Behavior under Risk and in Social Dilemmas: Theoretical and Experimental Approaches

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

2011

vorgelegt von
Johannes Karl-Georg Maier

Referent: Prof. Dr. Klaus M. Schmidt
Korreferent: Prof. Dr. Martin G. Kocher
Promotionsabschlussberatung: 8. Februar 2012

Datum der mündlichen Prüfung: 30. Januar 2012
Namen der Berichterstatter: Klaus M. Schmidt, Martin G. Kocher, Fabian Herweg

Acknowledgements

First and foremost, I would like to thank my first supervisor Klaus M. Schmidt for his invaluable support during the time of my dissertation. His guidance was challenging and stimulating. To me it provided the perfect compromise of indicating eligible directions without setting too narrow guidelines. I would also like to thank Martin G. Kocher, my second supervisor, for always having an open door. The numerous comments I received were really helpful, often in deciding which path to take concerning specific issues that emerged. Moreover, I am grateful to Fabian Herweg, who agreed to join the thesis committee as third supervisor for the oral examination. I demanded his opinion on several issues and related discussions were truly valuable and encouraging.

More than anything, I would like to thank my family, in particular my mother, and those close to me for their constant and immense support. It is good to know that the same support would have been given to me regardless of the path I was choosing. From my colleagues in Munich, of whom many became friends, I did not only receive expedient comments at all stages of this dissertation, but very useful assistance in the somewhat stressful final stage. They really made the last couple of years a joyful experience for which I am very thankful. Moreover, I would like to thank colleagues I had the chance to meet at seminars, workshops, summer schools, and conferences. They contributed to this dissertation by providing excellent comments and, further, motivated me by being interested in my work.

This dissertation would not exist in its present form without my two co-authors to whom I am deeply indebted. From my first day in the *Munich Graduate School of Economics*, Jan Schikora has been a great colleague, office mate, and friend. While his all-round talents go far beyond research, the empirical part of the fourth chapter profited very much from his efficient but thoughtful way of doing research. The first three chapters of this dissertation root in joint work with Maximilian Rüger. I guess none of us could have imagined more than 15 years ago, when we were sitting next to each other in school, that we would be co-authoring our dissertations. Fortunately, it turned out to be different. Max accompanied me during my entire career, was helpful, reliable, and always there as an extraordinary friend and colleague. He inspired me for the field of economics at the beginning of my studies. I enjoyed not only the time when we were working together, but also the moments when we were relaxing from it. His dedication and perfection with which he acquired and imparted knowledge in many different areas were truly admirable. I learned a lot from him and our lively discussions were fruitful in many respects. I am really thankful that I had the pleasure to know him, while still asking why he had to leave so early.

Contents

List of Tables

List of Figures

Preface

Often, standard economic theory seems not to provide correct predictions of individuals' behavior, both when they act individually and when they interact with each other. This becomes especially problematic if there are *systematic* deviations in the behavior of people and the way this behavior is usually modeled in economics. Behavioral economics as a way of thinking constitutes the belief that economists should aspire to make assumptions about humans that are as realistic as possible. If these more realistic assumptions then lead to models which are able to better explain observed phenomena (Task A), there is a need to modify the classical approach to economics. Experiments can guide us towards these models. On the one hand, they provide a controlled environment which allows to test isolated hypotheses derived from the models aimed to be validated (Task B). On the other hand, experiments can disclose phenomena that represent empirical facts urging an explanation by theoretical models (Task C). Reliable conclusions can, however, only be drawn if experimental methods are robust towards various underlying models that can potentially explain the observed behavior (Task D).[1]

This dissertation sheds light on these various tasks that, among others, behavioral and experimental economics aim to fulfill.[2] In doing so, I focus on two areas which have received enormous attention in the literature. First, behavior under risk (Chapters 1, 2, and 3) and second, behavior in social dilemmas (Chapter 4).

Decision making in the face of risk is an area in which standard economic theory has long been challenged for its predictive power. Several models have been proposed that deviate from the canonical model of expected utility theory (EUT), most prominently

[1]For instance, in experiments it has become common to elicit certain preferences which are possibly responsible for observed behavior, and hence need to be controlled for, with specific experimental methods. These methods (or at least their outcomes) must be robust towards various underlying models that have been identified to explain such preferences.

[2]Tasks A to D are by no means exhaustive and are not meant to represent an exclusive list of tasks to be fulfilled by the fields of behavioral and experimental economics. They are only intended to guide the reader through the contributions this dissertation aims to make. While Task A is addressed to behavioral economics and can be understood as 'theory-building' task, Tasks B, C, and D are rather addressed to experimental economics and can be respectively interpreted as 'theory-testing', 'facts-generating', and 'method-validating' tasks.

models which base utility not only on absolute wealth levels as in EUT, but in which utility (additionally) depends on how these possible wealth levels compare to reference levels of wealth. This prominent strand of literature evolved in order to resolve various empirical puzzles concerning second-order risk preferences, namely risk aversion. The corner stone of such models is Kahneman and Tversky's (1979) prospect theory (PT). In addition to the reference-dependent nature of utility, which has already been proposed earlier by Markowitz (1952) and will be discussed in more detail later, PT provided an alternative to another central element of EUT which is that probabilities enter linearly into the evaluation of choices.

This feature of PT led to the development of so-called non-EUT models which allow for non-linear influences of probabilities and thereby try to realistically capture individuals' perception of these probabilities. An important advancement has been made since earlier models, like PT in its original formulation, violated stochastic dominance criteria because they applied the same weighting function to all probabilities of the corresponding states of nature. Rank-dependence (see Quiggin, 1982; Kahneman and Tversky, 1992) solved the problem of stochastic dominance violations and still allows for probability weighting, depending on the outcome level of corresponding states of nature. While in EUT, only the value function, which transforms outcomes into utilities, can generate risk-averse behavior, in non-EUT models the probability weighting function can generate risk aversion as well.

In experiments, behavior may often be (partly) driven by individuals' risk attitudes. In order to control for individual-specific risk preferences, experiments often employ the multiple-price list method of Holt and Laury (2002). While having several remarkable advantages, this method requires EUT in the classification of subjects as more or less risk-averse. Nowadays, it is a widely established fact that some individuals behave consistent with EUT while others rather follow non-EUT models (see e.g. Harrison, Humphrey, and Verschoor, 2010). Since risk aversion in these models is generated by the value and the probability weighting function, it is hard to compare individuals' risk attitude across different models based on their parameter estimates. Assuming the same underlying model for all individuals and thereby disregarding their heterogeneity is inappropriate as well and may lead to a misclassification. However, risk aversion has been defined in a model-independent, purely behavioral sense, namely as an aversion to mean-preserving spreads (see Rothschild and Stiglitz, 1970).[3] An individual is then defined as being less risk-averse than another individual if she accepts more such in-

[3]A lottery is less risky than another lottery if the latter is a mean-preserving spread of the former, describing that the latter can be generated from the former by adding mean-zero noise. In its simplest form, a mean-preserving spread is the lottery rather than this lottery's expected value.

creases in risk in exchange for a fixed compensation. This model-independent definition of being more risk-averse has found broad acceptance and has been applied to EUT (see Diamond and Stiglitz, 1974) as well as non-EUT models (see e.g. Machina, 1982, 1987; Chew, Karni, and Safra, 1987, or Röell, 1987).

In Chapter 1 (which is joint work with Maximilian Rüger), we use this definition of risk aversion and propose a modification of the Holt and Laury (2002) (HL) method, that does not require EUT for the classification of individuals as more or less risk-averse. In fact, our method can consistently be applied even if individuals follow different models of risky choice. The only requirement for the underlying models is that they are consistent with the behavioral definition of risk aversion in terms of disliking mean-preserving spreads. This work therefore pursues Task D as mentioned in the beginning. In order to see whether the methodological problem of assuming a specific underlying model is in fact empirically relevant, we directly compare the two methods in an experiment. Our findings suggest that the two elicitation methods yield identical results concerning the classification of subjects as risk-averse, -neutral, or -seeking. However, with respect to classifications as more or less risk-averse the two methods diverge substantially. The average subject in our experiment changes her relative standing of being more, less, or equally risk-averse to about half of the other subjects across the two elicitation methods. Thus, assuming EUT for such a classification can be highly misleading.

When measuring the absolute level of risk aversion intensity via the class of constant relative risk aversion utility functions, we further find that our method yields higher intensities of risk aversion than the HL method, and only with our method are these estimates robust towards increasing stakes. Consistent with the results of Holt and Laury (2002), the measure of risk aversion is increasing with stake size using the HL method. A cognitive effort explanation is provided that accounts for these results. This conjecture is based on the fact that our method uses variations in outcomes but holds the probabilities of these outcomes constant at 50%, while the HL method uses variations in probabilities which may be harder to conceive for subjects in the experiment.

Chapter 1 is entirely concerned with second-order risk preferences. However, not all important behavior under risk can be captured by risk aversion alone. It is well known that higher-order risk preferences, such as prudence (third order, also called downside risk aversion) or temperance (fourth order, also called outer risk aversion), are crucial for various economically important phenomena of risky decision making in areas such as saving, insurance, auctions, bargaining, prevention, or global commons problems. Among these various applications, higher-order risk preferences have first and foremost been identified to play *the* crucial role for so-called precautionary savings which describe the part of present saving warranted to compensate the riskiness of the future. Already Leland (1968) and Sandmo (1970) identified the importance of a positive third derivative of the value function for motives of precautionary saving. Since Kimball (1990) the feature of preferences that is necessary and sufficient (given preferences are monotone and risk-averse) for such precautionary behavior is termed prudence. Recently, Eeckhoudt and Schlesinger (2008) showed that while prudence is necessary and sufficient for precautionary saving to increase when the future becomes more risky in terms of second-order risk increases (mean-preserving spreads), temperance is necessary and sufficient for precautionary saving to increase when the downside risk of the future increases, which is rather a third-order increase in risk (or loosely speaking, a change in the skewness of the income distribution).

In addition to this large theoretical literature, there are also numerous empirical studies investigating the relevance of precautionary saving. As reported for instance in an overview article by Carrol and Kimball (2008), in these studies precautionary saving is often observed but to a much lesser extent than what would be optimal under standard EUT predictions using reasonable degrees of risk aversion. A result in similar vein has been found in a laboratory experiment by Ballinger, Palumbo, and Wilcox (2003). Closely related is what Gollier (2003) termed the 'insurance puzzle'. He shows within a dynamic life-cycle EUT model that allowing for accumulated precautionary savings leads theoretically to a crowding out of costly insurance demand. Of course, this is in stark contrast to empirical observations and hence a puzzle.

So far, models analyzing higher-order risk preferences were constrained to EUT. Within this framework, higher-order risk preferences correspond to the derivatives of the value function. Similar to second-order risk preferences, where risk aversion is equivalent to a negative second derivative, prudence has been identified to be equivalent to a positive third derivative, and temperance to a negative fourth derivative of the value function. But what do higher-order risk preferences represent from a model-independent and rather behavioral perspective? Similar to the definition of risk aversion in terms of an aversion to mean-preserving spreads, Eeckhoudt and Schlesinger (2006) defined lotteries representing higher-order risk attitudes independently of a specific model and showed that within EUT their gamble definitions correspond to the definitions in terms of derivatives used so far. These model-independent definitions draw the curtain to analyzing higher-order risk preferences outside EUT, just as mean-preserving spreads made it possible to analyze risk aversion outside EUT.

As mentioned earlier, the most prominent alternatives to standard EUT models recognize the importance of a reference point for utility. This seems a quite natural extension since our brain often evaluates situations relative to other situations. "When we respond to attributes such as brightness, loudness, or temperature, the past and present context of experience defines an adaptation level, or reference point, and stimuli are perceived in relation to this reference point. Thus, an object at a given temperature may be experienced as hot or cold to the touch depending on the temperature to which one has adapted. The same principle applies to non-sensory attributes such as health, prestige, and wealth." (Kahneman and Tversky, 1979, p277) While PT assumed the reference point to be a single point, often advocated to be the status quo, more recent developments capture the idea that individuals make multiple comparisons when evaluating choices under risk and thereby take the uncertainty of the choice into account. In these models, the reference point is assumed to be a distribution to which each potential outcome is compared to (see Sugden, 2003; Delquié and Cillo, 2006; Kőszegi and Rabin, 2006, 2007, 2009). In doing so these models abstract from issues of probability weighting and retain the expected utility structure, meaning that probabilities enter linearly into the choice evaluation. Another important advancement, made by Kőszegi and Rabin (2006, 2007, 2009), was to endogenize the reference point. In their model, the reference point is shaped by *recent* expectations, so that the chosen alternative's distribution becomes the reference distribution in a so-called 'personal equilibrium'.[4] This step closes the model and replies to advocates of classical EUT who criticized reference-dependent models for introducing the reference point as an additional degree of freedom.

Another feature of reference-dependent models has always been the specific shape of the value function.[5] Earlier models assumed an S-shaped value function that was concave (and thus risk-averse) for outcomes above the reference point, the gain domain, but convex (and thus risk-seeking) for outcomes below the reference point, the loss domain. Further, loss aversion captured the notion that losses loom larger than equally-sized gains, represented by a kink of the value function at the reference point. More recent applications of reference-dependent preferences à la Kőszegi and Rabin often abandon the S-shape and retain loss aversion when they assume a piece-wise linear

[4]In the model of Kőszegi and Rabin (2006, 2007, 2009) it is not entirely clear what recent actually means. They note that in out-of-'personal equilibrium' surprise situations, where the reference point does not have sufficient time to adjust to the upcoming choice, an individual may still expect to stay at the status quo making it the relevant reference point, even if this is not an option that can be chosen.

[5]This roots in the fact that with the reference point being a simple point, a reference-dependent model is equivalent to an EUT model if the shape of the value function does not depend on a reference point.

value function. The evidence on the exact shape of the value function is rather mixed. While several experiments have shown that most individuals are risk-averse in the gain domain, this debate is not settled in the loss domain because institutional constraints usually impede that subjects in experiments are exposed to real losses. Thus, most evidence is based only on hypothetical choices or subject to 'house-money' effects (see Thaler and Johnson, 1990).

In contrast, the evidence is less mixed when it comes to the question of what shapes the reference point. Numerous field as well as carefully designed laboratory studies have recently been conducted that suggest that it is individuals' expectations that determine the reference point. In pursuing Task A as outlined in the beginning, there is also a rapidly growing theoretical literature which applies Kőszegi and Rabin's model of expectation-based reference-dependent preferences to explain important phenomena in areas of industrial organization, optimal contracts, or auctions. The common theme of these studies is that they are usually concerned with phenomena that rely on first- and/or second-order preferences. To date, there has been no systematic treatment of higher-order risk preferences under reference dependence.

Chapter 2 (which is joint work with Maximilian Rüger) closes this gap. More specifically, we use Eeckhoudt and Schlesinger's (2006) model-independent gamble definitions in order to analyze higher-order risk preferences in Kőszegi and Rabin's (2007) reference-dependent model. Our findings suggest that risk preferences of higher orders are fundamentally different under reference dependence than in EUT. Standard formulations of EUT imply that the derivatives of the value function alternate in sign, leading to risk aversion, prudence, temperance, etc.. Under expectation-based reference dependence, individuals exhibit even- but never odd-order risk attitudes. This result is not due to a specific formulation of the value function and holds completely independent of its functional form. Since Chapter 2 is also devoted to Task A, it is further shown how these results can explain the empirical patterns of seemingly sub-optimal behavior concerning precautionary saving and insurance demand.

As a robustness analysis, we additionally show that alternative assumptions of reference points that have been proposed in the literature cannot explain these patterns. For instance, disappointment models (see Bell, 1985; or Loomes and Sugden, 1986) use the mean of expectations as the reference point rather than the full distribution. This close relationship to Kőszegi and Rabin's (2006, 2007) model has made it impossible for recent experimental work aimed at testing expectation-based reference dependence, and thereby performing Task B, to actually distinguish between the model of Kőszegi and Rabin (2006, 2007) and disappointment models. Our analysis suggests that this is due to being constrained to the first two orders. Under disappointment, we do not observe the absence of odd-order risk preferences while even-order preferences are similar to those of expectation-based reference dependence. This finding suggests a new way, namely using higher orders, to differentiate between these two models when executing Task B. Another popular form of reference dependence, although having emerged quite separately, focuses on the anticipation of ex-post regret feelings in the choice evaluation (see Bell, 1982, 1983; or Loomes and Sugden, 1982). In these regret models, the reference point is the alternative that was *not* chosen. We show that decisions taken to avoid regret are unaffected by even-order effects, but odd-order effects are still present. This result suggests that regret influences higher-order risk preferences in the exactly opposite way as expectation-based reference dependence does. As a result, risk preferences of a particular order are only affected by either expectations or regret as the reference point but never by both. Our analysis of alternative reference points, be they endogenous (as under disappointment or regret) or exogenous (like the status quo), shows that they imply different results than expectation-based reference dependence with respect to higher-order risk attitudes. Therefore, these alternatives cannot similarly explain the empirical puzzles observed.

Until recently, the empirical literature on higher-order risk preferences has exclusively focused on derived behavior from such preferences rather than on the preferences themselves. But it is important to ask whether individuals in fact exhibit those preferences that drive such behavior. Deck and Schlesinger (2010) were the first who provided direct evidence on higher-order risk attitudes. Using the model-independent gamble definitions of Eeckhoudt and Schlesinger (2006), they found that individuals are prudent but intemperate. However, more studies are warranted to derive at reliable conclusions when the aim is to fulfill Task C. In Chapter 3 (which is joint work with Maximilian Rüger) we contribute to this task and provide further experimental evidence on higher-order risk attitudes. Deck and Schlesinger (2010) interpreted their findings to be consistent with PT but not with standard EUT models. As pointed out before, a crucial feature of PT is the distinction between gains and losses. As in most other experiments that investigate the loss domain, Deck and Schlesinger (2006) imposed losses by a specific framing of the choice situation. Their results may therefore be subject to 'house-money' effects. Bosch-Domènech and Silvestre (2006) proposed an experimental design where the same subjects participate in two experiments separated from each other by several weeks. Their joint earnings from both experiments cannot be negative, but they can be negative for the second experiment. Bosch-Domènech and Silvestre (2006) only studied second-order risk attitudes but, interestingly, found that a majority of subjects displayed risk aversion in both gains and *real* losses (for sufficiently high stakes).

In Chapter 3, we use such a two-date design for our experiment in order to impose real losses. Although being unable to control for, we most likely implemented the status quo (and not expectations) as the reference point as we pointed out that subjects can make gains just as well as they can make losses at the second date and gave no further information about the choices to be made. We find that subjects are risk-averse, prudent, and temperate. Preferences are not affected by the domain so that this pattern holds in gains as well as losses. Further, risk preferences of different orders are highly correlated. Although Chapter 3 has its focus in contributing to Task C, it is consistent with Task D as we also used Eeckhoudt and Schlesinger's (2006) model-independent gamble definitions. [6] With respect to Task B, our results are consistent with standard EUT models. The derivatives of the value function do not change their sign across the gain and loss domain. Also, the high correlation between orders would be predicted by standard EUT models. Although our evidence speaks against an S-shaped value function as promoted in PT, it does not speak against the concept of a reference point or loss aversion. And as shown in Chapter 2, reference dependence can have effects on higher-order risk preferences quite independently from the shape of the value function. Therefore, our experimental evidence on higher-order risk attitudes does not undermine the findings of our theoretical analysis.

One of the most crucial contributions for individual behavior under risk has been von Neumannn and Morgenstern's (1944) axiomatic foundation of EUT. This approach still justifies EUT as the appropriate *normative* concept. In the same work, von Neumannn and Morgenstern (1944) further introduced main features of game theory which describes behavior when individuals (being it humans, firms, nations, etc.) interact with each other. Later, the concepts of equilibrium and fully rational (and selfish) individuals have become standard tools in the analysis of games. Sometimes and especially for humans, however, these concepts seem out of place. Several experiments, using a multitude of games, have been conducted that suggest that individuals often violate these concepts. Behavioral game theory, as a branch of behavioral economics, has the goal to capture such violations. "Behavioral game theory is about what players *actually* do. It expands analytical theory by adding emotion, mistakes, limited foresight, doubts about how smart others are, and learning to analytical game theory. [...] behavioral game theory can explain what people do more accurately by extending analytical

[6] These lotteries are especially suited for experimental purposes as they have the virtue to rigorously define higher-order risk attitudes in an accessible way.

game theory to include how players feel about the payoffs other players receive, limited strategic thinking, and learning." (Camerer, 2003, p3/7)

The last part of this dissertation focuses on two of these avenues that have enriched standard game theory, caring about others and limited strategic thinking. In doing so, I focus on one of the most popular games capturing social dilemmas, namely public goods games. The leading textbook introductory example for game theory is the prisoners dilemma game. It captures a social dilemma as, from the equilibrium point of view (in which neither player cooperates), both players would be better off if they both cooperated, but given that the other player cooperates, are better off by not cooperating. The famous public goods game is, loosely speaking, a continuous version of this game where players can cooperate at different degrees. Still, full cooperation is socially desirable, but each player's dominant strategy is not to cooperate. The literature on testing the standard theoretical prediction is vast. However, most field as well as laboratory evidence falsifies the standard prediction of no cooperation not only in one-shot but also in repeated versions of the public goods game.

"Experiments on public goods began with the aim of testing the received theory, but experimenters quickly turned to investigating and trying to understand the decay of overcontribution quite independently from the existing of an alternative theoretical framework (albeit with the aim of developing one)." (Guala, 2005, p40) In other words, experiments started out with Task B, then turned to Task C, but always had the goal to propose a model satisfying Task A. The latest important advancement with respect to Task C has been the experimental work of Fischbacher and Gächter (2010) on dynamic contribution behavior in repeated public goods games. They used a two-part experimental design. In the first part, they elicited subjects' contribution preferences conditional on the averages of others' contributions by employing the elicitation method of Fischbacher, Gächter, and Fehr (2001) (which is based on Selten's, 1967 strategy method). These preferences were then used to empirically predict contributions in the second part of the experiment which implemented a repeated public goods game. In the first part, the authors showed that people are 'imperfect conditional cooperators' who match contributions of others only partly. With the help of data-based simulation techniques, they could show that it is this preference that drives the decay of contributions in the second part of the experiment. Furthermore, and in contrast to other models focusing on the interaction of different types of players (see e.g. Ambrus and Pathak, 2011), they showed that the presence of purely selfish types is not needed for contributions to be declining. This is an interesting result from a theoretical point of view, because the preference of *imperfect* conditional cooperation would predict *zero* contributions in equilibrium, both in (simultaneous) one-shot and all periods of repeated public goods games.

Chapter 4 shows that a preference for imperfect conditional cooperation also *theoretically* explains the dynamics of contribution behavior in a repeated public goods game. The presented model accounts for limited strategic reasoning since it is based on the non-equilibrium concept of level-k thinking. The standard game-theoretic notion of strategic thinking, represented by Nash equilibrium, postulates that each player forms *correct* beliefs about other players' behavior and best responds to these beliefs. In contrast, level-k models of strategic thinking are only rational in that players best respond to their beliefs, but these beliefs are derived from simplified non-equilibrium models of others.[7] Level-k thinking is closely related to so-called cognitive hierarchy models (see Camerer, Ho, and Chong, 2004) where level-k players best respond to an estimated mixture of lower-level players. However, in a level-k model a player of level-type k rather best responds to level-$k-1$ players. This form of limited strategic thinking was introduced by Nagel (1995) and Stahl and Wilson (1994, 1995). Since then, it has been further developed and applied to explain behavior in a variety of settings, including guessing games, auctions, hide-and-seek games, or coordination games. It has also been used to answer mechanism design questions. Empirical evidence on level-k thinking mainly concentrates on data from laboratory experiments, but there is also some field evidence supporting such non-equilibrium behavior (see Crawford, Costa-Gomes, and Iriberri, 2010 for an overview of the literature on strategic thinking).

The model of Chapter 4 applies level-k thinking to a repeated public goods game, predicts common patterns of public goods contribution behavior, and thereby contributes to Task A. The theoretical analysis of Chapter 4 is complemented by a two-part experiment (which is joined work with Jan Schikora), similar to the one of Fischbacher and Gächter (2010) but with design variations that allow us to make further contributions with respect to Tasks B, C, and D. The first part of our experiment elicits subjects' contribution preferences conditional on others' *individual* contributions in homogeneously as well as heterogeneously endowed groups. With this specific design we can answer two open questions. First, we show that eliciting contribution preferences conditional on the *averages* of others' contributions can well be justified. Hence, the often used elicitation method of Fischbacher, Gächter, and Fehr (2001) seems robust with respect to inequalities in others' contributions. This is an important insight from the perspective of Task D as leading models of outcome-based social preferences would predict that this variation matters for own contributions. Second, conditional coop-

[7] Interestingly, level-k thinking respects k-rationalizability (see Bernheim, 1984; Pearce, 1984).

eration has previously been investigated only under homogeneous endowments where it is impossible to distinguish whether own contributions are conditional on absolute or relative contributions of others. We show that the average preference pattern is consistent with imperfect *relative* rather than absolute conditional cooperation. This result contributes to Task C and should be considered in future approaches to model the foundations of conditional cooperation.[8]

Based on these findings, the preference of imperfect relative conditional cooperation is then incorporated into the level-k model in order to analyze a repeated public goods game. Besides allowing to derive predictions (and hypotheses thereof) on contribution dynamics that are in line with the results of Fischbacher and Gächter (2010) and would also obtain under imperfect absolute conditional cooperation, the model offers further predictions on the effects of heterogeneous and uncertain endowments that only obtain under imperfect relative conditional cooperation.[9] In performing Task B, all these predictions are tested and *validated* in the second part of our experiment which implements a repeated public goods game, with treatments capturing heterogeneous and uncertain endowments. More specifically, we show that subjects strategically over-contribute, meaning that they contribute more in the repeated version of the game than what is predicted from their elicited (one-shot) preferences. In line with the results of Fischbacher and Gächter (2010), these over-contributions as well as total contributions are declining over time. With respect to our treatment variations, we find that heterogeneous groups are less efficient than homogeneous groups in providing the public good. However, this is not the case under uncertain endowments. Here, poor groups reduce their contributions compared to certainty while rich groups' contributions increase. Thus, uncertainty causes the inequality of the income distribution over all subjects to increase. Regarding further possible policy implications, our experimental results also suggest that homogeneous groups with endowments being common knowledge are most efficient in the public goods provision.

