in our setting. For intertemporal choices, the literature agrees that the joint elicitation of stochastic risk and time preferences is crucial since both preferences are closely intertwined (Keren & Roelofsma, 1995; Anderhub et al., 2001; Andersen et al. 2008). Also cooperation in a social dilemma involves trusting the other persons and therefore is closely associated with the willingness to take risk (Ben-Ner & Puttermann, 2001; Camerer, 2003; Cook & Cooper, 2003; Kocher et al., 2011). This type of risk, however, is a social risk since it originates from the decisions of other human beings.[33]

We implement three treatments, each of which consists of two rounds. All three treatments offer exactly the same monetary payoff for the decision maker. Thus, if decision makers only maximize their payoffs and the probabilities of securing them, there should be no behavioral differences between the treatments. However, the context varies. Our main treatment is an intertemporal trust game. Choices made in the intertemporal trust game are compared to an intertemporal decision problem, where the payoffs and time frames are the same but no other subject is involved, as in a standard experiment for intertemporal preferences. The two treatments differ in two domains. Firstly, in the intertemporal trust game, the decision determines not only one's own payoff but also the payoff of another person. Hence, other-regarding preferences, such as inequality aversion, altruism or efficiency concerns might play a role. Secondly, in the intertemporal trust game another person – not nature – is the main source of uncertainty. Hence, next to the stochastic risk inherent in the lottery, some social risk might be added. If the bad outcome realizes, some additional psychological costs – so-called *betrayal costs* (Bohnet & Zeckhauser, 2004) – might have an influence on the decisions. To account for these elements, a third treatment – an intertemporal risky dictator game – is added, which has the same payoffs and probabilities as the intertemporal trust game, but the outcome is determined by nature. As in the trust game, however, another person is involved and payoffs depend on the decision maker's choices.

The main experimental findings are as follows. First, we can confirm the presence of betrayal aversion in our intertemporal setting. It seems that the importance of social risk

32 A self-control conflict occurs when time preferences are present-biased, which means that when considering trade-offs between two future moments, the relative weight given to the earlier moment is stronger when the moment gets closer (O'Donoghue & Rabin, 1999a).

33 Kocher et al. (2011) find that trust elicited by a trust game is significantly associated with public good contributions, whereas stochastic risk preferences elicited in a standard risk attitude elicitation experiment (i.e. Holt & Laury, 2002) do not correlate with contributions to the public good.

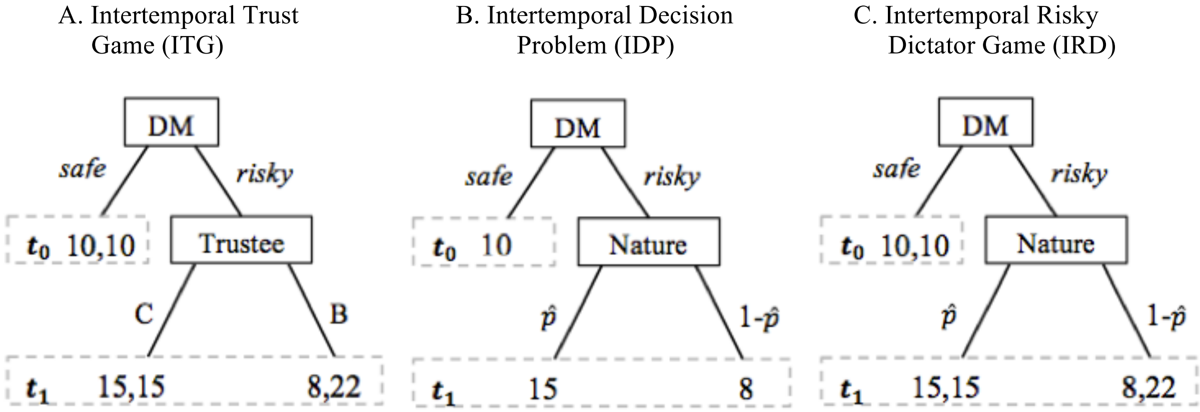
remains for intertemporal decisions. We do not, however, find that social comparison per se does play a role. The mere presence of a passive dependent does not make a difference in the decision maker's willingness to take risk. Second, we do not find any differences in behavior between present-biased individuals and those with time-consistent preferences in the intertemporal trust game. When making an intertemporal choice in interaction with another person and both dimensions are elicited jointly, individuals that display self-control problems when deciding only for themselves do not cooperate less than those individuals who behaved time-consistently in a socially detached environment. This result suggests that the social component outweighs the temporal aspect. Third, the share of present-biased choices decreases in the intertemporal dictator game compared to the individual intertemporal decision problem but not in the intertemporal trust game. A possible explanation is that responsibility has an influence on intertemporal decisions.

The rest of the paper is structured as follows. In section 3.2 the experiment is described in more details. Section 3.3 comprises the hypothesis development. In section 3.4 experimental results are presented. Section 3.5 concludes.

3.2 Experimental Design and Procedures

We consider the same binary-choice task in three decision situations: The intertemporal trust game (ITG), the intertemporal decision problem (IDP), and the intertemporal risky dictator game (IRD). In each treatment, the decision maker (DM) has a choice between a safe amount and a risky outcome. The safe strategy results in a safe outcome of 10, whereas the payoffs of the risky choice can be either higher (15) or lower (8) than the safe outcome. Figure 3.1 presents the three decision trees with the payoff structure employed in each.

FIGURE 3.1: DECISION TREES ALL TREATMENTS



We calibrate risk acceptance in the three decision situations by asking the DM about his minimum acceptable probability (MAP) of receiving the high payoff in order to prefer the risky choice to the safe amount. Subjects know that prior to the experiment a probability, $\hat{p}$, of receiving the high payoff has been determined, which is the same in all treatments. If their MAP is higher than $\hat{p}$, they earn the safe payoff. If their MAP is lower or equal to $\hat{p}$, they choose the risky choice. This implies that the less the DM likes the gamble, the higher his MAP should be.[34]

Our main treatment is the intertemporal trust game, where the risky choice consists of the trust game employed in Bohnet & Zeckhauser (2004). If the DM's MAP exceeds $\hat{p}$, the safe outcome is realized and the DM as well as the trustee earn (10, 10). If the DM's MAP is lower than or equal to $\hat{p}$, the DM participates in a trust game, where he allows the trustee to determine the final payoffs. Simultaneously, the trustee can either choose option C (cooperate) yielding outcomes (15, 15) or option B (betray) yielding outcomes (8, 22) for the DM and the trustee, respectively (see Figure 3.1, panel A). In case the trust game is played, the two final payoffs are then decided by the decision of the trustee with whom the DM has been randomly paired.

Decisions made in the intertemporal trust game are compared to the intertemporal decision problem. Payoffs are the same for the DM but no other person is involved. As before, if the DM's MAP exceeds $\hat{p}$, he earns the safe amount of 10. Yet, if the DM's MAP is lower than or equal to $\hat{p}$, the outcome of the risky option is determined by a random draw with a $\hat{p}$–chance of receiving 15 and a $(1 - \hat{p})$–chance of receiving 8 (see Figure 3.1, panel B).

The two treatments differ in two domains. First, in treatment ITG, another person – not nature – is the main source of uncertainty. Second, the decision also determines the payoff of another person. To control for these aspects, we run a third treatment: The intertemporal risky dictator game. In this treatment, as in treatment ITG, another person is affected by the DM's choices. The critical difference is that – as in treatment IDP – the outcome of the risky choice is determined by chance (see Figure 3.1, panel C).

As already pointed out, our experimental design introduces a novel element: The treatments have an intertemporal structure, in the sense that the timing of the realization of the decisions differs. The safe outcome is realized at an earlier date in time. The outcome of the risky choice, on the other hand, is paid out one day later. Every treatment consists of 2 decisions. In the first decision, the safe outcome is paid out one day after the experimental

[34] This mechanism is incentive compatible. The reported MAP of a rational decision maker should be the probability that makes him indifferent between the safe outcome and the risky outcome since individuals cannot affect the probability of the risky choice.

session (t_0= tomorrow).[35] The risky outcome is paid out one day later (t_1= in 2 days). To test for preference reversals, the second decision consists of the same choices at a later point in the future, where the safe outcome is paid out in 30 days (t_0 = in 30 days) and the risky outcome is paid out one day later (t_1 = in 31 days). Payoffs and their timing were presented to subjects in a neutral form (see Appendix 3.6.3).

The value of $\hat{p}$ is established from the trustees' decisions in the intertemporal trust game. While DM's indicate their required minimum probability of trustworthy behavior in order to trust, trustees simultaneously choose between option C and B.[36] The value of $\hat{p}$ then is the fraction of trustees who choose to reward trust in the trust game, with a distinct $\hat{p}$ calculated in each session. Thus, $\hat{p}$ is the probability that if the DM trusts, his trust is rewarded, or put differently, the probability of receiving 15 in the risky option.[37] The DMs are informed about this procedure. Trustees are not informed about the specifics of the MAP-procedure and do not know that their decision help to determine $\hat{p}$. In treatments IDP and IRD, we inform DMs that $\hat{p}$ has been determined before they make their decisions. The derivation of $\hat{p}$ is not revealed. Table 3.1 gives an overview of the three treatments.

TABLE 3.1: OVERVIEW TREATMENTS

Treat-ment	**Payoffs**		**Payoff Realization Round 1**	**Payoff Realization Round 2**	**Determinant of Risky Outcome**
	DM	Recipient/Trustee			
ITG	10 (15, 8)	10 (15; 22)	t_0 = tomorrow t_1 = in 2 days	t_0 = in 30 days t_1 = in 31 days	**Trustee's Choice**
IDP	10 (15, $\hat{p}$; 8, 1-$\hat{p}$)	- -	t_0 = tomorrow t_1 = in 2 days	t_0 = in 30 days t_1 = in 31 days	Nature
IRD	10 (15, $\hat{p}$; 8, 1-$\hat{p}$)	10 (15, $\hat{p}$; 22, 1-$\hat{p}$)	t_0 = tomorrow t_1 = in 2 days	t_0 = in 30 days t_1 = in 31 days	Nature

Treatment IDP allows us to measure people's (individual) time preferences. The difference in the MAPs between the short run future trade-off and the long run future trade-off reflects the compensation for delaying the monetary reward. We call this compensation *time compensation*. When comparing the MAP in treatment IRD to the MAP in treatment

[35] The earliest payment date (t_0) was chosen to be the next day to eliminate the problem of front-end delay.

[36] Thus, we applied the strategy method for trustees. The literature agrees that results in investment games obtained by the strategy method or with an experimental choice task do not differ (i.e. Ashraf et al., 2006).

[37] The elicitation procedure is similar to the Becker-DeGroot-Marshak mechanism (Benhabib et al., 2010). The main difference is that $\hat{p}$ is not generated randomly from a uniform distribution.

IDP, any additional compensation on top of the time compensation results from the presence of another person. We call this the *social compensation*. Finally, any difference between the MAP in treatment ITG and the MAP in treatment IRD reflects the extra compensation for potentially being betrayed by another person. We call this *betrayal compensation*.

Pairs were matched at the beginning, but payoffs of all three treatments were only shown at the end of the experiment to avoid players' behavior being influenced by experience.[38] At the end of the experiment, each player received an overview of all choices made in the experiment and the corresponding tokens earned in each round. Only then the participants could deduce the choices made by their partners. We used a within-subject design to be able to analyze potential behavioral changes depending on the different contexts. To avoid any order effects, the order of the treatments was randomized. The participants got new instructions after each round. The preference reversal literature suggests that people shall not see both decisions at the same time to avoid making any choices based on average calculations. The distribution of instructions was used to insure that there is some time between the decisions. Moreover, it gave us the possibility to apply both within as well as between treatment analyses. Since the former potentially could have been sensitive to order effects, the distribution of instructions after each round insured that players did not know that a larger-later question would come in the first round. If there were any order effects, we could have used only the first part of each session as an independent observation.

One round was randomly drawn to be payoff relevant. The payoff was transferred to each participant's bank account at the specific date to ensure that transaction costs were unitized among all participants. Moreover, the chair handed out a transferal guarantee to remove uncertainty about payments (see Appendix 3.6.4). Participants had the chance to leave before the experiment started getting the show up fee of 2.50 Euro if they did not want to or could not give us their bank account details. In total, only two people chose this option.

The experiments were run at the Cologne Laboratory for Economic Research. The experiment was programmed and conducted with the software z-tree (Fischbacher, 2007). In total, 86 subjects recruited via the organizational software Orsee (Greiner, 2004) participated in the experiment; 30 subjects in the first and third session and 26 subjects in the second session.[39] Subjects were randomly assigned to role A (DM) and B (recipient/ trustee) and then pairs of two were randomly matched (partner matching). In treatment IDP, all participants –

[38] Moreover, only after all players have participated in all treatments (and especially in treatment ITG), we could deduce $\hat{p}$ and thus derive final payoffs.
[39] Initially, we had a forth session with 30 more participants. These data had to be removed since the software broke down in the middle of the experiment.

irrespective of their role – were DMs. Thus, we have 15, 13, and 15 independent observations in session 1, 2, and 3, respectively, yielding 43 independent observations in total. Each treatment consisted of three sessions. Each session lasted about 60 minutes. On average a subject earned 7.50 Euro (about \$10.20) in the experiment.

3.3 Hypotheses

To capture the phenomenon of self-control problems, behavioral economists suggest various specifications of discounting (i.e. Loewenstein & Prelec, 1992; Laibson, 1997; O'Donoghue & Rabin, 1999a). In this paper, we adopt the quasi-hyperbolic framework to model intertemporal preferences due to its simplicity and tractability and predominant use in the literature. Moreover, it is perfectly in line with our behavioral assumptions and is compatible with present-biased preferences but also allows for time-consistent behavior.[40] Let u_t be a person's cardinal instantaneous utility function in period t and $U^t(u_t, u_{t+1}, \dots, u_T)$ represent a person's intertemporal preferences from the perspective of period t. Then, the hyperbolic structure assumes the following functional form for a person's intertemporal utility function:

$$U^t(u_t, u_{t+1}, \dots, u_T) = u_t + \beta \sum_{T=t+1}^{T} \delta^T u_T, \text{ for all } t \qquad (3.1)$$

where $\beta > 0$ and $\delta \leq 1$. In this formulation, δ^T represents the long run discount function, and β is the present-biased parameter. The quasi-hyperbolic discounting model assumes that the per-period discount rate between now and the next period is $\frac{1-\beta\delta}{\beta\delta}$, whereas the per-period discount rate between any two future periods is $\frac{1-\delta}{\delta} < \frac{1-\beta\delta}{\beta\delta}$ since $\beta > 0$. Hence, the (β, δ)-formulation assumes a declining discount rate between this period and next but a constant discount rate thereafter. This captures exactly the problem of self-control we consider here.

Applying the quasi-hyperbolic discount framework to our setting, a subject has present-biased preferences if the MAP for the question between tomorrow and in 2 days (short run future) is higher than for the equivalent question between 30 days and 31 days (long run future). On the contrary, a subject has time-consistent preferences if the MAP is the same in both time frames. In treatment IDP, we ask for the minimum probability, $\hat{p}_I$, for the high outcome (15) to arise in order to play the lottery in $t+1$ against the alternative to get 10 at date t. Then, the probabilities in the short and long run future for the DM are such that

[40] There also exist other models, such as models of willpower and temptation (i.e. Benabou & Tirole, 2004) or multiple selves' models (i.e. Fudenberg & Levine, 2006) to capture present-biased preferences.

$$u(10) = \hat{p}_I^S \beta\delta u(15) + (1 - \hat{p}_I^S)\beta\delta u(8) \tag{3.2}$$

$$\beta\delta^{30} u(10) = \hat{p}_I^L \beta\delta^{31} u(15) + (1 - \hat{p}_I^L)\beta\delta^{31} u(8) \tag{3.3}$$

Rewriting equations (3.2) and (3.3), the short run and long run future probabilities become

$$\hat{p}_I^S = \frac{u(10) - \beta\delta u(8)}{\beta\delta[u(15) - u(8)]} \tag{3.4}$$

$$\hat{p}_I^L = \frac{u(10) - \delta u(8)}{\delta[u(15) - u(8)]} \tag{3.5}$$

It can be easily seen that the two equations only coincide when $\beta = 1$. If instead $\beta < 1$, then the MAP asked in the short run future is larger than in the long run future. Let's call the difference in MAPs *time compensation* and define it as $\varepsilon_{time} \equiv \hat{p}_I^S - \hat{p}_I^L$. Then

$$\varepsilon_{time} = \frac{u(10) - \beta\delta u(8)}{\beta\delta[u(15) - u(8)]} - \frac{u(10) - \delta u(8)}{\delta[u(15) - u(8)]} = \frac{(1-\beta)}{\beta} \frac{u(10)}{\delta[u(15) - u(8)]} \tag{3.6}$$

Since $\frac{\partial \varepsilon_{time}}{\partial \beta} < 0$,[41] the time compensation increases with the extent of the present bias. Thus, the larger the bias for immediate gratification, i.e. the lower β, the higher the compensation a person asks for in order to participate in the lottery in the short run future. Note that our main focus is on the presence of present-biased preferences, and not on its intensity. Since we link self-control (not impatience) and cooperation, we opt for a simple design, abstaining from a more complicated preference elicitation procedure. Notice, however, that $\hat{p}_I^s$ and $\hat{p}_I^l$ are functions of other parameters than β, such as the risk attitude, i.e. the curvature of the utility function, and the impatience parameter δ. Because we are not interested in these parameters, we consider the difference in the MAPs, which is informative only on β, as the other factors affecting choice are held constant by design.[42] We can now deduce the first hypothesis:

HYPOTHESIS 3.1: Players with present-biased preferences ask for a higher MAP in the short run than in the long run future. Thus: $\varepsilon_{time} > 0$.

The second hypothesis assumes the existence of betrayal aversion. This implies that players ask for a higher premium when they have to trust another person compared to the case, where

[41] $\frac{\partial \varepsilon_{time}}{\partial \beta} = \frac{-u(10)*[\beta\delta(u(15)-u(8))]-[(\delta(u(15)-u(8)))*(u(10)-\beta u(10))]}{[\beta\delta(u(15)-u(8)]^2} < 0.$

[42] More precisely, the observed choices inform us only about whether $\beta < 1$, i.e. the existence of a present bias, not its intensity. In general, we cannot say how much of ε_{time} is due to β, δ and the risk attitude of the subject.

nature determines the outcome. This additional premium compensates for the psychological costs of (potentially) being betrayed. Let's define this cost of betrayal as $1 - \lambda$, where $0 < \lambda \leq 1$. Then, the probabilities for the short run and long run future decisions in the intertemporal trust game are such that

$$u(10) = \hat{p}_T^S \beta\delta u(15) + \left(1 - \hat{p}_T^S\right)\beta\delta\lambda u(8) \tag{3.7}$$

$$\beta\delta^{30}u(10) = \hat{p}_T^L \beta\delta^{31}u(15) + (1 - \hat{p}_T^L)\beta\delta^{31}\lambda u(8) \tag{3.8}$$

with short run and long run future probabilities

$$\hat{p}_T^S = \frac{u(10) - \beta\delta\lambda u(8)}{\beta\delta[u(15) - \lambda u(8)]} \tag{3.9}$$

$$\hat{p}_T^L = \frac{u(10) - \delta\lambda u(8)}{\delta[u(15) - \lambda u(8)]} \tag{3.10}$$

Comparing equations (3.9) and (3.10) to (3.4) and (3.5), shows that the additional costs of betrayal aversion further increase the MAPs. Thus:

HYPOTHESIS 3.2: Players ask for a higher premium in the intertemporal trust game compared to the intertemporal decision problem ($\hat{p}_T^S > \hat{p}_I^S$ and $\hat{p}_T^L > \hat{p}_I^L$) as well as the intertemporal risky dictator game ($\hat{p}_T^S > \hat{p}_D^S$ and $\hat{p}_T^L > \hat{p}_D^L$), both in the short run future and in the long run future.

The intertemporal risky dictator game may differ from the intertemporal decision problem by an additional *social compensation*. However, a clear hypothesis about the direction of this compensation is difficult to make. If players care about inequality aversion, the MAP should increase when moving from the one-person choice context to the social context of the dictator game. However, altruism or efficiency preferences would work in the opposite direction, increasing the attractiveness of the gamble and thus would lead to a reduction in the MAPs. Which of the two effects dominates is unclear ex-ante and might depend on other factors, such as the size of payoffs. However, there is ample evidence that people seem to care about disparities in payoffs, i.e. are inequality averse (Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000). If we incorporate social preferences into our utility function, using the inequity-aversion model suggested by Fehr and Schmidt (1999), the level of MAPs should increase in the intertemporal risky dictator game, both in the short run future and in the long run future compared to the intertemporal decision problem. To see this, the probabilities including inequality aversion are such that $u(10) = \hat{p}_D^S \beta\delta u(15) + (1 - \hat{p}_D^S)\beta\delta[u(8) - \propto (u(15) - u(8))]$ in the short run future and $\beta\delta^{30}u(10) = \hat{p}_D^L \beta\delta^{31}u(15) + (1 - \hat{p}_D^L)\beta\delta^{31}[u(8) - \propto (u(15) - u(8))]$ in the long run future. The parameter $\propto$ measures the decision maker's disadvantageous inequality aversion.[43] Following Fehr and Schmidt (1999), it is assumed that $\propto > 0$. Thus, we have $\hat{p}_D^S = \frac{u(10) - \beta\delta[u(8) + \propto(u(15) - u(8))]}{\beta\delta[u(15) - u(8) + \propto(u(15) - u(8))]}$ and $\hat{p}_D^L = \frac{u(10) - \delta[u(8) + \propto(u(15) - u(8))]}{\delta[u(15) - u(8) + \propto(u(15) - u(8))]}$, respectively. Comparing $\hat{p}_D^S$ and $\hat{p}_D^L$ to equations (3.4) and (3.5) immediately reveals that the probabilities to compensate for inequality aversion in the dictator game are larger than the time compensation in the individual decision problem. Since $\frac{\partial \hat{p}_D^S}{\partial \propto} < 0$ and $\frac{\partial \hat{p}_D^L}{\partial \propto} < 0$, the probability increases with an increase in the disadvantageous inequality aversion parameter $\propto$.[44] Therefore, we derive a hypothesis about the general ordering of the premiums. Let's define the social compensation as $\varepsilon_{social}^S \equiv \hat{p}_D^S - \hat{p}_I^S$ and $\varepsilon_{social}^L \equiv \hat{p}_D^L - \hat{p}_I^L$ and the betrayal compensation as $\varepsilon_{betrayal}^S \equiv \hat{p}_T^S - \hat{p}_D^S$ and $\varepsilon_{betrayal}^L \equiv \hat{p}_T^L - \hat{p}_D^L$. This leads to our third hypothesis:

HYPOTHESIS 3.3: Players, who are betrayal and inequality averse, ask for the highest premium in the intertemporal trust game, followed by the intertemporal risky dictator game and ask for the lowest premium in the intertemporal decision problem. Hence: $\varepsilon_{betrayal} > \varepsilon_{social} > \varepsilon_{time}$ *both in the short run and in the long run future.*

We can now move to the set of behavioral hypotheses on the interaction of time and social preferences. As already mentioned before, our design allows us to distinguish between present-biased individuals and time-consistent individuals. If present-biased preferences lead to less cooperation, subjects displaying these preferences should give up to temptation and trust less often (DM) and reward trust less often (trustees) in the intertemporal trust game. This implies that the MAPs of time-inconsistent subjects are higher than those of time-consistent ones. Moreover, decision maker and trustees with present-biased preferences should trust/ reward trust more often in the long run than in the short run future.

---

43 Note that, in our setting, a DM can never be better off than the recipient and thus there is no advantageous inequality aversion. The β used in our equation is the present-bias parameter, not the parameter for being better off as used in Fehr & Schmidt (1999).

44 $\frac{\partial \hat{p}_D^S}{\partial \propto} = \frac{-\beta\delta(u(15) - u(8)) * [\beta\delta(u(15) - u(8) + \propto(u(15) - u(8))] - [\beta\delta(u(15) - u(8)) * (u(10) - \beta\delta(u(8) + \propto(u(15) - u(8))]}{[\beta\delta(u(15) - u(8) + \propto(u(15) - u(8))]^2} < 0$ and $\frac{\partial \hat{p}_D^L}{\partial \propto} = \frac{-\delta(u(15) - u(8)) * [\delta(u(15) - u(8) + \propto(u(15) - u(8))] - [\delta(u(15) - u(8)) * (u(10) - \delta(u(8) + \propto(u(15) - u(8))]}{[\delta(u(15) - u(8) + \propto(u(15) - u(8))]^2} < 0.$

HYPOTHESIS 3.4: Decision makers with present-biased preferences trust less often than decision makers with time-consistent preferences and reward trust less often as trustees. Moreover, decision makers/ trustees with present-biased preferences trust less/ reward trust less in the short run future than in the long run future.

If, however, the interplay between intertemporal preferences and social preferences is such that the social element outweighs the intertemporal dimension, there should be little or no difference in behavior between present-biased and time-consistent individuals in the intertemporal trust game.

HYPOTHESIS 3.5: If the social dimension outweighs the intertemporal dimension, then there should be little or no difference in the choice to trust by decision makers who have present-biased preferences and those who have time-consistent preferences. The same holds for the decision to reward trust by the trustees.

3.4 Experimental Results

Before starting our analysis, a preliminary step is required. Since our subjects make choices in all treatments, and behavior in a treatment might change depending on the previous treatment, we check for order effects. However, we do not find significant order effects (Appendix 3.6.1). Moving to the analysis of subject behavior in the treatments, our evidence confirms the presence of present-biased preferences in the intertemporal decision problem. However, this difference disappears when intertemporal choices are made in interaction with another person, i.e. in a social environment.

RESULT 3.1: Overall, MAPs for the short run future are higher than the MAPs for the long run future in the IDP treatment. In treatments IRD and ITG, however, there is no difference in the MAPs between the short run and long run future.

Support for result 3.1 is presented in Figure 3.2, which shows the cumulative distribution of MAPs per treatment. It can be seen that there is a difference between the average short run future MAP (MAP_S) and the long run-future MAP (MAP_L) in the intertemporal decision problem, but neither in the intertemporal trust game nor the intertemporal risky dictator game. Table 3.2 (upper part) summarizes the average MAPs in the short run and long run future per treatment. The average MAP_S and MAP_L in the IDP treatment are around 57 and 55 percent, respectively. This difference is significant (Wilcoxon signed-rank test, p=0.0471). The average MAP_S and MAP_L are 76.09, 74.77 and 58.23, 58.02 for treatments ITG and IRD, respectively. This difference is not significant (Wilcoxon signed-rank test, ITG: p=0.6696; IRD: p=0.6566). Hence, we do find present-biased preferences in the decision problem and thus Hypothesis 3.1 cannot be rejected. However, choices are more time consistent in the other two treatments and the present-bias disappears.

FIGURE 3.2: CUMULATIVE DISTRIBUTION MAPs SHORT AND LONG

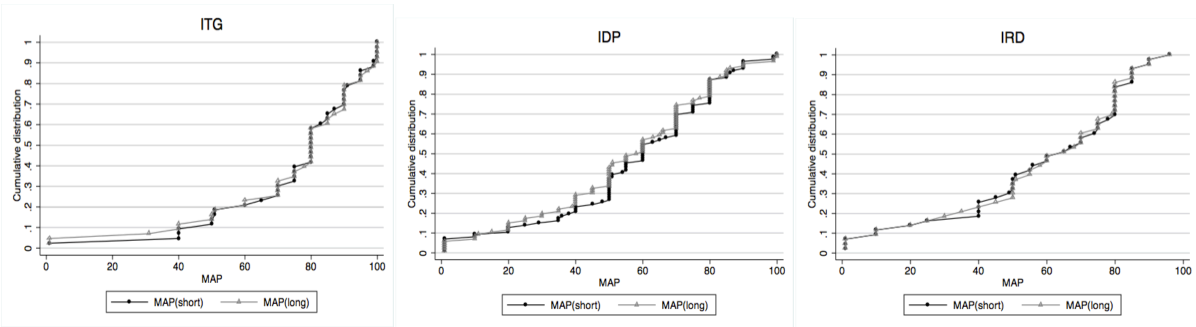


TABLE 3.2: AVERAGE MINIMUM ACCEPTABLE PROBABILITIES

	Treatment								
		ITG			IDP			IRD	
Average MAPs									
Short Run		76.1			57.3			58.2	
Long Run		74.8			54.9			58.0	
	future	cons.	present	future	cons.	present	future	cons.	present
Average MAP DM									
Short Run	74.8	74.3	83.6	48.3	55.8	66.1	67.2	53.3	71.0
Long Run	74.2	73.4	80.0	65.2	55.8	47.2	65.6	52.9	72.5
Reward Trust Trustees									
Short Run	0.20	0.42	0.33						
Long Run	0.40	0.38	0.25						

Note: The numbers depict percentages. Furthermore, future = future-biases, cons. = time-consistent, and present = present-biases.

Next, we consider the role of betrayal aversion. Our result confirms the finding of Bohnet & Zeckhauser (2004) and suggests that betrayal aversion also plays a crucial role in an intertemporal context.

RESULT 3.2: The premiums asked for in the intertemporal trust game are significantly higher than in the intertemporal decision problem as well as in the intertemporal risky dictator *game, both in the short run and in the long run future. Betrayal aversion persists in an intertemporal context.*

Consider Table 3.2 (upper part). Average MAP levels are around 57, 76, and 58 percent for the short run future in treatments IDP, ITG, and IRD, respectively. The average MAP_S is significantly higher in treatment ITG compared to both the IDP and IRD treatments (Wilcoxon signed-rank test, ITG vs. IDP: p=0.0000; ITG vs. IRD: p=0.0002) for the short run future. The same holds for the long run future, where the average MAPs are around 55, 75, and 58 percent in treatments IDP, ITG, and IRD, respectively. Again, the average MAP_L is significantly higher in treatment ITG compared to both the IDP and IRD treatments (Wilcoxon signed-rank test, ITG vs. IDP: p=0.0000; ITG vs. IRD: p=0.0002). Hence, we cannot reject Hypothesis 3.2.

In the following, we analyze whether social comparison affects behavior in our experiment. Results show that subjects do not behave differently in the risky dictator game than in the individual decision problem, neither in the short nor in the long run future. However, as pointed out in the previous section, we only observe the net result of opposing forces: inequality aversion, altruism or efficiency and their individual effect cannot be disentangled with our design. Hence, we cannot say whether the different social motives might cancel each other out or social comparison per se is not important in our environment.

RESULT 3.3: There is no significant difference between the premiums asked in the intertemporal decision problem and the intertemporal risky dictator game. Thus, on net, behavior is not influenced by social comparison.

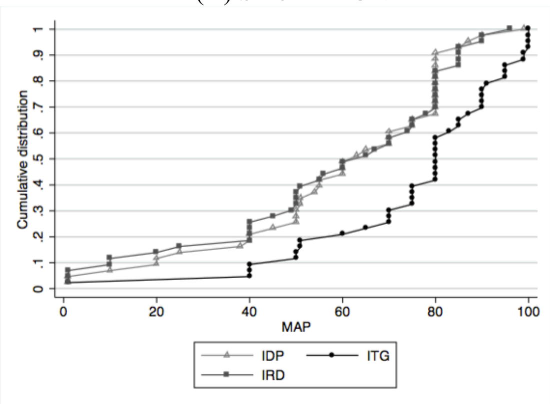
Again, consider Table 3.2 (upper part). The average MAP_S's are 57.31 and 58.09 in IDP and IRD, and the average MAP_L's are 54.87 and 58.02 in IDP and IRD, respectively. Thus, the premiums asked in IRD are slightly higher than in the IDP treatment, especially in the long run future. However, this difference is not significant (Wilcoxon signed-rank test, short: p=0.2236; long: p=0.1887). This finding is in line with the findings of Bohnet & Zeckhauser (2004) and Bolton & Ockenfels (2010).[45] Result 3.2 and 3.3 are summarized in Figure 3.3, which depicts the cumulative distribution of short and long run MAPs per treatment. It shows

[45] Bolton & Ockenfels (2010) find that social comparison plays an important role in risk-taking but only under certain conditions. When inequality results from the risky choice (as in our setting), risk-taking is not affected by social comparison.

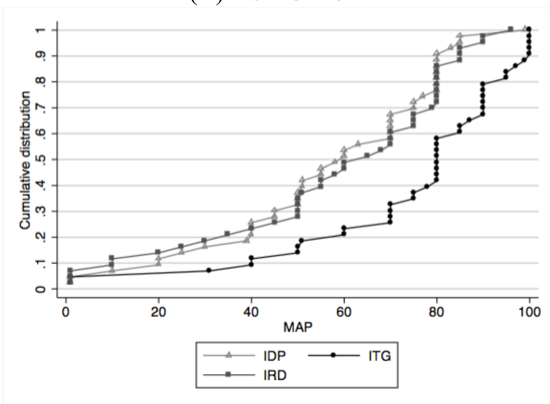
that the premium is similar in the decision problem and the dictator game. However, when there is social risk, premiums are much higher.

FIGURE 3.3: CUMULATIVE DISTRIBUTION MAPs ALL TREATMENTS

(A) SHORT RUN



(B) LONG RUN



Next, we analyze the order of premiums. Result 3.3 has shown that the difference of the MAPs between the intertemporal decision problem and the intertemporal risky dictator game is not significant. However, MAPs are slightly higher in the intertemporal risky dictator game than in the intertemporal decision problem (see Table 3.2). Indeed, we find that the order of premiums are as predicted both in the short as well as the long run future: $\varepsilon^s_{betrayal} > \varepsilon^s_{social} > \varepsilon_{time}$ and $\varepsilon^l_{betrayal} > \varepsilon^l_{social} > \varepsilon_{time}$. This ascending order is highly significant (Jonkheere Terpstra test for ascending order, short: p=0.9915; long: p=0.9998).[46] Thus, Hypothesis 3.3 cannot be rejected as well.