All four chapters of this dissertation can be read independently from each other. They all have their own introductions and appendices (covering extensions, proofs, additional information, and experimental instructions), but a joint bibliography at the end of the dissertation.

[8]This is not a question we approach. Our aim is rather to show how this preference is able to explain contribution behavior in a repeated game.

[9]The model makes additional predictions with respect to variations that are not testable with the specific design of the second part of our experiment. Nevertheless, these predictions seem to be in line with previous accounts of the literature.

Chapter 1

Measuring Risk Aversion Model-Independently[0]

1.1 Introduction

In order to measure individuals' risk attitudes, the multiple price-list method[1] of Holt and Laury (2002) has become the industry standard in experimental economics. Major advantages that led to the popularity of the Holt and Laury (HL) tables include its transparency to subjects (easy to explain and implement), its incentivized elicitation, and that it can be easily attached to other experiments where risk aversion may have an influence. Nevertheless, the HL method has also its drawbacks. The major disadvantage is that it requires a specific utility framework such as expected utility theory (EUT) in order to classify subjects as more or less risk-averse.[2] If individuals' risk preferences are heterogeneous in the way that some act according to EUT while others rather act according to non-EUT, it becomes problematic to use the HL tables in order to classify subjects' risk attitudes. The reason is that the HL method is not based on a general notion of increasing risk which is satisfied by EUT *and* non-EUT models.

To account for this disadvantage, we propose a modification of the HL tables. This new method is based on the well-known increasing risk definitions of Rothschild and Stiglitz (1970). These definitions are solely in terms of attributes of the distribution

[0]This chapter is based on joint work with Maximilian Rüger.

[1]To our knowledge, a multiple price-list format was first used by Miller, Meyer, and Lanzetta (1969).

[2]Holt and Laury (2002) use specific parametric forms of EUT in order to classify subjects. Following the approach of structural estimation, one could also estimate parameters of a specific non-EUT model based on the HL tables. However, the crucial point is that one would have to assume that all subjects follow a common model.

function and are therefore independent of a utility framework. Moreover, they have been used to characterize risk aversion in both EUT and non-EUT models (e.g. Machina, 1982, 1987; Chew, Karni, and Safra, 1987; Röell, 1987; Yaari, 1987; or Schmidt and Zank, 2008). By just imposing 'duality' asserting that less risk-averse individuals accept riskier gambles (see Diamond and Stiglitz, 1974), our method enables us to classify subjects as more or less risk-averse without assuming a specific utility framework. It is therefore applicable with heterogeneous risk preferences, which is a desired feature for studies that need to control for risk aversion in behavior. Furthermore, our approach does not require subjects to handle various and rather complex probabilities, since it uses variations in outcomes (i.e. mean preserving spreads) and holds probabilities of outcomes constant at 50%.

In a laboratory experiment we directly compare the HL method and our method using low and high stakes. This is of interest because it sheds light on whether the methodological differences are of empirical relevance. We find that both methods yield the same classification of individuals concerning the *direction* of risk attitudes (i.e. risk-averse, risk-neutral, or risk-seeking). However, we also find that both methods yield diverging results concerning the *intensity* of risk attitude. The classification of individuals as being more or less risk-averse than others is quite different between both methods. Moreover, we find that our method yields higher levels of risk aversion intensity that are much closer to what is observed in the field. These estimates of risk aversion intensity are robust towards multiplying the stakes by five only when our method is used. For the HL method we can confirm the result of Holt and Laury (2002) that increasing the stakes increases risk aversion. It is also shown that these results are robust towards possible confounds like certain switching preferences (e.g. switching in the middle of a table) or order effects (with respect to elicitation methods and stakes).

This chapter is structured as follows. In Section 1.2 we review the HL tables and note further advantages and disadvantages. We then propose a new method that shares the advantages but not the disadvantages of the HL tables in Section 1.3. The experiment used to directly compare both methods is explained in Section 1.4. The results of our experiment as well as their robustness are discussed in Section 1.5 and the conclusion can be found in Section 1.6. In Appendix A we generalize our new method. This allows for a customization and extends the range of potential applications in future experiments. Appendix B presents the instructions of our experiment.

1.2 The Holt and Laury Method

Measuring the intensity of risk preferences is very important for theoretical predictions. Also, in experiments individuals' decisions are often (partly) driven by their risk preferences. In order to control for these individual-specific characteristics, the multiple price-list method of Holt and Laury (2002) is commonly used in experiments nowadays. Table 1.1 presents the original HL design.

Table 1.1: The Holt and Laury Method

	Option A		Option B		RRA if row was		
Row No.	Outcome A1 = \$2.00	Outcome A2 = \$1.60	Outcome B1 = \$3.85	Outcome B2 = \$0.10	last choice of A and below all B	EV[A]− EV[B]	Var[A]− Var[B]
1	Prob. 1/10	Prob. 9/10	Prob. 1/10	Prob. 9/10	$[-1.71\ ;\ -0.95]$	1.17	-1.25
2	Prob. 2/10	Prob. 8/10	Prob. 2/10	Prob. 8/10	$[-0.95\ ;\ -0.49]$	0.83	-2.22
3	Prob. 3/10	Prob. 7/10	Prob. 3/10	Prob. 7/10	$[-0.49\ ;\ -0.14]$	0.50	-2.92
4	Prob. 4/10	Prob. 6/10	Prob. 4/10	Prob. 6/10	$[-0.14\ ;\ 0.15]$	0.16	-3.34
5	Prob. 5/10	Prob. 5/10	Prob. 5/10	Prob. 5/10	$[0.15\ ;\ 0.41]$	-0.18	-3.84
6	Prob. 6/10	Prob. 4/10	Prob. 6/10	Prob. 4/10	$[0.41\ ;\ 0.68]$	-0.51	-3.34
7	Prob. 7/10	Prob. 3/10	Prob. 7/10	Prob. 3/10	$[0.68\ ;\ 0.97]$	-0.85	-2.92
8	Prob. 8/10	Prob. 2/10	Prob. 8/10	Prob. 2/10	$[0.97\ ;\ 1.37]$	-1.18	-2.22
9	Prob. 9/10	Prob. 1/10	Prob. 9/10	Prob. 1/10	$[1.37\ ;\ \infty)$	-1.52	-1.25
10	Prob. 10/10	Prob. 0/10	Prob. 10/10	Prob. 0/10	non-monotone	-1.85	0.00

An individual makes a decision between option A and option B in each of the ten rows. Option A as well as option B can have two different outcomes (A1 or A2 and B1 or B2) with varying probabilities over the ten rows. The expected outcome of option A is higher for the first four rows and lower for the last six rows (as indicated by the second to last column in Table 1.1). So, a risk-neutral subject should choose option A in row 1 to 4 and then switch over and choose option B in row 5 to 10. However, as option B has a higher variance (indicated by the last column in Table 1.1), there is a trade-off when to switch to option B. Clearly, by row 10 everybody should have switched to option B as it yields the higher outcome with certainty. An individual who switches to option B between row 6 and row 10 is classified as being risk-averse and the more risk-averse individual will switch later as she needs a higher expected value to choose the more variable option. Someone who switches earlier to option B (between row 1 and row 4) is classified as risk-seeking by similar arguments. In column 6 of Table 1.1 we report the risk preference intensity measured by the amount of relative risk aversion (RRA) that is induced from the switching behavior if we assume the class of constant relative risk-averse (CRRA) utility functions.[3]

[3] Then, utility of income is $u(x) = \frac{x^{1-r}}{1-r}$ where x is the lottery outcome and r the RRA parameter. Note that the bound in rows 3 and 4 of $r = -0.15$ as reported in the original article of Holt and Laury (2002) is in fact according to our calculation $r = -0.14$. Also, if the subject chooses always option B, her relative risk aversion is $r \in (-\infty; -1.71]$.

The advantages of the HL method are due to its design. It is very easy to explain to subjects since they only have to choose between option A and option B in each row.[4] It is incentivized and usually one of the ten rows is randomly selected and paid out for real. And because it is so easy to implement, the HL tables can be attached to other experiments where risk aversion may play a role.

Nevertheless, the HL method also has its disadvantages. One disadvantage is that there is no flexibility in adjusting the ranges of RRA without affecting the round-numbered probabilities. So, for instance, if one would want to decrease the ranges of the RRA intervals in row 4 to 6 in order to better classify most subjects' risk attitudes (according to Holt and Laury, 2002, 75% of the subjects fall into this category), one would have to give up the round-numbered probabilities in Table 1.1. One way to circumvent this problem is proposed by Andersen et al. (2006) using a complex more-stage procedure and thereby loosing the advantages of the simple HL design mentioned above.

The use of variations in probabilities (whereas outcomes are held constant) makes the HL tables sensitive to probability weighting. For instance, by using the standard parametric Prospect Theory assumptions (Tversky and Kahneman, 1992) on the probability weighting function, we obtain the result that a subject with a linear utility function should choose A only for the first three and not for the first four rows. Such an individual would be classified as risk-seeking in HL. This makes it difficult to draw conclusions about the shape of the utility function. Furthermore, varying probabilities require subjects to have some imagination of what these different probabilities mean.

The major disadvantage of the HL tables is, however, that they are not based on a general notion of increasing risk. They need a specific utility framework, namely EUT, in order to classify subjects as more or less risk-averse.[5] However, evidence rather suggests that risk preferences are heterogeneous and subjects follow different models of risky choice.[6] It is then not only problematic to impose EUT in order to classify subjects, but also to assume the very same choice model over all subjects. Hence, a more *general* measure of risk aversion intensity is needed that allows for a classification across

[4]A variant of the design is to induce a single switching point, i.e. to ask for the row where the subjects would want to switch from A to B.

[5]In order to discriminate between intensities of risk aversion, HL use EUT and the specific class of CRRA functions. Using the HL method and assuming for instance Tversky and Kahneman's (1992) Prospect Theory would require a 'trade-off' between the curvatures of the utility function and the probability weighting function since both simultaneously influence the level of risk aversion. However, even if such a 'trade-off' could be made, there is no way to compare subjects in case they follow different models of risky choice.

[6]The evidence that many individuals behave according to non-EUT models is vast. For instance, a recent study by Harrison, Humphrey, and Verschoor (2010) finds a 50/50 share of EUT and Prospect Theory.

different underlying models. In the next section, we therefore propose a modification of the HL method that is based on a general 'behavioral' notion of risk aversion, namely an aversion to mean preserving spreads.

Other elicitation methods (for a comprehensive recent review see Harrison and Rutström, 2008) where choices can also be used to directly deduce the intensity of risk aversion suffer from the same problem, i.e. they also need the assumption of EUT to classify subjects.[7] For instance, another popular alternative to the multiple price-list design is the ordered lottery selection design as used by Binswanger (1980, 1981) or Eckel and Grossman (2008). Here, subjects choose one out of eight (or five) different 50/50 gambles where higher expected values come at the cost of higher standard deviations (see also Jacobson and Petrie, 2009).

There is one alternative elicitation method that does not impose EUT for the classification of subjects but which has another problem. Asking subjects to state a certain amount that makes them indifferent to a given lottery (implemented e.g. via the Becker-DeGroot-Marschak mechanism as done for instance by Kachelmeier and Shehata, 1992, or via a price-list design as for instance done by Dohmen et al., 2010, 2011) can classify subjects as more or less risk-averse by their certainty equivalents.[8] The problem here is the so-called certainty effect which describes the widely observed phenomenon that certain alternatives are perceived in a fundamentally different way than risky alternatives, even if the risk is negligible. "The certainty effect introduces

[7] An alternative approach is to derive preference functionals that can be used to make statements about risk aversion. Here, one way is to use structural estimation techniques, which Holt and Laury (2002) additionally perform without the CRRA assumption. In fact, there is no restriction to EUT in this case but still classification as more or less risk-averse are not necessarily possible (even if subjects follow the same model of risky choice). The most prominent alternative to the multiple price-list design of Holt and Laury in this case is the random lottery pair design of Hey and Orme (1994), where subjects are asked to answer a battery of 100 choices with different probabilities. Another way is a deterministic approach, where chained responses are used to construct a utility function either with the assumption of EUT as in the certainty or probability equivalent methods, or without the assumption of EUT as in the trade-off method of Wakker and Deneffe (1996). The chained design of these methods suffers from being not incentive compatible and classifications as more or less risk-averse still require EUT. The closely related lottery equivalent method of McCord and de Neufville (1986) avoids the chained design (and the certainty effect) in order to obtain the utility function, but also requires EUT, not only for the construction of utility but also for a classification as more or less risk-averse via the uncertainty equivalent as elicited by Andreoni and Sprenger (2010).

[8] Closely related but somewhat reversed, Bruner (2009) varies the lottery and holds the certain amount fixed in order to investigate whether changing the probability or changing the reward, for a given increase in the mean of the lottery, is preferred. Consistent with EUT, he finds some evidence that subjects are more risk-averse (i.e. they choose more often the safe option) when the winning outcome instead of the winning probability is varied (such that the means of the lottery options across treatments are identical). We control for such effects as our RRA ranges are exactly matched across methods. This is further explained in Section 1.4. Note that varying the lottery instead of the fixed amount does not reduce certainty effect problems.

systematic errors into any method based on certainty equivalents." (McCord and de Neufville, 1986, p57)

1.3 A Model-Independent Method

In this section we propose a new method that shares the advantages of the HL table but not its disadvantages (and those of the alternative methods) as mentioned above. Table 1.2 presents our new approach.

Table 1.2: Our Elicitation Method

Row No.	Option A Prob. 1/2 Outcome A1	 Prob. 1/2 Outcome A2	Option B Prob. 1/2 Outcome B1	 Prob. 1/2 Outcome B2	RRA if row was first choice of A and above all B	RRA if row was last choice of A and below all B
1	0.05	4.95	2.65	2.75		$[-0.51 ; -0.13]$
2	1.10	3.90	2.65	2.75		$(-\infty ; -0.51]$
3	**2.40**	**2.60**	**2.65**	**2.75**		non-monotone
4	2.40	2.60	2.00	3.40	$[2.27 ; \infty)$	
5	2.40	2.60	1.90	3.50	[1.70 ; 2.27]	
6	2.40	2.60	1.75	3.65	[1.18 ; 1.70]	
7	2.40	2.60	1.60	3.80	[0.86 ; 1.18]	
8	2.40	2.60	1.45	3.95	[0.65 ; 0.86]	
9	2.40	2.60	1.05	4.35	[0.36 ; 0.65]	
10	2.40	2.60	0.20	5.20	[0.13 ; 0.36]	

Again, subjects choose in each row between option A and option B. As in the HL table, option A as well as option B has two possible outcomes. However, instead of varying the probabilities and keeping the outcomes constant over all rows as in the HL table, we rather vary the outcomes and keep the probabilities constant (i.e. all probabilities are equal, namely 50%). First note that an individual with monotone preferences will always prefer option B over option A in row 3 of Table 1.2 (this is similar to row 10 in Table 1.1) as here option B first-order stochastically dominates (or more specifically, state-wise dominates) option A.

We now compare options in row 4 to those in row 3. While option A is identical to the one in row 3, option B in row 4 is a mean preserving spread of the one in row 3. We can therefore say that option B becomes more risky in the sense of the very general increasing risk definition of Rothschild and Stiglitz (1970), while option A stays the same. In row 5 option A is again unaltered whereas option B is a further mean preserving spread of the one in row 4 and thus a further increase in risk. This continues until row 10. By just imposing 'duality' stating that less risk-averse individuals should take riskier gambles, we can say that someone (call her j) who preferred option B in the first four rows and option A in the last six rows is more risk-averse than someone (call her i) who preferred option B in the first five rows and option A in the last five rows. Such a statement can be made without referring to any particular utility framework.

The property of 'duality' is based on the concept of mean utility preserving spreads first proposed by Diamond and Stiglitz (1974).[9] To see how, note that there exists a hypothetical gamble that is a mean preserving spread of option B in row 4, a mean preserving contraction of B in row 5, and leaves j just indifferent between A and B. This hypothetical gamble then is a mean utility preserving spread of A for j. At this point i still prefers B, so her hypothetical gamble representing a mean preserving spread of B in row 5, a mean preserving contraction of B in row 6, and leaving i just indifferent between A and B is a mean preserving spread of j's hypothetical gamble. It follows that j is more risk-averse than i. To illustrate how our table relates to the one of Holt and Laury (2002), we state in the last two columns of Table 1.2 how our method would elicit measures of RRA if we would also assume CRRA.

Risk seeking is identified through switches of choices in the first two rows of Table 1.2. Consider again the options in row 3, but now compare them to those in row 2. Now the 'less attractive' option A is altered by a mean preserving spread when going from row 3 to row 2, while option B stays the same. Only a very risk-seeking individual would like this spread so much that she would now prefer option A in row 2. In row 1 option A is a further mean preserving spread. Now, also less extreme risk seekers, who in row 2 were still choosing option B, are lured by the further increase in risk towards choosing option A in row 1. An individual who is risk-neutral, or is very close to being risk-neutral, will always choose option B in Table 1.2 since its expected value is higher than the one of option A in all rows.

Both options in Table 1.2, option A and option B, are always risky. This avoids the 'certainty effect', a well-known problem of any elicitation method using certainty equivalents. As the HL method, our method therefore "has the virtue of comparing gamble-versus-gamble choices (rather than gamble-versus-sure amount) to control for the possibility that differences in gamble complexity are themselves part of preference (e.g. Huck and Weizsäcker, 1999; Sonsino et al., 2002) or that fundamentally different outcome values are associated with risky and sure outcomes (Keller, 1985; Andreoni and Sprenger, 2009)." (Wang, Filiba, and Camerer, 2010, p3)

[9]The concept of mean utility preserving spreads can be viewed as a generalization of the concept of acceptance sets by Yaari (1969). The acceptance set of an outcome captures the set of lotteries for which an individual prefers to gamble rather than to obtain the outcome for sure. An individual is then more risk-averse than another one if her acceptance set is contained in the one of the other individual. Thus, Yaari (1969) defines an individual as being less risk-averse than another individual if she is ready to accept more risky gambles, starting from an initial situation that is risk-less. Diamond and Stiglitz (1974) allow the initial situation to be risky and define someone as being less risk-averse if she accepts more increases in risk (i.e. mean preserving spreads) in exchange for a fixed compensation. Diamond and Stiglitz (1974) study this concept solely within EUT. By referring to simple compensated spreads, Machina (1982, 1987), Chew, Karni, and Safra (1987), or Röell (1987) used it to analyze risk aversion in non-EUT models.

The concept of riskiness we use in our table follows the established theoretical literature. "Clearly riskiness is related to dispersion, so a good riskiness measure should be monotonic with respect to second-order stochastic dominance. Less well understood, perhaps, is that riskiness should also relate to location and thus be monotonic with respect to first-order stochastic dominance, in particular, that a gamble that is sure to yield more than another should be considered less risky. Both stochastic dominance criteria are uncontroversial [. . .]." (Aumann and Serrano, 2008, p811) In Table 1.2 we use both criteria. Option B first-order stochastically dominates option A in row 3 and can therefore be considered less risky. Going downward from row 3 option A stays unaltered whereas option B gets worse in terms of second-order stochastic dominance. Going upward from row 3 option A gets worse in terms of second-order stochastic dominance whereas option B stays the same. Individuals who switch from option B to A after row 3 are risk-averse (the earlier the more risk-averse). And individuals who switch from option A to option B before row 3 are risk-seeking (the later the more risk-seeking).

Note that mean preserving contractions of option A (option B) could be applied in addition to the mean preserving spreads of option B (option A) when going downward (upward) from row 3. This variant of the design could be useful if one wanted to induce similar changes in both options over all rows of our table. However, since it may come at the cost of increased complexity for the subjects, we chose not to do so.

Although the two concepts we use for our elicitation method, mean preserving spreads and mean utility preserving spreads, were originally analyzed within EUT, it has become common in other models as well to understand risk aversion in terms of this more behavioral definition, namely as an aversion to mean preserving spreads. Subsequently, it is this definition that is used when risk aversion is analyzed in non-EUT models (see e.g. Machina, 1982, 1987; Chew, Karni and Safra, 1987; Röell, 1987; Yaari, 1987; or Schmidt and Zank, 2008). And as Machina (2008, p80) notes, most non-EUT models "are capable of exhibiting first-order stochastic dominance preference, [and] risk aversion [. . .]." This shows that our method can classify subjects as more or less risk-averse across various models of risky choice.

Using variations of outcomes (i.e. mean preserving spreads) not only makes it easy for subjects to compute expected values but also allows us quite some flexibility in designing the range of the intervals to elicit estimates of relative risk aversion if we adopt the CRRA framework of Holt and Laury (2002). In principle, this could also be achieved in the HL table, but only at the price of stating odd probabilities. By contrast, in our table probabilities stay always at 50% and only outcomes vary. We believe that subjects are more experienced in dealing with odd outcomes (such as price tags) than with odd probabilities.[10] More importantly, constant probabilities of 1/2 do not require subjects to have experience in dealing with chances other than those for instance imposed by a coin toss. Also, such probabilities are much less sensitive to probability weighting. Already Quiggin (1982) uses 1/2 as plausible fixed point in his theory. "The claim that the probabilities of 50-50 bets will not be subjectively distorted seems reasonable, and [...] has proved a satisfactory basis for practical work [...]." (Quiggin, 1982, p328) This becomes, however, only important when decisions taken in the table are interpreted solely in terms of the curvature of a utility function.[11]

1.4 The Experiment

The experiment was computer-based and was conducted in November 2009 at the experimental laboratory MELESSA of the University of Munich. It used the experimental software z-Tree (Fischbacher, 2007) and the organizational software Orsee (Greiner, 2004). 232 subjects (graduate students were excluded) participated in 10 sessions and earned 11 euros (including 4 euros show-up fee) on average (with a maximum (minimum) of 30 (4.10) euros) for a duration of approximately one hour.

In the beginning of the experiment subjects received written instructions that were read privately by them. At the end of these instructions they had to answer test questions that showed whether everything was understood. There was no time limit for the instructions and subjects had the opportunity to ask questions in private. The experiment started on the computer screen only after everybody answered the test questions correctly and subjects had no further questions.

The remaining procedure of the experiment was the following. Each subject made decisions in four tables.[12] Again, they could take as much time as they wanted to make their decisions. After all subjects had made their decisions, an experimental instructor

[10]In Appendix A we provide a more general treatment of our method. This shows how our method can be easily modified to meet different requirements on the elicitation ranges.

[11]Note that besides the problem of not being incentive compatible, the trade-off method of Wakker and Deneffe (1996) is also well suited for this purpose. If, in contrast, the primary interest is in estimating a probability weighting function, our specific design will be of limited use. Only single-parameter weighting functions, as proposed by Tversky and Kahneman (1992), can be estimated with our design. However, the concept of mean preserving spreads could also be applied to construct a similar multiple price-list format with more complex probabilities than 50/50 bets. Such a modified table could then also be used to estimate probability weighting functions with more than one parameter.

[12]Eight treatments varied which tables in which order a subject received. The treatments are further explained below.

came to each subject to let them randomly determine their payoff from the tables.[13] Before they saw what their payoff from the experiment was they could again see how they actually decided in the randomly determined relevant table. At the end of the experiment all subjects further answered a questionnaire about their socio-economic characteristics. As soon as everybody had answered the questionnaire they were paid in private (not by the experimenter) and could leave.

Table 1.3: Our Adjusted Elicitation Method

	Option A		Option B		RRA if row was	RRA if row was
Row No.	Prob. 1/2 Outcome A1	Prob. 1/2 Outcome A2	Prob. 1/2 Outcome B1	Prob. 1/2 Outcome B2	first choice of A and above all B	last choice of A and below all B
1	0.03	4.89	2.62	2.72		[−0.49 ; −0.14]
2	1.01	3.91	2.62	2.72		[−0.96 ; −0.49]
3	1.40	3.52	2.62	2.72		[−1.70 ; −0.96]
4	1.65	3.27	2.62	2.72		(−∞ ; −1.70]
5	**2.36**	**2.56**	**2.62**	**2.72**		non-monotone
6	2.36	2.56	1.77	3.57	[1.37 ; ∞)	
7	2.36	2.56	1.61	3.73	[0.97 ; 1.37]	
8	2.36	2.56	1.42	3.92	[0.68 ; 0.97]	
9	2.36	2.56	1.09	4.25	[0.41 ; 0.68]	
10	2.36	2.56	0.26	5.08	[0.15 ; 0.41]	

As mentioned above, each subject made decisions in four tables. Two of the tables were low-stakes tables and two of them were high-stakes tables, where all low-stake outcomes were multiplied by five. In total, we used eight different tables. One of them was the original HL table (HLol) as outlined in Section 1.2 (Table 1.1) and another was the original HL table but with all outcomes multiplied by five (HLoh). In order to be able to directly compare the HL method and our method, we adjusted our tables to the exact same ranges of RRA that were used by Holt and Laury (2002). A third table therefore used our method but adjusted to the original low-stake outcomes of the HL table (MRal) as outlined in Table 1.3. And a fourth table used our method adjusted to the high-stake version of the original HL table (MRah), where all outcomes in Table 1.3 are multiplied by five.

Subjects further received our table (Table 1.2) from Section 1.3 (MRol). In designing this table we employed criteria mentioned by Holt and Laury (2002). There is an approximately symmetric range of RRA around 0, 0.5, 1, and 2. Based on the experimental results of Holt and Laury (2002), our table has only two risk seeking ranges and therefore more ranges for reasonable degrees of risk aversion. There was also a high-stakes version of this table where all outcomes are multiplied by five (MRoh). Again,

[13]Each subject had to role four dice. First, a four-sided die determined which of the four tables was payoff-relevant. Second, a ten-sided die determined which row in the payoff-relevant table was selected. And lastly, two ten-sided dice determined whether the amount A1 or A2 (if A was chosen in the relevant table and row) or whether the amount B1 or B2 (if B was chosen in the relevant table and row) was paid out to them (in addition to the show-up fee of 4 euros).

in order to directly compare both methods, we also adjusted the HL tables to the exact same ranges of RRA that were used in our tables. Table 1.4 shows the adjusted HL table for low stakes (HLal). Again, the high-stakes version of Table 1.4 (HLah) multiplied all outcomes by five.

Table 1.4: The Adjusted Holt and Laury Method

	Option A		Option B		RRA if row was
Row No.	Outcome A1 = \$2.00	Outcome A2 = \$1.60	Outcome B1 = \$3.85	Outcome B2 = \$0.10	last choice of A and below all B
1	Prob. 29/100	Prob. 71/100	Prob. 29/100	Prob. 71/100	[−0.53 ; −0.14]
2	Prob. 40/100	Prob. 60/100	Prob. 40/100	Prob. 60/100	[−0.14 ; 0.12]
3	Prob. 49/100	Prob. 51/100	Prob. 49/100	Prob. 51/100	[0.12 ; 0.36]
4	Prob. 58/100	Prob. 42/100	Prob. 58/100	Prob. 42/100	[0.36 ; 0.65]
5	Prob. 69/100	Prob. 31/100	Prob. 69/100	Prob. 31/100	[0.65 ; 0.85]
6	Prob. 76/100	Prob. 24/100	Prob. 76/100	Prob. 24/100	[0.85 ; 1.19]
7	Prob. 86/100	Prob. 14/100	Prob. 86/100	Prob. 14/100	[1.19 ; 1.70]
8	Prob. 95/100	Prob. 5/100	Prob. 95/100	Prob. 5/100	[1.70 ; 2.37]
9	Prob. 99/100	Prob. 1/100	Prob. 99/100	Prob. 1/100	[2.37 ; ∞)
10	Prob. 100/100	Prob. 0/100	Prob. 100/100	Prob. 0/100	non-monotone

Each of the eight different tables was received by 116 subjects and all 232 subjects were in either of eight different treatments. The treatments were designed to control for order effects, not only whether subjects answered low- or high-stakes tables first, but also whether HL tables or our tables (adjusted and original) were answered first. The eight treatments ensured that every subject had the same ex-ante expected income.[14]

In the comparison of the HL method and our method we will ask several questions. Firstly, whether both methods yield the same classification of individuals concerning both the direction and the intensity of risk attitude. Secondly, whether both methods yield similar levels of risk aversion intensity. Thirdly, what the effect is of increasing the stakes (i.e. multiplying all outcomes by five) on these RRA estimates. And lastly, how robust our results are.

1.5 Results

1.5.1 Directions of Risk Attitudes

Before analyzing the intensities of risk attitudes, we can ask how many of the subjects that can be classified are risk-averse, risk-neutral, and risk-seeking. Under low stakes, we find that 79% are risk-averse, 11% are risk-neutral, and 10% are risk-seeking

[14]The eight treatments were: 1. HLol, MRal, HLoh, MRah; 2. MRol, HLal, MRoh, HLah; 3. MRal, HLol, MRah, HLoh; 4. HLal, MRol, HLah, MRoh; 5. HLoh, MRah, HLol, MRal; 6. MRoh, HLah, MRol, HLal; 7. MRah, HLoh, MRal, HLol; 8. HLah, MRoh, HLal, MRol. We chose to let each subject answer only four and not all eight different tables, because eight tables would have required too many decisions for them.

in the HL tables. The respective numbers for our tables are 81%, 9%, and 10%. Under high stakes, 88% are risk-averse, 7% are risk-neutral, and 5% are risk-seeking in the HL tables whereas the respective numbers are 88%, 6%, and 6% in our tables. This suggests that both methods yield identical classifications of subjects concerning the *direction* of risk attitude.

This finding is confirmed when we look at within subject classifications. Among the subjects that can be classified, we find that 84% have the same direction of risk attitude across both methods under low stakes, where 76% are risk-averse, 4% are risk-neutral, and 4% are risk-seeking in both methods. Under high stakes, 90% have the same direction of risk attitude in both methods with 84% being risk-averse, 4% risk-neutral, and 2% risk-seeking.

For each relevant comparison of both methods (i.e. HLol vs. MRal, HLoh vs. MRah, HLal vs. MRol, and HLah vs. MRoh) we perform sign tests in order to see whether the HL tables yield a different classification on the direction of risk attitude than our tables. We find that none of the comparisons is significant.[15] We can now state our first result.

Result 1.1 *The HL method and our method do not yield a different classification of individuals concerning the direction of risk attitude (risk-averse, risk-neutral, or risk-seeking).*

1.5.2 Intensities of Risk Attitudes

While Result 1.1 shows that both methods yield the same classification concerning the *direction* of risk attitude, another question is whether both methods also yield the same classification concerning the *intensity* of risk attitude. Are individuals similarly classified as more or less risk-averse across both methods? This is an important question as it answers whether the methodological problem of using the HL method (due to the required assumption of EUT) is indeed an empirically relevant one. If both methods yielded the same classification of subjects concerning the intensity of risk attitude, results of experimental studies that used the HL method (and thereby assumed EUT) in order to classify subjects as more or less risk-averse would not be flawed. If, however, subjects are differently classified as more or less risk-averse in both methods, the methodological problem of assuming EUT for the classification in the HL method is in fact empirically relevant.

[15]The results (p-values, number of observations N) of two-sided Sign tests are: HLol vs. MRal ($p = 1.0000$, $N = 76$), HLoh vs. MRah ($p = 1.0000$, $N = 71$), HLal vs. MRol ($p = 0.1094$, $N = 66$), and HLah vs. MRoh ($p = 1.0000$, $N = 71$).

In order to answer this question we take the following approach. We first take a subject in an HL table and determine whether she is classified as more, less, or equally risk-averse than any other subject. We then take the same subject in our table and determine as well whether she is classified as more, less, or equally risk-averse than each of the other subjects. When we now compare this subjects *pair-wise* comparisons (to all other subjects) in both tables, we can identify for this subject whether she is classified the same or differently (against each of the other subjects) across both tables. Since we perform this task not only for one specific but for all subjects, we can identify against how many of the other subjects a subject changes her pair-wise comparison across methods on average.

This method is exemplified in Table 1.5. In Table 1.5A we consider the HL table with the original RRA ranges and high stakes (HLoh). If a cell displays the sign '$>$', it means that the row subject is measured to be more risk-averse than the column subject. Similarly, an entry of the sign '$<$' denotes that the column subject was found to be more risk-averse than the row subject and '=' indicates an identical amount of risk aversion of both subjects. An entry of '$-$' means that at least one of the two subjects made inconsistent choices in this elicitation table. In Table 1.5B the same procedure is applied to the elicitation table of our method that is the relevant comparison to the table of Table 1.5A. MRah is in this case the relevant comparison, since it exhibits identical RRA ranges and stakes as HLah. Finally, in Table 1.5C it is displayed whether the relative comparisons of the two subjects of a cell are identical ('=') or different ('$\neq$') for the two measurement devices of Table 1.5A and Table 1.5B. An entry of '$-$' means that in at least one of the two elicitation tables at least one subject made inconsistent choices. Such an analysis of changes in pair-wise comparisons cannot only be made for HLoh vs. MRah, but also for HLol vs. MRal, HLal vs. MRol, and HLah vs. MRoh.

Table 1.5: Relative Comparison of Risk Aversion

Panel A:
HLoh

Subject	1	2	3	⋯	116
1					
2	-				
3	>	-			
⋮	⋮	⋮	⋮	⋱	
116	>	-	=	⋯	

Panel B:
MRah

Subject	1	2	3	⋯	116
1					
2	-				
3	>	-			
⋮	⋮	⋮	⋮	⋱	
116	<	-	<	⋯	

Panel C:
HLoh vs. MRah

Subject	1	2	3	⋯	116
1					
2	-				
3	=	-			
⋮	⋮	⋮	⋮	⋱	
116	≠	-	≠	⋯	

When we compare the HL tables to our tables we find that on average subjects change their pair-wise comparison to 49% of the other subjects.[16] So, the average subject has a different standing towards almost half of the other subjects across methods.[17]

Result 1.2 *The HL method and our method yield a different classification of individuals concerning the intensity of risk aversion relative to other individuals.*

Results 1.1 and 1.2 are of great interest for other experimental studies that use the HL method to control for risk aversion in observed behavior but which are not specifically interested in the absolute level of risk aversion intensity. However, there are other studies where the absolute level of risk aversion is important in order to derive quantitative predictions of a theoretical model. In the following we therefore investigate what the levels of risk aversion intensity are in both methods and how robust these RRA estimates are towards increasing the stakes.

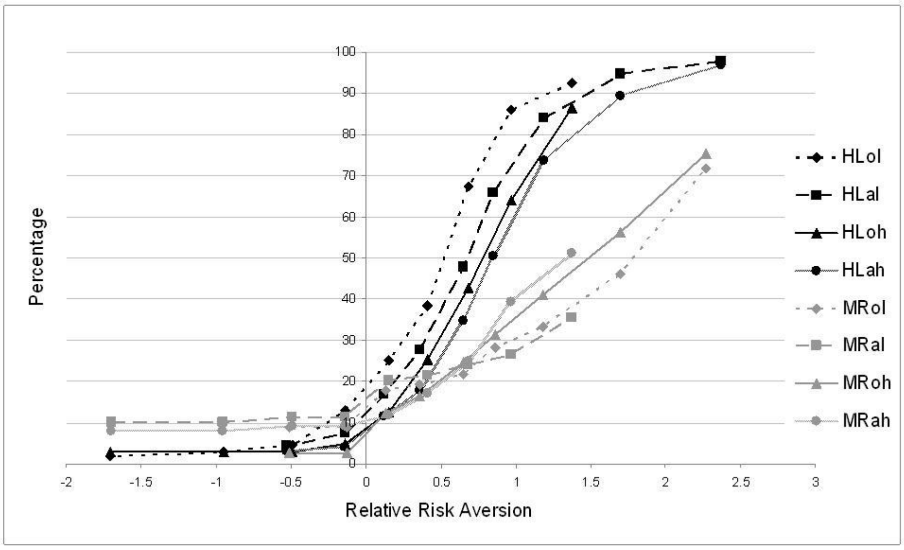
Concerning the level of the intensity of risk attitude we again find systematic differences between both methods. Figure 1.1 shows the cumulative distributions of RRA for all eight different elicitation tables (using uniform distributions within the RRA ranges). The cumulative distributions of relative risk aversion of all four HL tables lie above those of our four tables. While almost none of the subjects lies in the highest RRA range in the HL tables, many subjects fall into the highest RRA range when our method is used.[18] The medians of RRA using the HL method are all below the medians when our method is used. In the HL tables the medians lie in RRA ranges below one (HLol: $[0.41; 0.68]$; HLal: $[0.65; 0.85]$; HLoh: $[0.68; 0.97]$; HLah: $[0.65; 0.85]$) but they lie

[16]The comparisons we make yield the following results. The average subject changes her relative standing (of being more, less, or equally risk-averse) to 54%, 52%, 47%, and 42% of the other subjects in the respective comparisons of HLol vs. MRal, HLoh vs. MRah, HLal vs. MRol, and HLah vs. MRoh. Overall, we employed 9965 pairs of pair-wise comparisons in which both subjects were consistent in both elicitation tables.

[17]Note that subjects also change their standing to other subjects across stakes (but within methods). However, we can test whether subjects change their standing across methods more than across stakes. Holding the stakes effect constant, we find that subjects increase their standing to other subjects on average by 36% when the method in addition to the stakes changes. When performing two-sided Wilcoxon signed-rank tests we find that subjects change their ranking significantly more across methods than across stakes (HLol vs. HLoh against HLol vs. MRah ($z = -5.262$, $p = 0.0000$, $N = 68$), HLal vs. HLah against HLal vs. MRoh ($z = -6.366$, $p = 0.0000$, $N = 67$), MRol vs. MRoh against MRol vs. HLah ($z = -2.431$, $p = 0.0151$, $N = 63$), and MRal vs. MRah against MRal vs. HLoh ($z = -3.169$, $p = 0.0015$, $N = 57$)). Results do not change when using two-sided Mann Whitney U tests instead (HLol vs. HLoh against HLol vs. MRah ($p = 0.0000$, $N = 170$), HLal vs. HLah against HLal vs. MRoh ($p = 0.0000$, $N = 157$), MRol vs. MRoh against MRol vs. HLah ($p = 0.0273$, $N = 135$), and MRal vs. MRah against MRal vs. HLoh ($p = 0.0000$, $N = 133$)).

[18]Note that as the highest range goes to infinity, the cumulative distribution functions do not end at 100 for the displayed values of RRA. Similarly, as the lowest range goes to minus infinity, the cumulative distribution functions do not start at 0 for the reported RRA values. This is also the reason why we cannot investigate the means of RRA but only the medians.

Figure 1.1: Cumulative Distributions of RRA for All Tables



in RRA ranges above one with our tables (MRol: $[1.70; 2.27]$; MRal: $[1.37; \infty)$; MRoh: $[1.18; 1.70]$; and MRah: $[0.97; 1.37]$). While Figure 1.1 shows the overall picture, Figures 1.2 and 1.3 show the specific distributions that need to be compared in the analysis of our data.[19]

Figure 1.2 compares the HL method and our method. For low and high stakes we can compare the original to the adjusted tables since the adjustment was such that the ranges of RRA were identical in both tables. In all four panels the HL method yields clearly lower measures of RRA. This is the case no matter whether the adjustment took place for the HL method (Figure 1.2C and 1.2D) or for our method (Figures 1.2A and 1.2B) or whether we look at low (Figures 1.2A and 1.2C) or high stakes (Figures 1.2B and 1.2D).

Figure 1.3 shows the effect of increasing the stakes in both methods. Since the cumulative distributions of the high-stakes HL tables lie below those of the low-stakes HL tables (Figures 1.3A and 1.3C), increasing the stakes seems to increase relative risk aversion. This is in contrast to our method where increasing the stakes does not cause risk aversion to increase. The cumulative distributions of our high-stakes tables rather cross those of our low-stakes tables (Figures 1.3B and 1.3D). This picture seems not to be affected by the fact whether we compare original (Figures 1.3A and 1.3B) or adjusted tables (Figures 1.3C and 1.3D).

[19]In Figures 1.2 and 1.3 we omitted naming the axis, but since all distribution lines are taken from Figure 1.1 and just considered in isolation, the notation of Figures 1.2 and 1.3 is of course the same as in Figure 1.1.

Figure 1.2: Cumulative Distributions of RRA: Comparing Methods

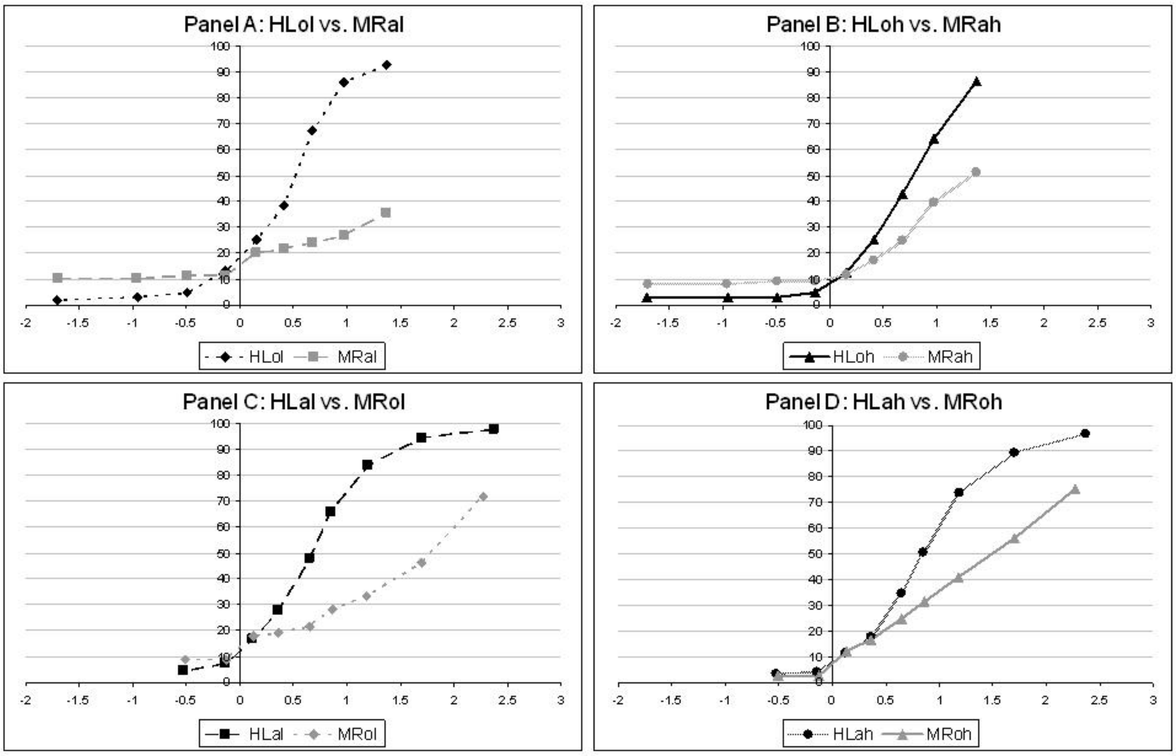
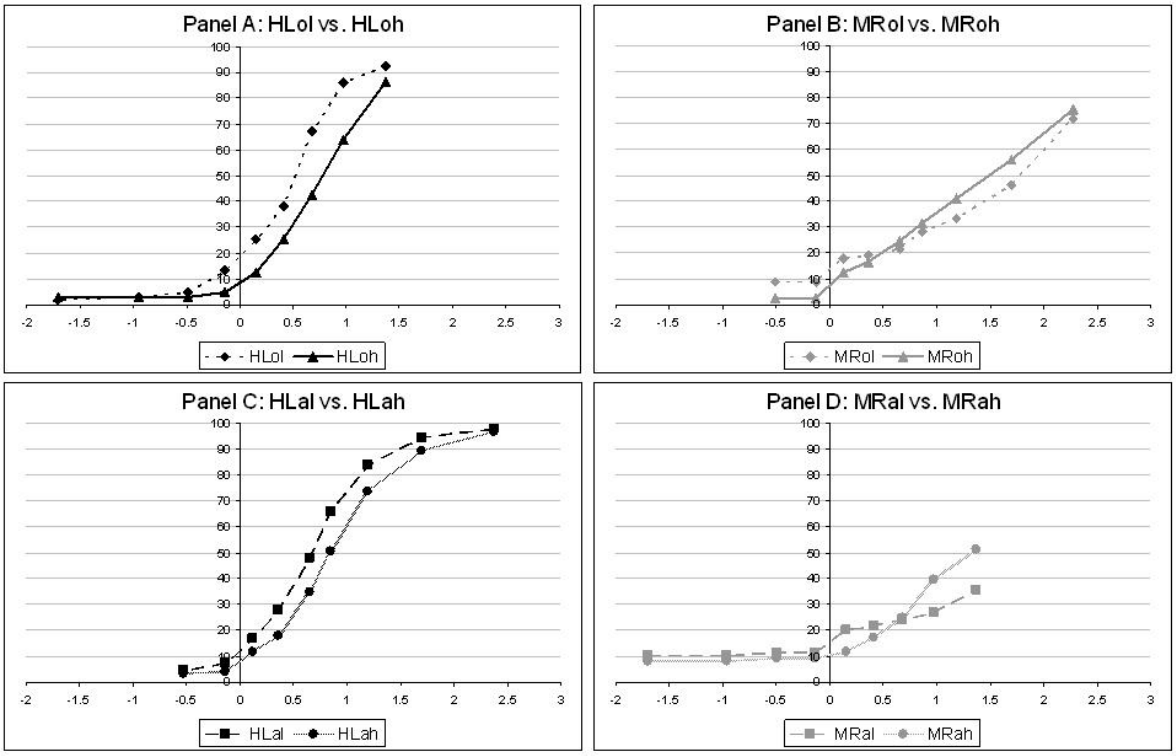


Figure 1.3: Cumulative Distributions of RRA: Comparing Stakes



Since each subject made decisions in four tables, we also test differences between methods and stakes using matched pairs. Table 1.6 relates the original HL tables to our adjusted tables, such that RRA ranges are identical and can be directly compared. And Table 1.7 relates the adjusted HL tables to our original tables such that all comparisons have identical RRA intervals. Reported are the number of observations (N), the z-values, and the p-values of two-sided Wilcoxon signed-rank tests.[20] In both tables, Table 1.6 and Table 1.7, in each cell it is tested if the elicitation method to the left yields different measures of RRA than the elicitation method above. Consider an example from Table 1.6. If HLol is compared to MRal, then a z-value of $z = -4.922$ indicates that the left table (HLol) yields lower measures of RRA than the table above (MRal).

Table 1.6: Wilcoxon Signed-Rank Tests (HLo and MRa)

	(N) / **z-value** / **p-value**		**MRah**
		MRal	(61) 0.488 0.6257
	HLoh	(72) -4.000 0.0001	(71) -2.580 0.0099
HLol	(99) -5.081 0.0000	(76) -4.922 0.0000	(71) -4.815 0.0000

Comparing first the HL method and our method, we observe significantly higher measures of RRA when our method is used (HLol vs. MRal and HLoh vs. MRah in Table 1.6; and HLal vs. MRol and HLah vs. MRoh in Table 1.7). This holds at the 1%-level for all four comparisons of Figure 1.2. We can therefore state the following result.

Result 1.3 *Our method yields a higher intensity of risk aversion than the HL method.*

Looking at the effect of increasing the stakes, we see that there is a significantly positive effect on the RRA measure in the HL tables (HLol vs. HLoh in Table 1.6; and HLal vs. HLah in Table 1.7). In contrast, there is no such effect observed when our

[20] Our results do not change when using two-sided Kolmogorov-Smirnov tests instead. The p-values are all $p = 0.000$ for the comparisons of Figure 1.2 (with $N = 186, 179, 172, 168$ respectively for Figures 1.2A, 1.2B, 1.2C, 1.2D). Thus, compared to the HL method our method yields significantly higher measures of RRA. For the comparisons of Figure 1.3, we get $p = 0.798$ ($N = 151$) and $p = 0.227$ ($N = 155$) for Figures 1.3B and 1.3D, respectively. And we get $p = 0.002$ ($N = 210$) and $p = 0.167$ ($N = 189$) for the respective Figures 1.3A and 1.3C. Hence, while measures of RRA are not significantly different between low and high stakes when our method is used, this is not the case with the HL method. Here, increasing the stakes significantly increases measures of RRA in the original (two-sided at 1%-level) as well as in the adjusted (one-sided at 10%-level) HL tables.

Table 1.7: Wilcoxon Signed-Rank Tests (HLa and MRo)

(N) **z-value** **p-value**	**HLah**	**MRol**	**MRoh**
MRol			(65) -0.270 0.7875
HLah		(70) -3.215 0.0013	(71) -4.293 0.0000
HLal	(90) -4.957 0.0000	(66) -3.756 0.0002	(67) -4.654 0.0000

method is used (MRal vs. MRah in Table 1.6; and MRol vs. MRoh in Table 1.7). So, for the comparisons of Figure 1.3, we see that increasing the stakes by a factor of five increases risk aversion significantly at the 1%-level only when the HL method is used. With our method, there is no significant effect of increasing stakes.[21]

Result 1.4 *While increasing the stakes by a factor of five increases the intensity of risk aversion with the HL method, increasing the stakes has no effect on the intensity of risk aversion with our method.*

Results 1.3 and 1.4 show that our method not only yields higher risk aversion estimates than the HL method, but also that our estimates are robust towards multiplying all outcomes by five (thereby indicating CRRA). This is not the case for the HL estimates. Here, we find increasing relative risk aversion (IRRA). Our findings for the HL method are completely in line with the findings of Holt and Laury (2002). Nevertheless, the results for our method are much closer to what is observed in the empirical literature. Several empirical studies indicate a measure of RRA roughly between one and two (e.g. Tobin and Dolde, 1971; Friend and Blume, 1975; Kydland and Prescott, 1982; Hildreth and Knowles, 1982; Szpiro, 1986; Chetty, 2006; or Bombardini and Trebbi, 2010) and Mehra (2003, p59) notes that "most studies indicate a value for α that is close to 2." (Here, α is the measure of RRA) An experimental study by Levy (1994) rejects the existence of IRRA. Other empirical studies (e.g. by Szpiro, 1986; Friend and Blume, 1975; Brunnermeier and Nagel, 2008; or Calvet and Sodini, 2010) find supportive evidence for CRRA. Fehr-Duda et al. (2010) show in their experimental study that IRRA is entirely driven by transformations of the probability weighting function as stakes increase. In contrast to the HL method, our method is invariant to probability

[21] As can be seen from Tables 1.6 and 1.7, even the signs of the z-values are positive (adjusted tables) *and* negative (original tables).

weighting. Hence, this may be the reason why there is a stakes effect only with the HL method.

Another possible explanation for the results of Holt and Laury (2002) and ours is that subjects need higher incentives when they have to exert more cognitive effort. In the HL tables, subjects need to exert a higher amount of cognitive effort than in our tables since the varying probabilities are difficult to handle. By contrast, in our tables all probabilities are one half and 50/50 odds seem easy to work with. When subjects have too little incentives to exert the cognitive effort that is required to reveal their true level of risk aversion, it seems reasonable that they anchor their decision on the 50/50 choice (i.e. row 5 in the original HL tables). Note that this is also the modal switching point under low stakes in Holt and Laury (2002). Our results and the results of Holt and Laury (2002) suggest that increasing the stakes increases risk aversion with the HL method. So, as incentives increase subjects are more willing to exert such effort and thereby show their true level of risk aversion.[22] This might be the reason that Holt and Laury (2002) do not observe such a stake effect with hypothetical payoffs. With our method, subjects do not need much cognitive effort to handle 50/50 chances. Already our low-stakes tables give sufficient incentives to show subjects' true level of risk aversion. Increasing the stakes has therefore no effect on risk aversion with our method. It is interesting to note that increasing the stakes by a factor of five seemed not to be sufficient to show subjects' true level of risk aversion in the HL tables since risk aversion was still lower than in our tables. This is, however, consistent with the findings of Holt and Laury (2002) who increased stakes by factors of 20, 50, and 90. They showed that risk aversion increased with these higher stakes and that the RRA of one third of subjects was in the highest RRA range when stakes are increased by a factor of 90. Harrison et al. (2005) use a design similar to Holt and Laury (2002) but control for order effects. They find that stakes effects are only significant between a factor of 20 and 50 (or 90), but not between 50 and 90. Of course, if incentives increase even further beyond a factor of 50, only few subjects need such high incentives to exert their required cognitive effort and we should expect that the increase in risk aversion is not significant anymore. Although this cognitive effort explanation may drive the results of our experiment, we did not design the experiment to test this explanation. Further research should therefore investigate this potential explanation more closely in order to answer the question what the true level of risk aversion is.

[22]Note that in Holt and Laury (2002) there are also less inconsistent subjects under high stakes and more inconsistencies under comparable hypothetical payoffs.