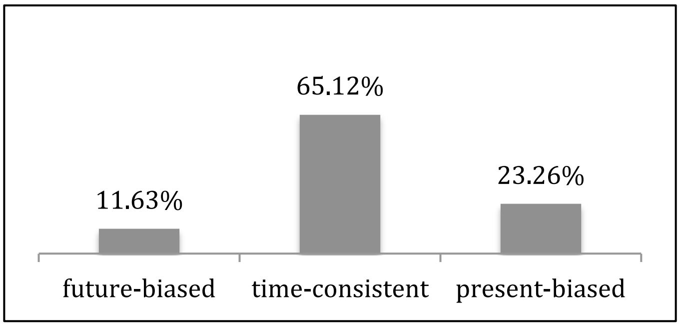
RESULT 3.4: Players ask for the highest compensation in the intertemporal trust game, followed by the intertemporal dictator game and the intertemporal decision problem. Thus, the betrayal compensation is higher than the social compensation, is higher than the time compensation.

For the following analysis, where time preferences are linked to cooperative behavior, we categorize individuals as present-biased ($\varepsilon_{time} > 0$), time-consistent ($\varepsilon_{time} = 0$) and future-biased ($\varepsilon_{time} < 0$). First, note that the distribution of time preferences in our subject pool is in line with general findings in the literature (Frederick et al., 2002). In our subject pool, about

[46] The Jonkheere Terpstra test for ascending order cannot reject the hypothesis that the order of alternatives is ascending von ε_{time} over ε_{social} to $\varepsilon_{betrayal}$ both for the short and long future horizon at the 1% level.

23 percent of participants have a bias for immediate gratification, 65 percent are time-consistent, and around 12 percent are future-biased (see Figure 3.4).

FIGURE 3.4: DISTRIBUTION OF TIME PREFERENCES



Now, we analyze the core question of our study: How are time preferences and behavior in the cooperation game interlinked. Remember, we hypothesized that if time preferences predominate the social component, DMs who have present-biased preferences should trust less often compared to people with time-consistent preferences and trustees should reward trust less often (Hypothesis 3.4). If, however, the social context outweighs the individual time preferences, there should be no difference in cooperative behavior between time-inconsistent and time-consistent individuals (Hypothesis 3.5). Our data suggest the following result:

RESULT 3.5: Players with present-biased time preferences and those with time-consistent preferences do not behave differently in the social context. We do not find a correlation between time preferences and cooperation behavior in the trust game.

Table 3.2 (lower part) summarizes average MAPs per treatment categorized by their time-preferences. It can be seen, that present-biased decision makers ask for higher MAPs and thus trust less often in the trust game than consistent individuals. Time-consistent individuals ask, on average, for a MAP of 74 percent in the short run versus 84 percent of time-inconsistent individuals. In the long run, average MAPs of consistent individuals lies at 74 percent and around 80 percent for people with present-biased preferences. This difference is not significant, neither in the short nor the long run future (Wilcoxon signed-rank, short: p=0.2563; long: p=0.9426). If we now look at the behavior of trustees, a similar picture arises. Trustees with present-biased preferences do reward trust less often both in the short (33 percent of trustees reward trust vs. 42 percent of time-consistent ones) and in the long (25 percent vs. 38 percent) but this difference fails to be significant (Wilcoxon signed-rank test, short: p=0.6037; long: p=0.4224). Hence, according to our evidence, Hypothesis 3.4 is rejected but Hypothesis 3.5 cannot be rejected.

Summarizing, our results suggest that – when using joint elicitation of time and social preferences – the social context outweighs individual problems of self-control. However, due to the small sample size, our results have to be treated with caution. The lack of significance could also be a result of lacking power. To check for the robustness of our results, we implemented a data simulation. Even with a simulated sample size of more than 500 subjects, the difference in the cooperation rate between subjects that displayed present-biased preferences in the individual decision problem and those who behaved time-consistently would become only marginally significant (see Appendix 3.6.2). This strengthens the robustness of our results.

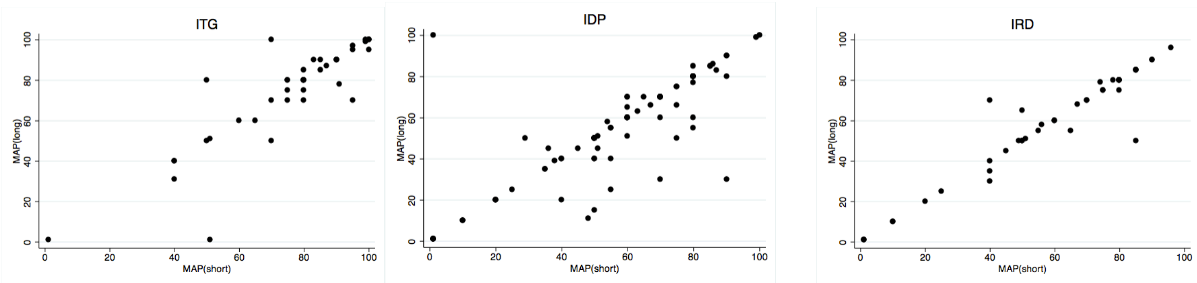
Finally, we have another interesting descriptive finding. We have seen that, on average, both in the intertemporal risky dictator game as well as in the intertemporal trust game, there is no significant difference between the short run and long run MAPs. However, when looking at the proportion of consistent and inconsistent choices, we find that choices become more consistent in the intertemporal risky dictator game compared to the intertemporal decision problem but not in the intertemporal trust game. Table 3.3 summarizes the percentage of consistent and inconsistent choices among treatments; 72 percent of choices are time-consistent in the IRD treatment compared to 65 and 55 percent in IDP and ITG, respectively. Most importantly, the proportions of present-biased choices are around 23 percent in IDP and ITG, but only 12 percent in the IRD treatment. This difference is marginally significant (One-sample test of proportion: p= 0.0355), whereas the proportions are the same between ITG and IDP (One-sample test of proportions, p=0.9943). Figure 3.5 depicts the short run and long run average MAPs per treatment. The forty-five-degree line indicates than the MAPs are the same and thus is the reference line for consistency. As can be seen, in the IRD game, choices seem to be more consistent than in the other two treatments. Choices are much more present-biased (below the forty-five degree line) in the IDP and do not change much in the ITG treatment.

It seems that people are more (time-) consistent when a passive dependent bears the consequences of their choices. However, in the trust game where the other person actively takes part in the decision process, behavior does not change. A possible explanation might be that responsibility influences subjects' behavior when facing intertemporal choices.

TABLE 3.3: PROPORTION CONS./ INCONS. INDIVIDUALS PER TREATMENT

	Treatment								
	IDP			ITG			IRD		
	future	cons.	present	future	cons.	present	future	cons.	present
Percentage of choices	11.63	65.12	23.26	20.93	55.81	23.26	16.28	72.09	11.63
Absolute Number	(10)	(56)	(20)	(9)	(24)	(10)	(7)	(31)	(5)

FIGURE 3.5: MINIMUM ACCEPTABLE PROBABILITIES SHORT AND LONG RUN



3.5 Conclusion

In the framework of this study, we investigate the interplay between self-control and cooperative behavior. Former studies find that there is a positive correlation between self-control and contributions to a public good (Curry et al., 2008; Fehr & Leibbrandt, 2011, Kocher et al., 2012; Martinsson et al., 2012). However, such correlation might arise due to the particular experimental setting. In these studies time preferences are elicited in a socially detached environment and linked to cooperative behavior thereafter. However, many intertemporal choices, such as policy decisions, simultaneously adhere a temporal as well as a social dimension. We address this problem by offering a novel experimental setting that is based on a joint elicitation of both preferences. We modify a standard time preference experiment to put it into a social context. To capture the social component, we replace the usual lotteries of time experiments with the trust game employed in Bohnet & Zeckhauser (2004). The results show that betrayal aversion persists in an intertemporal setting. The premiums are, on average, significantly higher when a trustee is involved. However, when another person silently bears the consequences of the decision maker's choices, premiums do not differ from the individual context. This is in line with previous findings that social comparison per se does not change behavior (Bohnet & Zeckhauser, 2004; Bolton & Ockenfels, 2010). Moreover, we do not find any difference in the trusting behavior between present-biased individuals and those with time-consistent preferences when eliciting social and time preferences simultaneously. People that display problems of self-control when only deciding for themselves, behave similar to time-consistent people in the intertemporal trust game. This result suggests that the social component outweighs individual time preferences.

We have an additional interesting descriptive result. The data suggest that choices become more time-consistent in the intertemporal dictator game but not in the intertemporal trust game. The core difference between the two treatments is that in the dictator game the decision maker decides for himself and a passive dependent who does not have any active role. In the trust game, however, the trustee actively participates in the decision process. This result opens the question of how responsibility might influence intertemporal choices. There are many studies that investigate the role of responsibility in variant contexts, such as risk aversion (Song, 2008; Charness & Jackson, 2009; Reynolds et al., 2009; Bolton & Ockenfels, 2010) and cooperation (Charness et al., 2007; Sutter, 2009; Engel & Rockenbach, 2011; Humphrey & Renner, 2013). However, as to our knowledge, there is no study that investigates the influence of responsibility on intertemporal choices. This remains an interesting question for future research.

The experimental setting suggested in this study goes beyond the existing literature by bringing a standard time preference experiment into a social dimension. It supplements the existing literature with the suggestion of an experimental design for the joint elicitation of time and social preferences in order to understand behavioral pattern that might govern the choices in an intertemporal context with social interaction. Our results suggest that people behave differently when both preferences interact than when choices are made separately. Hence, for several decisions a joint elicitation of time and social preferences might be more appropriate.

By eliciting time preferences in a social context, our experimental design reflects many real life decisions. Important examples are policy decisions, where the decision maker (politician) has to make intertemporal decisions while interacting with the members of the parliament and voters (social interaction). Our finding that the social aspect outweighs the intertemporal dimension in a cooperation game has important implications for policy decisions, i.e. with respect to the provision of public goods. Contradictory to previous findings in the literature, it suggests that the probability to contribute to a public good does not depend on the timing of the decision.

3.6 Appendix

3.6.1 Tests For Order Effects

TABLE 3.4: OVERVIEW OF ORDER EFFECT ANALYSIS

		Kolmogorov-Smirnov	Mann-Whitney U Test
		Intertemporal Decision Problem (IDP)	
IDP First vs. Not First	ε_{time}	p=0.781	p=0.3091
	$\hat{p}_I^s$	p=0.303	p=0.2354
	$\hat{p}_I^L$	p=0.336	p=0.1424
IDP First vs. IDP Second	ε_{time}	p=0.636	p=0.0524
	$\hat{p}_I^s$	p=0.467	p=0.1449
	$\hat{p}_I^L$	p=0.631	p=0.6408
IDP First vs. IDP Third	ε_{time}	p=0.855	p=0.3942
	$\hat{p}_I^s$	p=0.088	p=0.0096
	$\hat{p}_I^L$	p=0.320	p=0.0553
		Intertemporal Risky Dictator Game (IRD)	
IRD First vs. Not First	ε_{time}	p=0.921	p=0.3238
	$\hat{p}_D^s$	p=0.486	p=0.3958
	$\hat{p}_D^L$	p=0.328	p=0.2709
		Intertemporal Trust Game (ITG)	
ITG First vs. Not First	ε_{time}	p=0.563	p=0.4158
	$\hat{p}_T^s$	p=0.209	p=0.0886
	$\hat{p}_T^L$	p=0.209	p=0.0562
		$\varepsilon_{social} + \varepsilon_{betrayal}$	
IDP First vs. Not First	$(\varepsilon_{social} + \varepsilon_{betrayal})^S$	p=0.777	p=0.6360
	$(\varepsilon_{social} + \varepsilon_{betrayal})^L$	p=0.617	p=0.3375
IRD First vs. Not First	$(\varepsilon_{social} + \varepsilon_{betrayal})^S$	p=0.328	p=0.1302
	$(\varepsilon_{social} + \varepsilon_{betrayal})^L$	p=0.878	p=0.5863
ITG First vs. Not First	$(\varepsilon_{social} + \varepsilon_{betrayal})^S$	p=0.132	p=0.0534
	$(\varepsilon_{social} + \varepsilon_{betrayal})^L$	p=0.310	p=0.1379
		ε_{social}	
IDP First vs. Not First	ε_{social}^s	p=0.481	p=0.3091
	ε_{social}^L	p=0.481	p=0.3807
IRD First vs. Not First	ε_{social}^s	p=0.948	p=0.9551
	ε_{social}^L	p=0.535	p=0.0667
ITG First vs. Not First	ε_{social}^s	p=0.577	p=0.2840
	ε_{social}^L	p=0.563	p=0.3733
		$\varepsilon_{betrayal}$	
IDP First vs. Not First	$\varepsilon_{betrayal}^s$	p=0.966	p=0.8979
	$\varepsilon_{betrayal}^L$	p=0.586	p=0.3373
IRD First vs. Not First	$\varepsilon_{betrayal}^s$	p=0.230	p=0.0587
	$\varepsilon_{betrayal}^L$	p=0.148	p=0.0892
ITG First vs. Not First	$\varepsilon_{betrayal}^s$	p=0.269	p=0.0903
	$\varepsilon_{betrayal}^L$	p=0.080	p=0.0094

Note: Neither the epsilons nor the probabilities are normally distributed. The order of treatments has been IDP, IRD, ITG; ITG, IRD, IDP, and IRD, IDP, ITG.

3.6.2 Data Simulation

TABLE 3.5: AVERAGE MAPs SIMULATED DATA

	Treatment		
	ITG		Wilcoxon Signed Rank Test
	cons.	present	
	N=172		
Average MAP DM			
Short Run	74.3	83.6	p=0.1061
Long Run	73.5	80.0	p=0.9184
Reward Trust Trustee			
Short Run	0.42	0.33	p=0.4598
Long Run	0.38	0.25	p=0.2534
	N=258		
Average MAP DM			
Short Run	74.3	83.6	p= 0.1001
Long Run	73.5	80.0	p=0.8999
Reward Trust Trustee			
Short Run	0.42	0.33	p=0.3643
Long Run	0.38	0.25	p=0.1609
	N=344		
Average MAP DM			
Short Run	74.3	83.6	p= 0.0973
Long Run	73.5	80.0	p=0.8844
Reward Trust Trustee			
Short Run	0.42	0.33	p=0.2943
Long Run	0.38	0.25	p=0.1051
	N=430		
Average MAP DM			
Short Run	74.3	83.6	p= 0.0912
Long Run	73.5	80.0	p=0.8707
Reward Trust Trustee			
Short Run	0.42	0.33	p=0.2407
Long Run	0.38	0.25	p=0.0998
	N=516		
Average MAP DM			
Short Run	74.3	83.6	p= 0.0521*
Long Run	73.5	80.0	p=0.8585
Reward Trust Trustee			
Short Run	0.42	0.33	p=0.1985
Long Run	0.38	0.25	p=0.0874*

Note: Table 3.5 illustrates the minimum acceptable probabilities for the realization of the high outcome in the lottery in order to choose the gamble over the safe amount. This implies that the numbers depict percentages. Furthermore, cons. = time-consistent, and present = present-biases. *significant at 10%. Wilcoxon signed rank test: Difference between MAPs asked by present-biased and time-consistent individuals in the intertemporal trust game.

3.6.3 Experimental Instructions

General Information

Welcome to this experiment and thank you very much for your participation. Please read these instructions carefully. Please switch off your mobile phone and do not communicate with any other participant anymore. If you have any questions, please raise your hand. One of the experimenters will then come to your table and answer your question personally.

You can earn money in this experiment. The amount earned depends on your decisions as well as the decisions of other participants made during the experiment. All decisions made during the experiment as well as your final earnings remain anonymous.

In this experiment the currency Experimental Currency Unit (ECU) is used. This experiment consists of three parts with two rounds each. After each participant has made his decisions in all three parts, the individual earnings from all rounds in all three parts are shown to the corresponding participant on the screen. Then, *one of the rounds* is randomly chosen to be payoff-relevant; thereby, the probability of being chosen is exactly the same for all rounds. The ECU amount of the randomly drawn round will be converted to Euro, using the following exchange rate:

1 ECU = 0.50 Euro.

The timing of the payment depends on the choices you have made in the round that has been randomly chosen. Additionally, each participant gets an amount of 2.50 Euro, which will be paid independently of the decisions made during the experiment. The whole Euro amount will be transferred to your bank account. We assure you the certainty of your future payments.

All payment dates in this experiment are the date, where the bank transfer is *disposed.*

You will get the instructions for the second round of this part after having finished the first round. The instructions of parts 2 and 3 will then be distributed sequentially.

You find the exact description of the first round of part I on the following pages.

IDP: Round 1 [Round 2]

In this round, you have the choice between options O [Q] and P [R].

If you choose option O [Q], you will earn a safe amount of 10 ECU, which will be transferred to you **tomorrow** [**in 30 days**].

If you choose option P [R], you participate in a lottery, where you earn with probability p an income of 15 ECU (alternative 1) and with probability (1-p) an income of 8 ECU (alternative 2). Your earnings from option P will be transferred to you **in 2 days** [**in 31 days**].

Hence, the individual earning possibilities of each participant in this round are:

Option	Nature of Choice		Probability	Your Earnings	Payment
O	Certainty			10	tomorrow
[Q]					[in 30 days]
P	Lottery	1	p	15	in 2 days
[R]		2	(1-p)	8	[in 31 days]

The value of probability p that the lottery produces alternative 1 has been fixed in advance. Neither you nor any other participant of the experiment learns the true value of p. Your choice does not influence the value of p. We kindly ask you to answer the following question:

How large does the probability p of the lottery producing alternative 1 minimally have to be for you to pick option P [Q] over option O [R]?
(Please indicate the minimum probability in %!)

If your minimal value for the probability is **larger** than the actual probability, you automatically choose option O [Q] and earn 10 ECU in this round, which will be transferred to you tomorrow [in 30 days]. If your minimal value for the probability is **equal to or smaller** than the actual probability, you automatically participate in the lottery. Your earnings then depend on the outcome of the lottery, where you earn with a given probability of p 15ECU and with a given probability of (1-p) 8 ECU. The earnings from option P[R] will be transferred to you in 2 days [in 31 days].

After you have made your decision, the second round of part I [part II of the experiment] will begin.

When you have finished reading the instructions and all your questions are clarified, please press the button "forward" on your screen.

IRD: Round 1 [Round 2] [*Role B]

In this part of the experiment, you interact with another person. At the beginning of the experiment, you were randomly assigned to role A [*B]. Each participant with role A [*B] is randomly matched with a participant with role B [*A]. You will never learn the identity of the participant that you have been matched with. The matching remains constant during the whole experiment.

In this round, you have [*participant A has] the choice between options S [U] and T [V|. With your choice, you decide about your own earnings as well as the earnings of the participant with role B, which has been randomly assigned to you. [*With his choice, participant A decides about his own earnings as well as your earnings.] The participant with role B does not make any active decision. [*You do not make any active decision].

If you choose [*participant A chooses] option S [U], you as well as participant B [*A] will earn a safe amount of 10 ECU in this round, which will be transferred to both of you **tomorrow** [**in 30 days**].

If you choose [*participant A chooses] option T [V], you participate in a lottery, where you as well as participant B [*A] earn with probability p an income of 15 ECU (alternative 1) and with probability (1-p) you earn an income of 8 ECU [*22 ECU] and participant B [*A] earns an income of 22 ECU [*8 ECU] (alternative 2). The earnings from option T will be transferred to you as well as participant B [*A] **in 2 days** [**in 31 days**].

Hence, the individual earning possibilities of each participant in this round are:

Option	Nature of Choice		Probability	Your Earnings	Earnings Person B [*A]	Payment
S [U]	Certainty			10	10	tomorrow [in 30 days]
T	Lottery	1	p	15	15	in 2 days
[V]		2	(1-p)	8 [*22]	22 [*8]	[in 31 days]

The value of probability p that the lottery produces alternative 1 has been fixed in advance. Neither you nor any other participant of the experiment learns the true value of p. Your choice does not influence the value of p. We kindly ask you to answer the following question [*In the following, we kindly ask you to wait until participant A has answered the question]:

How large does the probability p of the lottery producing alternative 1 minimally have to be for you to pick option S [U] over option T [V]?
(Please indicate the minimum probability in %!)

If your [the] minimal value for the probability [*chosen by participant A who has been randomly assigned to you] is **larger** than the actual probability, you [*he] automatically choose [chooses] option S [U] and you as well as participant B [*A] earn 10 ECU in this round, which will be transferred to both of you tomorrow [in 30 days]. If your [the] minimal value for the probability [*chosen by participant A who has been randomly assigned to you] is **equal to or smaller** than the actual probability, you automatically participate in the lottery. Your earnings as well as the earnings of participant B [*A] then depend on the outcome of the lottery, where you as well as participant B [*A] earn with a given probability of p 15ECU and with a given probability of (1-p) you earn 8 ECU [*22 ECU] and participant B [*A] earns 22 ECU [*8 ECU]. The earnings from option T[V] will be transferred to you as well as participant B [*A] in 2 days [in 31 days].

After you [*participant A has] have made your [*his] decision, the second round of part II [part III of the experiment] will begin.

When you have finished reading the instructions and all your questions are clarified, please press the button "forward" on your screen.

ITG: Round 1 [Round 2] [*Role B]

In this part of the experiment, you interact with another person. At the beginning of the experiment, you were randomly assigned to role A [*B]. Each participant with role A [*B] is randomly matched with a participant with role B [*A]. You will never learn the identity of the participant that you have been matched with. The matching remains constant during the whole experiment.

In this round, you have [*participant A has] the choice between options W [Y] and X [Z].

If you choose [*participant A chooses] option W [Y], you as well as participant B [*A] will earn a safe amount of 10 ECU, which will be transferred to both of you **tomorrow [in 30 days].**

If you choose [*participant A chooses] option X [Z], your earnings depend on the choice made by participant B. [*your earnings as well as the earnings of participant A depend on your choice]. Participant B has [*You have] the choice between alternative 1 and 2. If participant B chooses [*you choose] alternative 1, you as well as participant B [*A] earn 15 ECU. If participant B chooses [*you choose] alternative 2, you earn 8 ECU [*22 ECU] and participant B [*A] earns 22 ECU [*8 ECU]. The earnings from option X will be transferred to both of you **in 2 days [in 31 days]**.

Hence, the individual earning possibilities of each participant in this round are:

Option	Nature of Choice	Your Earnings	Earnings Person B [*A]	Payment
W [Y]	Certainty	10	10	tomorrow [in 30 days]
X [Z]	Person B chooses 1 [*You choose] 2	15 8 [*22]	15 22 [*8]	in 2 days [in 31 days]

We kindly ask you to answer the following question:

How large would the probability p of being paired with a participant B who chooses alternative 1 minimally have to be for you to pick option X [Y] over option W [Z]?
(Please indicate the minimum probability in %!) *[*Which alternative, 1 or 2, do you choose in option X [Z]]*

Based on the decisions of all participants with role B, the actual probability that alternative 1 is chosen, will be calculated. The actual probability with which participant B choosing alternative 1 will not be revealed to you. Neither you nor any other participant of the experiment learns the true value of p. Your choice does not influence the value of p.

[*At the same time, all participants with role A indicate their minimal value for the probability that they are paired with a participant B who chooses alternative 1, in order to pick option X[Y] over option W[Z].]

If your [the] minimal value for the probability [*chosen by participant A who has been randomly assigned to you] is **larger** than the actual probability, you [*he] automatically choose [chooses] option W [Y] and you as well as participant B [*A] earn 10 ECU in this round, which will be transferred to both of you tomorrow [in 30 days]. If your [the] minimal value for the probability [*chosen by participant A who has been randomly assigned to you] is **equal to or larger** than the actual probability, your earnings as well as the earnings of particiant b [*A] depend on the choice of participant B [*your choice]. If participant B chooses [*you choose] alternative 1, you as well as participant B [*A] earn 15 ECU. If participant B chooses [*you choose] alternative 2, you earn 8 ECU [*22 ECU] and participant B [*A] earns 22 ECU [*8 ECU]. The earnings from option X [Z] will be transferred to you as well as participant B [*A] in 2 days [31 days].

[After you have made your decision, the second round of part III will begin.]

End of the experiment

When the experiment is finished, please remain seated and wait until the experimenter will call your seat number. Please fill in your bank account details in the destined slip of paper and take it with you when you come to the front. The team will hand out a certificate for your participation and the amount you earned during the experiment.

Thank you very much for your participation!

When you have finished reading the instructions and all your questions are clarified, please press the button "forward" on your screen.

3.6.4 Certificate for Transferal of Earnings

Universität zu Köln
Wirtschafts- und Sozialwissenschaftliche Fakultät
Staatswissenschaftliches Seminar
Prof. Dr. Bettina Rockenbach

Kölner Laboratorium
für Wirtschaftsforschung

Hiermit versichern wir Ihnen, dass wir die Überweisung in Höhe von

_______________ EURO am ______________ auf das von

Ihnen angegebene Konto veranlassen werden.

Dabei versichern wir, dass Ihre Bankdaten nur zur einmaligen Transaktion genutzt und danach uneingeschränkt gelöscht werden. Die erfassten Bankdaten dienen ausschließlich zur Auszahlung dieses wissenschaftlichen Experimentes.
Die Bankdaten werden nicht an Dritte weitergegeben.

Wir übernehmen keine Verantwortung für zeitliche Verzögerungen durch die Empfängerbank.

____
Datum, Unterschrift

Chapter 4

Responsibility Effects in Intertemporal Decision Making

*The content of the following chapter was produced in collaboration with Mattia Nardotto and Bettina Rockenbach**

4.1 Introduction

Intertemporal choices – decisions comprising trade-offs among costs and benefits occurring at different points in time – are among the most important choices made by economic agents. Whenever costs and benefits are spread over time, the possibility of meaningful comparison requires the calculation of present-value equivalents, which depends on the respective discount rate. Then, the discount rate used by an individual is a reflection of his subjective time preference (Harrison et al., 2002). In the standard experimental method to elicit time preferences (i.e. Thaler, 1981; Tversky & Kahnemann, 1986; Shelley, 1993; Green et al., 1994), subjects state their preferences over (a set of) trade-offs between a smaller-sooner and larger-later payment. By varying time frames and monetary amounts, the implicit individual discount rate can be inferred. The majority of people thereby display a tendency towards impatience (for an overview see Frederick et al., 2002). Moreover, enormous evidence suggests that people have present-biased preferences, which means that they assign especially high value to *immediate* gratification when choices are tempting compared to any future time (i.e. Rachlin et al., 1991; Kirby & Herrnstein, 1995; Myerson & Green, 1995; Kirby & Maracovic, 1996; Kirby, 1997; O'Donoghue & Rabin, 1999b, 2001; Harrison et al., 2002; Benhabib et al., 2010).[47] In all these studies, the typical choice task to elicit time preferences consists of decisions, where the decision maker's choices exclusively affect his own payoff.

The aim of this study is to analyze behavior when someone has to make an intertemporal decision while being responsible for another person, i.e. when the decision has payoff consequence not only for himself but also for someone else, whose time preferences are unknown to the decision maker. This is relevant because intertemporal decisions taken with responsibility are ubiquitous. For example, managers make intertemporal choices as the representative of a firm that affects all employees. Politicians make decisions that affect their constituents as well as society. Moreover, if responsibility has an impact on intertemporal

* Financial support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

[47] This implies, at the same time, that a person puts off an onerous activity more than he would like from a prior perspective (Fischer, 2001; O'Donoghue & Rabin, 1999b, 2001).

choices, it can affect the performance of a broad range of economic and social institutions, the depletion of non-renewable resources, and its further consequences even determines the economic prosperity of nations (Frederick et al., 2002).

Responsibility can possibly affect two aspects that influence intertemporal decision-making. First, the general level of patience – how much the person discounts rewards in the future – might be affected when someone else is involved. For example, managers may be more patient when making long run investment decisions for the firm to insure the financial stability and thus the employment security of their employees. Or, politicians display more impatience in their savings decisions because voters expect real-time outputs within the current legislative period. Second, being responsible for the welfare of another person may affect the time-consistency (degree of present bias) of choices. For example, managers are less likely to postpone tedious tasks if this endangers the salaries of their employees. Or, politicians are less likely to give in their present-bias if this harms their constituents. All the examples have a common feature: Neither a manager nor a politician knows the exact time preferences of their dependents. Hence, decisions have to be made based on the *estimation* of the preferences of the other person(s). This is in contrast to intertemporal choices made i.e. by parents who are responsible for their children, where the degree of patience of the dependents can be easily estimated.

To analyze the impact of responsibility on intertemporal choices, we implement two treatments. In our main treatment, we introduce responsibility into a standard intertemporal decision problem where a principal chooses between a smaller-sooner and larger-later payment, with such payments being also the payoff of a passive dependent who is not taking part in the decision process. The control treatment is an intertemporal decision problem, where only the decision maker is affected, as in a standard experiment for intertemporal preferences. Both treatments consist of the same intertemporal trade-offs but only differ by the presence of responsibility. Thus, if decision makers maximize their payoffs without being affected by responsibility, there should be no differences between the treatments.

The necessary condition that responsibility has an influence on choices is that it is *salient* to the decision maker that another person is affected by his choices, i.e. group saliency is the driving force for the difference in behavior between individual tasks and tasks with responsibility (Charness et al., 2007; Sutter, 2009). Group saliency can be created via payoff commonality and further increased by making the decision maker's choices accountable to the dependent by adding a feedback mechanism (Charness et al., 2007; Sutter, 2009).[48] To ensure group saliency in the main treatment, we allow the passive dependent to observe the principal's behavior by getting feedback about his choices after each round.

The psychological literature argues that responsibility would lead to more patient behavior, irrespective of whether the dependent is acquainted with the time preferences of the principal or not, because the principal wants to appear as rational agent (Charness & Jackson, 2009; Sutter, 2009). We offer an alternative explanation for the effect of responsibility on patience if the time preferences of the recipient are unknown to the principal. If the principal rationally maximize his payoff, being responsible for another person will not change behavior. If however, the principal's utility is influenced by concerns for other players (i.e. Rabin, 1993; Levine, 1998; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000), he aggregates preferences by a weighted utility function. If preferences are the same for the principal and the recipient, behavior does not change with responsibility compared to the individual decision problem. However, agents might have heterogeneous discount factors (Jackson & Yariv, 2014). To ensure that the principal does not know the time preferences of the dependent, we choose a homogeneous subject pool of students to impede obvious signals about intertemporal consumption preferences such as age, income and education, which have been shown to be closely correlated with the level of patience (i.e. Laibson, 1997, 1998; Harrison et al., 2002). Then, the principal has to make his decisions based on his expectations about the dependent's preferences. Depending on whether the principal expects the dependent to be more or less patient than himself, the aggregation of a weighted utility function will then lead to either more impatient or more patient choices. Moreover, in line with the argumentation of the psychological literature that consistent behavior is an important signal of personal and intellectual strength (Falk & Zimmermann, 2013), we expect to find more time-consistency with responsibility.

The main experimental findings are as follows. First, people decide more impatiently when they are responsible for the payoff of someone else. This is in line with our argument that principals use a weighted utility function of their own preferences and the unknown preferences of the recipient. Building on vast evidence from the psychological literature that a similar peer group leads to the better-than-average effect (Brown, 1986; Giladi & Klar, 2002; Gregg & Sedikides, 2004; Alicke & Govorun, 2005), the majority of principals seems to be overconfident about their own degree of patience, i.e. they expect the recipient to be more impatient than himself. Second, we find that choices are very time-consistent already in the

[48] Payoff commonality leads to group saliency not only in a strategic environment (Charness et al., 2007) but also in a non-strategic task without an out-group (Sutter, 2009).

baseline, such that an increase in the measure is difficult to induce. Hence, our results confirm a preference for consistency, as suggested by Falk & Zimmermann (2013).

The rest of the paper is structures as follows. Section 4.2 embeds the research question into the related literature. Section 4.3 encompasses the model and behavioral hypotheses. Section 4.4 presents the experiment in more detail. In section 4.5 the experimental results are analyzed. Section 4.6 concludes.

4.2 Literature Review

In recent years, various models of decisions assume a wider notion of self-interest (Sobel, 2005; Fehr & Schmidt, 2006), where utility is not exclusively driven by monetary incentives but also by concerns for other players (Rabin, 1993; Levine, 1998; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000; Charness & Rabin, 2002; Dufwenberg & Kirchsteiger, 2004). Considerable evidence shows that responsibility influences behavior. For example, laboratory evidence documents more conservative behavior in decisions under risk when one's choices affect other's welfare (Song, 2008; Charness & Jackson, 2009; Reynolds et al., 2009; Bolton & Ockenfels, 2010). Also, responsibility leads to a decrease in contributions to public goods (Charness et al., 2007; Sutter, 2009; Engel & Rockenbach, 2011; Humphrey & Renner, 2013). Sutter (2009) finds that salient group membership – where decisions are independent of others but influence them via payoff commonality – results in the same behavior as team membership, where several subjects have to agree on a joint decision and concludes, "both can be considered as substitutes" (p.2249).

The literature about team behavior is very rich and agrees that decisions made in a group are typically closer to standard game theoretic predictions than individual decisions in a broad variety of strategic and non-strategic tasks (for a survey of the literature see Cooper & Kagel, 2005). For example, groups are more selfish in a dictator game (Luhan et al., 2009), send and accept smaller transfers in an ultimatum game (Bornstein & Yaniv, 1998), play more strategically than individuals in signaling games (Cooper & Kagel, 2005), exhibit greater compliance with the risk-return principle of the portfolio-selection-theory (Rockenbach et al., 2007), trust less in a trust game (Kugler et al., 2007), cooperate less in a one-shot prisoner's dilemma game (see Bornstein, 2003), exit significantly earlier than individuals in a centipede game (Bornstein et al., 2004), and groups outperform individuals with respect to effectiveness in a beauty contest game (Kocher & Sutter, 2005). Thus, it seems that responsibility, in general, lead to more "rational" behavior. The psychological literature explains this "cautious shift" by the desire of the decision maker to avoid being seen as an irrational agent (Charness

& Jackson, 2009). Applied to an intertemporal framework, the (psychological) literature therefore predicts responsibility to increase patience among intertemporal choices.