1.5.3 Robustness

One may be concerned that the differences in the RRA estimates between the HL method and our method may be due to a preference for switching in the same row across tables. For instance, if a subject always switches after row 5 (from A to B in the HL tables and from B to A in our tables) we would measure a higher RRA in our tables. When comparing the HL tables to our tables, there are always the same number of rows where identical switching behavior would induce a higher RRA with our method as there are where identical switching leads to a higher RRA with the HL method.[23] Nevertheless, if a majority of subjects switched in exactly those rows where we would measure a higher RRA for our tables, Result 1.3 may simply be explained by a preference for identical switching behavior.

For all four comparisons of Figure 1.2, we find that only a minority of subjects switches in the same row in both tables. The respective numbers are 19.7%, 9.9%, 10.6%, and 12.7% for Figures 1.2A, 1.2B, 1.2C, and 1.2D.[24] Nevertheless, of those subjects that switch in the same row a majority switches in those rows that induce a higher RRA with our method. For the respective Figures 1.2A, 1.2B, 1.2C, and 1.2D 80%, 85.7%, 71.4%, and 55.6% of the 'same row switchers' switch in rows where our tables yield a higher RRA, and only 13.3%, 0%, 28.6%, and 11.1% switch in rows where the HL tables yield a higher RRA. It may therefore be possible that a preference for identical switching behavior drives Result 1.3.

However, we can test whether Result 1.3 still holds for those subjects that switch in different rows. If we exclude the 'same row switchers', we can test whether Result 1.3 is driven by a preference for switching in the same row across both types of tables. For the comparisons of Figure 1.2, we again perform two-sided Wilcoxon signed-rank tests where we exclude 'same row switchers'.[25] For all comparisons we observe that differences are still significant. For the comparison of HLol vs. MRal (Figure 1.2A) we

[23]When comparing original HL tables to our adjusted tables (HLol vs. MRal and HLoh vs. MRah), switching after row 1, 5, and 6 yields a higher RRA in our tables, switching after row 3, 4, 8, and 9 yields a higher RRA in the HL tables, and switching after row 2 and 7 yields the same RRA in both tables. Never switching (choosing always B in both tables) also leads to a higher RRA in our tables. When comparing our original to the adjusted HL tables (HLal vs. MRol and HLah vs. MRoh), switching after row 3, 4, and 5 yields a higher RRA in our tables, switching after row 2, 7, 8, and 9 yields a higher RRA in the HL tables, and switching after row 1 and 6 leads to the same RRA in both tables. Again, never switching and choosing always B yields a higher RRA in our tables.

[24]Note that all subjects answered not only two but four tables. And there are only two subjects that switch in the same row in all four tables.

[25]When performing two-sided Kolmogorov-Smirnov tests instead, we observe similar results. All comparisons of Figure 1.2 stay significant at the 1%-level ($p = 0.000$, $N = 122$ for Figure 1.2A; $p = 0.008$, $N = 128$ for Figure 1.2B; $p = 0.000$, $N = 118$ for Figure 1.2C; and $p = 0.000$, $N = 124$ for Figure 1.2D).

get $z = -4.540$, $p = 0.0000$, $N = 61$. For the comparison of HLoh vs. MRah (Figure 1.2B) we get $z = -1.787$, $p = 0.0740$, $N = 64$. The comparison of HLal vs. MRol (Figure 1.2C) yields $z = -3.908$, $p = 0.0001$, $N = 59$, and the comparison of HLah vs. MRoh (Figure 1.2D) yields $z = -3.982$, $p = 0.0001$, $N = 62$. Thus, the differences of Figures 1.2A, 1.2C, and 1.2D are still significant at the 1%-level, and the difference of Figure 1.2B is still significant at the 10%-level. We can now state the following result.

Result 1.5 *Result 1.3 is not due to a preference for switching in the same row across both methods.*

Since we had eight different treatments in our experiment, we can also test whether the order in which the tables were presented to the subjects had an effect on the elicited intensity of risk aversion. This could be the case if subjects are primed to think about all tables in terms of the first table they completed. To control for such order effects, we first test for each of the comparisons of Figure 1.2 whether it made a difference if a specific table was completed *before* or *after* its corresponding 'other-type' table. As an example, consider the HLol table. Here, we test whether the RRA distribution of HLol in treatment 1 and 5 (where it is completed before MRal) is different to the RRA distribution of Hlol in treatment 3 and 7 (where it is completed after MRal). For each of the eight different tables we perform a two-sided Kolmogorov-Smirnov test on the equality of RRA distributions and find no significant difference in either table.[26]

To control for order effects in the comparisons of Figure 1.3, we also test for each table whether it made a difference if it was completed *before* or *after* its corresponding 'other-stake' table. Consider again as an example the HLol table. Here, we test whether the RRA distribution of HLol in treatment 1 and 3 (where it is completed before HLoh) is different to the RRA distribution of HLol in treatment 5 and 7 (where it is completed after HLoh). Again, none of the two-sided Kolmogorov-Smirnov tests shows a significant difference.[27] We are now able to summarize these results on order effects. The order effects that are captured by Result 1.6 test whether high-stakes tables are completed before or after corresponding low-stakes tables (and vice versa) as well as whether original (HL and MR) tables are completed before or after corresponding adjusted (MR and HL) tables (and vice versa).

[26]The results of the two-sided Kolmogorov-Smirnov tests are: HLol ($p = 0.199$, $N = 107$), HLal ($p = 0.822$, $N = 94$), HLoh ($p = 1.000$, $N = 103$), HLah ($p = 0.949$, $N = 95$), MRol ($p = 0.424$, $N = 78$), MRal ($p = 0.896$, $N = 79$), MRoh ($p = 0.545$, $N = 73$), and MRah ($p = 0.763$, $N = 76$).

[27]The results of the two-sided Kolmogorov-Smirnov tests are: HLol ($p = 0.188$, $N = 107$), HLal ($p = 0.949$, $N = 94$), HLoh ($p = 0.990$, $N = 103$), HLah ($p = 0.940$, $N = 95$), MRol ($p = 0.796$, $N = 78$), MRal ($p = 0.979$, $N = 79$), MRoh ($p = 0.775$, $N = 73$), and MRah ($p = 0.957$, $N = 76$).

Result 1.6 *The order in which tables were completed by subjects does not influence the elicited intensities of risk aversion.*

The analysis of Sections 1.5.1 and 1.5.2 excluded subjects that switched multiple times from one option to another, or chose the first-order dominated option, and thus were inconsistent in completing a table. Over all eight different tables, we find that on average 24% of subjects are inconsistent which is in line with other studies.[28] However, Holt and Laury (2002) treated those inconsistent subjects differently than we did. They simply counted how often subjects chose each option (even if the option was not only chosen in subsequent rows) and assumed the inconsistency away. If we treat our inconsistent subjects in the same way, we can test whether the inconsistent subjects show a different pattern of risk aversion than the consistent subjects. When performing Kolmogorov-Smirnov tests, we do not find any significant difference in the distributions of consistent and inconsistent subjects.[29]

1.6 Conclusion

The multiple price-list method of HL has become *the* standard way to measure the intensity of individuals' risk attitudes in experiments. Several other methods which also involve choices between lotteries have been proposed to accomplish a similar task. None of these approaches uses mean preserving spreads *among* different choice situations in a way similar to our approach. Since it is exactly this feature that allows us to classify subjects as more or less risk-averse independently of a specific model, other elicitation methods need to impose the same utility structure on all individuals in order to classify

[28]Of those 24%, 18% switched multiple times and 6% chose the first-order dominated option. Bruner et al. (2008), for instance, find 30% inconsistent subjects where 25% are multiple switchers and 5% have non-increasing utility. Jacobson and Petrie (2009) or Prasad and Salmon (2010) even find that more than half of the subjects behave inconsistently. As noted by Andersen et al. (2006, p386) "it is quite possible that [multiple] switching behavior is the result of the subject being indifferent between the options. The implication here is that one simply use a "fatter" interval to represent this subject in the data analysis, defined by the first row that the subject switched at and the last row that the subject switched at." When allowing for indifference, Andersen et al. (2006) find that multiple switching is only 5.8% and 24.3% choose the indifference option. When imposing a single switching point, choosing indifference increases to 30%, which again indicates that multiple switching behavior is caused by 'thicker' indifference curves. In contrast, Bruner (2007) suggests an instructional variation emphasizing that only *one* decision will determine earnings (and further emphasizing incentive compatibility of the payment rule) and finds multiple switching reduced, from 25.8% to 6.7% and from 13.3% to 2.3% in two different elicitation formats (see also Andreoni and Sprenger, 2010). This would rather suggest that multiple switching is caused by errors.

[29]The results of the two-sided Kolmogorov-Smirnov tests are: HLol ($p = 0.916$, $N = 116$), HLal ($p = 0.390$, $N = 113$), HLoh ($p = 0.328$, $N = 116$), and HLah ($p = 0.180$, $N = 116$). Subjects were only excluded in case they never chose option B (3 subjects in HLal), since then even reshuffling their choices cannot make them consistent.

them. None of the methods is model-independent in that it can rank individuals as more or less risk-averse when they have heterogeneous risk preferences, such as EUT *and* non-EUT. This seems especially problematic in light of existing evidence.

After modifying the HL method and thereby proposing a new model-independent multiple price-list method to elicit the intensity of individuals' risk attitudes, we further compared our proposed method to the HL method in a laboratory experiment. Our results for the HL method replicate the findings of Holt and Laury (2002). Furthermore, concerning the direction of risk attitude (i.e. risk-averse, risk-neutral, or risk-seeking) we find that individuals are classified the same across both methods.

However, with our method we found systematic differences concerning the intensity of risk attitude. The classification of individuals as being more or less risk-averse than other individuals is quite different across methods. This is important for experimental studies that use the HL method (and thereby assume EUT for the classification) in order to control for risk aversion in observed behavior, because it shows that the methodological problem is in fact empirically relevant. Concerning the level of individuals' intensity of risk aversion we found that our method yields higher measures of risk aversion that are much closer to what is observed in the field. Furthermore, while increasing the stakes increases risk aversion with the HL method, our method is robust towards such stakes effects. We also offered a cognitive effort explanation of our results that needs to be tested in future research. This may help to answer the important question of what subjects' true level of risk aversion is after all.

1.7 Appendix A: The General Model-Independent Method

Define row i as consisting of two alternatives A_i and B_i. Each alternative consists of two possible outcomes, $A1_i$ and $A2_i$ of alternative A_i and $B1_i$ and $B2_i$ of alternative B_i. Given row i is the row played (if only one row of the table is randomly selected to be played), each outcome of each alternative is realized with probability 1/2. Choose a row $i = m$ and define $A1_m \equiv a, A2_m \equiv b, B1_m \equiv c, B2_m \equiv d$, where $a, b, c, d \in \mathbb{R}$. Without loss of generality, choose a and b such that $a < b$, and c and d such that $c < d$.[30] Our method of elicitation requires that $a \leq c$, $b \leq d$, and at least one relation holds strictly. It follows that

$$A_m \prec_{FSD} B_m. \tag{1.1}$$

(1.1) means that B_m first-order stochastically dominates (FSD) A_m.[31] Every individual with strictly increasing utility prefers B_m over A_m.

Then define values of row $m+1$ as follows: $A1_{m+1} \equiv a, A2_{m+1} \equiv b, B1_{m+1} \equiv c - k_1, B2_{m+1} \equiv d + k_1$. Further define values of row $m+2$ in the following way: $A1_{m+2} \equiv a, A2_{m+2} \equiv b, B1_{m+2} \equiv c - k_1 - k_2, B2_{m+2} \equiv d + k_1 + k_2$. Continue these mean preserving spreads (MPS's) of option B until the last row is reached. Generally, for $\check{n} > 0$ we can define $A1_{m+\check{n}} \equiv a, A2_{m+\check{n}} \equiv b, B1_{m+\check{n}} \equiv c - k_1 - k_2 - \ldots - k_{\check{n}}, B2_{m+\check{n}} \equiv d + k_1 + k_2 + \ldots + k_{\check{n}}$, where $k_1, k_2, \ldots \in \mathbb{R}^+$.

In a similar way, we can define values of row $m-1$ as follows: $A1_{m-1} \equiv a - k^1, A2_{m-1} \equiv b + k^1, B1_{m-1} \equiv c, B2_{m-1} \equiv d$. Now define values of row $m-2$ as follows: $A1_{m-2} \equiv a - k^1 - k^2, A2_{m-2} \equiv b + k^1 + k^2, B1_{m-2} \equiv c, B2_{m-2} \equiv d$. Again, these MPS's of option A can be continued until the first row is reached. Generally, for $\hat{n} > 0$ define $A1_{m-\hat{n}} \equiv a - k^1 - k^2 - \ldots - k^{\hat{n}}, A2_{m-\hat{n}} \equiv b + k^1 + k^2 + \ldots + k^{\hat{n}}, B1_{m-\hat{n}} \equiv c, B2_{m-\hat{n}} \equiv d$, where $k^1, k^2, \ldots \in \mathbb{R}^+$.

With these definitions it follows that for all i the expected values are $\mathbb{E}[A_i] = \frac{a+b}{2}$ and $\mathbb{E}[B_i] = \frac{c+d}{2}$. Table 1.8 illustrates our generalized elicitation method.

From $k_1, k_2, \ldots > 0$ and $k^1, k^2, \ldots > 0$ it follows that for all $i \geq m$ we can state $B1_i > B1_{i+1}$ and $B2_i < B2_{i+1}$. Also, for all $i \leq m$ it holds that $A1_i > A1_{i-1}$ and $A2_i < A2_{i-1}$.

Let us first consider all rows with $i \geq m$. Since all $A1_i$ and $A2_i$ are identical, it follows that

$$\begin{aligned} & A_m = A_{m+1} = \ldots = A_{m+\check{n}} \\ \Rightarrow \quad & A_m \sim A_{m+1} \sim \ldots \sim A_{m+\check{n}}. \end{aligned} \tag{1.2}$$

[30] We chose to make these inequalities strict in order to avoid 'certainty effect' issues.

[31] In order to make this more salient, we even used state-wise dominance as a special case of FSD in the experiment.

Table 1.8: Our Generalized Elicitation Method

Row No. i	Option A_i Prob. 1/2 Outcome $A1_i$	Option A_i Prob. 1/2 Outcome $A2_i$	Option B_i Prob. 1/2 Outcome $B1_i$	Option B_i Prob. 1/2 Outcome $B2_i$
...	...	...	...	...
$m-\hat{n}$	$a-k^1-k^2-\ldots-k^{\hat{n}}$	$b+k^1+k^2+\ldots+k^{\hat{n}}$	c	d
...	...	...	...	...
$m-2$	$a-k^1-k^2$	$b+k^1+k^2$	c	d
$m-1$	$a-k^1$	$b+k^1$	c	d
m	a	b	c	d
$m+1$	a	b	$c-k_1$	$d+k_1$
$m+2$	a	b	$c-k_1-k_2$	$d+k_1+k_2$
...	...	...	...	...
$m+\check{n}$	a	b	$c-k_1-k_2-\ldots-k_{\check{n}}$	$d+k_1+k_2+\ldots+k_{\check{n}}$
...	...	...	...	...

Since $B1_{m+\check{n}} = B1_{m+\check{n}-1} - k_{\check{n}}$ and $B2_{m+\check{n}} = B2_{m+\check{n}-1} + k_{\check{n}}$ it follows that $B_{m+\check{n}}$ is a MPS of $B_{m+\check{n}-1}$.[32] We use the standard (behavioral) definition of risk aversion and say that an individual is risk-averse if she dislikes increases in risk, i.e. if she dislikes MPS's. We thus employ the increasing risk definitions of Rothschild and Stiglitz (1970). For every risk-averse individual

$$\begin{aligned} & B_m \prec_{MPS} B_{m+1} \prec_{MPS} \ldots \prec_{MPS} B_{m+\check{n}} \\ \Rightarrow \quad & B_m \succ B_{m+1} \succ \ldots \succ B_{m+\check{n}}. \end{aligned} \tag{1.3}$$

Combining (1.1), (1.2), and (1.3) we can derive that the choice between $A_{m+\check{n}}$ and $B_{m+\check{n}}$ for every $\check{n} > 0$ is a trade-off between the advantage of the FSD-improvement from B_m over A_m ($= A_{m+\check{n}}$) and the disadvantage of the increase(s) in risk (MPS('s)) when going from B_m to $B_{m+\check{n}}$.[33]

Suppose an individual chooses $A_{m+\check{n}}$ over $B_{m+\check{n}}$, but chooses $B_{m+\check{n}-1}$ over $A_{m+\check{n}-1}$. Then, the MPS's from row m until row $m+\check{n}-1$ were not enough to distract him from the FSD-improvement of B_m over A_m. In contrast, the MPS's from row m until row $m+\check{n}$ were enough to outweigh the FSD-improvement of B_m over A_m. So, we can define an 'intermediate' hypothetical option $\bar{B}_{m+\check{n}}$ with $\bar{B}1_{m+\check{n}} = B1_{m+\check{n}-1} - \kappa k_{\check{n}}$ and $\bar{B}2_{m+\check{n}} = B2_{m+\check{n}-1} + \kappa k_{\check{n}}$. The $\kappa \in [0;1]$ is chosen such that $A_{m+\check{n}} \sim \bar{B}_{m+\check{n}}$. Put differently, κ is the fraction of $k_{\check{n}}$ that would make the individual indifferent between $A_{m+\check{n}}$ and $B_{m+\check{n}}$ if $\kappa k_{\check{n}}$ instead of $k_{\check{n}}$ would

[32] In going from $B_{m+\check{n}-1}$ to $B_{m+\check{n}}$, general MPS's could be applied. However, since we defined each option as having two possible outcomes only (where each occurs with probability 1/2) a MPS can only be attained by adding and subtracting $k_{\check{n}}$. In giving up that all outcomes of the table are equally likely, one would not only require that subjects understand situations other than certainty and equally likely, but one would also introduce complications such as probability weighting.

[33] Note that one could also use mean preserving contractions (i.e. decreases in risk) when going from A_m to $A_{m+\check{n}}$, either instead or in addition to the MPS's that are used when going from B_m to $B_{m+\check{n}}$. This would be useful if one wanted to induce similar changes between the two alternatives over the rows of the table.

have been used. For this individual $\bar{B}_{m+\check{n}}$ is then a mean utility preserving spread of $A_{m+\check{n}}$ in the sense of Diamond and Stiglitz (1974).

Now, suppose there is another individual who is less risk-averse. This individual would then still strictly prefer $\bar{B}_{m+\check{n}}$ over $A_{m+\check{n}}$. It follows, that this individual would choose at least as many times B over A as the more risk-averse individual.[34] The smaller the k_i's, the finer are the differences in risk aversion that can be observed. Any observed differences in behavior must be due to (sufficiently strong) differences in risk aversion.

In order to distinguish risk-seeking individuals, consider all rows with $i \leq m$. Since all $B1_i$ and $B2_i$ are identical, it follows that

$$\begin{aligned} & B_m = B_{m-1} = \ldots = B_{m-\hat{n}} \\ \Rightarrow \quad & B_m \sim B_{m-1} \sim \ldots \sim B_{m-\hat{n}}. \end{aligned} \tag{1.4}$$

Since $A1_{m-\hat{n}} = A1_{m-\hat{n}+1} - k^{\hat{n}}$ and $A2_{m-\hat{n}} = A2_{m-\hat{n}+1} + k^{\hat{n}}$ it follows that $A_{m-\hat{n}}$ is a MPS of $A_{m-\hat{n}+1}$. Of course, every risk-seeking individual likes MPS's and thus

$$\begin{aligned} & A_m \prec_{MPS} A_{m+1} \prec_{MPS} \ldots \prec_{MPS} A_{m-\hat{n}} \\ \Rightarrow \quad & A_m \prec A_{m-1} \prec \ldots \prec A_{m-\hat{n}}. \end{aligned} \tag{1.5}$$

Combining (1.1), (1.4) and (1.5) we can derive that the choice between $A_{m-\hat{n}}$ and $B_{m-\hat{n}}$ for every $\hat{n} > 0$ is a trade-off between the advantage of the FSD-improvement from B_m ($= B_{m-\hat{n}}$) over A_m and the advantage of the increase(s) in risk when going from A_m to $A_{m-\hat{n}}$. By similar arguments as before, it follows that a less risk-seeking individual would choose at least as many times B over A as the more risk-seeking individual.

As in the HL tables, for risk-averse as well as for risk-seeking individuals it is optimal to switch options only once. Risk-averse individuals switch from B to A after row m (the earlier the more risk-averse) and risk-seeking individuals switch from A to B before row m (the earlier the less risk-seeking).[35] For risk-neutral individuals it is optimal to choose option B throughout since it offers the higher expected payoff.

[34] Strictly speaking, an individual who chooses as many times B over A as another individual could still be slightly more or slightly less risk-averse. In this case her κ would be lower or higher, respectively.

[35] Of course, if one wanted to induce switching from A to B for most (i.e. risk-averse) subjects as in the HL tables, one could achieve this by just exchanging alternatives A and B.

1.8 Appendix B: Instructions (translated from German)

Instructions for the Experiment
(please read carefully!)

Dear participant,
thank you very much for your attendance. You are now participating in an experiment at the MELESSA laboratory. From now on you are not allowed to communicate with the other participants. If you have any questions, please alert one of the experimenters by raising your hand. The experimenter will then come to your seat to answer your question. It is prohibited to use cell phones or to open other programs on your computer. Should you have any personal belongings (books, magazines, etc.) on your desk, please store them in your bag now. If you violate any of these rules, we will have to exclude you from the experiment without compensation.

This experiment consists of **two parts**.
In the **first part** of the experiment you will have to make decisions in four tables. These decisions and chance will determine how large your payment (in addition to your show-up fee of 4 euros) will be for this experiment.
After the determination of your payment, we will ask you in the **second part** of the experiment, to fill out a questionnaire about personal characteristics. Then, the experiment will be finished, you will receive your payment, and can leave.

The **course of events** during the experiment will be the following:

1. First, please read the provided instructions for this experiment carefully. Control question will show if you understood them.

2. Then, you will complete the first part of the experiment (**four tables**) at the computer.

3. After all participants have completed the first part, chance and the decisions you made will determine your payment for the experiment.

4. Subsequently, you will complete the second part of the experiment (**questionnaire**) at the computer.

5. When your seat number is called, you will leave the room and receive your payment. The experiment is finished and you can leave.

General Instructions

In the first part of the experiment you will have to make decisions in **four tables**. The tables will appear one after another on your computer screen. Once you have completed a table (by clicking the OK button on the screen) you cannot go back to change your decisions.

The following holds for all four tables:
A table consists of ten rows and you have to decide in each row between two alternatives (alternative A and alternative B). Both alternative A and alternative B have two possible outcomes.
The two possible outcomes of alternative A are called A1 and A2. The two possible outcomes of alternative B are called B1 and B2. If you choose alternative A, chance will determine whether you will receive outcome A1 or outcome A2. Likewise, if you choose alternative B, chance will determine whether you will receive outcome B1 or outcome B2.
There are no objectively right or wrong decisions. However, you should in every row think through whether you prefer the possible outcomes of alternative A (A1 and A2) or the possible outcomes of alternative B (B1 and B2).

There are **two different types** of tables. You will make decisions in two tables of type 1 and in two tables of type 2. All tables have different values of payments. This means that neither tables of type 1 nor tables of type 2 are identical.

The **difference between both types of tables** is the following:
In tables of type 1, the probabilities of the possible outcomes are identical in every row while the amounts of these outcomes change across rows.
In tables of type 2, the amounts of the possible outcomes are identical in every row while the probabilities of these outcomes change across rows.
In the following both types of tables will be described in more detail.

In both types of tables the probabilities are always stated in parentheses.

Instructions for Tables of Type 1

Please have a look at the following sample table (this is not one of the tables you will complete in the experiment).

Screen 1: **Sample Table of Type 1**

Beispieltabelle Typ 1

	Alternative A			Alternative B		
Zeile	**A1:** Wenn die blauen Würfel einen Wert ergeben von **1** bis **50** (50 %) eine Auszahlung (in €) von ...	**A2:** Wenn die blauen Würfel einen Wert ergeben von **51** bis **100** (50 %) eine Auszahlung (in €) von ...	Ihre Wahl	**B1:** Wenn die blauen Würfel einen Wert ergeben von **1** bis **50** (50 %) eine Auszahlung (in €) von ...	**B2:** Wenn die blauen Würfel einen Wert ergeben von **51** bis **100** (50 %) eine Auszahlung (in €) von ...	Zeile
1.	0,10	9,90	A ○ ○ B	5,30	5,50	1.
2.	2,20	7,80	A ○ ○ B	5,30	5,50	2.
3.	4,80	5,20	A ○ ○ B	5,30	5,50	3.
4.	4,80	5,20	A ○ ○ B	4,00	6,80	4.
5.	4,80	5,20	A ○ ○ B	3,80	7,00	5.
6.	4,80	5,20	A ○ ○ B	3,50	7,30	6.
7.	4,80	5,20	A ○ ○ B	3,20	7,60	7.
8.	4,80	5,20	A ○ ○ B	2,90	7,90	8.
9.	4,80	5,20	A ○ ○ B	2,10	8,70	9.
10.	4,80	5,20	A ○ ○ B	0,40	10,40	10.

OK

In every row (1 to 10) you have to decide between alternative A and alternative B. The two possible outcomes of alternative A (A1 and A2) and alternative B (B1 and B2) are always equally likely. This means for instance, if you choose alternative A in a row, you will receive outcome A1 with 50% probability and outcome A2 also with 50% probability.
While the probabilities of the possible outcomes are identical in every row, the amounts of these outcomes differ.

Instructions for Tables of Type 2

Please have a look at the following sample table (this is not one of the tables you will complete in the experiment).