Next, we consider the effect of responsibility on time-consistency. The literature agrees, even though individuals are often aware that their dynamic inconsistent time preferences lead to outcomes that they will regret in the future, myopic players have difficulties to overcome their present bias. However, it seems that people are well aware of the fact that their present-biases hinder them to make the "right" choices and see the accomplishment to overcome present-biased preferences as rational goal. A suggested solution is to offer people to bind their behavior (Strotz, 1956; Thaler & Shefrin, 1981; Schelling, 1992; Wertenbroch, 1998).[49] So far, empirical studies on self-control strategies have found that people, when giving the opportunity, use pre-commitment devices strategically in the anticipation of self-regulation failures (Wertenbroch, 1998; Trope & Fishbach, 2000; Ariely & Wertenbroch, 2002). Moreover, evidence from social psychology shows that consistent behavior is generally associated with personal and intellectual strength (Cialdini, 1993). Inconsistent behavior, on the other hand, is linked to undesirable personal characteristics (i.e. Asch, 1956; Allgeier et al., 1979). Falk & Zimmermann (2013) find that a key driver of consistent behavior is the role of signaling positive traits since consistent behavior is valued as a signal of ability. Hence, the presence of observers creates a particularly strong desire to appear consistent (Falk & Zimmermann, 2013). Based on these discourse insights, the literature assumes that responsibility leads to more time-consistency.

4.3 Model and Hypotheses

To derive predictions on the consequences of responsibility for decision-making in an intertemporal setting, we start from a model of choice that puts together both the elements we have discussed, namely patience and present bias. Such model is the quasi-hyperbolic decision model that we prefer to other models because of its tractability and relevance in the literature.[50] The model is also directly testable and allows us to separate, both theoretically and empirically, the implications on impatience from those on present-bias.

Decision is modeled as follows. Consider a decision-maker i who maximizes a time-additive discounted utility function, where u_t is his cardinal instantaneous utility function in period t and $U^t(u_t, u_{t+1}, \dots, u_T)$ represent his intertemporal preferences from the perspective

[49] Binding behavior is characterized by a strategic attempt to resist future temptations by voluntarily imposing (costly) constraints on future choices (Ariely & Wertenbroch, 2002).

[50] There also exist other models, such as models of willpower and temptation (Benabou & Tirole, 2004) or multiple selves' models (i.e. Fudenberg & Levine, 2006) to capture present-biased preferences.

of period t. The decision-maker has a discount factor $\delta \in (0,1)$, a present-bias parameter $\beta \leq 1$ and an increasing twice continuously differentiable utility function u from consumption, such that

$$U^t(u_t, u_{t+1}, \dots, u_T) = u_t + \beta \sum_{T=t+1}^{T} \delta^T u_T, \text{ for all } t \tag{4.1}$$

In this formulation, δ^T represents the long run discount function and β is the present-bias parameter. If $\delta^T = 1$, the person is perfectly patient. If instead $\delta^T < 1$, then future utility is discounted, and in particular a lower δ^T implies greater impatience. Moreover, $\beta = 1$ means that a person is time-consistent. If, however $\beta < 1$, the person's per-period discount rate between now and next period is larger than the per-period discount rate between any two future periods. This captures exactly the problem of present-bias discussed above. Moreover, this discount structure results in preference reversals as function of time.

Applying the quasi-hyperbolic discount framework to our setting, consider a binary decision problem in a close future framework, with $u(m)$ representing the utility the decision-maker gets when choosing the smaller-sooner amount available immediately, and $u(M)$ the utility for the larger-later amount (where $M > m$) available in t periods. Consider also the same decision shifted forward by τ periods (into the far future), where the delay of receiving the larger-later amount is denoted by T. Then, the decision-maker chooses the larger-later payment M in the close and far future if

$$u(m) < \beta\delta^t u(M) \tag{4.2}$$

and

$$\beta\delta^\tau u(m) < \beta\delta^{\tau+T} u(M) \tag{4.3}$$

Rewriting equations (4.2) and (4.3), the close and far future condition for delta in order to choose the larger-later payment becomes

$$\delta^t > \frac{u(m)}{\beta u(M)} \tag{4.4}$$

$$\delta^T > \frac{u(m)}{u(M)} \tag{4.5}$$

We refer to these thresholds as *patience thresholds*. Comparing equations (4.4) and (4.5), it can be easily seen that the two equations only coincide when $\beta = 1$, that is the case of time-consistency. If instead we have $\beta < 1$, then the close future patience threshold is larger than the far future threshold, i.e. $\delta^t > \delta^T$, which implies to choose less often the larger-later reward in the close future than in the far future. Since $\frac{\partial \delta^t}{\partial \beta} < 0$,[51] the smaller β, the more heavily the larger-later payment is discounted in the close future, so the more likely is the decision-maker to choose the smaller-sooner payment.

Now consider the case, where the decision-maker is a principal (denoted by P) who has to make a collective decision over streams of common consumption for himself and a dependent (denoted by D). The principal and the dependent have parameters $\delta_P, \delta_D \in (0,1)$, $\beta_P, \beta_D \leq 1$, and utility functions $u_P(m), u_D(m)$ and $u_P(M), u_D(M)$, respectively. The definition of having social preferences is that the principal's utility is influenced by concerns for the dependent, i.e. that u_P depends on u_D. Hence, the principal makes his decision based on a weighted utility function, with weights $\omega_P, \omega_D \in (0,1)$, such that $\omega_P + \omega_D = 1$. Moreover, we assume that some additional psychological costs enter the utility function if someone behaves inconsistently, in line with the presence of a preference for consistency. Let λ denote this cost of inconsistency, with $\lambda = 0$ if $\beta_P = 1$ and $\lambda > 0$ if $\beta_P < 1$. Then, the principal chooses the larger-later payment in the close and far future trade-off if

$$\omega_P u_P(m) + \omega_D u_D(m) < \beta_P \delta_P^t \omega_P u_P(M) + \beta_D \delta_D^t \omega_D u_D(M) - \lambda \tag{4.6}$$

$$\beta_P \delta_P^\tau \omega_P u_P(m) + \beta_D \delta_D^\tau \omega_D u_D(m) < \beta_P \delta_P^{\tau+T} \omega_P u_P(M) + \beta_D \delta_D^{\tau+T} \omega_D u_D(M) - \lambda \tag{4.7}$$

Rewriting equations (4.6) and (4.7), the principal's patience thresholds in the close and far future become

$$\delta_P^t > \frac{u_P(m)}{\beta_P u_P(M)} + \frac{\omega_D\left(u_D(m) - \beta_D \delta_D^t u_D(M)\right) + \lambda}{\beta_P \omega_P u_P(M)} \tag{4.8}$$

$$\delta_P^T > \frac{u_P(m)}{u_P(M)} + \frac{\omega_D\left(u_D(m) - \delta_D^T u_D(M)\right) + \lambda}{\omega_P u_P(M)} \tag{4.9}$$

Let`s first focus on the impatience parameter, δ_P^t, and assume $\beta_i = \beta_j = 1$ (and thus $\lambda = 0$). Note, that the general level of patience is independent of whether the trade-off is asked in the close or far future. Comparing equations (4.8) and (4.9) to equations (4.4) and (4.5), it can be easily seen that the patience threshold increases with a decrease in δ_D^t, which means that the

---

51 $\frac{\partial \delta^t}{\partial \beta} = -\frac{u(m)}{\beta^2 u(M)} < 0$.

larger the degree of impatience of the dependent, the higher the patience needed for the principal to opt for the larger-later payment. There are three cases. If $\delta_P^t = \delta_D^t$, the condition to choose the larger-later payment becomes $\delta_P^t > \frac{u_P(m)}{u_P(M)}$, the same as in the individual case. Hence, if the preferences of the two agents are the same, then behavior with responsibility does not differ from individual choices. If instead $\delta_P^t < \delta_D^t$, i.e. the passive dependent is more patient than the principal, then the patience threshold, δ_P^t, decreases and behavior becomes more patient. If, however, $\delta_P^t > \delta_D^t$, the opposite is true: The patience threshold increases, resulting in more impatient choices by the principal. We can now deduce the first hypothesis:

HYPOTHESIS 4.1: If the principal's utility is influenced by concerns for the dependent, responsibility leads to an increase in impatience among intertemporal choices if the principal expects himself to be more patient than the dependent, i.e. $\delta_P^t > \delta_D^t$ or a decrease in impatience if the principal expects himself to be more impatient than the dependent, i.e. $\delta_P^t < \delta_D^t$. If the principal expects the dependent to have the same time preferences as himself, i.e. $\delta_P^t = \delta_D^t$, behavior with responsibility does not differ from individual choices.

Next, we analyze the effect of weighting utilities on time-consistency. Consider equation (4.8). Since $\frac{\partial \delta_P^t}{\partial \beta_D} < 0$,[52] the patience threshold of the principal increases with the extent of the present-bias of the dependent. Thus, the larger the dependents bias for immediate gratification, i.e. the lower β_D, the higher the patience threshold in the close future and the more likely is the principal to choose the smaller-sooner payment, which would lead to time-inconsistent choices. However, if the principal has a preference for consistency, the psychological cost of displaying time-inconsistent behavior work in the opposite direction: The larger λ, the more the principal wants to behave consistently. If the principal suffers enough from behaving inconsistently, he will behave more consistently when being responsible for someone else.

HYPOTHESIS 4.2: Responsibility and accountability lead to an increase in time-consistency among intertemporal choices if the principal has a preference for consistency.

[52] $\frac{\partial \delta_P^t}{\partial \beta_D} = -\frac{\omega_D \delta_D^t}{\beta_P \omega_P} < 0.$

4.4 Experimental Design and Procedures

Our goal is to analyze how responsibility affects intertemporal choices. Our main treatment is an intertemporal choice task with responsibility (ICR), where we modify a standard time experiment. The decision maker (DM) is a principal whose choice is also the payoff of a passive dependent. The dependent bears the consequences of the choice without being able to actively take part in the decision process but can silently observe the DM's choices. We then compare behavior with responsibility to a standard intertemporal choice task (ICI), where the DM's choices only affect his own payoff. Moreover, no one observes the DM's choices to be able to reveal his true preferences and therefore get his upper bound of impatience. Both treatments offer exactly the same monetary payoffs for the DM. However, the treatments vary by the presence of responsibility: In treatment ICR the DM is responsible for the payoff of a passive dependent, whereas in treatment ICI no other person is affected by his choice.

To estimate the impatience parameter, δ_P, we employ 40 trade-offs between a smaller-sooner and a larger-later payment with varying money amounts and moments of payments.[53] To be able to identify present-biased preferences, we divide the choice trial into two blocks. 20 trade-offs belong to the close future block, where the smaller-sooner payment is paid out one day after the experimental session (t_1= tomorrow)[54] and the larger-later amount with varying days of delay (t_{1+D}= tomorrow+delay). We then add the same 20 trade-offs shifted forward by 30 days that depict the far future block, where the smaller-sooner payment is paid out in 30 days (t_{30}= in 30 days) and the larger-later payment with varying days if delay (t_{30+D}= in 30 days+delay). In our experiment, provided that t_1 is close enough in time and t_{30} is far enough in time, subjects with present-biased preferences should choose the smaller-sooner payment if the trade-off is between t_1 and t_{1+D} (display a present-bias in the trade-offs of the close future block), while they should reverse their preferences if the tradeoff is between t_{30} and t_{30+D} (wait for the same questions in the far future block). Hence, our design allows us to estimate jointly the impatience measure, δ_P, and the present-bias parameter, β_P.

Additionally, we calculate the patience threshold for the principal, δ_P^t, which equalizes the present value of the delayed reward (PV) and the value for the smaller-sooner reward according to equation (4.1). Thus, we calculate $PV = \frac{A}{1+\rho D}$, where A the amount of the delayed reward, D the delay and ρ the discount rate parameter with $\delta^t = \frac{1}{1+\rho D}$ representing the

[53] With this wide range of questions we reduce the problem of noise which data from time-preference elicitation in the laboratory are especially sensitive to (Kirby & Herrnstein, 1995; Kirby, 1997).

[54] The earliest payment date was chosen to be the next day to eliminate the problem of front-end delay.

hyperbolic discount function. Table 4.1 presents all questions and the corresponding patience thresholds.

TABLE 4.1: CHOICE TRIALS AND THEIR ASSOCIATED PATIENCE THRESHOLDS

	Choice Trial Close Future	Choice Trial Far Future	δ_P^t
1.	10.00€ tomorrow or 10.10€ in 21 days	10.00€ in 30 days or 10.10€ in 50 days	0.9900
2.	8.00€ tomorrow or 8.10€ in 27 days	8.00€ in 30 days or 8.10€ in 56 days	0.9877
3.	7.00€ tomorrow or 7.10€ in 6 days	7.00€ in 30 days or 7.10€ in 35 days	0.9859
4.	10.00€ tomorrow or 10.20€ in 3 days	10.00€ in 30 days or 10.20€ in 32 days	0.9804
5.	6.00€ tomorrow or 6.20€ in 13 days	6.00€ in 30 days or 6.20€ in 42 days	0.9675
6.	11.00€ tomorrow or 11.50€ in 29 days	11.00€ in 30 days or 11.50€ in 58 days	0.9571
7.	8.00€ tomorrow or 8.50€ in 11 days	8.00€ in 30 days or 8.50€ in 40 days	0.9469
8.	5.00€ tomorrow or 5.50€ in 18 days	5.00€ in 30 days or 5.50€ in 47 days	0.9117
9.	9.00€ tomorrow or 10.00€ in 31 days	9.00€ in 30 days or 10.00€ in 60 days	0.9000
10.	7.00€ tomorrow or 8.20€ in 19 days	7.00€ in 30 days or 8.20€ in 48 days	0.8540
11.	6.00€ tomorrow or 7.70€ in 5 days	6.00€ in 30 days or 7.70€ in 34 days	0.7792
12.	11.00€ tomorrow or 14.20€ in 5 days	11.00€ in 30 days or 14.20€ in 34 days	0.7747
13.	9.00€ tomorrow or 11.70€ in 15 days	9.00€ in 30 days or 11.70€ in 44 days	0.7695
14.	4.00€ tomorrow or 5.60€ in 12 days	4.00€ in 30 days or 5.60€ in 41 days	0.7180
15.	9.00€ tomorrow or 13.50€ in 31 days	9.00€ in 30 days or 13.50€ in 60 days	0.6662
16.	7.00€ tomorrow or 10.60€ in 5 days	7.00€ in 30 days or 10.60€ in 34 days	0.6603
17.	7.00€ tomorrow or 11.50€ in 12 days	7.00€ in 30 days or 11.50€ in 41 days	0.6089
18.	5.00€ tomorrow or 8.60€ in 21 days	5.00€ in 30 days or 8.60€ in 50 days	0.5814
19.	6.00€ tomorrow or 11.00€ in 11 days	6.00€ in 30 days or 11.00€ in 40 days	0.5453
20.	4.00€ tomorrow or 7.70€ in 8 days	4.00€ in 30 days or 7.70€ in 37 days	0.5196

Note: δ_P^t captures the minimum value of the hyperbolic discount function $\delta_P(t)$ at which the immediate and delayed rewards are of equal value to the principal. Trials are shown in decreasing order of δ_P^t.

In the table, the trade-offs are organized such that the patience threshold is decreasing. Note that the higher the discount function δ_P^t implied by the observed choice, the more patient a person is. For the first trade-off, the patience threshold is very high, which implies that choosing the larger-later payment requires a very high degree of patience ($\delta_P^t \geq 0.99$). However, consider the last question (row 20). For this trade-off, choosing the larger-later payment requires less patience ($\delta_P^t \geq 0.5196$). The average discount factor lies at around $\delta(t) = 0.8$ (for a meta analysis, see Frederick et al., 2002). Hence, we call trade-offs with a patience threshold above this value (choice 1-10 [21-30]) *high patience threshold* and expect less people to choose the larger-later payment than for trade-offs with a patience threshold below this value (choice 11-20 [31-40]), which we call *low patience threshold*.[55]

The order of all 40 trade-offs was randomized and each choice was asked separately to avoid any order effects. Subjects were informed about the structure and number of trade-offs

[55] Of course, it is also possible that subjects simply use heuristics for absolute monetary amounts ("I only wait for at least one Euro more") as well as the waiting time ("I do not wait for more than 10 days").

but not that the same trade-offs are present in the close future and in the far future (see instructions, Appendix 4.7.1). We opted for this procedure, instead of displaying the whole choice trial at once, to avoid that subjects could potentially bias their choices (maybe towards more consistent behavior). One round was randomly drawn to be payoff relevant. Earnings were transferred to each participant's bank account at the specific date to unitize transaction costs among all participants. Moreover, the chair handed out a transferal guarantee to remove uncertainty about payments (see Appendix 4.7.2). Participants had the chance to leave before the experiment and get the show-up fee of 2.50 Euro if they did not want to or could give us their bank account details. In total, only one subject left the experiment.

The experiments were run at the Cologne Laboratory for Economic Research. It was programmed and conducted with the software z-tree (Fischbacher, 2007). In total, 182 students recruited via the organizational software Orsee (Greiner, 2004) participated in the experiment; 64 subjects in treatment ICI and 118 subjects in treatment ICR. We used a between-subject design. In the ICR treatment, subjects were randomly assigned to role A (principal) and B (recipient) and then pairs of two were randomly formed (partner matching). In treatment ICI, all participants were principals. Thus, we have 64 independent observations in the control treatment and 59 independent observations in the responsibility treatment, yielding 123 independent observations in total. Each treatment consisted of three sessions. Each session lasted about 60 minutes. On average, subjects earned 8.93 Euro (about $12.16) in the experiment.

4.5 Experimental Results

4.5.1 Descriptive Results

The result section is structured as follows. First, we address the question of how responsibility influences patience. Next, we investigate whether present-biased time preferences change with responsibility. To complement the analysis, we then estimate the individuals' discount function with maximum likelihood (MLE).

First, people behave according to delta. With a decrease in the patience threshold, the share of people that opt for the larger-later payment (choice Y) increases, such that for the high patience threshold trade-offs, around 53 (48) percent of people wait in treatments ICI (ICR), whereas for the low patience threshold trade-offs, over 94 (91) percent wait (see Table 4.2, upper part; see also Table 4.3). Second, the majority of people act highly consistently (see Table 4.2, lower part). Since we made it difficult for people to recognize that the same trade-offs are present both in the close and in the far future, it is unlikely that this time-consistent behavior is subject to potential biases towards consistent behavior induced by the design.

TABLE 4.2: DECISIONS PER TREATMENT/ DISTRIBUTION OF TIME PREFERENCES

	Treatment					
	ICI			ICR		
	X	Y		X	Y	
Overall Choices	25.90	74.10		29.15	70.85	
High Threshold	47.42	52.58		51.95	48.05	
Low Threshold	5.97	94.03		8.40	91.60	
	Future	Consistent	Present	Future	Consistent	Present
Overall Choices	4.22	90.08	5.70	3.90	90.68	5.42

Note: The upper part of Table 4.2 shows the percentages of X or Y chosen. The lower part of Table 4.2 illustrates the distribution of time preferences per treatment. The numbers in both parts depict percentages; future=future-biased, consistent=time-consistent, and present=present-biased. N=32 for ICI and N=27 for ICR.

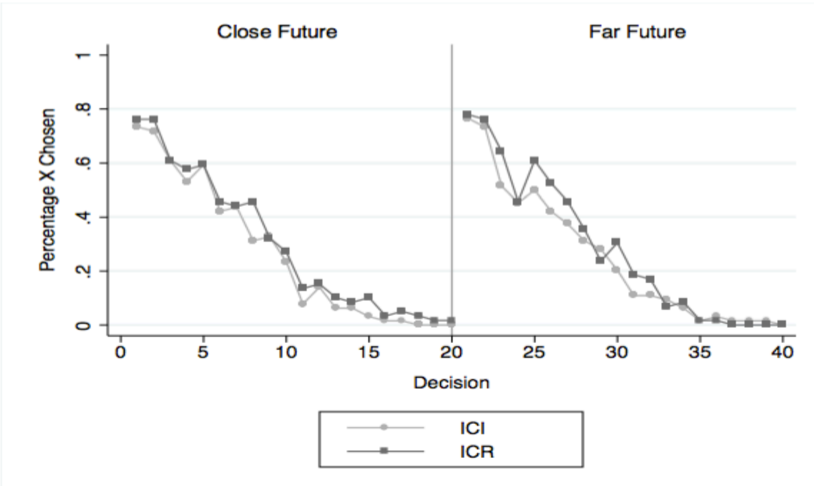
Moving to the analysis of subject behavior in the treatments, we first analyze the effect of responsibility on the impatience parameter, δ. Our evidence shows that, on average, choices are more impatient when made with responsibility.

RESULT 4.1: People decide more impatiently when being responsible for someone else.

Table 4.2 (upper part) depicts the shares of people that chose the smaller-sooner (choice X) or larger-later payment (choice Y). As can be seen, 74 percent choose Y in treatment ICI compared to 71 percent in ICR. This difference is significant (Two-sample test of proportion: p=0.0106). This result is also shown in Figure 4.1, which depicts impatient behavior, by showing the share of people that chose X per treatment. Both, in the close and in the far future, the share of people that chose the smaller-sooner payment is for all choices always larger in treatment ICR than in treatment ICI.

As argued above, the principal does not know the discount parameter of the dependent and therefore has to base his choice on his expectations about δ_D^t. The weighted discount function is lower than the individual discount function of the principal if the principal expects the discount parameter of the dependent to be smaller than his own, i.e. if he expects the dependent to be more impatient than himself. Hence, our result suggests that the majority of principals is overconfident about their own degree of patience and expect the dependent to be more impatient than themselves. This is in line with the empirical evidence from social psychology that, when comparing oneself to the aggregate, any single member of a group tends to evaluate oneself above the group's statistical mean (Giladi & Klar, 2002). The empirical evidence suggests that this "Better-than-average heuristic" is one of the most robust of all self-enhancement phenomena (Brown, 1986; Gregg & Sedikides, 2004), which especially arises when the characteristics of the peer group are similar (Alicke & Govorun, 2005).

FIGURE 4.1: PERCENTAGE OF X CHOSEN BOTH TREATMENTS



Next, we focus on the present-biased parameter, β. Thus, we analyze subjects' behavior in the two treatments, focusing on the time-consistency of choices. Result 4.2 shows that, on average, there is no difference in behavior between the treatments, i.e. subjects do not behave more time-consistently when another person's payoff is affected.

RESULT 4.2: We find a strong preference for consistency in both treatments. Overall, the time-consistency of choices is not increased by the introduction of responsibility.

Support for result 4.2 is presented in Table 4.2 (lower part), which summarizes the distribution of time preferences in both treatments. The categorization is as follows. A choice is consistent if the subject chooses the same answer (X or Y) both in the close and in the far future. Instead, a choice is present-biased if a person chooses the sooner-smaller amount (choice X) in the close future whereas he opts for the larger-later reward (choice Y) in the far future. It is also possible to observe future-biased decisions. This is the case when a subject opts for the larger-later reward (choice Y) in the close future but chooses the sooner-smaller amount (choice X) in the far future time frame.

TABLE 4.3: DECISIONS PER CHOICE NUMBER

	ICI				ICR			
	X		Y		X		Y	
	Nr.	%	Nr.	%	Nr.	%	Nr.	%
Choice Number								
1	47	73.44	17	26.56	45	76.27	14	23.73
2	46	71.88	18	28.12	45	76.27	14	23.73
3	39	60.94	25	39.06	36	61.02	23	38.98
4	34	53.13	30	46.87	34	57.63	25	42.37
5	38	59.38	26	40.62	35	59.32	24	40.68
6	27	42.19	37	57.81	27	45.76	32	54.24
7	28	43.75	36	56.25	26	44.07	33	55.93
8	20	31.25	44	68.75	27	45.76	32	54.24
9	21	32.81	43	67.19	19	32.20	40	67.80
10	15	23.44	49	76.56	16	27.12	43	72.88
11	5	7.81	59	92.19	8	13.56	51	86.44
12	9	14.06	55	85.94	9	15.25	50	84.75
13	4	6.25	60	93.75	6	10.17	53	89.83
14	4	6.25	60	93.75	5	8.47	54	91.53
15	2	3.13	62	96.87	6	10.17	53	89.83
16	1	1.56	63	98.44	2	3.39	57	96.61
17	1	1.56	63	98.44	3	5.08	56	94.92
18	0	0.00	64	100.00	2	3.39	57	96.61
19	0	0.00	64	100.00	1	1.69	58	98.31
20	0	0.00	64	100.00	1	1.69	58	98.31
21	49	76.56	15	23.44	46	77.97	13	22.03
22	47	73.44	17	26.56	45	76.27	14	23.73
23	33	51.56	31	48.44	38	64.41	21	35.59
24	29	45.31	35	54.69	27	45.76	32	54.24
25	32	50.00	32	50.00	36	61.02	23	38.98
26	27	42.19	37	57.81	31	52.54	28	47.46
27	24	37.50	40	62.50	27	45.76	32	54.24
28	20	31.25	44	68.75	21	35.59	38	64.41
29	18	28.13	46	71.87	14	23.73	45	76.27
30	13	20.31	51	79.69	18	30.51	41	69.49
31	7	10.94	57	89.06	11	18.64	48	81.36
32	7	10.94	57	89.06	10	16.95	49	83.05
33	6	9.38	58	90.62	4	6.78	55	93.22
34	4	6.25	60	93.75	5	8.47	54	91.53
35	1	1.56	63	98.44	1	1.69	58	98.31
36	2	3.13	62	96.87	1	1.69	58	98.31
37	1	1.56	63	98.44	0	0.00	59	100.00
38	1	1.56	63	98.44	0	0.00	59	100.00
39	1	1.56	63	98.44	0	0.00	59	100.00
40	0	0.00	64	100.00	0	0.00	59	100.00

In treatment ICI over 90 percent of choices are time-consistent, around 4 percent are future-biased and around 6 percent are present-biased. A similar picture arises in treatment ICR, where around 91 percent of choices are time-consistent, 4 percent are future-biased, and 5 percent are present-biased. The difference in time-consistent, and present-biased choices between treatments is not significant (Two-sample test of proportion, consistent: p=0.6144; present-biased: p=0.7627). Since already in the baseline treatment over 90 percent of people behaves perfectly consistently, a further increase in this measure is difficult to induce.[56] This

[56] This is a very untypical result compared to general findings in the literature, where normally 30 to 40 percent of subjects do have present-biased preferences (for an overview see Frederick et al., 2002).

high share of consistent choices - even though the randomization of all 40 choices induced by our design aggravated consistent behavior – confirms the findings of Falk & Zimmermann (2013) that players seem to have a (considerable) preference for consistency.

4.5.2 Statistical Specification

To further test the descriptive results, we estimate the individuals' discount functions by maximum likelihood. Under exponential discounting, it is given by:

$$u_{m,t} = m\delta^{t} \tag{4.10}$$

where m is the monetary amount, δ is the discount factor, and t is the number of days of delay until the payment is received. Based on the descriptive results presented in the former section, the vast majority of individuals behave time-consistently in our experiment. To capture these behavioral patterns, we start with the exponential discounting model, which is, by construction, time-consistent. Including an effect of the treatment condition on the discount parameter, the model simply becomes:

$$u_{m,D} = m(\delta_{ICI} + \Delta^{\delta}\mathrm{ICR})^{t} \tag{4.11}$$

where $\Delta^{\delta}\mathrm{ICR}$ captures the change in the discount factor due to the treatment, i.e. the introduction of responsibility (indicated with ICR). The descriptive results indicate a tendency towards less patient choices, which would mean that the change in the discount factor should be negative, i.e. $\Delta^{\delta} < 0$.

The model can be easily extended to accommodate a non-linear functional form of utility of income:

$$u_{m,D} = m^{\propto}(\delta_{ICI} + \Delta^{\delta}\mathrm{ICR})^{t} \tag{4.12}$$

Now, we estimate this function in one step with full maximum likelihood. To estimate the model and test for the main hypotheses, we pool all 59x40 choices between the smaller-sooner and the larger-later payments. To account for the fact that the choices by the 59 individuals are not independent of each other, we compute cluster-robust standard errors. Moreover, we compute confidence intervals using bootstrap instead of relying on the normality assumption. Normality seems to be unlikely to hold in our case. Bootstrap works even if the normality assumption is relaxed but is also consistent with a normal distribution. Moreover, in many cases bootstrap confidence intervals have proved to be better than asymptotic normal ones. Table 4.4 reports the estimated coefficient of model (4.12) employing clusters at the individual level and confidence intervals computed with bootstrap.

As can be seen, subjects have, on average, a pretty high discount factor already in the baseline treatment ICI, being $\delta_{ICI} = 0.99689$. This means they are very patient. Second, $\propto$ is significantly lower than 1, so the utility is increasing in money but at a decreasing rate. Next, consider the change in δ induced by the introduction of responsibility, Δ^{δ}. We have $\Delta^{\delta} < 0$, so subjects in the responsibility treatment have a lower discount factor, i.e. behave more impatiently in the ICR treatment than in the ICI treatment. Since the baseline is roughly 1, it is easy to get an idea of the magnitude of this effect. The change in δ induced by responsibility is -0.293%. Despite the relatively small effect on the general level of impatience, the bootstrap confidence interval shows that the coefficient is significant at 5% level. Hence, result 4.1 can be fortified: Responsibility leads to an increase in impatience among choices. Thus, hypothesis 4.1 cannot be rejected.

TABLE 4.4: ESTIMATES EXPONENTIAL MODEL, POOLED DATA AND CLUSTERS

Estimates	**(4.12)**	**(4.13)**	**(4.14)**
Discount factor (δ_{ICI})	.99689**	.99661**	.99644**
	(.00134)	(.00132)	(.00126)
	[.99604 .99768]	[.99562 .99748]	[.99539 .9974]
Responsibility (Δ^{δ})	-.00293**	-.00298**	-.00259**
	(.00213)	(.00216)	(.00179)
	[-.00418 -.00171]	[-.00427 -.00175]	[-.00406 -.00111]
Curvature of u(·) ($\propto$)	.82502**	.82469**	.82476**
	(.02822)	(.02848)	(.02835)
	[.79611 .84714]	[.79526 .84668]	[.79549 .84679]
Time Inconsistency factor $\bar{\beta}$		.01314	.02058**
		(.0133)	(.01762)
		[-.0006 .0234]	[.00239 .03428]
Responsibility on TI factor ($\Delta^{\bar{\beta}}$)			-.01623
			(.02731)
			[-.04443 .005]
Number of iterations:	5	7	8
Log likelihood:	-1091.2246	-1090.40935	-1090.10238
No. of. Obs.:	2360	2360	2360

Note: Standard errors are given in round brackets. The bootstrap confidence intervals are given in square brackets. ** denote significance on the 5% level.

Next, we analyze time-consistency. We simply include an additional parameter to model the present bias. We therefore adopt an empirical version of the quasi-hyperbolic discounting model to be able to separate the impatience parameter from the present-bias. Then, the empirical model is:

$$u_{m,t} = m^{\propto}(1 - \bar{\beta} I_{t>1})(\delta_{ICI} + \Delta^{\delta} ICR)^{t} \tag{4.13}$$

Note that we parameterize the present-bias parameter, β, as $\beta = 1 - \bar{\beta}$. If a subject has present-biased time preferences, we expect to find $\bar{\beta} > 0$. In case of time-consistency we have $\bar{\beta} = 0$. The second element of the present-bias parameter is the indicator function $I_{t>1}$, which takes the value 1 if the payment is made in one day. This captures the present-bias in a way that is consistent with the quasi-hyperbolic framework.[57]

Table 4.4 also reports the estimated coefficients of model (4.13). Note that the baseline impatience parameter, δ_{ICI}, the treatment effect on impatience, Δ^{δ}, as well as the curvature parameter, ∝, are almost unaffected by the introduction of the new parameter. The new parameter, $\bar{\beta}$, is positive (as expected) but quite small in size and not significant.

Finally, we test the effect of responsibility on time-consistency. We introduce a model that puts both treatment effects together:

$$u_{m,t} = m^{\propto}(1 - (\bar{\beta}_{ICI} + \Delta^{\bar{\beta}} ICR) I_{t=1})(\delta_{ICI} + \Delta^{\delta} ICR)^{t} \tag{4.14}$$

The estimated coefficients for (4.14) are reported in the last column of Table 4.4. Note that δ_{ICI}, Δ^{δ}, and ∝ do not change much compared to the former models. However, when we control for the effect of responsibility on time-consistency, we find that the confidence interval for the baseline parameter, $\bar{\beta}$, suggests that our representative decision maker is actually present-biased (even though the absolute value of the present bias is very small). The coefficient of $\Delta^{\bar{\beta}}$ is negative but not significantly different from 0. Hence, we do not find a significant increase in time-consistency with responsibility and therefore can reject hypothesis 4.2 based on the estimate of $\Delta^{\bar{\beta}}$. However, the direction of the coefficients is in line with the hypothesis that social image concerns increase the taste for consistency.

[57] Note that the definition of what is exactly the close future is subject to the specific application. In our setting, the immediate date is tomorrow to (a) unitize transaction costs among individuals and (b) remove any problem of front-end delay.

4.6 Conclusion

This study investigates the effect of responsibility on intertemporal decision-making. Intertemporal choices, i.e. individual savings and investment decisions are ever-present in life and some of them have severe consequences for the well being of a person. However, many of such choices do not only have consequences for the decision maker but also affect the welfare of other people who do not actively take part in the decision process. If this is the case than the intertemporal choice is made with responsibility. In this study, we address such choices made with responsibility for an unknown person with similar demographic characteristics, i.e. age and income, such that the time preferences of the dependent are not easy to infer but have to be estimated by the decision maker. Even though these choices are ubiquitous, as to our knowledge, there is no study that analyzes the influence of responsibility in intertemporal decision-making. We address this loophole in the literature by offering a novel experimental setting, where we introduce responsibility in a standard intertemporal decision problem. A principal chooses between payments due at different times, with such payments being also the payoff of a passive dependent who is not taking part in the decision problem but (silently) observes the principal's behavior. We then compare behavior to the same choices made without responsibility. The results show that responsibility leads to a shift towards more impatience among choices. This can be explained by social preferences, where utility is influenced by concerns for other players (i.e. Rabin, 1993; Levine, 1998; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000). Hence, to incorporate both his own as well as the dependent's time preferences, the principal aggregates utility by a weighted utility function. Since the principal does not know the time preferences of the dependent, and there are no obvious signals about the other person's utility functions, the principal has to estimate the dependent's discount factor. The better-than-average effect infers that the principals are overconfident about their own degree of patience, i.e. they expect the recipient to be more impatient than themselves and, hence, the aggregation of a weighted utility function leads to more impatient choices. Our results are further strengthened by the fact that, in our main treatment, the passive dependent observes the principals' choices. Social image concerns would work in the opposite direction and should induce the principal to behave more patiently. Yet, we still find the effect of a shift towards more impatience. This indicates that the utility weighting outweighs social image concerns.