Screen 2: **Sample Table of Type 2**

Beispieltabelle Typ 2

	Alternative A			Alternative B		
Zeile	**A1:** Eine Auszahlung von **4,00 €** wenn die blauen Würfel einen Wert ergeben von ...	**A2:** Eine Auszahlung von **3,20 €** wenn die blauen Würfel einen Wert ergeben von ...	**Ihre Wahl**	**B1:** Eine Auszahlung von **7,70 €** wenn die blauen Würfel einen Wert ergeben von ...	**B2:** Eine Auszahlung von **0,20 €** wenn die blauen Würfel einen Wert ergeben von ...	**Zeile**
1.	**1** bis **10** (10 %)	**11** bis **100** (90 %)	A ○ ○ B	**1** bis **10** (10 %)	**11** bis **100** (90 %)	1.
2.	**1** bis **20** (20 %)	**21** bis **100** (80 %)	A ○ ○ B	**1** bis **20** (20 %)	**21** bis **100** (80 %)	2.
3.	**1** bis **30** (30 %)	**31** bis **100** (70 %)	A ○ ○ B	**1** bis **30** (30 %)	**31** bis **100** (70 %)	3.
4.	**1** bis **40** (40 %)	**41** bis **100** (60 %)	A ○ ○ B	**1** bis **40** (40 %)	**41** bis **100** (60 %)	4.
5.	**1** bis **50** (50 %)	**51** bis **100** (50 %)	A ○ ○ B	**1** bis **50** (50 %)	**51** bis **100** (50 %)	5.
6.	**1** bis **60** (60 %)	**61** bis **100** (40 %)	A ○ ○ B	**1** bis **60** (60 %)	**61** bis **100** (40 %)	6.
7.	**1** bis **70** (70 %)	**71** bis **100** (30 %)	A ○ ○ B	**1** bis **70** (70 %)	**71** bis **100** (30 %)	7.
8.	**1** bis **80** (80 %)	**81** bis **100** (20 %)	A ○ ○ B	**1** bis **80** (80 %)	**81** bis **100** (20 %)	8.
9.	**1** bis **90** (90 %)	**91** bis **100** (10 %)	A ○ ○ B	**1** bis **90** (90 %)	**91** bis **100** (10 %)	9.
10.	**1** bis **100** (100 %)	**niemals** (0 %)	A ○ ○ B	**1** bis **100** (100 %)	**niemals** (0 %)	10.

OK

Again, in every row (1 to 10) you have to decide between alternative A and alternative B. The difference from type 1 tables is that now the amounts of the two possible outcomes of alternative A (A1 and A2) and alternative B (B1 and B2) are identical in every row. However, the probabilities of these possible outcomes are different in each row.

Instructions for the Determination of Your Payment

Having made all decisions by completing all four tables, chance will determine the payment you will receive. You will have to roll four dice. As soon as all participants have completed the tables, an experimenter will come to your seat. Then, you will roll the four dice and, **supervised by the experimenter**, enter the results of the dice rolls into the respective computer screen. Note that you are not allowed to enter any numbers on this screen until the experimenter is at your seat. A violation of this rule will lead to the exclusion from the experiment without compensation. The four dice look like the following:

The red four-sided die (numbered from 1 to 4) decides which of the four tables is relevant for your payment.

The white ten-sided die (numbered from 0 to 9) then decides which row in the diced table is relevant. Note that the roll of a 0 counts as a 10.

Finally, the two blue ten-sided dice decide which of the two possible outcomes you will in fact receive.

For instance, if you chose alternative B in the relevant table and row, the two blue ten-sided dice would decide whether you received outcome B1 or B2. The sum of the two blue dice yields a number between 1 and 100. One blue die is numbered with 0, 10, 20, 30, 40, . . . , 90, and the other blue die is numbered with 0, 1, 2, 3, 4, . . . , 9. If the sum of both dice yields 0, it will count as 100. Therefore, every number between 1 and 100 has the same probability of being rolled.

In a **table of type 1** you will receive the first outcome if the rolled number is between 1 and 50, and you will receive the second outcome if the rolled number is between 51 and 100. For instance, if you chose alternative A in the relevant table and row, you would receive outcome A1 if you rolled a number between 1 and 50 (which happens with 50% probability), and you

would receive outcome A2 in case you rolled a number between 51 and 100 (which happens with 50% probability). Therefore, in tables of type 1 both outcomes of an alternative are always equally likely.
Example: Suppose the sample table of type 1 (screen 1) would be the relevant table for your payment. Further, suppose you chose alternative B in rows 1 to 7 and alternative A in rows 8 to 10.
Now, if row 3 is the relevant row, you will receive B1 = 5.30 euros if you roll a number between 1 and 50 (which happens with 50% probability) and you will receive B2 = 5.50 euros if you roll a number between 51 and 100 (which happens with 50% probability).
By contrast, if row 4 is the relevant row you will receive B1 = 4.00 euros if you roll a number between 1 and 50 (which happens with 50% probability) and you will receive B2 = 6.80 euros if you roll a number between 51 and 100 (which happens with 50% probability).

In a **table of type 2** the relevant row determines with which rolled number you will receive which outcome.
Example: Suppose the sample table of type 2 (screen 2) would be the relevant table for your payment. Further, suppose you chose alternative A in rows 1 to 6 and alternative B in rows 7 to 10.
Now, if row 2 is the relevant row you will receive A1 = 4.00 euros in case you roll a number between 1 and 20 (which happens with 20% probability), and you will receive A2 = 3.20 euros if you roll a number between 21 and 100 (which happens with 80% probability).
By contrast, if row 6 is the relevant row you will receive A1 = 4.00 euros if you roll a number between 1 and 60 (which happens with 60% probability), and you will receive A2 = 3.20 euros if you roll a number between 61 and 100 (which happens with 40% probability).

Control Questions

The following two control questions will help you to check whether you understood the instructions. Please write down the respective solutions and raise your hand when you have answered all questions.

Control Question 1:
Suppose, after having completed the four tables, you roll a 3 with the red four-sided die. Further, suppose that the sample table of type 1 (screen 1) was the third (and therefore relevant) table you completed. Suppose that you chose alternative B in rows 1 to 5 and alternative A in rows 6 to 10 in this table.

How large will your payment be if you roll a 2 with the white ten-sided die and a 89 with the two blue dice? ______

How large will your payment be if you roll a 7 with the white ten-sided die and a 62 with the two blue dice? ______

How large will your payment be if you roll a 5 with the white ten-sided die and a 41 with the two blue dice? ______

Control Question 2:
Suppose, after having completed the four tables, you roll a 1 with the red four-sided die. Further, suppose that the sample table of type 2 (screen 2) was the first (and therefore relevant) table you completed. Suppose that you chose alternative A in rows 1 to 7 and alternative B in rows 8 to 10.

How large will your payment be if you roll a 2 with the white ten-sided die and a 89 with the two blue dice? ______

How large will your payment be if you roll a 7 with the white ten-sided die and a 62 with the two blue dice? ______

How large will your payment be if you roll a 5 with the white ten-sided die and a 41 with the two blue dice? ______

If you have answered both control questions, please raise your hand. An experimenter will come to your seat to ensure that your solutions are correct. In case you have any further questions, they will be answered as well.

As soon as all participants have answered the control questions and no further questions arise, the experiment will start on the computer screen.

Chapter 2

Reference-Dependent Risk Preferences of Higher Orders[0]

2.1 Introduction

While the well-known second-order effect of risk aversion describes a preference for less uncertainty, higher-order risk preferences characterize how this preference changes under different circumstances. For instance, the third-order effect of prudence plays an important role for changes in risk aversion due to variations in wealth, and the fourth-order effect of temperance is crucial for changes in risk aversion in the face of a background risk. The theoretical literature on these higher orders is well-developed (building on Kimball, 1990, 1992) since the importance of prudence for phenomena like precautionary saving was recognized early (Leland, 1968). While numerous empirical studies find supportive evidence for the presence of precautionary saving, they also show that it is usually too low to be consistent with those theoretical models and common assumptions about risk aversion.[1] Also, contrary to observation, optimal precautionary saving should theoretically lead to substantial crowding out of insurance demand (Gollier, 2003).

Until now, theoretical models analyzing higher-order risk preferences were constrained to the framework of expected utility theory (EUT) in which they correspond to the derivatives of the utility function. In contrast, in order to resolve various empir-

[0] This chapter is based on joint work with Maximilian Rüger.

[1] For instance, the test of Dynan (1993, p1104) "[...] yields a fairly precise estimate of a small precautionary motive; in fact, the estimate is too small to be consistent with widely accepted beliefs about risk aversion." In an overview article Carroll and Kimball (2008, pp583-584) conclude that "[...] estimates of relative risk aversion imply precautionary saving motives much stronger than those that have been used empirically to match observed wealth holdings. This discrepancy remains unresolved."

ical puzzles concerning second-order risk preferences, a prominent strand of literature (starting with Kahneman and Tversky, 1979) evolved that stresses the dependence of preferences on a reference point. The latest advancement in this literature was made by Kőszegi and Rabin (2006, 2007) who endogenize recent expectations as the reference point. We analyze higher-order risk preferences within their model of reference-dependent preferences. We further show that our results can explain seemingly sub-optimal levels of precautionary saving and insurance demand. As a robustness analysis, it is also shown that alternative models, like disappointment, regret, or exogenous reference points yield different results and cannot resolve these empirical puzzles concerning higher-order risk preferences.

Section 2.2 presents the two concepts that are merged in this chapter and gives the theoretical background that is needed in order to understand how we derive our results. In Section 2.2.1 we follow Eeckhoudt and Schlesinger (2006) and define lotteries that represent higher-order risk preferences independently of a specific model. It is their gamble representation that allows us to analyze higher orders in the model of Kőszegi and Rabin (2007) which is presented in Section 2.2.2. In order to make our results comprehensible as well as comparable to each other, we transform choices such that they reflect preference relations as if the reference point was zero. This method of transformation is explained in Section 2.2.3.

Our results are presented in Section 2.3. We show that individuals with expectation-based reference-dependent preferences are neither prudent nor imprudent while second- and fourth-order effects are still present. More specifically, using common functional forms of gain-loss utility we show that individuals are risk-averse and intemperate, but they are indifferent toward the third order *independent* of the functional form of gain-loss utility. While risk aversion is also predicted by classical EUT models, third- and fourth-order behavior differs substantially under reference dependence. We generalize these results for risk preferences of arbitrary order n and show that those of odd orders (except order one) are absent while those of even orders are still present. The intuition for this result is based on the fact that odd-order risk preferences can always be thought of as a decision on the location of risk(s), whereas even-order risk preferences are concerned with the aggregation of risk(s). Due to the anticipation of choices, the location does not matter for sensations of gains and losses.

In Section 2.4 some of the possible economic implications of our results are discussed. In general our results show that it may be highly problematic to derive optimal behavior that rests on preferences of higher orders from measures of risk aversion. With reference-dependent preferences combinations of certain risk attitudes are possible that cannot be consistently derived within EUT. More specifically, we show that empirical findings on seemingly sub-optimal amounts of precautionary saving under classical EUT can be attributed to the optimal behavior of individuals with expectation-based reference-dependent preferences. The reason is that reference dependence contributes positively to risk aversion, but leaves prudence unaffected. Within classical EUT models optimal precautionary saving is directly derived from measures of risk aversion. However, if individuals have reference-dependent preferences instead, measured risk aversion will indeed be higher but not the optimal amount of precautionary saving. Also, the theoretically puzzling existence of costly insurance when capital markets are well developed and individuals can accumulate buffer-stock wealth is less surprising for such individuals. Under reference dependence second-order effects become relatively more important than third-order effects.

In Section 2.5 we consider several other behavioral models in order to show that the same results do not obtain under alternative specifications of the reference point. Section 2.5.1 considers disappointment models where the reference point is the expectational mean of the chosen alternative (Bell, 1985; Loomes and Sugden, 1986). Here, we do not observe the absence of odd-order risk preferences while even-order risk preferences are similar to those of expectation-based reference dependence. In recent empirical work it has been difficult to distinguish between these two models of reference dependence. Our third-order result therefore suggests a new way how to differentiate between expectation-based reference dependence and models of disappointment. In Section 2.5.2 we analyze higher orders in models of regret (Bell, 1982, 1983; Loomes and Sugden, 1982) where the reference point is the alternative that was *not* chosen. It is shown that decisions taken to avoid regret are unaffected by even-order effects, but odd-order effects still exist. This suggests that regret influences higher-order risk preferences in the opposite way as expectation-based reference dependence. Risk preferences of any order (except order one) differ between these two models, but regret induces similar odd-order risk preferences as disappointment models. Models with reference points shaped by expectations and regret are usually analyzed separately rather than jointly in a unified framework. However, our analysis shows the close relationship they share. Risk preferences of a particular order are only affected by either expectations or regret as the reference point, but never by both. Lastly, we analyze exogenous reference points, such as the status quo, in Section 2.5.3. Here, results vary with the various reference points considered and attitudes concerning gains and losses. However, just as in the case of disappointment and regret, exogenous reference points do not yield the same results as expectation-based reference dependence on higher-order risk preferences. Our robustness analysis of Section 2.5 therefore shows that alternative specifications of the reference point imply different results and cannot explain the considered empirical puzzles that have been found with respect to higher orders.

Section 2.6 is devoted to the conclusion where we summarize our results and discuss further extensions. All proofs appear in the Appendix.

2.2 Theoretical Background

2.2.1 Higher Orders

Within EUT it has long been recognized that risk preferences are not exhaustively described via the concept of risk aversion alone. Already Leland (1968) and Sandmo (1970) identified the importance of a positive third derivative of the utility function for motives of precautionary saving. Since Kimball (1990) the feature of preferences that is necessary and sufficient for such precautionary behavior is termed prudence.[2] An equivalent concept to prudence, termed downside risk aversion, was established by Menezes, Geiss, and Tressler (1980). It describes an aversion to mean-variance preserving transformations that shift risk from the right to the left of a wealth distribution. Afterward, an extensive literature developed that discussed also higher derivatives of utility functions and their implications for economically important decisions under uncertainty. For instance, negative fourth derivatives of the utility function were shown to be crucial for risk aversion to increase with the existence of an independent zero-mean background risk (see Gollier and Pratt, 1996; or Eeckhoudt, Gollier, and Schlesinger, 1996). This fourth-order property of the utility function was termed temperance by Kimball (1992) and outer risk aversion by Menezes and Wang (2005). Eeckhoudt and Schlesinger (2008) show that temperance is necessary and sufficient for precautionary saving to increase as the *downside* risk of future income rises.

Outside standard EUT, preferences of higher orders have not been considered explicitly in the literature yet. Building on the concept of n^{th}-degree risk by Ekern (1980), Eeckhoudt and Schlesinger (2006), however, provided definitions of all higher-order ef-

[2] Despite being not necessary *and* sufficient, prudence has also been shown to play an important role in other economic applications, like for instance rent-seeking games (Treich, 2010), auctions (Eső and White, 2004), bargaining (White, 2008), principal-agent monitoring (Fagart and Sinclair-Desagné, 2007), global commons problems (Bramoullé and Treich, 2009), inventory management (Eeckhoudt, Gollier, and Schlesinger, 1995), or optimal prevention (Eeckhoudt and Gollier, 2005; or Courbage and Rey, 2006). It has also been used in a more normative context to derive an optimal ecological discount rate (Gollier, 2010) or to justify the so-called precautionary principle (Gollier, Jullien, and Treich, 2000; and Gollier and Treich, 2003).

fects in terms of preferences over simple lotteries.[3] This representation is quite general since it is independent of a specific model. The authors showed that when applied in an EUT framework their definitions correspond to the common definitions in terms of derivatives used so far. Because we are interested in higher-order risk preferences under reference dependence, we employ these gamble definitions in our analysis.

Denote initial wealth by $y \in \mathbb{R}$.[4] Let $k \in \mathbb{R}$ be a sure reduction in wealth with $k > 0$. $\tilde{\varepsilon}_i$ are symmetric random variables which are non-degenerate, independent of all other random variables that affect wealth, and have $\mathbb{E}[\tilde{\varepsilon}_i] = 0$.[5] Now, we define the following standard gambles, where each element denotes an outcome and each outcome of a specific gamble is realized with equal probability.[6] Such defined outcomes can be either deterministic or stochastic themselves.

$$
\begin{aligned}
B_1 &\equiv [y] & A_1 &\equiv [y-k] \\
B_2 &\equiv [y] & A_2 &\equiv [y+\tilde{\varepsilon}_1] \\
B_3 &\equiv [y-k; y+\tilde{\varepsilon}_1] & A_3 &\equiv [y; y-k+\tilde{\varepsilon}_1] \\
&= [A_1; B_1+\tilde{\varepsilon}_1] & &= [B_1; A_1+\tilde{\varepsilon}_1] \\
B_4 &\equiv [y+\tilde{\varepsilon}_1; y+\tilde{\varepsilon}_2] & A_4 &\equiv [y; y+\tilde{\varepsilon}_1+\tilde{\varepsilon}_2] \\
&= [A_2; B_2+\tilde{\varepsilon}_2] & &= [B_2; A_2+\tilde{\varepsilon}_2] \\
B_n &\equiv [A_{n-2}; B_{n-2}+\tilde{\varepsilon}_{\mathrm{int}(\frac{n}{2})}] & A_n &\equiv [B_{n-2}; A_{n-2}+\tilde{\varepsilon}_{\mathrm{int}(\frac{n}{2})}].
\end{aligned}
$$

Intuitively, B_1 vs. A_1 is equivalent to the question whether more is preferred to less. B_2 vs. A_2 is the comparison between a certain level of wealth and a risky alternative with an identical mean. B_3 vs. A_3 is equivalent to the question whether one prefers to add an unavoidable random variable to the higher or lower outcome. It is also equivalent to the question whether one prefers to accept a sure reduction in wealth in the certain or uncertain state. B_4 vs. A_4 can be seen as the question whether one prefers to aggregate or disaggregate two independent random variables.

[3] These lotteries have also been used in recent experiments to investigate higher-order risk preferences empirically (see Deck and Schlesinger, 2010; Ebert and Wiesen, 2011; or Maier and Rüger, 2010).

[4] The non-randomness of initial wealth is only used for simplicity.

[5] Except the assumption of symmetry, all $\tilde{\varepsilon}_i$ are identically defined in Eeckhoudt and Schlesinger (2006).

[6] We defined all gambles having a mean of y if $k = 0$ instead of 0 as in Eeckhoudt and Schlesinger (2006).

Following Eeckhoudt and Schlesinger (2006), preferences are then

$$\begin{array}{llllll} \text{monotone} & \Leftrightarrow & B_1 \succsim A_1 & \forall y, k & (\Leftrightarrow u'(x) \geq 0 \text{ in EUT models}) \\ \text{risk-averse} & \Leftrightarrow & B_2 \succsim A_2 & \forall y, \tilde{\varepsilon}_1 & (\Leftrightarrow u''(x) \leq 0 \text{ in EUT models}) \\ \text{prudent} & \Leftrightarrow & B_3 \succsim A_3 & \forall y, \tilde{\varepsilon}_1, k & (\Leftrightarrow u'''(x) \geq 0 \text{ in EUT models}) \\ \text{temperate} & \Leftrightarrow & B_4 \succsim A_4 & \forall y, \tilde{\varepsilon}_1, \tilde{\varepsilon}_2 & (\Leftrightarrow u''''(x) \leq 0 \text{ in EUT models}) \end{array}$$

or, more generally

$$\text{risk apportioning of order } n \;\Leftrightarrow\; B_n \succsim A_n \quad \forall y, \tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_{\mathrm{int}(\frac{n}{2})}, k$$
$$\big(\Leftrightarrow u^n(x) \leq (\geq)\ 0 \text{ if } \tfrac{n}{2} \in (\notin)\ \mathbb{N} \text{ in EUT models}\big)$$

where $u^n(\cdot)$ denotes the n^{th} derivative of $u(\cdot)$. Note that all commonly used utility functions in EUT, like for instance those exhibiting constant relative or absolute risk aversion, have derivatives that alternate in signs with $u'(\cdot)$ being positive (see Brockett and Golden, 1987), a property termed mixed risk aversion by Caballé and Pomansky (1996).

2.2.2 Reference Dependence

Reference-dependent preferences were established in economics via Kahneman and Tversky's (1979) prospect theory. It introduced four novel aspects: reference dependence, loss aversion, diminishing sensitivity concerning gains and losses and non-linear probability weighting. Our main focus is on the first feature. Some results will be completely independent of whether the second and third features are present while other results will vary with their existence. The fourth feature of probability weighting will be absent from the whole analysis. In this respect we follow the models we discuss in this chapter.

Focusing on reference dependence immediately leads to the question of what the reference point is. In the traditional literature on prospect theory the reference point is a real numbered variable that is completely exogenous. It is a simple point which every potential outcome is compared to, and it is not influenced by the decision-maker in any way. Often, the status quo was advocated as a candidate for this reference point (e.g. Kahneman and Tversky, 1979; or Tversky and Kahneman, 1991). However, even this early literature noted that this might not be the relevant comparison in many important situations. "For example, an unexpected tax withdrawal from a monthly pay check is experienced as a loss, not as a reduced gain." (Kahneman and Tversky, 1979, p286). Later, considerations that expectations may play a role were explicitly included into formal analysis (see Bell, 1985; Loomes and Sugden, 1986; Sugden, 2003; Delquié and Cillo, 2006; and Kőszegi and Rabin, 2006, 2007, 2009).

In early models of disappointment (see Bell, 1985; and Loomes and Sugden, 1986) individuals compare potential outcomes to the mean of the choice. Every option that can be chosen has therefore its specific reference point which is thus endogenous. An advancement of disappointment models has been proposed by Kőszegi and Rabin (2006, 2007). Here, the reference point is not simply the mean but rather the full distribution of outcomes resulting from the chosen alternative. We refer to such reference points as expectational references. This specification is motivated by the observation that individuals can make multiple comparisons to evaluate an outcome. It also accounts for the fact that individuals realize uncertainties and incorporate them into their reference point.

Expectational references have since been applied in order to explain various important phenomena. For instance, Heidhues and Kőszegi (2008) use them in a model of price competition to explain sticky prices. Herweg and Mierendorff (forthcoming) shows that the flat-rate bias when choosing optimal tariffs can be explained when consumers have expectational references. Heidhues and Kőszegi (2010) use expectational references in order to explain why "sale" prices in addition to regular prices only exist in certain environments. Herweg, Müller, and Weinschenk (2010) analyze a moral hazard setting with expectational references in order to explain binary payment schemes. Closely related, Macera (2010) shows in a two-period model that contracts deferring all present incentives into future payments, like e.g. yearly productivity bonuses combined with a present fixed wage, are optimal with expectational references. Lange and Ratan (2010) study bidding behavior of agents with expectational references in different auction formats and explain why in laboratory experiments but not necessarily in the field overbidding should be observed.

Empirical work also points toward expectations as the relevant reference point. Meng (2009) uses expectations as the reference point in order to explain the disposition effect, i.e. the observation that stock market investors tend to keep their loosing assets for too long and sell their winning ones too soon, and further finds expectations as best estimate of investors' reference point from individual trading data. Card and Dahl (2011) show that the rate of family violence when the local professional football team loses depends on the extent to which losing the game was expected. Mas (2006) shows that the larger the difference is between requested (by the union) and received wages the more police performance declines. Pope and Schweitzer (2011) show that expectation-based loss aversion is present among professional golfers suggesting that it is a bias which survives experience, competition and large stakes. Post et al. (2008) analyze game show data and find that behavior of contestants is consistent with reference points shaped by expectations. Crawford and Meng (2011) (building upon previous work by Camerer et al., 1997; and Farber, 2005, 2008) propose and estimate a labor supply model of New York City cabdrivers with reference-dependent preferences where income- as well as working hour-targets are both determined by rational expectations. Using new data of New York City cabdrivers and using a natural experiment Doran (2009) finds that a permanent wage increase causes hours worked to remain constant, a finding consistent with expectations as the reference point.

Labor supply decisions have also been tested in a laboratory experiment by Abeler et al. (2011). They manipulate subjects' rational expectations and find that effort provision is affected in the way predicted by expectation-based reference-dependent preferences. Ericson and Fuster (2010) show in an exchange and in a valuation experiment that the reference point is determined by expectations rather than by the status quo and discuss why some researchers have not found endowment effects in different settings. Knetsch and Wong (2009) also conduct exchange experiments and suggest that expectations shape the reference point. A recent experiment by Gill and Prowse (forthcoming) finds that expectations are the relevant reference point when subjects compete in a real effort tournament. Finally, Loomes and Sugden (1987), Choi et al. (2007), or Hack and Lammers (2009) find supportive evidence for expectations as the reference point in risky choice experiments.

All the above mentioned literature is concerned with first- or second-order effects of reference-dependent preferences. In this chapter we are rather interested in higher-order effects. To our knowledge there has been no attempt to consider higher orders under reference dependence. We follow Kőszegi and Rabin (2007) and define overall decision utility $V(\cdot)$ as

$$\begin{aligned} V(F|G) &= M(F) + U(F|G) \\ &= \int m(x)dF(x) + \iint u(x|r)dG(r)dF(x). \end{aligned} \tag{2.1}$$

The outcome $x \in \mathbb{R}$ and the reference point $r \in \mathbb{R}$ are independently drawn from the probability distributions F and G, respectively. Pure consumption utility depending solely on the outcome x is denoted by $m(x)$ and thus the expected pure consumption utility is $M(F) = \int m(x)dF(x)$. The function

$$u(x|r) = u(m(x) - m(r))$$

captures sentiments of gains and losses that occur if the outcome x is realized and the reference point was r. The expected level of gain-loss utility is denoted by $U(F|G) = \iint u(x|r)dG(r)dF(x)$. The assumption of independently distributed x and r captures the notion that the evaluation of all possible wealth outcomes is based on comparing each of them to all possibilities in the support of the reference lottery. What we refer to as expectational references (Kőszegi and Rabin, 2006, 2007, 2009) equates the reference and the outcome distribution and therefore $G = F$.[7]

Some of our results in the following section hold independently of the functional form of gain-loss utility $u(\cdot)$. The remaining results depend on its specific shape. For the latter cases we will consider u functions that are usually proposed in the literature. This includes standard (weakly) concave as well as loss-averse S-shaped u functions. S-shaped u functions are defined as being (weakly) concave in gains and (weakly) convex in losses (i.e. $u''(\tau) \leq 0\ \forall\tau > 0$ and $u''(\tau) \geq 0\ \forall\tau < 0$). Following Köbberling and Wakker (2005) we refer to loss aversion as a kink at the reference point (i.e. $\lim_{\tau\to 0} \frac{u'(-|\tau|)}{u'(|\tau|)} \equiv \lambda > 1$). Furthermore, when S-shaped utility is assumed we refer to loss-averse u functions as the standard case where utility is diminishing faster or equally in gains than in losses (i.e. $u''(\tau) \leq -u''(-\tau)\ \forall\tau > 0$). This includes the case where $u(\cdot)$ is piecewise linear with loss aversion. Often results differ for strict formulations of the functional characteristics above and the limiting cases. A limiting case of S-shaped u functions without loss aversion are symmetric S-shaped u functions. A linear $u(\cdot)$ is a limiting case of both weakly concave and S-shaped functions without loss aversion. Both limiting cases have in common that they are (two-fold rotational) symmetric around the reference point. This property is crucial in explaining why their results differ compared to their strict (and asymmetric) counterparts.[8] Note that both strictly concave and loss-averse S-shaped u functions share the common feature that losses loom larger than corresponding gains (i.e. $u'(\tau) < u'(-\tau)\ \forall\tau > 0$).[9] Throughout it is assumed that $u(\tau)$ is strictly

[7]Kőszegi and Rabin (2006, 2007, 2009) further develop an equilibrium concept that restricts which choices can be rationally expected. We will discuss this at the end of Section 2.3.