Another possible explanation for the shift towards more impatience could be that responsibility is used as an excuse to behave impatiently. Since the time preferences of the recipient are unknown, the principal can justify (to himself as well as to others) falling into the temptation of behaving impatiently by shifting the blame onwards to the recipient. The two explanations are similar but the behavioral motivation is different; the former explanation is based on an altruistic assumption, whereas the latter is not. Yet, in the individual intertemporal choice task, neither an observer nor a dependent is present. Thus, people state their preferences without any external influence, such as social image concerns or social preferences. Since the mechanism is incentive compatible, behavior in our control treatment should therefore already display the upper bound of impatience. Hence, we can exclude the argument that principals use the dependent as an excuse for the disclosure of their true degree of impatience.

Moreover, our results confirm a preference for consistency, as suggested by Falk & Zimmermann (2013). In both treatments, over 90 percent of choices are consistent. The psychological literature argues that consistent behavior is associated with intellectual and personal strength (Cialdini, 1993), whereas inconsistent behavior is an undesirable personal characteristic (i.e. Asch, 1956; Allgeier et al., 1979). Falk & Zimmermann (2013) find that the key driver of consistent behavior is to signal positive traits. However, in our experiment, already in the baseline treatment, where social image concerns do not play a role, we find that choices are very time-consistent even if the experimental design made it especially difficult to behave consistently. Since consistency is already at its upper bound, an increase in this parameter is difficult to infer. However, the coefficient goes in the right direction, which is in line with the argument that social image concerns increase the taste for consistency.

The experimental setting suggested in this study extends the existing literature by incorporating responsibility into a typical intertemporal choice task. By suggesting an experimental design that analyzes time preferences when another person is affected by the decision maker's choice, it enhances the comprehension of behavioral regularities that might govern decision-making in an intertemporal context.

4.7 Appendix

4.7.1 Experimental Instructions

General Information [*PCFeed]

Welcome to this experiment and thank you very much for your participation. Please read these instructions carefully. Please switch off your mobile phone and do not communicate with any other participant anymore. If you have any questions, please raise your hand. One of the experimenters will then come to your table and answer your question personally.

You can earn money in this experiment. The amount earned depends on your decisions as well as the decisions of other participants made during the experiment. All decisions made during the experiment as well as your final earnings remain anonymous.

This experiment consists of 40 rounds. In each round, you [*or another participant has] have to make a decision. After all participants have made their decisions, *one of the rounds* is randomly chosen to be payoff-relevant at the end of the experiment; thereby, the probability of being chosen is exactly the same for all rounds.

The timing of the payment depends on the choices you have made [* that has been made] in the round that has been randomly chosen. The payoff from the randomly selected round will be transferred to your bank account. We assure you the certainty of your future payments. Additionally, each participant gets a show-up fee of 2.50 Euro, which is independent of the decisions made during the experiment and will be paid out to you cash at the end of the experiment.

All payment dates in this experiment are the date, where the bank transfer is *disposed.*

You find the exact description of the experiment on the following pages.

The Experiment

[*At the beginning of the experiment, you are randomly assigned to role A or B. Each participant with role A is randomly matched with a participant with role B. You will never learn the identity of the participant that you have been matched with. The matching remains constant during the whole experiment.]

This experiment consists of 40 rounds. In each round, you have [*participant A has] the choice between option X and option Y. The options offer different payoff alternatives, where the monetary amount as well as the timing of the payment, vary. In each round, you have [*participant A has] the choice between an amount, which will be transferred to you [*him as well as participant B] at an earlier date, and a larger amount, which will be transferred to you [*him as well as participant B] at a later date. [*This means that participant A decides about his *own* payoff as well as *the payoff of participant B*. Participant B does not make any active decision. However, participant B is informed about participant A's choices after each round.] [Only T1: Please think carefully about each choice and choose the option you prefer.]

Here is an example:

Option X	Option Y	Your Choice
[A:] 9 Euro tomorrow	10 Euro in 5 days	X Y
[B: 9 Euro tomorrow]	[10 Euro in 5 days]	

If you choose [participant A chooses] option X, you prefer [he prefers] 9 Euro tomorrow to 10 Euro in 5 days. If this round is randomly selected to be payoff-relevant, you [participant A as well as participant B] earn 9 Euro in this experiment, which will be transferred to your [both] bank account [accounts] tomorrow.

If you choose [participant A chooses] option Y, you prefer [he prefers] 10 Euro in 5 days to 9 Euro tomorrow. If this round is randomly selected to be payoff-relevant, you [participant A as well as participant B] earn 10 Euro in this experiment, which will be transferred to your [both] bank account [accounts] in 5 days.

End of the Experiment

At the end of the experiment, the team will hand out a questionnaire. Please fill in the answers truthfully. When the experiment is finished, please remain seated and wait until the experimenter will call your seat number. Please fill in your bank account details, the amount you earned as well as the date of transferal in the destined remittance slip and take it with you when you come to the front. The team will hand out a certificate for your participation and the amount you will get transferred in the future and additionally pay you 2.50 Euro in cash.

Thank you very much for your participation!

4.7.2 Certificate for Transferal of Earnings

Universität zu Köln
Wirtschafts- und Sozialwissenschaftliche Fakultät
Staatswissenschaftliches Seminar
Prof. Dr. Bettina Rockenbach

Kölner Laboratorium
für Wirtschaftsforschung

Hiermit versichern wir Ihnen, dass wir die Überweisung in Höhe von

_______________ EURO am ______________ auf das von

Ihnen angegebene Konto veranlassen werden.

Dabei versichern wir, dass Ihre Bankdaten nur zur einmaligen Transaktion genutzt und danach uneingeschränkt gelöscht werden. Die erfassten Bankdaten dienen ausschließlich zur Auszahlung dieses wissenschaftlichen Experimentes.
Die Bankdaten werden nicht an Dritte weitergegeben.

Wir übernehmen keine Verantwortung für zeitliche Verzögerungen durch die Empfängerbank.

____
Datum, Unterschrift

Chapter 5

Conclusions

This thesis presents three studies that examine the impact of non-standard preferences on consumption decisions in several economic settings. To be more precise, we analyze potential influences of non-standard preferences on contribution behavior in a social dilemma (chapter 2), cooperation in intertemporal decision-making (chapter 3) and the effect of responsibility on intertemporal choices (chapter 4). The starting point for each research project is the assumption that non-standard preferences alter economic decision-making. In each study, a laboratory experiment is implemented to analyze behavioral regularities and explore their relevance in the different economic contexts.

Chapter 2 analyzes the endogenous formation of a sanction institution in a public goods game. The aim is to address potential enforcement problems by – instead of letting sanctions be executed by a central authority ex-post – allowing players to regulate their own punishment ex-ante in a decentralized way. To do so, subjects are offered to pre-commit to contributions before the actual public goods game is played. The results show that the vast majority of players voluntarily pre-commit to contributions and the possibility of institution formation positively influences cooperation and group welfare. Even though the pre-commitment device is designed such that players can only punish themselves for contributing to less and the punishment is independent of group contributions, players repudiate free-riding behavior on the institution. The results therefore stress the importance of fairness consideration within the institution formation process. This is in line with considerable evidence that a taste for fairness and equality affects economic behavior in many important areas (for an overview, see Fehr & Gächter, 2000b; Camerer, 2003).

Our results has important implications for the design of institutions; it is good news that fairness leads to higher efficiency levels than if fairness motives were irrelevant. However, fairness considerations also have to be considered as a limitation on equilibria and might impede the institution formation process by increasing the likelihood that the implementation of the institution fails. Therefore, fairness motives should be taken into account ex ante.

The experimental setting complements the existing literature by combining the endogenous formation of institutions in a public goods game with a pre-commitment mechanism. Based on the recent literature that suggests that a social dilemma also presents a self-control dilemma (Curry et al., 2008; Fehr & Leibbrandt, 2011; Kocher et al., 2012; Martinsson et al., 2012), the mechanism presented in this study can help players to overcome self-control problems in the anticipation of self-regulatory failure.

The idea that a social dilemma also represents self-control dilemma, where the individual has to trade off the objective of acting pro-socially and the temptation to act selfishly has been recently discussed in the enforcement literature. The experimental evidence reported so far suggests that a high level of self-control leads to significantly higher contributions in a public goods game if this is perceived as a self-control conflict (Kocher et al., 2012; Martinsson et al., 2012) and that impatient individuals contribute less to the public good than patient ones (Curry et al., 2008; Fehr & Leibbrandt, 2011, Kocher et al., 2012). However, all these studies measure the individual degree of self-control independently from the contribution decision in the public good game, where the choice tasks to elicit time preferences consist of lotteries that embrace payoffs that differ very much from those in the cooperation game. Most importantly, the typical choice task to elicit time preferences does not involve a second person. Rather, the decision maker makes choices that only affect his own payoff and therefore his decisions are made in a socially detached environment. However, if social dilemmas represent a self-control conflict, any choice involves *both* a temporal as well as a social dimension and it is possible that the choice differs profoundly for the decision maker if the intertemporal decision is made in interaction with another person.

To approach this problem, chapter 3 focuses on the interplay between time and social preferences based on joint elicitation. The experimental setting introduces a novel element to the existing literature by bringing a standard time preference experiment into a social dimension. The results suggest that – when both preferences are elicited jointly – the social component outweighs individual self-control problems in a cooperation game. This finding has implications for policy decisions, i.e. with respect to the provision of public goods. Contradictory to previous findings in the literature, it suggests that the probability to contribute to a public good does not depend on the timing of the decision.

Moreover, the study opened the question of how responsibility might influence intertemporal choices. Chapter 4 addresses this question by studying the influence of responsibility on intertemporal decision-making. To do so, we introduce responsibility into a standard intertemporal choice task. Our results show that the decisions made with responsibility reflect more impatience. This is consistent with a model where a weighted utility function results in a shift towards impatience if the time preferences of the recipients are unknown to the principal and the latter is overconfident about his own degree of patience.

Chapter 3 and 4 both focus on intertemporal decision-making in (different aspects of) a social context. By suggesting an experimental design for the joint elicitation of time and social preferences based on a cooperation (trust) game, as well as for the integration of responsibility in an intertemporal choice task, both studies contribute to the understanding of behavioral regularities that might govern the choices in an intertemporal context with social interaction.

Summing up, all three studies suggest that other-regarding preferences influence behavior. This finding is in line with various models of decisions that assume that utility is not solely influenced by monetary gains but also by concerns for other players (Rabin, 1993; Levine, 1998; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000; Charness & Rabin, 2002; Dufwenberg & Kirchsteiger, 2004; Sobel, 2005; Fehr& Schmidt, 2006). Specifically, our results show that fairness considerations are important for public goods provision (Chapter 2), concerns for other players can outweigh individual problems of self-control (Chapter 3), and being responsible for the welfare of another person influences behavior (Chapter 4).

6 Bibliography

Alicke, Mark, and Olesya Govorun. 2005. *Self in Social Judgment. Studies in Self and Identity*. ed. Mark Alicke, David Dunning, and Joachim Krüger. 85-106. NY, New York: Psychology Press.

Allgeier, A. R., Donn Byrne, Barbara Brooks, and Diane Revnes. 1979. "The Waffle Phenomenon: Negative Evaluations of Those Who Shift Attitudinally." *Journal of Applied Social Psychology*, 9(2): 170-182.

Anderhub, Vital, Uri Gneezy, Werner Güth, and Doron Sonsino. 2001. "On the Interaction of Risk and Time Preferences: An Experimental Study." *German Economic Review,* 2(3): 239-253.

Andersen, Steffen, Glenn Harrison, Morten Lau, and Elisabeth Rutström. 2008. "Eliciting Risk and Time Preferences." *Econometrica,* 76(3): 583-618.

Andreoni, James, William Harbaugh, and Lise Vesterlund. 2003. "The Carrot or the Stick: Rewards, Punishments, and Cooperation." *American Economic Review,* 93(3): 893-902.

Ariely, Dan, and Klaus Wertenbroch. 2002. "Procrastination, Deadlines and Performance: Self-control by Precommitment." *Psychological Science,* 13(3): 219-224.

Asch, Solomon. 1956. "Studies of Independence and Conformity: A Minority of One Against a Unanimous Majority." *Psychological Monographs: General and Applied,* 70(9): 1-70.

Ashraf, Nava, Iris Bohnet, and Nikita Piankov. 2006. "Decomposing Trust and Trustworthiness." *Experimental Economics*, 9(3): 193-208.

Bagnoli, Mark, and Barton Lipman. 1989. "Provision of Public Goods: Fully Implementing the Core through Private Contributions." *Review of Economic Studies,* 56(4): 583-601.

Baland, Jean-Marie, and Jean-Philippe Platteau. 1996. *Halting Degradation of Natural Resources*. Oxford, U.K.: Oxford University Press.

Barrett, Scott. 1994. "Self-Enforcing International Environmental Agreements." *Oxford Economic Papers*, 46: 878-894.

Benabou, Roland, and Jean Tirole. 2004. "Willpower and Personal Rules." *Journal of Political Economy,* 112(4): 848-886.

Benhabib, Jess, Alberto Bisin, and Andrew Schotter. 2010. "Present-Bias, Quasi-Hyperbolic Discounting, and Fixed Costs." *Games and Economic Behavior*, 69(2): 205-223.

Benzion, Uri, Amnon Rapoport, and Joseph Yagil. 1989. "Discounting Rates Inferred From Decisions: An Experimental Study." *Management Science*, 35: 270-284.

Ben-Ner, Avner, and Louis Puttermann. 2001. "Trusting and trustworthiness." *Boston University Law Review*, 81: 523-551.

Bohnet, Iris, and Richard Zeckhauser. 2004. "Trust, Risk and Betrayal." *Journal of Economic Behavior and Organization*, 55: 467-484.

Bohnet, Iris, Fiona Greig, Benedikt Herrmann, and Rochard Zeckhauser. 2008. "Betrayal Aversion: Evidence from Brazikl, China, Oman, Switzerland, Turkey, and the United States." *American Economic Review*, 98(1): 294-310.

Brosig, Jeannette, Joachim Weimann, and Axel Ockenfels. 2003. "The Effect of Communication Media on Cooperation." *German Economic Review*, 4(2): 217-241.

Bolton, Gary. 1991. "A Comparative Model of Bargaining: Theory and Evidence." *American Economic Review*, 81(5): 1096-1136.

Bolton, Gary, and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1): 166–93.

Bolton, Gary, and Axel Ockenfels. 2010. "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States: Comment." *American Economic Review*, 100(1): 628-633.

Bornstein, Gary. 2003. "Intergroup Conflict: Individual, Group, and Collective Interests." *Personality and Social Psychology Review*, 7: 129-145.

Bornstein, Gary, and Ilan Yaniv. 1998. "Individual and Group Behavior in the Ultimatum Game: Are Groups More 'Rational' Players?" *Experimental Economics,* 1(1): 101-108.

Bornstein, Gary, Tamar Kugler, and Anthony Ziegelmeyer. 2004. "Individual and Group Decisions in the Centipede Game: Are Groups More 'Rational' Players?" *Journal of Experimental Social Psychology,* 40(5): 599-605.

Botelho, Anabela, Glenn W. Harrison, Lígia Pinto, and Elisabet E. Rutström. 2005. "Social Norms and Social Choice." Working Papers 30, Núcleo de Investigacao em Microeconomia Aplicada (NIMA), Universidade do Minho.

Brown, Jonathan. 1986. "Evaluations of Self an Others: Self-Enhancement Biases in Social Judgments." *Social Cognition*, 4(4): 353-376.

Buchanan, James, and Roger Congleton. 1998. *Politics by Principle, Not Interest.* Cambridge, U.K.: Cambridge University Press**.**

Camerer, Colin. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, New York: Princeton University Press/Russell Sage Foundation.

Camerer, Colin, and George Loewenstein. 2004. "Behavioral Economics: Past, Present, and Future." In: *Advances in Behavioral Economics*, ed. Colin Camerer, George Loewenstein, and Matthew Rabin. 3-52. Princeton and Oxford: Princeton University Press.

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics,* 117: 817-869.

Charness, Gary, Luca Rigotti, and Aldo Rustichini. 2007. "Individual Behavior and Group Membership." *American Economic Review*, 97(4): 1340-1352.

Charness, Gary, and Matthew Jackson. 2009. "The Role of Responsibility in Strategic Risk-Taking." *Journal of Economic Behavior and Organization*, 69(3): 241-247.

Chaudhuri, Ananish. 2011. "Sustaining Cooperation in Laboratory Public Good Experiments: A Selective Survey of the Literature." *Experimental Economics*, 14: 47-83.

Cialdini, Robert. 1993. *Influence- the Psychology of Persuasion.* New York: Quill William Morrow.

Cook, Karen, and Robin Cooper. 2003. "Experimental Studies of Cooperation, Trust, and Social Exchange." In: *Trust and Reciprocity*, ed. Elionor Ostrom, and James Walker. 209–244. New York: Russell Sage.