[8]A loss-averse S-shaped u function is the typical assumption in the literature on reference dependence. The reason why we additionally discuss concave as well as (two-fold rotational) symmetric u functions is the following. In Section 2.5 we consider alternative models, like disappointment or regret theory, in order to show that the results on expectational references cannot be derived using such alternative models. These alternative models can also be captured by (2.1) but with a different specification concerning the reference point. A common assumption in the regret literature is a concave u function. In their disappointment theories, Loomes and Sugden (1986) assume a u function that is convex in gains and concave in losses but has the property of being (two-fold rotational) symmetric, while Bell (1985) assumes a piecewise linear $u(\cdot)$. Since there exists no a priori reason to assume that a specific functional form is limited to a specific reference scenario we rather consider all possibilities throughout.

[9]While evidence for this feature in the literature on reference dependence is vast, for evidence that regret looms larger than rejoicing, see e.g. Inman, Dyer, and Jia (1997) and Camille et al. (2004).

increasing, continuous, and sufficiently often differentiable for all τ ($\forall \tau \neq 0$ if $\lambda > 1$), and that $u(0) = 0$.

While we consider various general functional forms of $u(\cdot)$, we again follow Kőszegi and Rabin (2007) in that we restrict $m(\cdot)$ to be linear (i.e. $m'(\cdot) > 0, m''(\cdot) = 0$).[10] The reason for this assumption is twofold. First, in models where $m(\cdot)$ is assumed to be concave instead usually strong restrictions on the functional form of $u(\cdot)$ are needed to keep the model tractable.[11] Since we are interested in the effect of reference dependence on higher-order risk attitudes, we rather consider different functional forms that directly affect gain-loss utility $u(\cdot)$. Second and more importantly, higher orders are already well analyzed within classical EUT models. Assuming a concave m function would simply reintroduce higher-order effects of EUT into the gain-loss function and would therefore dilute the pure effect of reference dependence on higher-order risk attitudes.[12] Moreover, reference dependence is especially considered as important in situations where classical consumption utility is almost linear. In order to generate reasonable risk taking behavior, Rabin (2000) shows that consumption utility is compatible with risk neutrality for amounts as high as a monthly salary.

2.2.3 Higher Orders under Reference Dependence

For notational convenience we will use $F|G \succsim F'|G'$ interchangeably to $V(F|G) \geq V(F'|G')$ throughout the chapter. Since we are interested in higher-order risk preferences under reference dependence, we apply the gamble definitions of Section 2.2.1 to the outlined model (2.1) in the following way.

Definition 2.1 *A reference-dependent individual is said to be risk apportioning of order n (its opposite, neither nor) if and only if $B_n|G \succ (\prec, \sim)\ A_n|G'$, or synonymously if and only if $V(B_n|G) > (<, =)\ V(A_n|G')$, for all $y, k, \tilde{\varepsilon}_i$.*

Expectational references equate the reference distribution to the distribution of the chosen alternative, and we can therefore further specify the reference point.

[10] With the exception of their proposition 9, this assumption is made throughout by Kőszegi and Rabin (2007) and for some results in Kőszegi and Rabin (2006). Other more applied models in the literature impose this assumption as well (see e.g. Heidhues and Kőszegi, 2010; Lange and Ratan, 2010; Meng, 2009; or Gill and Prowse, forthcoming). In the context of regret, Bleichrodt, Cillo, and Diecidue (2010) show in an experiment that $m(\cdot)$ is *not* significantly different from linear. In the disappointment models of Bell (1985) and Loomes and Sugden (1986) $m(\cdot)$ is assumed to be linear as well.

[11] These models restrict attention to piecewise linear gain-loss utility (see Heidhues and Kőszegi, 2008, 2010; Crawford and Meng, 2011; Meng, 2009; Herweg, Müller, and Weinschenk, 2010; Herweg and Mierendorff, forthcoming; Lange and Ratan, 2010; Macera, 2010; and some results in Kőszegi and Rabin, 2006, 2007).

[12] For instance, Macera (2010) needs to restrict consumption utility to be not too concave such that the reference-dependent effect still dominates the consumption-utility effect.

Definition 2.2 *If the reference point is determined by expectational references, $G = B_n$ and $G' = A_n$.*

Because we only have to consider pairwise choices in our analysis, it is often useful to transform the comparison of $B_n|G$ vs. $A_n|G'$ into a comparison that is easier to comprehend but is behaviorally equivalent to the original comparison. We will systematically use two transformations that simplify comparisons. First, we will transform every gamble given a specific reference point distribution into an equivalent gamble with a hypothetical reference point of 0. Second, we will reduce both gambles that are compared to the components in which these two gambles differ. Both transformations make it much easier to derive behavioral predictions in cases where our results depend on the gain-loss utility u, since its functional characteristics are formulated in relation to the reference point 0. They further ensure that our results are comparable to each other.

For an illustration consider a situation in which either B_2 or A_2 has to be chosen. In this hypothetical situation the assumed reference point distribution is $[y + \tilde{\varepsilon}'_1 + \tilde{\varepsilon}'_2]$ if B_2 is chosen and the assumed reference point distribution is $[y; y + \tilde{\varepsilon}'_2]$ if A_2 is chosen.[13] Under which conditions in the original comparison, which is

$$\begin{aligned} & B_2|[y + \tilde{\varepsilon}'_1 + \tilde{\varepsilon}'_2] && \succ (\prec, \sim) && A_2|[y; y + \tilde{\varepsilon}'_2] \\ \Leftrightarrow \quad & [y]|[y + \tilde{\varepsilon}'_1 + \tilde{\varepsilon}'_2] && \succ (\prec, \sim) && [y + \tilde{\varepsilon}_1]|[y; y + \tilde{\varepsilon}'_2], \end{aligned} \tag{2.2}$$

a certain preference relation holds may be hard to conceive intuitively. However, we can restate (2.2) by an equivalent comparison in which the reference point of both alternatives is [0]. This makes an intuitive assessment possible. Using the fact that $\tilde{\varepsilon}_i$ is symmetric we can say that

$$M([y]) + U([\tilde{\varepsilon}'_1 + \tilde{\varepsilon}'_2]|[0]) \quad > (<, =) \quad M([y + \tilde{\varepsilon}_1]) + \tfrac{1}{2}U([\tilde{\varepsilon}_1]|[0]) + \tfrac{1}{2}U([\tilde{\varepsilon}_1 + \tilde{\varepsilon}'_2]|[0]) \tag{2.3}$$

[13] $\tilde{\varepsilon}'_i$ and $\tilde{\varepsilon}_i$ are identically and independently distributed random variables. Note that this is a direct implication of the assumption that reference and outcome distributions are independent. We denote all random variables stemming from the reference point with a $'$ to clarify the fact that $\tilde{\varepsilon}_i + \tilde{\varepsilon}'_i$ denotes the convolution of two random variables rather than the sum of outcomes.

is equivalent to (2.2). Since $m(\cdot)$ is linear, $\mathbb{E}[\tilde{\varepsilon}_i] = 0$, and hence $M([y]) = M([y + \tilde{\varepsilon}_1])$ and $M([\tilde{\varepsilon}_1]) = M([\tilde{\varepsilon}_1 + \tilde{\varepsilon}_2])$, (2.3) in turn is equivalent to

$$\begin{aligned} M([\tilde{\varepsilon}'_1 + \tilde{\varepsilon}'_2]) + U([\tilde{\varepsilon}'_1 + \tilde{\varepsilon}'_2]|[0]) \quad &> (<, =) \quad \tfrac{1}{2}M([\tilde{\varepsilon}_1]) + \tfrac{1}{2}U([\tilde{\varepsilon}_1]|[0]) \\ &\qquad + \tfrac{1}{2}M([\tilde{\varepsilon}_1 + \tilde{\varepsilon}'_2]) + \tfrac{1}{2}U([\tilde{\varepsilon}_1 + \tilde{\varepsilon}'_2]|[0]) \quad (2.4) \\ \Leftrightarrow \quad [\tilde{\varepsilon}'_1 + \tilde{\varepsilon}'_2]|[0] \quad &\succ (\prec, \sim) \quad [\tilde{\varepsilon}_1; \tilde{\varepsilon}_1 + \tilde{\varepsilon}'_2]|[0]. \end{aligned}$$

Although this comparison is already easier to conceive than the original comparison, we can further simplify (2.4) by eliminating components that are common in both gambles and can therefore not contribute to a preference of one gamble over the other, so that

$$\begin{aligned} M([\tilde{\varepsilon}'_1 + \tilde{\varepsilon}'_2]) + U([\tilde{\varepsilon}'_1 + \tilde{\varepsilon}'_2]|[0]) \quad &> (<, =) \quad M([\tilde{\varepsilon}_1]) + U([\tilde{\varepsilon}_1]|[0]) \\ \Leftrightarrow \quad [\tilde{\varepsilon}'_1 + \tilde{\varepsilon}'_2]|[0] \quad &\succ (\prec, \sim) \quad [\tilde{\varepsilon}_1]|[0]. \end{aligned}$$

Clearly, the question whether an individual prefers the convolution of $\tilde{\varepsilon}_1$ and $\tilde{\varepsilon}_2$ over $\tilde{\varepsilon}_1$ given the choice-independent reference point of 0 is more intuitive than the original comparison in (2.2). For most commonly used functional forms of $u(\cdot)$ an answer could be derived easier when the problem is stated in such residual gambles, whose preference relation is nevertheless equivalent to that of the original gambles.

We denote these residual gambles by $[B_n|G_{\langle A_n|G'\rangle}]\,|[0]$ and $[A_n|G'_{\langle B_n|G\rangle}]\,|[0]$, respectively, and apply them to the comparisons of B_n and A_n for all $n \geq 2$, where $M(B_n) = M(A_n)$ and thus only the reference-dependent part differs. Generally, $[B_n|G_{\langle A_n|G'\rangle}]\,|[0]$ contains all components that occur in $B_n|G$ with higher probability than in $A_n|G'$. Similarly, $[A_n|G'_{\langle B_n|G\rangle}]\,|[0]$ only contains those components that have a higher probability in $A_n|G'$ than in $B_n|G$. We derive the densities of the residual gambles by applying the following procedure. The probability density function of $[B_n|G_{\langle A_n|G'\rangle}]\,|[0]$ is $f_{BA}(x) = \max\{[f_B(x) - f_A(x)]/f; 0\}$ where $f_B(x)$ is the density of $B_n|G$, $f_A(x)$ is the density of $A_n|G'$, and $f = \int \max\{f_B(x) - f_A(x); 0\}dx$. Similarly, the density of $[A_n|G'_{\langle B_n|G\rangle}]|[0]$ is $f_{AB}(x) = \max\{[f_A(x) - f_B(x)]/f; 0\}$. These transformations ensure that the probability mass of the residual gambles is one without changing relative densities. While such definitions may seem notationally cumbersome they precisely indicate how a residual gamble was derived. A special case arises if both gambles contain the same components with identical probabilities. The relation in the original comparison is then equivalent to the comparison of any arbitrary gamble C with itself, formally $C|[0]$ is compared to $C|[0]$. Any such situation is then characterized by indifference without any further assumptions concerning the functional form of $u(\cdot)$.

2.3 Results

Our first result states that preferences are monotone. Note that the component of gain-loss utility does not influence this preference in any way, so the result is solely driven by pure consumption utility. Intuitively, in risk-less decisions, no feelings of gain or loss can arise when the outcome was expected.

Proposition 2.1 *If preferences depend on reference points formed by expectations, individuals have monotone preferences, that is* $B_1|B_1 \succ A_1|A_1$.

In contrast to Proposition 2.1, our results for all orders higher than one will be driven solely by the gain-loss utility component of overall decision utility (as formally expressed in Lemma 2.16 in the appendix). Intuitively, because of the linearity of $m(\cdot)$ the expectational mean of two options is all that matters in terms of pure consumption utility, and it is easily verified that $\mathbb{E}[B_n] = \mathbb{E}[A_n]$ for all $n \geq 2$ holds.

Our second result states the sufficient and necessary condition for second-order risk preferences. It replicates findings of the existing literature.

Proposition 2.2 *If preferences depend on reference points formed by expectations, individuals are risk-averse (risk-seeking, risk-neutral), that is* $B_2|B_2 \succ (\prec, \sim)\ A_2|A_2$, *if and only if* $[0]|[0] \succ (\prec, \sim)\ [\tilde{\varepsilon}_1 + \tilde{\varepsilon}_1']|[0]$.

Proposition 2.2 specifies risk aversion in a framework of expectational references. Since in the comparison of $B_2|B_2$ vs. $A_2|A_2$ densities of similar elements can be reduced as described in the previous section, we can derive the last equivalence. Thus, $[B_2|B_{2\langle A_2|A_2\rangle}]|[0]$ equals $[0]|[0]$ and $[A_2|A_{2\langle B_2|B_2\rangle}|[0]$ is equal to $[\tilde{\varepsilon}_1 + \tilde{\varepsilon}_1']|[0]$. Our result resembles the classical definition of risk aversion if the choice is transformed such that the hypothetical reference point is $[0]$. An individual who dislikes the convolution of any symmetric and zero-mean $\tilde{\varepsilon}_1$ is classified as being risk-averse.

As an illustrative example, suppose $\tilde{\varepsilon}_1 = [-\varepsilon_1; \varepsilon_1]$ with $\varepsilon_1 > 0$. Then, if an individual chooses the gamble B_2 and this is also her expectation, the choice would result in the utility of $M([y]) + U([0]|[0])$ since the deterministic outcome of y is always correctly anticipated. However, if this individual chooses the gamble A_2 and this is also her expectation, four combinations of outcomes and reference points will be considered. With probability $1/2$ the outcome is $y + \varepsilon_1$. But only in half of these cases this outcome was also anticipated and reduces to 0. In the other half of these cases $y - \varepsilon_1$ was expected instead, which yields the sensation of a gain of $2\varepsilon_1$. Also, with probability $1/2$ the outcome is $y - \varepsilon_1$. In half of these cases this bad outcome was anticipated and reduces to 0. In the other half of these cases the good outcome $y + \varepsilon_1$ was anticipated, yielding the sensation of a loss of $2\varepsilon_1$. Thus, the choice of A_2 would deliver a utility of $M([y-\varepsilon_1; y+\varepsilon_1]) + U([0;0;-2\varepsilon_1;2\varepsilon_1]||[0])$. Since for a linear m function it follows that $M([y]) = M([y-\varepsilon_1; y+\varepsilon_1])$, this individual prefers B_2 over A_2 if and only if $u(0) > \frac{1}{2}u(0) + \frac{1}{4}u(-2\varepsilon_1) + \frac{1}{4}u(2\varepsilon_1)$. This can be further simplified to the condition $u(0) > \frac{1}{2}u(-2\varepsilon_1) + \frac{1}{2}u(2\varepsilon_1)$. In gamble notation, $B_2|B_2 = [0]||[0] \succ [0;0;-2\varepsilon_1;2\varepsilon_1]||[0] = A_2|A_2$ reduces to $[B_2|B_{2\langle A_2|A_2\rangle}]||[0] = [0]||[0] \succ [2\tilde{\varepsilon}_1]||[0] = [A_2|A_{2\langle B_2|B_2\rangle}]||[0]$.

It directly follows from Proposition 2.2 that for u functions which are usually proposed in the literature individuals are risk-averse.

Corollary 2.3 *If preferences depend on reference points formed by expectations, individuals with*

(i) $u''(\cdot) < (>,=)\ 0$ *over the whole range are risk-averse (risk-seeking, risk-neutral);*

(ii) loss-averse S-shaped $u(\cdot)$ *are risk-averse;*

(iii) $u(\cdot)$ *being (two-fold rotational) symmetric around the reference point are risk-neutral.*

The next result is concerned with third-order risk preferences.

Proposition 2.4 *If preferences depend on reference points formed by expectations, individuals are never prudent or imprudent, that is* $B_3|B_3 \sim A_3|A_3$*. This holds for all formulations of* $u(\cdot)$*.*

Proposition 2.4 constitutes a crucial result of our analysis. Since $B_3|B_3$ always contains the same elements with identical probabilities as $A_3|A_3$, both choices are equally valued regardless of the shape of $u(\cdot)$. Therefore, reference-dependent utility does not contribute to third-order risk preferences despite its effect on the second order.

For an illustration consider again an example where $\tilde{\varepsilon}_1 = [-\varepsilon_1; \varepsilon_1]$. If the choice of an individual is B_3 and she was expecting this choice, her utility will be determined by comparing every possible outcome to every possible reference point. Thus, $U(B_3|B_3) = \frac{3}{8}u(0) + \frac{1}{8}u(-k-\varepsilon_1) + \frac{1}{8}u(-k+\varepsilon_1) + \frac{1}{8}u(k-\varepsilon_1) + \frac{1}{8}u(k+\varepsilon_1) + \frac{1}{16}u(-2\varepsilon_1) + \frac{1}{16}u(2\varepsilon_1)$. However, if the individual chooses A_3 and she was expecting this to be her choice, her utility will be identical to the expression above and hence $U(B_3|B_3) = U(A_3|A_3)$. Since also $M(B_3) = M(A_3)$ holds, we can conclude that $V(B_3|B_3) = V(A_3|A_3)$ for all functional forms of $u(\cdot)$.

The result of Proposition 2.4 is in stark contrast to the usual properties of models in EUT. As an example, functional forms exhibiting constant or decreasing absolute risk aversion imply prudence in EUT. Within EUT only rarely used utility functions, like those with quadratic utility, never exhibit prudence or imprudence. In contrast, assuming dependence on expectational references eliminates any third-order risk preference. Since this holds for arbitrary u functions, Proposition 2.4 has very general implications that are discussed in Section 2.4. We will show that under reference dependence precautionary saving is lower and insurance demand higher than in a classical EUT model of pure consumption utility.

In the next proposition we consider fourth-order risk preferences.

Proposition 2.5 *If preferences depend on reference points formed by expectations, individuals are temperate (intemperate, neither nor), that is* $B_4|B_4 \succ (\prec, \sim)\ A_4|A_4$*, if and only if* $[\tilde{\varepsilon}_1 + \tilde{\varepsilon}_1'; \tilde{\varepsilon}_2 + \tilde{\varepsilon}_2']|[0] \succ (\prec, \sim)\ [0; \tilde{\varepsilon}_1 + \tilde{\varepsilon}_1' + \tilde{\varepsilon}_2 + \tilde{\varepsilon}_2']|[0]$.

Proposition 2.5 specifies temperance when preferences depend on expectational references. It is similar to Proposition 2.2 in that it resembles the general definition of temperance in case the choice is transformed such that the reference point is [0]. Since densities of identical elements in $B_4|B_4$ and $A_4|A_4$ can be reduced in comparison, $[B_4|B_{4\langle A_4|A_4\rangle}]|[0]$ equals $[\tilde{\varepsilon}_1 + \tilde{\varepsilon}_1'; \tilde{\varepsilon}_2 + \tilde{\varepsilon}_2']|[0]$ and $[A_4|A_{4\langle B_4|B_4\rangle}]|[0]$ is equal to $[0; \tilde{\varepsilon}_1 + \tilde{\varepsilon}_1' + \tilde{\varepsilon}_2 + \tilde{\varepsilon}_2']|[0]$.

Unlike in Propositions 2.2 and 2.4, fourth-order risk preferences differ for various shapes of $u(\cdot)$ that have been proposed in the literature.

Corollary 2.6 *If preferences depend on reference points formed by expectations, individuals with*

(i) $u''''(\cdot) < (>, =)\ 0$ *over the whole range are temperate (intemperate, neither nor);*

(ii) loss-averse S-shaped $u(\cdot)$ *are intemperate;*

(iii) $u(\cdot)$ *being (two-fold rotational) symmetric around the reference point are neither temperate nor intemperate.*

As case *(ii)* of Corollary 2.6 is the standard assumption in the literature on reference dependence, we find a clear difference in the predictions of standard EUT with commonly used utility functions and loss-averse models concerning temperance. The implication of Proposition 2.5, that typical loss-averse u functions always induce intemperance, is in line with Kőszegi and Rabin (2007, p1060) who show that in their model with a piecewise linear loss-averse u function and a linear m function "[...] behavior also approaches risk neutrality with even moderate amounts of background risk."[14]

[14] Kőszegi and Rabin (2007, p1052) "[...] identify common ways in which the decision maker becomes less risk-averse if she had been expecting, or is now facing, more risk." This is in line with our

To receive some intuition for case *(ii)* of Corollary 2.6 consider the example where $\tilde{\varepsilon}_1 = [-\varepsilon_1; \varepsilon_1]$, $\tilde{\varepsilon}_2 = [-\varepsilon_2; \varepsilon_2]$ and $\varepsilon_1 = \varepsilon_2$. Then, if the choice is B_4 and this was also expected, the individual will experience with probability 1/4 a gain of $2\varepsilon_1$ and with probability 1/4 a loss of the same size. With probability 1/2 she will experience neither a gain nor a loss. If the individual chooses A_4 instead and also expected this choice, she will face a gain of $2\varepsilon_1$ with probability 1/8 and a gain of $4\varepsilon_1$ with probability 1/32. However, with probability 1/8 she will also face a loss of $2\varepsilon_1$ and with probability 1/32 a loss of $4\varepsilon_1$. With probability 11/16 she will experience neither a gain nor a loss. Note that choosing B_4 results in feelings of loss with a higher probability than choosing A_4. At a first glance, A_4 may still seem less attractive because the maximal potential loss is higher. However, because of the weakly decreasing sensitivity in losses, a loss of $4\varepsilon_1$ can have at most twice the effect as a loss of $2\varepsilon_1$. Hence, the danger of a larger loss with A_4 is always overcompensated by the smaller probability of experiencing a loss.

Propositions 2.4 and 2.5 are special cases of the following theorem that identifies risk preferences of arbitrary order n when reference points are formed by expectations.

Theorem 2.7 *If preferences depend on reference points formed by expectations, individuals are,*

(i) for even orders ($\frac{n}{2} \in \mathbb{N}$) and $n \geq 4$, risk apportioning of order n (its opposite, neither nor), that is $B_n|B_n \succ (\prec, \sim)\ A_n|A_n$, if and only if
$$[B_n|B_{n\langle A_n|A_n\rangle}]|[0] = \left[A_{n-2}|A_{n-2\langle B_{n-2}|B_{n-2}\rangle}; B_{n-2}|B_{n-2\langle A_{n-2}|A_{n-2}\rangle} + \tilde{\varepsilon}_{(\frac{n}{2})} + \tilde{\varepsilon}'_{(\frac{n}{2})}\right]|[0]$$
$$\succ (\prec, \sim)$$
$$\left[B_{n-2}|B_{n-2\langle A_{n-2}|A_{n-2}\rangle}; A_{n-2}|A_{n-2\langle B_{n-2}|B_{n-2}\rangle} + \tilde{\varepsilon}_{(\frac{n}{2})} + \tilde{\varepsilon}'_{(\frac{n}{2})}\right]|[0] = [A_n|A_{n\langle B_n|B_n\rangle}]|[0];$$

(ii) for odd orders ($\frac{n}{2} \notin \mathbb{N}$) and $n \geq 3$, never risk apportioning of order n or its opposite, that is $B_n|B_n \sim A_n|A_n$. This holds for all formulations of $u(\cdot)$.

In order to get an intuition for the general structure of the results of Propositions 2.1 to 2.5 and Theorem 2.7, consider first the case where $n = 1$. If a certain outcome is expected, it will not generate any feelings of gains or losses. Hence, any first-order effect is solely driven by the difference in pure consumption utility. However, for $n \geq 2$ the expected value of the pure consumption utility of B_n and A_n is identical because of the

results not only for surprise situations (comparing their proposition 1 with our Table 2.3 in Section 2.5.3), but also for expectational references. More specifically, in their proof of proposition 5 (p1068) they show that, for some lotteries G and F, $U(F|F) \leq U(G+F|G+F) - U(G|G)$ holds. Rewriting this inequality as $U(F|F) + U(G|G) \leq U(G+F|G+F)$ offers the direct interpretation that a person satisfying this inequality would rather aggregate than disaggregate the two lotteries. This is exactly the interpretation of intemperance. Moreover, by applying $F = y + \tilde{\varepsilon}_1$, $G = y + \tilde{\varepsilon}_2$, and $u(0) = 0$ it can be shown that this inequality corresponds to (weakly) choosing $[A_4|A_{4\langle B_4|B_4\rangle}]|[0]$ over $[B_4|B_{4\langle A_4|A_4\rangle}]|[0]$.

linearity of $m(\cdot)$. Therefore, for $n \geq 2$ only the gain-loss component of utility determines the assessment of the two gambles. With $n = 2$ (see Proposition 2.2), expecting a risk does not eliminate the risk in outcomes. This is because both random variables are realized independently. Thus, it is possible that a high expectation coincides with a low outcome (generating a sensation of loss) or that a low expectation is exceeded by a high outcome (generating a sensation of gain). In contrast to the second order, for $n = 3$ the question is not whether to take or avoid a risk but rather whether to add an unavoidable risk in a high- or low-wealth state. Since the decision is anticipated, location is irrelevant for feelings of gains and losses. Therefore, individuals do not care where to add the risk. Unlike the third order, with $n = 4$ the question is not whether to add risk either to a high or to a low state, but rather whether to add it to a certain or uncertain state. Generally, aggregating or disaggregating risk(s) affects gain-loss utility and this is captured by even-order risk preferences. By contrast, odd-order risk preferences of orders higher than one can always be thought of as a decision on the location of risk(s). With expectational references the choice of location becomes irrelevant.