Cooper, David, and John Kagel. 2005. "Are Two Heads Better Than One? Team Versus Individual Play in Signaling Games." A*merican Economic Review*, 95(3): 477-509.

Cox, James, Daniel Friedman, and Steven Gjerstad. 2007. "A Tractable Model of Reciprocity and Fairness." *Games and Economic Behavior*, 59(1): 17–45.

Croson, Rachel, and Melanie Marks. 2000. "Step Returns in Threshold Public Goods: A Meta- and Experimental Analysis." *Experimental Economics,* 2(3): 239-259.

Curry, Oliver, Michael Price, and Jade Price. 2008. "Patience is a Virtue: Cooperative People Have Lower Discount Rates." *Personality and Individual Differences,* 44: 780-785.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman. 2010. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review,* 100(5): 2205-2229.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence From The Field*." Journal of Economic Literature*, 47: 315-372.

Denant-Boemont, Laurent, David Masclet, and Charles Noussair. 2007. "Punishment, Counterpunishment and Sanction Enforcement in a Social Dilemma Experiment." *Economic Theory,* 33: 145-167.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, 47: 268-298.

Earl, Peter. 1990. "Economics and Psychology." *Economic Journal*, 100: 718-755.

Egas, Martijn, and Arno Riedl. 2008. "The Economics of Altruistic Punishment and the Maintenance of Cooperation." *Proceedings of the Royal Society B: Biological Science,* 275(1637): 871-878.

Engel, Christoph, and Bettina Rockenbach. 2011. “We Are Not Alone: The Impact of Externalities in Public Good Provision.” *Working paper.*

Falk, Armin, and Urs Fischbacher. 2006. “A Theory of Reciprocity.” *Games and Economic Behavior*, 54: 293-315.

Falk, Armin, and Florian Zimmermann. 2013. “A Taste for Consistency and Survey Response Behavior.” *CESifo Economic Studies*, 59(1): 181-193.

Falkinger, Josef, Ernst Fehr, Simon Gächter, and Rudolf Winter-Ebmer. 2000. “A Simple Mechanism for the Efficient Provision of Public Goods: Experimental Evidence.“ *American Economic Review,* 90(1): 247-64.

Fehr, Ernst, and Simon Gächter. 2000a. “Cooperation and Punishment in Public Goods Experiments.” *American Economic Review*, 90(4): 980-994.

Fehr, Ernst, and Simon Gächter. 2000b. “Fairness and Retaliation: The Economics of Reciprocity.” *Journal of Economic Perspective*, 14(3): 159-181.

Fehr, Ernst, and Simon Gächter. 2002. “Altruistic Punishment in Humans.” *Nature*, 415: 137-140.

Fehr, Ernst, and Klaus Schmidt. 1999. “A Theory of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics,* 114(3): 817-868.

Fehr, Ernst, and Klaus Schmidt. 2006. “The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories.” In: *Handbook on the Economics of Giving, Reciprocity and Altruism,* ed. Serge-Christophe Kolm and Jean Ythier. 615-691. North-Holland: Elsevier.

Fehr, Ernst, and Andreas Leibbrandt. 2011. “A Field Study on Cooperativeness and Impatience in the Tragedy of the Commons.“ *Journal of Public Economics*, 95: 1144-1155.

Fischbacher, Urs. 2007. “z-Tree: Zurich Toolbox for Ready-Made Economic Experiments.” *Experimental Economics*, 10(2): 171-178.

Fischer, Carolyn. 2001. “Read this Paper Later: Procrastination with Time-Consistent Preferences.” *Journal of Economic Behavior and Organization*, 46(3): 249-269.

Frederick, Shane, George Loewenstein, and Ted O’Donoghue. 2002. “Time Discounting and Time Preference: A Critical Review.” *Journal of Economic Literature*, 40(2): 351-401.

Fudenberg, Drew, and David Levine. 2006. “A Dual Self Model of Impulse Control.” *American Economic Review*, 96(5): 1449-1476.

Giladi, Eilath, and Yechiel Klar. 2002. “When Standards are Wide of the Marks: Nonselective Superiority and Inferiority Biases in Comparative Judgments of Objects and Concepts.” *Journal of Experimental Psychology*, 131(4): 538-551.

Gächter, Simon. 2006. "Conditional Cooperation: Behavioral Regularities from the Lab and the Field and their Policy Implications." *CeDEx Discussion Paper No. 2006-03.*

Gächter, Simon. 2007. "Conditional Cooperation: Behavioral Regularities from the Lab and the Field and their Policy Implications." In: *Economics and Psychology: A Promising New Cross-Disciplinary Field,* ed. Bruno S. Frey and Alois Stutzer. 19-50. Cambridge: MIT Press.

Gerber, Anke, and Philipp Wichardt. 2009. "Providing Public Goods in the Absence of Strong Institutions." *Journal of Public Economics,* 93: 429-439.

Green, Leonard, Irving Fisher, Steven Perlow, and Lisa Sherman. 1981. "Preference Reversal and Self-Control: Choice as a Function of Reward Amount and Delay." *Behavior Analysis Letter*, 1: 43-51.

Green, Leonard, Nathanael Fristoe, and Joel Myerson. 1994. "Temporal Discounting and Preference Reversals in Choice Between Delayed Outcomes." *Psychonomic Bulletin & Review,* 1(3): 383-389.

Gregg, Aiden, and Constantine Sedikides. 2004. "Is Social Psychological Research Really So Negatively Biased?" *Behavioral and Brain Science*, 27(3): 340-341.

Greiner, Ben. 2004. "An Online Recruitment System for Economic Experiments." In *Forschung und wissenschaftliches Rechnen*, ed. Kurt Kremer, 79-93. Göttingen: GWDG Bericht 63.

Grosse, Stefan, Louis Putterman, and Bettina Rockenbach. 2011. "Monitoring in Teams: Using Laboratory Experiments to Study a Theory oft he Firm." *Journal of the European Economic Association,* 9: 785-816.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach. 2006. "The Competitive Advantage of Sanctioning Institutions." *Science*, 312: 108-111.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*, 3(4): 367-388.

Harris, Andrew, and Gregory Madden. 2002. "Delay Discounting and Performance on the Prisoner`s Dilemma Game." *The Psychological Record,* 52: 429-440.

Harrison, Glenn, Morten Lau, and Melonie Williams. 2002. "Estimating Individual Discount Rates for Denmark: A Field Experiment." *American Economic Review*, 92(5): 1606-1617.

Harrison, Glenn, and John List. 2004. "Field Experiments." *Journal of Economic Literature*, 42(4): 1009-1055.

Heuermann, Karen. 2011. "The Endogenous Formation of Institutions in a Public Good Game with a Decentralized Sanction Mechanism." *Unpublished Master Thesis*. Maastricht University.

Holt, Charles, and Susan Laury. 2002. “Risk Aversion and Incentive Effects.” *American Economic Review,* 92(5): 1644-1655.

Humphrey, Steven, and Elke Renner. 2013. “The Social Cost of Responsibility.” *Working paper.*

Jackson, Matthew, and Leeat Yariv. 2014. “Present Bias and Collective Dynamic Choice in the Lab.” *American Economic Review (forthcoming).*

Kahnemann, Daniel. 2003. “Maps of Bounded Rationality: Psychology for Behavioral Economics.” *American Economic Review*, 93(5): 1449-1475.

Keren, Gideon, and Peter Roelofsma. 1995. “Immediacy and Certainty in Intertemporal Choice.” *Organizational Behavior and Human Decision Processes*, 63(3): 287-297.

Kirby, Kris. 1997. “Bidding on the Future: Evidence Against Normative Discounting of Delayed Rewards.” *Journal of Experimental Psychology: General,* 126(1): 54-70.

Kirby, Kris, and Richard Herrnstein. 1995. “Preference Reversals Due to Myopic Discounting of Delayed Rewards.” *Psychological Science*, 6: 83-89.

Kirby, Kris, and Nino Marakovic. 1996. “Delay-Discounting Probabilistic Rewards: Rates Decrease as Amounts Increase.” *Psychonomic Bulletin and Review*, 3(1): 100-104.

Kocher, Martin, and Matthias Sutter. 2005. “The Decision Maker Matters: Individual versus Group Behavior in Experimental Beauty-Contest Games.” *Economic Journal*, 115(500): 200-223.

Kocher, Martin, Peter Martinsson, Dominik Matzat, and Conny Wollbrant. 2011. “The Role of Beliefs, Trust, and Risk in Contributions to a Public Good.” *Working Paper No. 482, University of Gothenburg.*

Kocher, Martin, Peter Martinsson, Kristian Myrseth, and Conny Wollbrant. 2012. “Strong, Bold, and Kind: Self-Control and Cooperation in Social Dilemmas.” *ESMT Working Paper No. 12-01.*

Kosfeld, Michael, Akira Okada, and Arno Riedl. 2009. “Institution Formation in Public Goods Games.” *American Economic Review*, 99(4): 1335-1355.

Kroll, Stephan, Todd L. Cherry, and Jason F. Shogren. 2007. “Voting, Punishment, and Public Goods.” *Economic Inquiry*, 45(3): 557-70.

Kugler, Tamar, Gary Bornstein, Martin Kocher, and Matthias Sutter. 2007. “Trust Between Individuals and Groups: Groups Are Less Trusting Than Individuals But Just as Trustworthy.” *Journal of Economic Psychology*, 28(6): 646-657.

Laibson, David. 1997. “Golden Eggs and Hyperbolic Discounting.” *Quarterly Journal of Economics*, 112: 443-477.

Laibson, David. 1998. “Life-Cycle Consumption and Hyperbolic Discount Functions.” *European Economic Review*, 42: 861-871.

Ledyard, John. 1995. "Public Goods: A Survey of Experimental Research." In: *The Handbook of Experimental Economics,* ed. John Kagel and Alvin Roth. 111-94. Princeton: Princeton University Press.

Levine, David. 1998. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics*, 1: 593-622.

Levitt, Steven, and John List. 2009. "Field Experiments in Economics: The Past, The Present, and The Future." *European Economic Review*, 53(1): 1-18.

Loewenstein, George. 1996. "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes,* 65: 272-292.

Loewenstein, George. 1999. "Experimental Economics From the Vantage-Point of Behavioral Economics." *Economic Journal*, 109(453): 25-34.

Loewenstein, George. 2000. "Preferences, Behavior and Welfare: Emotions in Economic Theory and Economic Behavior." *American Economic Review,* 90(2): 426-432.

Loewenstein, George, and Drazen Prelec. 1992. "Anomalies in Intertemporal Choice: Evidence and an Interpretation." *Quarterly Journal of Economics*, 107(2): 573-597.

Luhan, Wolfgang, Martin Kocher, and Matthias Sutter. 2009. "Group Polarization in the Team Dictator Game Reconsidered." *Experimental Economics*, 12(1): 26-41.

Martinsson, Peter, Kristian Myrseth, and Conny Wollbrant. 2012. "Reconciling Pro-Social vs. Selfish Behavior: On the Role of Self-Control." *Judgment and Decision Making,* 7(3): 304-315.

McKean, Margaret. 2000. "Common Property: What Is It, What Is It Good For, and What Makes It Work?" In: *People and Forests: Communities, Institutions, and Governance*, ed. Clark Gibson, Margaret McKean, and Elinor Ostrom. 27-55. Cambridge, MA: MIT Press.

Mongin, Philippe. 1995. "Consistent Bayesian Aggregation." *Journal of Economic Theory*, 66(2): 313-351.

Myerson, Joel, and Leonard Green. 1995. "Discounting of Delayed Rewards: Models of Individual Choice." *Journal of Experimental Analysis of Behavior*, 64: 263–276.

Nikiforakis, Nikos. 2008. "Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?" *Journal of Public Economics*, 92: 91-112.

Nikiforakis, Nikos, and Hans-Theo Normann. 2008. "A Comparative Statics Analysis of Punishment in Public Good Experiments." *Experimental Economics*, 11: 358-369.

O'Donoghue, Ted, and Matthew Rabin. 1999a. "Doing It Now or Later." *American Economic Review*, 89(1): 103-124.

O'Donoghue, Ted, and Matthew Rabin. 1999b. "Incentives for Procrastinators." *Quarterly Journal of Economics*, 114(3): 769-816.

O'Donoghue, Ted, and Matthew Rabin. 2001. "Choice and Procrastination." *Quarterly Journal of Economics*, 116(1): 121-160.

Oliver, Pamela. 1980. "Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations." *American Journal of Sociology*, 85(6): 1356-1375.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge, U.K.: Cambridge University Press.

Ostrom, Elinor. 1999. "Coping with the Tragedy of the Commons." *Annual Review of Political Science*, 2: 493-535.

Ostrom, Elinor. 2000. "Collective Action and the Evolution of Social Norms." *Journal of Economic Perspectives*, 14(3): 137-158.

Ostrom, Elinor, James Walker, and Roy Gardner. 2002. "Covenants With and Without a Sword: Self Governance is Possible." *American Political Science Review,* 86: 404-417.

Pender, John. 1996. "Discount Rates and Credit Markets: Theory and Evidence from Rural India." *Journal of Development Economics,* 50: 257-296.

Plous, Scott. 1995. "A Comparison of Strategies for Reducing Interval Overconfidence in Group Judgment." *Journal of Applies Psychology*, 80(4): 443-454.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review,* 83(5): 1281-1302.

Rabin, Matthew. 1998. "Psychology and Economics." *Journal of Economic Literature*, 36: 11-46.

Rabin, Matthew. 2002. "A Perspective on Psychology and Economics." *European Economic Review*, 46(5): 657-685.

Rachlin, Howard, Andreas Raineri, and David Cross. 1991. "Subjective Probability and Delay." *Journal of the Experimental Analysis of Behavior*, 55: 233-244.

Reynolds, Douglas, Jacob, Joseph and Reuben Sherwood. 2009. "Risky Shift versus Cautious Shift: Determining Differences in Risk Taking Between Private and Public Management Decision-Making." *Journal of Business and Economic Research*, 7(1): 63-77.

Rockenbach, Bettina, Abdolkarim Sadrieh, and Barbara Mathauschek. 2007. "Teams Take The Better Risks." *Journal of Economic Behavior & Organization*, 63: 412-422.

Roth, Alvin, Michael Malouf, and J. Keith Murnighan. 1981. "Sociological versus Strategic Factors in Bargaining." *Journal of Economic Behavior and Organization*, 2(2): 153-177.

Schelling, Thomas. 1984. "Self-Command in Practice, in Policy, and in a Theory of Rational Choice." *American Economic Review,* 74(2): 1-11.

Schelling, Thomas. 1992. "Self-command: A New Discipline." In *Choice over time,* ed. George Loewenstein and Jon Elster, 167-76. New York: Russell Sage Foundation.

Sutter, Matthias, Stefan Haigner, and Martin Kocher. 2010. "Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations." *Review of Economic Studies,* 77: 1540-1566.

Sefton, Martin, Robert Shupp, and James Walker. 2007. "The Effect of Reward and Sanctions in the Provision of Public Goods." *Economic Inquiry,* 45: 671-690.

Shelley, Marjorie. 1993. "Outcome Signs, Question Frames and Discount Rates." *Management Science,* 39(7): 806-815.

Sniezek, Janet, and Rebecca Henry. 1989. "Accuracy and Confidence in Group Judgment." *Organizational Behavior and Human Decision Processes*, 43(1): 1-28.

Sobel, Joel. 2005. "Interdependent Preferences and Reciprocity." *Journal of Economic Literature*, 43: 392-436.

Song, Fei. 2008. "Trust and Reciprocity Behavior and Behavioral Forecasts: Individuals versus Group-Representatives." *Games and Economic Behavior*, 62(2): 675-696.

Strotz, Robert. 1956. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies,* 23: 165-180.

Sutter, Martin. 2009. "Individual Behavior and Group Membership: Comment." *American Economic Review*, 99(5): 2247-2257.

Thaler, Richard. 1981. "Some Empirical Evidence on Dynamic Inconsistency." *Economic Letters,* 8: 201-207.

Thaler, Richard, **and Hersh Shefrin.** 1981. "An Economic Theory of Self-Control." *Journal of Political Economy,* 89(2): 392-406.

Trope, Yaacov, and Ayelet. Fishbach. 2000. "Counteractive Self-Control in Overcoming Temptation." *Journal of Personality and Social Psychology,* 79: 493-506.

Tversky, Amos, and Daniel Kahneman. 1986. "Rational Choice and the Framing of Decisions." *The Journal of Business,* 59(4): 251–278.

Tyran, Jean-Robert, and Lars Feld. 2006. "Achieving Compliance when Legal Sanctions Are Non-Deterrent." *Scandinavian Journal of Economics*, 108(1): 135-56.

Walker, James, Roy Gardner, Andrew Herr, and Elinor Ostrom. 2000. "Collective Choice in the Commons: Experimental Results on Proposed Allocation Rules and Votes." *Economic Journal*, 110(460): 212-34.

Wertenbroch, Klaus. 1998. "Consumption Self-Control by Rationing Purchase Quantities of Virtue and Vice." *Marketing Science,* 17: 317-337.

Zelmer, Jennifer. 2003. “Linear Public Goods Experiments: A Meta-Analysis.“ *Experimental Economics*, 6: 299-310.