More formally, for even orders $-B_n \succ (\prec, \sim) \ -A_n$ if and only if $B_n \succ (\prec, \sim) \ A_n$ since all $\tilde{\varepsilon}_i$ are symmetric and there is no reduction of k. Thus, an individual prefers $B_n|B_n$ over $A_n|A_n$ if and only if she is risk apportioning of order n. For odd orders $\tilde{\varepsilon}_i$ are still symmetric, but there is now always a reduction of k which reverses into an additional k when B_n and A_n become negative. Thus, we get that $-B_n \succ (\prec, \sim) \ -A_n$ if and only if $B_n \prec (\succ, \sim) \ A_n$. Because the two effects with expectational references are of equal magnitude and opposing direction they cancel out and an individual is therefore always indifferent between $B_n|B_n$ and $A_n|A_n$ for odd orders.

While risk preferences of even orders with expectational references resemble those without reference dependence for a given functional form of $u(\cdot)$, they substantially differ for odd orders. In fact, risk preferences of odd orders are completely absent with expectational references. In EUT risk preferences of any order n depend on those of order $n-1$ and are therefore strongly related. If the $(n-1)^{th}$ derivative of $u(\cdot)$ is known, the n^{th} derivative can be directly computed. In contrast, Theorem 2.7 shows that with reference dependence this is no longer the case. Especially when the evaluation of an odd order delivered indifference, one cannot conclude that the next-higher order delivers indifference as well. This is of general importance for the empirical derivation of higher-order risk preferences via measures of risk aversion. We discuss such implications of Theorem 2.7 for the relationship between risk aversion and precautionary saving in Section 2.4.

Kőszegi and Rabin (2007) describe in which circumstances reference points that are formed by expectations seem appropriate. These can be separated into two broad categories. First, some decisions are anticipated well before they have to be made. Then, the individual also anticipates how she will decide. Therefore, at the time the decision is made the individual has already expected her actual choice. Consistent choices in these situations are *preferred personal equilibria* (PPE) as defined by the authors. An example for an economically interesting scenario would be the decision over savings. The individual may think about saving well before doing so and then has already incorporated this behavior into her reference point when conducting the transaction. Second, some decisions are real choices that cannot be reversed but whose actual outcome is realized much later. Here, individuals' references have time to adjust to the choice already made. Decision-makers anticipate the adjustment before deciding. According to Kőszegi and Rabin (2007) all choices are then *choice-acclimating personal equilibria* (CPE). This seems relevant in situations where one can commit to choices well in advance of events. Interesting examples include many insurance and pension decisions. Here, contracts are often binding and are agreed upon years before the risks are realized.

Both scenarios are analytically identical since the reference distribution is the same as the chosen outcome distribution. Thus, our results are applicable to both equlibrium concepts. However, the two concepts differ with respect to the existence of such personal equilibria. All choices are equilibria if commitment is possible and CPE is the relevant concept.[15] By contrast, if commitment is not possible but decisions have been anticipated not all decisions can rationally be expected and constitute a PPE. In these cases equilibria existence is an important issue because rational agents cannot trick themselves to expect a later choice they have no incentive to follow. In other words, if the agent wanted to deviate from her expectation once she makes the decision, this expectation could not sustain in the first place. All choices that can be rationally expected are defined as *unacclimated personal equilibria* (UPE) as introduced by Kőszegi and Rabin (2006, 2007). The UPE that maximize ex-ante decision utility $V(\cdot)$ are PPE and are therefore the chosen alternatives. In our context this translates into the following existence condition

$$\begin{aligned} B_n|B_n \text{ is an UPE} \quad &\Leftrightarrow \quad B_n|B_n \succsim A_n|B_n \quad \text{and} \\ A_n|A_n \text{ is an UPE} \quad &\Leftrightarrow \quad A_n|A_n \succsim B_n|A_n. \end{aligned}$$

[15] Recent experiments by Abeler et al. (2011) and Gill and Prowse (forthcoming) support CPE as the relevant concept. The latter even find that reference points adjust essentially instantaneously to choices made.

2.4 Implications

Results of the previous section may have strong implications in situations where decisions under uncertainty cannot solely be explained via risk aversion. Then, higher-order risk preferences are needed to explain phenomena like precautionary saving, precautionary labor, or the influence of background risk. Until now, in attempts to explain these phenomena only the effect of pure consumption utility in EUT models has been considered. Applying the preceding analysis allows us to investigate the influence of reference dependence in such settings. In the following, we will discuss two implications in more detail, the optimal amount of precautionary saving and the interaction of precautionary saving and insurance demand.

Often quantitative measures of one order were used to predict behavior concerning a higher order. For example, measures of risk aversion were thought to determine exactly the optimal amount of precautionary saving because both were solely derived from the functional form of consumption utility. As reported by Browning and Lusardi (1996) or Carroll and Kimball (2008) for instance, precautionary saving is often empirically observed but to a lesser extent than thought to be optimal. "Structural models that match broad features of consumption and saving behavior tend to produce estimates of the degree of prudence that are less than those obtained from theoretical models in combination with risk aversion estimates from survey evidence." (Carroll and Kimball, 2008, p584). Also, Ballinger, Palumbo, and Wilcox (2003) found in a laboratory experiment simulating life-cycle decisions that precautionary saving falls short of optimal theoretical predictions. When utility is partly dependent on expectational references, the reference-dependent component contributes positively to risk aversion but leaves prudence unaffected. Therefore, such a reference-dependent model would predict less precautionary saving than the standard model. This is consistent with the empirical and experimental literature.

Closely related and building on the well-known finding that saving and insurance are substitutes[16] is what Gollier (2003) called the 'insurance puzzle'. When financial markets are complete, individuals should prefer to cope with risk via buffer-stock saving rather than by purchasing costly insurance. In other words, allowing for precautionary saving leads theoretically to a crowding out of insurance demand in dynamic life-cycle models. "The bottom line is similar to the one underlying the literature on the equity premium puzzle: the theory cannot easily explain why people are so reluctant to ac-

[16]To our knowledge, the first formal analysis showing that these two are substitutes was by Moffet (1975) within a simple two-period model. For a more general treatment without additive separable utility, refer to Dionne and Eeckhoudt (1984).

cept risk." (Gollier, 2003, p21). Again, assuming preferences that depend partly on expectational references can serve as an alternative explanation to market imperfection. Intuitively, since insurance demand is driven by second-order effects but precautionary saving by third-order effects, the reference-dependent component generates additional insurance demand but does not affect precautionary saving decisions.

To illustrate these implications of our results we analyze a simple two-period setting ($T = 2$) without discounting and interest on savings. As *benchmark* model we consider a standard EUT model, where individuals maximize expected life-time utility $V^b = \sum_{t=1}^{T} V_t^b$ with $V_t^b = \int c(x_t)dF(x_t)$. x_t is consumption in period t and $c(\cdot)$ denotes pure consumption utility that is assumed to be strictly increasing and strictly concave.

To facilitate comparisons between implications of such a benchmark model and a reference-dependent model we specify the latter as a *hybrid* model. This hybrid model slightly deviates from the reference-dependent model before in that the pure consumption utility part has now the same properties as our standard EUT benchmark model while the gain-loss part stays unchanged and still has the properties of the previous analysis. More formally, individuals maximize $V^h = \sum_{t=1}^{T} V_t^h$ with $V_t^h = \int c(x_t)dF(x_t) + \iint u(x_t - r_t)dG(r_t)dF(x_t)$, where r_t denotes the reference point in t. The reason why we allow the pure consumption utility part to be concave in wealth in the hybrid model is that this leads to an equivalence of both models in risk-less situations. Within both models we will compare behavior in risky and risk-less situations. As we are interested in the implications of reference dependence for risk preferences we would like to keep behavior in risk-less situations constant between the two models. In order to isolate the pure effects of reference dependence and to avoid reintroducing classical EUT effects into gain-loss utility we still assume that the gain-loss part depends on absolute differences of outcomes and reference points.

Consumption in period 1 is income y which is potentially reduced by saving k^s and an insurance premium. The insurance premium consists of the premium of an actuarially fair insurance k^i plus proportional loading λ. Thus, the final consumption in period 1 is $x_1 = y - k^s - (1+\lambda)k^i$ and the distribution representing the beliefs over x_1 is simply $F(x_1) = [y - k^s - (1+\lambda)k^i]$. Consumption in period 2 is the same income y plus a random component $\tilde{\varepsilon}_1$ whose two realizations occur with equal probability, $\tilde{\varepsilon}_1 = [-\varepsilon_1; \varepsilon_1]$. Savings k^s increase outcomes in both states but insurance only in the bad state. Since the premium of an actuarially fair insurance was defined as k^i, the amount of indemnity is $2k^i$ when the probability of the bad state is $1/2$. Therefore, x_2 can take on two different values, $\underline{x}_2 = y - \varepsilon_1 + k^s + 2k^i$ and $\overline{x}_2 = y + \varepsilon_1 + k^s$. The beliefs over x_2 are then the distribution $F(x_2) = [\underline{x}_2; \overline{x}_2]$. In the following, we will restrict insurance to have either partial or full indemnity but rule out over-insurance, hence $k^i \leq \varepsilon_1$, which also implies $\underline{x}_2 \leq \overline{x}_2$.

To highlight the implications for the optimal amount of precautionary saving, consider first the case where no insurance is available, thus $k^i = 0$. In the benchmark case

$$V^b = c(y - k^s) + \tfrac{1}{2}c(y - \varepsilon_1 + k^s) + \tfrac{1}{2}c(y + \varepsilon_1 + k^s), \tag{2.5}$$

where the first term in (2.5) denotes consumption utility in period 1 and the last two terms are expected consumption utility in period 2. In the hybrid model

$$V^h = c(y - k^s) + \tfrac{1}{2}c(y - \varepsilon_1 + k^s) + \tfrac{1}{2}c(y + \varepsilon_1 + k^s) + \tfrac{1}{4}u(-2\varepsilon_1) + \tfrac{1}{4}u(2\varepsilon_1). \tag{2.6}$$

Again, the first three terms in (2.6) denote pure consumption utility in periods 1 and 2. There is no gain-loss utility in period 1 since there is no uncertainty resolved. Gain-loss utility in period 2 is captured by the last two terms. Since choosing the optimal k^s in (2.5) and (2.6) means solving identical first-order conditions we obtain the result that the decision for precautionary saving is not affected by the reference-dependent part of utility $V^h(\cdot)$. If preferences are reference dependent instead of reference independent, the extent of risk aversion will be higher[17] as long as the slope of $u(\cdot)$ is steeper in losses than in corresponding gains ($u'(-\tau) > u'(\tau)\ \forall \tau > 0$) which is met for strictly concave as well as loss-averse S-shaped u functions. But at the same time an identical amount of precautionary saving would be optimal. This shows that seemingly sub-optimal levels of precautionary saving may well be optimal if preferences are reference dependent. In contrast to classical models of EUT, this result shows that predicted behavior under reference dependence is in line with the empirical findings.

Only at first glance this results seems to be in contrast to a result of Kőszegi and Rabin (2009). In their proposition 8 they show that a piecewise linear gain-loss utility contributes to precautionary saving. However, their result rests on the assumption of concave consumption utility entering gain-loss utility and thereby reintroducing classical EUT effects into the gain-loss part. Indeed, with such a specification it can be shown that optimal precautionary saving is higher under reference dependence compared to the benchmark case as long as the slope of $u(\cdot)$ is steeper in losses than in corresponding gains ($u'(-\tau) > u'(\tau)\ \forall \tau > 0$).[18] However, since reference dependence increases risk

[17] For example, it can be easily verified that the risk premium of an individual with $V^h(\cdot)$ is higher than with $V^b(\cdot)$.

[18] At the optimal level of precautionary saving in the benchmark case $\partial V^h/\partial k^s > 0$ since the additional terms in the first-order condition of the hybrid model compared to the benchmark case are

aversion with this specification as well, this does neither contradict our result nor the empirical pattern observed.[19]

We next analyze precautionary saving and insurance jointly. Our analysis encompasses the case where insurance is costly. That is, it may have a positive loading factor λ. However, in the following we restrict λ to the more realistic cases where it is below 1, hence $0 \leq \lambda < 1$. Then, in the benchmark model

$$\begin{aligned} V^b &= c(y - k^s - (1+\lambda)k^i) + \tfrac{1}{2}c(y - \varepsilon_1 + k^s + 2k^i) + \tfrac{1}{2}c(y + \varepsilon_1 + k^s), \\ \partial V^b/\partial k^s &= -c'(y - k^s - (1+\lambda)k^i) + \tfrac{1}{2}c'(y - \varepsilon_1 + k^s + 2k^i) \\ &\quad + \tfrac{1}{2}c'(y + \varepsilon_1 + k^s) = 0, \qquad (2.7) \\ \partial V^b/\partial k^i &= -(1+\lambda)c'(y - k^s - (1+\lambda)k^i) + c'(y - \varepsilon_1 + k^s + 2k^i) = 0. \qquad (2.8) \end{aligned}$$

As shown in the appendix all second partial derivatives are strictly negative and the Hessian is negative definite. We denote the solution of (2.7) and (2.8) by $(\hat{k}^s, \hat{k}^i)$.

By combining (2.7) and (2.8) we get the following condition:

$$(1-\lambda)c'(y - \varepsilon_1 + \hat{k}^s + 2\hat{k}^i) = (1+\lambda)c'(y + \varepsilon_1 + \hat{k}^s). \qquad (2.9)$$

Although (2.9) cannot be solved explicitly, due to the strict concavity of $c(\cdot)$ we receive that $\hat{k}^i < \varepsilon_1$ for $\lambda > 0$ and $\hat{k}^i = \varepsilon_1$ for $\lambda = 0$.

Gollier (2003) simulated a life-cycle model for an infinite time horizon and found that demand for costly insurance completely vanishes with the opportunity of precautionary savings. He termed this result the 'insurance puzzle' because it is clearly at odds with empirical observations. His explanation for this puzzle rests on market imperfections. In the following, we show that it can also be explained by reference-dependent preferences, even if markets are complete. This is because saving and insurance are

$\frac{1}{4}u'[c(y+\epsilon_1+k^s) - c(y-\epsilon_1+k^s)][c'(y+\epsilon_1+k^s) - c'(y-\epsilon_1+k^s)] + \frac{1}{4}u'[c(y-\epsilon_1+k^s) - c(y+\epsilon_1+k^s)][c'(y-\epsilon_1+k^s) - c'(y+\epsilon_1+k^s)]$.

[19] Among the studies reviewed by Browning and Lusardi (1996) there are even numerous articles where no precautionary saving was measured at all. Another recent example is Jappelli, Padula, and Pistaferri (2008). Lee and Sawada (2007, p196) stress that "while a growing number of theoretical studies point out the importance of precautionary saving, the existing evidence suggests that precautionary saving motives may not be empirically important." Such empirical findings even point toward a reference-dependent model with linear consumption utility in both the reference-dependent and the reference-independent part. This is the model we analyzed in the previous section.

weaker substitutes under reference dependence than in the benchmark model. Consider our hybrid model where

$$\begin{aligned}
V^h &= c(y-k^s-(1+\lambda)k^i)+\tfrac{1}{2}c(y-\varepsilon_1+k^s+2k^i)+\tfrac{1}{2}c(y+\varepsilon_1+k^s) \\
&\quad +\tfrac{1}{4}u(-2\varepsilon_1+2k^i)+\tfrac{1}{4}u(2\varepsilon_1-2k^i), \\
\partial V^h/\partial k^s &= -c'(y-k^s-(1+\lambda)k^i)+\tfrac{1}{2}c'(y-\varepsilon_1+k^s+2k^i) \\
&\quad +\tfrac{1}{2}c'(y+\varepsilon_1+k^s)=0, \qquad (2.10)\\
\partial V^h/\partial k^i &= -(1+\lambda)c'(y-k^s-(1+\lambda)k^i)+c'(y-\varepsilon_1+k^s+2k^i) \\
&\quad -\tfrac{1}{2}u'(2\varepsilon_1-2k^i)+\tfrac{1}{2}u'(-2\varepsilon_1+2k^i)=0. \qquad (2.11)
\end{aligned}$$

Again, in the appendix it is shown that all elements of the Hessian are strictly negative and that the Hessian is negative definite. This holds for $u(\cdot)$ being (weakly) concave, (piecewise) linear, or (loss-averse) S-shaped. Combining (2.10) and (2.11) yields

$$\begin{aligned}
&(1-\lambda)c'(y-\varepsilon_1+k^s+2k^i) \\
=\ &(1+\lambda)c'(y+\varepsilon_1+k^s)-u'(2\varepsilon_1-2k^i)+u'(-2\varepsilon_1+2k^i). \qquad (2.12)
\end{aligned}$$

Using the assumptions that the slope of $u(\cdot)$ is steeper in losses than in corresponding gains, $k^i \leq \varepsilon_1$, and the strict concavity of $c(\cdot)$, we receive from (2.12) that in the optimum $k^i < \varepsilon_1$ for $\lambda > 0$ and $k^i = \varepsilon_1$ for $\lambda = 0$. In the following, we first analyze the case of positive loading ($\lambda > 0$) and then discuss actuarially fair insurance ($\lambda = 0$).

For a comparison of the solutions of both models consider the optimality conditions (2.10) and (2.11) at $(\hat{k}^s, \hat{k}^i)$ which would be the solution if preferences were not reference dependent. Indeed, $(\hat{k}^s, \hat{k}^i)$ is one possible solution of (2.10) because condition (2.10) equals condition (2.7). Using $u'(-\tau) > u'(\tau)\forall\tau > 0$ and $\hat{k}^i < \varepsilon_1$ we find that $(\hat{k}^s, \hat{k}^i)$ cannot be a solution of (2.11) since $\partial V^h(\hat{k}^s, \hat{k}^i)/\partial k^i > 0$. Our goal is to show that the optimal demand for insurance is higher and optimal saving is lower with reference-dependent preferences than without.[20] We denote by (dk^s, dk^i) the amounts that the optimal solution of (2.10) and (2.11) deviates from $(\hat{k}^s, \hat{k}^i)$.

As mentioned above, condition (2.10) is met at $(\hat{k}^s, \hat{k}^i)$. This implies that any change in k^s or k^i has to be compensated by a change of the other variable. By totally differentiating (2.10) we establish the explicit relationship between these changes as

$$dk^s = -\left(\frac{\partial^2 V^h}{\partial k^s \partial k^i}\Big/\frac{\partial^2 V^h}{(\partial k^s)^2}\right)dk^i. \qquad (2.13)$$

[20]There is no reason for a change in saving other than for precautionary motives since gain-loss utility vanishes without uncertainty. All changes in saving are therefore changes in precautionary saving.

We show in the appendix that the term in brackets is positive. Therefore, in order to fulfill condition (2.10) dk^s and dk^i need to have opposite signs as long as they are not both zero.

Our analysis of (2.11) showed that $\partial V^h(\hat{k}^s, \hat{k}^i)/\partial k^i > 0$. Hence, in the hybrid model $\partial V^h/\partial k^i$ has to be lower in the optimum than at $(\hat{k}^s, \hat{k}^i)$. Both an increase in k^s or k^i would decrease $\partial V^h/\partial k^i$, but (2.13) showed that their changes must have opposing signs. The overall effect of an increase in one variable and a decrease in the other has to be negative,

$$d\left(\frac{\partial V^h}{\partial k^i}\right) = \frac{\partial^2 V^h}{\partial k^s \partial k^i} dk^s + \frac{\partial^2 V^h}{(\partial k^i)^2} dk^i < 0. \tag{2.14}$$

In order to show that only an increase in k^i and a decrease in k^s can accomplish this we insert (2.13) in (2.14) and receive

$$\frac{dk^i}{\frac{\partial^2 V^h}{(\partial k^s)^2}} \left(\frac{\partial^2 V^h}{(\partial k^i)^2} \frac{\partial^2 V^h}{(\partial k^s)^2} - \left[\frac{\partial^2 V^h}{\partial k^s \partial k^i}\right]^2 \right) < 0. \tag{2.15}$$

In the proof of Proposition 2.8 we show that the term in brackets in (2.15), which is the determinant of the Hessian matrix, is positive and that $\partial^2 V^h/(\partial k^s)^2 < 0$. Hence, it follows that $dk^i > 0$ and in combination with (2.13) that $dk^s < 0$ in the optimum.

Proposition 2.8 *Individuals with preferences depending on reference points formed by expectations have a higher demand for insurance and save less than individuals without reference-dependent preferences.*

Proposition 2.8 holds for all u functions with a slope that is steeper in losses than in corresponding gains ($u'(-\tau) > u'(\tau)\ \forall \tau > 0$) and with sensitivity that is diminishing faster or equally in gains than in losses ($u''(\tau) \geq -u''(-\tau)\ \forall \tau > 0$). As mentioned before, these conditions are met for u functions that are strictly concave or loss-averse S-shaped, including the limiting case of $u(\cdot)$ being piecewise linear with loss aversion. Proposition 2.8 further shows that reference dependence can be an alternative explanation for the insurance puzzle that does not rest on market imperfections. We derived this result for costly insurance, i.e. $0 < \lambda < 1$. For actuarially fair insurance ($\lambda = 0$) it can be easily shown that both models yield the identical result of full insurance and the absence of a precautionary motive for saving.[21]

Although this section focused on implications concerning the third and second order, our results may have implications concerning higher orders as well. As an example, many financial decisions, such as demand for insurance or risky assets, are taken in the

[21] From $\lambda = 0$ it follows that $\hat{k}^i = \varepsilon_1$ and $\hat{k}^s = -\varepsilon_1$. Hence, $V^b = V^h$ in the optimum.

face of exogenous background risk. It is known that the fourth-order concept of temperance is necessary for risk aversion to increase with increasing background risk and all commonly used utility functions in EUT models exhibit this feature. If, by contrast, preferences are reference dependent, our analysis of the previous section showed that despite higher risk aversion, its sensitivity to background risk would be reduced under standard assumptions on gain-loss utility. A formalization of these ideas will be the subject of future work.

2.5 Alternative Models

It is important to analyze whether our results for Kőszegi and Rabin's (2007) model of expectational references can in fact *not also* be derived by using alternative reference points. Only then, we may be able to conclude that it is their concept of expectational references which is not only capable of resolving empirical puzzles concerning first and second orders, but also concerning higher-order risk preferences. As alternative models we consider disappointment models, models of regret, and reference-dependent models with exogenous reference points like the status quo. All these alternative models can be captured by (2.1) as well but are characterized by a different assumption concerning the reference point G.

Rather than using the full outcome distribution representing expectations as the reference point, individuals may also assess choices relative to the mean of these expectations. This is the assumption of disappointment models (Bell, 1985; Loomes and Sugden, 1986), where outcomes above the mean are perceived as elation and outcomes below as disappointment. So, in models of disappointment it is assumed that the reference point is the mean of the chosen alternative's distribution, thus $G = \mathbb{E}[F]$. Assuming such a reference point may seem appropriate in situations where information or framing focuses on the mean outcome. For instance, buyers of bonds may have clear expectations on their mean performance leading to the mean as a natural source of reference for the evaluation of their actual performance.

With disappointment as well as with expectational references, the alternatives not chosen do not have any impact on the final evaluation of a chosen option and thus every alternative's appeal can be described entirely by its own attributes. By contrast, in models of regret sentiments (see Bell, 1982, 1983; and Loomes and Sugden, 1982) possible outcomes of the chosen gamble are compared to possible outcomes of the gambles not chosen. This kind of behavior is motivated in the literature by an ex-ante anticipation of ex-post utility that is driven by regret (and rejoice) feelings. With the notion of regret employed in the following analysis all combinations of outcomes are evaluated by a regret function and weighted by their joint probabilities. This coincides with the initial models of Bell (1982, 1983) and Loomes and Sugden (1982) in case the individual anticipates prior to her decision that she will finally learn the true outcome that would have resulted had she decided differently than she has. This qualification is a direct consequence of the emphasis these models put on the resolution of foregone alternatives for the presence of regret feelings. In order for our representation (2.1) to be in line with these regret models an assumption needed is that stochastic components of alternatives are realized independently.[22] Since in models of regret the alternative *not* chosen is the relevant reference distribution and since in the definitions of higher orders only pairwise choices were applied, the relevant reference distribution under regret is unambiguous in our context. If we denote the distribution of the alternative not chosen by F', models of regret assume $G = F'$.

Although the theoretical literature has advanced toward endogenous reference points like regret, disappointment, and most recently, expectational references, first models of reference-dependent preferences (see Kahneman and Tversky, 1979) started with exogenous reference points and they still seem to be relevant for specific decision scenarios. For instance, in surprise situations, where it was not anticipated that a decision has to be made, the status quo may seem to be a relevant state of comparison. Also, in experiments subjects may not anticipate what will happen in the experiment and are unlikely to endogenize the rules of the game in the short time until they have to make their decisions. This is, however, not a necessary experimental feature. If the reference point is meant to be endogenous, this could be achieved by careful experimental design and a longer time horizon. With exogenous reference points, G can be chosen arbitrarily but we will focus on the most relevant ones in our analysis.

The following definition specifies how the reference point is formed in these alternative models and therefore complements (2.1) and Definition 2.1 for the analysis of this section.

Definition 2.3 *If the reference point is determined (i) by disappointment,* $G = \mathbb{E}[B_n]$ *and* $G' = \mathbb{E}[A_n]$*; (ii) by regret,* $G = A_n$ *and* $G' = B_n$*; and if the reference point is given (iii) exogenously, then* $G = G'$*.*

[22] For this reason our results do not carry over to cases where the outcome of the chosen alternative determines the counterfactual outcome (e.g. the toss of a coin). Nevertheless, despite our qualifications of stochastic independence and full resolution of information we perceive the remaining field of applicability as substantial. In particular, decisions made in a market environment are often characterized by independence and full information disclosure (e.g. the decision to buy a particular financial asset or a treasury bond).

If the reference point is given exogenously (case *(iii)* in Definition 2.3) and is stochastic rather than deterministic, our analysis is formally very similar to classical models of background risk (Ross, 1981; Kihlstrom, Romer, and Williams, 1981). In these models, preferences toward risk are analyzed in the presence of a given independent background risk. A stochastic, exogenous reference point is analytically equivalent to a background risk since it enters additively into the u function. This analogy does not carry over to endogenous reference points (Definition 2.2 and cases *(i)* and *(ii)* in Definition 2.3). Here, the actual choice influences the reference point. Thus, the reference point entering $u(\cdot)$ is in these cases not independent of the choice taken. Despite the fact that once the reference point is determined it is as if realized independently of the outcome of the choice, the ex-ante distribution of the reference point depends on the actual choice.

2.5.1 Disappointment

Just as in the case of expectational references, disappointment references are in principle endogenous since they depend on the chosen alternative. However, in our analysis for $n \geq 2$ the expectational mean of B_n is always identical to the expectational mean of A_n and it can therefore also be considered as exogenous. For the same reason, the results of this section also apply to models of regret in which the reference point is the expectational mean of the foregone alternative. This is the usual assumption in regret models where the foregone alternative is not realized or the individual does not receive feedback on its outcomes (see Bell, 1983; or Krähmer and Stone, 2008).[23] In the following, we report results for risk preferences when reference points are formed by the expectational mean. For the gambles used in our analysis the mean is y for even orders and $y - \frac{k}{2}$ for odd orders, except order one. For the first order the mean is identical to the full distribution since everything is certain and there is no risk, hence $B_1 = \mathbb{E}[B_1] = y$ and $A_1 = \mathbb{E}[A_1] = y - k$. We therefore receive for the first order the exact same result as with expectational references, namely that individuals with disappointment references have monotone preferences. Proposition 2.9 summarizes our results with disappointment references for all orders.

Proposition 2.9 *If preferences depend on reference points formed by the expectational mean, individuals*

(i) have monotone preferences, that is $B_1|\mathbb{E}[B_1] \succ A_1|\mathbb{E}[A_1]$*;*

[23]Note that in these models it is assumed that $U(F|G) = \int u(m(x) - \int m(r)dG(r))dF(x)$. This is also the assumption of the disappointment model of Loomes and Sugden (1986). However, with $m(\cdot)$ being linear, which is assumed by Loomes and Sugden (1986) as well, this is equivalent to our formulation and that of Bell (1985) where $U(F|G) = \int u(m(x) - m(\int r\ dG(r)))dF(x)$.

(ii) *are risk-averse (risk-seeking, risk-neutral), that is* $B_2|\mathbb{E}[B_2] \succ (\prec, \sim)\ A_2|\mathbb{E}[A_2]$*, if and only if* $[0]||[0] \succ (\prec, \sim)\ [\tilde{\varepsilon}_1]||[0]$*;*

(iii) *are prudent (imprudent, neither nor), that is* $B_3|\mathbb{E}[B_3] \succ (\prec, \sim)\ A_3|\mathbb{E}[A_3]$*, if and only if* $[-\frac{k}{2}; \frac{k}{2} + \tilde{\varepsilon}_1]||[0] \succ (\prec, \sim)\ [-\frac{k}{2} + \tilde{\varepsilon}_1; \frac{k}{2}]||[0]$*;*

(iv) *are (for* $n \geq 4$*) risk apportioning of order* n *(its opposite, neither nor), that is* $B_n|\mathbb{E}[B_n] \succ (\prec, \sim)\ A_n|\mathbb{E}[A_n]$*, if and only if*

$$\begin{aligned} & \left[A_{n-2}|\mathbb{E}[A_{n-2}]; B_{n-2}|\mathbb{E}[B_{n-2}] + \tilde{\varepsilon}_{int(\frac{n}{2})}\right] | [0] \\ \succ (\prec, \sim) \quad & \left[B_{n-2}|\mathbb{E}[B_{n-2}]; A_{n-2}|\mathbb{E}[A_{n-2}] + \tilde{\varepsilon}_{int(\frac{n}{2})}\right] | [0] \quad . \end{aligned}$$

Concerning orders higher than one, the mean is not equal to the full distribution and therefore results generally differ between expectational and disappointment references (cases *(ii)* – *(iv)* in Proposition 2.9). However, for the second order (case *(ii)* in Proposition 2.9) we find that if an individual is classified as being risk-averse (risk-seeking, risk-neutral) with disappointment references and a given u function, she will also be risk-averse (risk-seeking, risk-neutral) with the same u function and expectational references. As with expectational references an individual with disappointment references is therefore risk-averse if $u(\cdot)$ is strictly concave or loss-averse S-shaped and risk-neutral if $u(\cdot)$ is (two-fold rotational) symmetric. Corollary 2.3 also applies to disappointment references.

So, concerning the first two orders expectational and disappointment references yield the same behavior. This is one reason why it has been difficult to discriminate between these two forms of expectations as the relevant reference point in the empirical literature. Neither of the empirical studies reviewed in Section 2.2.2 has been able to distinguish them. Even recent controlled laboratory experiments that directly test expectations as the reference point, like for instance Abeler et al. (2011), Gill and Prowse (forthcoming), or Ericson and Fuster (2010), cannot differentiate between expectational and disappointment references.

However, with respect to the third order disappointment yields a different prediction than expectational references (case *(iii)* in Proposition 2.9).[24]

Corollary 2.10 *If preferences depend on reference points formed by the expectational mean, individuals with*

[24]Note that case *(ii)* in Corollary 2.10 can be further specified: Individuals are imprudent if $u(\cdot)$ is loss-averse S-shaped without the limiting case of piecewise linear, since then individuals like gambles in losses and dislike gambles in gains. If $u(\cdot)$ is piecewise linear with loss aversion, individuals are neither prudent nor imprudent.

(i) $u'''(\cdot) > (<,=)\ 0$ *over the whole range are prudent (imprudent, neither nor);*

(ii) loss-averse S-shaped $u(\cdot)$ *are never prudent.*

This result shows that by using the third order it may well be possible to discriminate between expectational and disappointment references. One advantage of the gamble definitions used to define higher-order risk preferences is that they are easily implemented into choice situations in experiments. Future experiments on expectations as the relevant reference point could therefore use this result in order to be able to distinguish between expectational and disappointment references. Moreover, this result further shows that the implications we derived for expectational references, i.e. the effects on precautionary saving and insurance demand, cannot be derived with disappointment references for all functional forms of $u(\cdot)$, since they heavily relied on the general absence of third-order risk preferences.

There is again no difference in predicted behavior between these two reference points for fourth-order risk preferences. More specifically, if an individual is temperate (intemperate, neither nor) with disappointment references for a given u function, she will also be temperate (intemperate, neither nor) with expectational references under this u function. In fact, individuals with disappointment references are temperate, that is $B_4|\mathbb{E}[B_4]$ is preferred to $A_4|\mathbb{E}[A_4]$ if and only if $[\tilde{\varepsilon}_1;\tilde{\varepsilon}_2]||[0]$ is preferred to $[0;\tilde{\varepsilon}_1+\tilde{\varepsilon}_2]||[0]$. This preference holds if $u(\cdot)$ globally satisfies $u''''(\cdot) < 0$. But individuals are intemperate if $u(\cdot)$ is loss-averse S-shaped. Moreover, individuals are neither temperate nor intemperate for $u(\cdot)$ being (two-fold rotational) symmetric around the reference point. Again, Corollary 2.6 not only holds with expectational but also with disappointment references. So, as in the case of second-order risk preferences, the same behavior is predicted for both reference points using common assumptions on $u(\cdot)$.

Our higher-order results under disappointment (cases *(ii)* – *(iv)* in Proposition 2.9) after our reference-point transformation resemble those of Eeckhoudt and Schlesinger (2006). They further show that common assumptions on $u(\cdot)$ generate contrary predictions on third-order effects compared to both standard EUT and expectational references, and on fourth-order effects compared to EUT. However, just as with EUT and expectational references, they predict risk aversion as well as monotone preferences. Furthermore, similar to our findings for expectational references, the significance of higher-order derivatives no longer applies in general. For strictly concave u functions we still need the third and fourth derivative in order to characterize prudence and temperance respectively. But in cases of $u(\cdot)$ being loss-averse S-shaped we only need the first two derivatives in order to describe the preference relations of those higher orders.

In all models where some form of expectations shape the reference point, we can additionally ask whether these expectations form rationally, because only choices followed through can reasonably be expected in the first place. We provided conditions in Section 2.3 that have to be fulfilled for a choice to be a UPE under expectational references. This is only relevant in decision scenarios where the CPE concept does not apply. If the CPE concept can be applied, any choice is in fact an equilibrium. This also holds for disappointment references. However, if CPE cannot be used, we call the disappointment counterpart of UPE *disappointment equilibria* (DE). The DE that maximize ex-ante decision utility $V(\cdot)$ are the *preferred disappointment equilibria* (PDE). Then,

$$\begin{aligned} B_n|\mathbb{E}[B_n] \text{ is a DE} \quad &\Leftrightarrow \quad B_n|\mathbb{E}[B_n] \succsim A_n|\mathbb{E}[B_n] \quad \text{and} \\ A_n|\mathbb{E}[A_n] \text{ is a DE} \quad &\Leftrightarrow \quad A_n|\mathbb{E}[A_n] \succsim B_n|\mathbb{E}[A_n]. \end{aligned}$$

Since $\mathbb{E}[B_n] = \mathbb{E}[A_n]$ (for $n \geq 2$), we can further say that if and only if $B_n|\mathbb{E}[B_n]$ is strictly preferred to $A_n|\mathbb{E}[A_n]$, choosing B_n is the only DE and thus PDE. Similarly, if and only if $A_n|\mathbb{E}[A_n]$ is strictly preferred to $B_n|\mathbb{E}[B_n]$, choosing A_n is the only DE and thus PDE. This holds for all but the first order. For the first order DE conditions are identical to the UPE conditions of expectational references since $\mathbb{E}[B_1] = B_1$ and $\mathbb{E}[A_1] = A_1$. Here, $B_1|B_1 = B_1|\mathbb{E}[B_1]$ is the only UPE and DE and thus PPE and PDE.

2.5.2 Regret

As already discussed, reference points need not necessarily be formed by the expectation of the choice made but can also be driven by regret feelings. Then, ex-ante expectations of the foregone alternative functions as the reference point for the choice. A prevalent question in the regret literature is how individuals evaluate choices when there are several alternatives initially. This is not discussed in our context because the definition of higher orders applied here is over binary alternatives only. In the following, we derive how regret references affect higher-order risk attitudes.

Proposition 2.11 *If preferences depend on reference points formed by regret, individuals have monotone preferences, that is* $B_1|A_1 \succ A_1|B_1$.

While Proposition 2.11 is a standard result which is satisfied by all other models as well, our next result differs substantially from previous findings.

Proposition 2.12 *If preferences depend on reference points formed by regret, individuals are always risk-neutral, that is $B_2|A_2 \sim A_2|B_2$. This holds for all formulations of $u(\cdot)$.*

This proposition applies to all u functions, since $M(B_2) = M(A_2)$ and $U(B_2|A_2)$ always contains the same elements with identical probabilities as $U(A_2|B_2)$. For instance, consider again the simple case where $\tilde{\varepsilon}_1 = [-\varepsilon_1; \varepsilon_1]$. If the individual chooses the gamble B_2 over A_2 she will receive y with certainty. Her foregone alternative would have given her $y - \varepsilon_1$ in half of the cases and $y + \varepsilon_1$ in the other half of the cases. With regret references, choosing B_2 over A_2 exposes her to the regret of not receiving ε_1 with probability 1/2. But it also gives her with probability 1/2 the chance of rejoicing having not lost the same amount. Therefore, choosing B_2 would give her the utility of $M([y]) + U([-\varepsilon_1; \varepsilon_1])$. By contrast, if the individual chooses A_2 over B_2, with equal probability she will receive $y - \varepsilon_1$ or $y + \varepsilon_1$ and will always compare it to the foregone certain alternative y. Thus, choosing A_2 would give her the utility $M([y - \varepsilon_1; y + \varepsilon_1]) + U([-\varepsilon_1; \varepsilon_1])$. Since $U(B_2|A_2) = U(A_2|B_2) = \frac{1}{2}u(-\varepsilon_1) + \frac{1}{2}u(\varepsilon_1)$ the individual is indifferent between B_2 and A_2 for all possible shapes of $u(\cdot)$.

Proposition 2.12 is a surprising result, so it may require some further explanation. One assumption that leads to this result is that $m(\cdot)$ is linear. Empirical evidence for this assumption in the context of regret can be found in Bleichrodt, Cillo, and Diecidue (2010). However, even if we allow $m(\cdot)$ to be concave or convex instead, it can be shown that second-order risk attitudes with regret references are solely driven by the shape of $m(\cdot)$. As long as $u(\cdot)$ and $m(\cdot)$ are strictly increasing, a strictly concave (convex, linear) m function implies risk aversion (risk seeking, risk neutrality).[25] This shows that regret references do in this case only contribute to risk aversion because EUT effects are reintroduced into the gain-loss function. Behavior under risk is still completely driven by the shape of pure consumption utility. Since we are interested in the pure effect of regret references without confounding effects of standard EUT preferences, Proposition 2.12 still constitutes an important result. The other assumption that leads to this result is that outcome and reference distributions are stochastically independent. This is also the reason why some models of regret, in which outcomes and foregone alternatives are compared state-wise, conclude that regret leads to risk-averse behavior (see e.g. Bleichrodt, Cillo, and Diecidue, 2010).

Unlike second-order effects, third-order effects are not eliminated by regret references.

[25] $B_2|A_2 \succ (\prec, \sim)\ A_2|B_2 \Leftrightarrow u(\delta_2) - u(-\delta_2) > (<, =)\ u(\delta_1) - u(-\delta_1)$ where $\delta_1 \equiv m(y+\varepsilon_1) - m(y)$ and $\delta_2 \equiv m(y) - m(y-\varepsilon_1)$. For all strictly increasing m functions $\delta_1, \delta_2 > 0$. Since $\delta_2 > (<, =)\ \delta_1 \Leftrightarrow m''(\cdot) < (>, =)\ 0$, for strictly increasing u functions it holds that $B_2|A_2 \succ (\prec, \sim)\ A_2|B_2 \Leftrightarrow m''(\cdot) < (>, =)\ 0$.

Proposition 2.13 *If preferences depend on reference points formed by regret, individuals are prudent (imprudent, neither nor), that is $B_3|A_3 \;\succ\; (\prec, \sim)\; A_3|B_3$, if and only if $[-k; k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}_1']|[0] \;\succ\; (\prec, \sim)\; [k; -k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}_1']|[0]$.*

Proposition 2.13 shows under which conditions individuals are prudent if their reference point is determined by regret. Since $M(B_3) = M(A_3)$ holds and in the comparison of $B_3|A_3$ vs. $A_3|B_3$ densities of similar elements can be reduced this choice is therefore equivalent to $[B_3|A_{3\langle A_3|B_3\rangle}]|[0] = [-k; k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}_1']|[0]$ vs. $[k; -k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}_1']|[0] = [A_3|B_{3\langle B_3|A_3\rangle}]|[0]$. Similar to the definition of prudence by Eeckhoudt and Schlesinger (2006), an individual in Proposition 2.13 is prudent if she prefers to face risk in the higher wealth rather than in the lower wealth state (given her reference point was zero).

Whether individuals with reference points formed by regret are in fact prudent or imprudent crucially depends on the functional form of $u(\cdot)$. If $u'''(\cdot) > (<, =)\ 0$ over the whole range, individuals are prudent (imprudent, neither nor). However, if $u(\cdot)$ is loss-averse S-shaped individuals are never prudent. Corollary 2.10 also applies to third-order risk preferences under regret. Moreover, if an individual is classified as being prudent (imprudent, neither nor) with disappointment references and a given u function, she will also be prudent (imprudent, neither nor) with regret references under this u function.

Since the usual assumption in the regret literature is that $u'''(\cdot) > 0$, we again observe a clear difference in predicted behavior between expectational and regret references concerning third-order risk preferences.[26] Moreover, since these two reference points yielded already different predictions concerning second-order risk preferences, the implications for precautionary saving and insurance demand could clearly not have been generated by regret references.

Unlike for third-order effects, we again receive a result that is independent of the functional form of $u(\cdot)$ for fourth-order effects.

Proposition 2.14 *If preferences depend on reference points formed by regret, individuals are never temperate or intemperate, that is $B_4|A_4 \sim A_4|B_4$. This holds for all formulations of $u(\cdot)$.*

Proposition 2.14 shows that reference-dependent preferences with reference points determined by regret can neither be temperate nor intemperate. Thus, there exists no

[26] Note that some authors (see e.g. Loomes and Sugden, 1982) use a function $Q(m(\alpha) - m(\beta)) \equiv m(\alpha) - m(\beta) + u(m(\alpha) - m(\beta)) - u(m(\beta) - m(\alpha))$ in order to state that the act is chosen where α obtains in a certain state of the world and β would have obtained by a different choice (if and only if the weighted value of $Q(\cdot)$ is positive). With this formulation, it is usually assumed that $Q(\cdot)$ is convex, which corresponds to $u(\cdot)$ being decreasingly concave (see Bleichrodt, Cillo, and Diecidue, 2010). This is equivalent to the assumption that $u'(\cdot) > 0, u''(\cdot) < 0, u'''(\cdot) > 0$.

preference to aggregate or disaggregate $\tilde{\varepsilon}_1$ and $\tilde{\varepsilon}_2$, two stochastically independent, symmetric and zero-mean random variables.

When looking at second- to fourth-order risk preferences, an interesting pattern can be observed when our results for different endogenous reference points are compared. Regret and expectational references do impose different risk attitudes for all these orders. However, disappointment references impose risk preferences that can be considered as in between expectational and regret references. For a given functional form of gain-loss utility, predicted behavior of disappointment references is generically different compared to expectational but not regret references with respect to the third order, and it is different compared to regret but not expectational references concerning the second and fourth order. This pattern generalizes to any arbitrary order n. By comparing case *(iv)* of Proposition 2.9 to Theorem 2.7 one can easily see that if an individual is risk apportioning of order n with $\frac{n}{2} \in \mathbb{N}$ (its opposite, neither nor) for a given u function under disappointment, she will also be risk apportioning of order n (its opposite, neither nor) under this u function with expectational references. This holds, however, only for even but not for odd orders. In contrast, a similar relationship between disappointment and regret only holds for odd but not for even orders. Comparing case *(iv)* of Proposition 2.9 to Theorem 2.15 shows that if an individual is risk apportioning of order n with $\frac{n}{2} \notin \mathbb{N}$ (its opposite, neither nor) for a given u function under disappointment, she will also be risk apportioning of order n (its opposite, neither nor) with this u function under regret.

Theorem 2.15 generalizes our results under regret for arbitrary order n.

Theorem 2.15 *If preferences depend on reference points formed by regret and $n \geq 4$, individuals are*

(i) for odd orders ($\frac{n}{2} \notin \mathbb{N}$) risk apportioning of order n (its opposite, neither nor), that is $B_n|A_n \succ (\prec, \sim)\ A_n|B_n$, if and only if
$$[B_n|A_{n\langle A_n|B_n\rangle}]\,|[0] = \left[A_{n-2}|B_{n-2\langle B_{n-2}|A_{n-2}\rangle}; B_{n-2}|A_{n-2\langle A_{n-2}|B_{n-2}\rangle} + \tilde{\varepsilon}_{(\frac{n-1}{2})} + \tilde{\varepsilon}'_{(\frac{n-1}{2})}\right]|[0]$$
$$\succ (\prec, \sim)$$
$$\left[B_{n-2}|A_{n-2\langle A_{n-2}|B_{n-2}\rangle}; A_{n-2}|B_{n-2\langle B_{n-2}|A_{n-2}\rangle} + \tilde{\varepsilon}_{(\frac{n-1}{2})} + \tilde{\varepsilon}'_{(\frac{n-1}{2})}\right]|[0] = [A_n|B_{n\langle B_n|A_n\rangle}]\,|[0];$$

(ii) for even orders ($\frac{n}{2} \in \mathbb{N}$) never risk apportioning of order n or its opposite, that is $B_n|A_n \sim A_n|B_n$. This holds for all formulations of $u(\cdot)$.

Considering $n = 1$ (Proposition 2.11) we showed that under regret first-order effects are present. The same result was obtained for expectational (Proposition 2.1) as well as disappointment references (case *(i)* in Proposition 2.9). However, with these reference points the result was solely driven by pure consumption utility. By contrast, under regret the gain-loss component of utility contributes to this result as well, since choosing a high and comparing it to a low outcome yields the sensation of rejoice, while choosing the low outcome and comparing it to the high outcome yields a feeling of regret. With $n = 2$, choosing an uncertain outcome and comparing it to a certain outcome with an equal mean can result in sensations of both rejoice and regret. However, choosing a certain outcome and comparing it to the risky alternative yields exactly the same feelings of rejoice and regret. Individuals with regret references therefore do not care whether to take or avoid a risk. Risk enters gain-loss utility either via the outcome or via the reference point. In general, questions of aggregating or disaggregating risk(s) become irrelevant with regret preferences since they enter the evaluation through either the avoided or the chosen alternative. This eliminates even-order risk preferences. Nevertheless, location still matters and makes odd-order effects relevant.

As for Theorem 2.7, a similar, more formal intuition also applies to Theorem 2.15. Since for even orders $-B_n \succ (\prec, \sim) \ -A_n$ if and only if $B_n \succ (\prec, \sim) \ A_n$ (due to the symmetry of $\tilde{\varepsilon}_i$ and the absence of a reduction of k) and with regret references the magnitude of both these opposing effects is equal, an individual is now always indifferent between $B_n|A_n$ and $A_n|B_n$ for even n. For odd orders note that $-B_n \succ (\prec, \sim) \ -A_n$ if and only if $B_n \prec (\succ, \sim) \ A_n$ (again due to the existence of a k reduction in addition to symmetric $\tilde{\varepsilon}_i$). Thus, an individual now prefers $B_n|A_n$ to $A_n|B_n$ if and only if she is risk apportioning of order n.

Comparing Theorem 2.15 to 2.7 highlights the fact that reference points formed by regret in a way induce opposite risk preferences as those formed by expectations. With regret, preferences of even orders vanish while those of odd orders may well be present. By contrast, with reference points formed by expectations preferences of odd orders are absent and those of even orders are still present. One implication is that reference-dependent risk preferences of any order are only affected by either expectations or regret. This suggests that both forms of endogenous reference points rather supplement than collide with each other. Our findings stress the advantage of jointly analyzing regret and expectational references in a unified framework. Until now, both streams of the literature have been analyzed separately despite the fact that they addressed similar questions.

2.5.3 Exogenous Reference Points

Unlike endogenous reference points like expectational, disappointment, or regret references, exogenous reference points are usually of an ad hoc nature. This is because almost every assumption on exogenous reference points can be supported by some reasoning. In the following, we concentrate on the intuitively most appealing and most frequently discussed reference points.

Assuming the status quo is not only very popular, but it is also easily assessed. Furthermore, it is undeniably exogenous at a given point in time. For these reasons the status quo is advantageous if limited information is available on processes prior to the decision at hand. Also, it does not require any assumption on how individuals form their beliefs over future states. Only the minimum assumption of an individual's awareness of her current state is necessary. Independently of these considerations the status quo may be the natural reference point when decisions come as a surprise. In our analysis wealth is normalized such that the status quo is 0.[27]

Although the status quo is often an obvious candidate for the reference point in experiments, it seems highly sensitive to framing effects. In general, many framing situations are possible and could support various exogenous reference points. If, for instance, an experiment was actually conducted over the gambles we used in our definitions, the reference point could be influenced in several ways. However, we focus firstly on situations where the framing of the choice reduces complexity. These are scenarios in which the subject decides in which situation she accepts an additional 'harm', i.e. $-k$ or $\tilde{\varepsilon}_i$ (see Eeckhoudt and Schlesinger, 2006). As in the experiment of Deck and Schlesinger (2010), which tests for prudence and temperance in the laboratory, a subject has to decide whether to take an unavoidable risk in a high- or low-wealth state (prudence) or in a certain or uncertain state (temperance). Formally, she has to add $\tilde{\varepsilon}_1$ to one of the outcomes in the gamble $[y; y-k]$ (prudence) or $[y; y+\tilde{\varepsilon}_2]$ (temperance) and these gambles may then act as the subject's reference point. This is, however, only the case if the reference point adjusts sufficiently fast. An equally complex framing to measure prudence, which has been used in the experiments of Deck and Schlesinger (2010) and Ebert and Wiesen (2011), would be the task of assigning a sure reduction of k to one of the outcomes in the gamble $[y; y+\tilde{\varepsilon}_1]$. Again, this gamble may then act as subjects' reference point.[28] Secondly, y is the single element that is present in all possible states of the world. It may therefore act as a mental anchor and thus be a natural point of comparison.

In Tables 2.1 to 2.3 we summarize our results for these reference scenarios. The first column defines the assumed exogenous reference point. We computed $B_n|\cdot$ and $A_n|\cdot$ given these reference points with $n \in \{2, 3, 4\}$. In the second column we state what the

[27]Other deterministic reference points could be generated by past outcomes (habit formation) or self-selected goals (aspiration levels). We do not analyze these scenarios in our static framework.

[28]Note that it is not required to implement the higher-order gamble definitions in such compound ways. In the experiment of Maier and Rüger (2010) all gambles were presented in a non-compound format which would suggest using 0 as the reference point even if the adjustment happens fast.

Table 2.1: Second-Order Effects with Exogenous References

Reference Point	Equivalent to Evaluating	Globally Linear	Linear w/ Loss Aversion	Globally Standard	Symmetric S-shape	S-shape (w/ Loss Aversion)
$\cdot\|[0]$	$[y]\|[0]$ vs. $[y+\tilde{\varepsilon}_1]\|[0]$	$B_2\|\cdot \sim A_2\|\cdot$	$B_2\|\cdot \succsim^1 A_2\|\cdot$	$B_2\|\cdot \succ A_2\|\cdot$	$B_2\|\cdot \overset{\succ}{\underset{\prec}{\sim}}^{2} A_2\|\cdot$	$B_2\|\cdot \overset{\succ}{\underset{\prec}{\sim}}^{3} A_2\|\cdot$
$\cdot\|[y]$	$[0]\|[0]$ vs. $[\tilde{\varepsilon}_1]\|[0]$	$B_2\|\cdot \sim A_2\|\cdot$	$B_2\|\cdot \succ A_2\|\cdot$	$B_2\|\cdot \succ A_2\|\cdot$	$B_2\|\cdot \sim A_2\|\cdot$	$B_2\|\cdot \succ A_2\|\cdot$

Notes:
1) $B_2|\cdot \succ A_2|\cdot$ if $-\varepsilon_1 < y < \varepsilon_1$; $B_2|\cdot \sim A_2|\cdot$ if $\varepsilon_1 \leq y$ or $y \leq -\varepsilon_1$; thus determined over the full range.
2) $B_2|\cdot \succ A_2|\cdot$ if $0 < y$; $B_2|\cdot \sim A_2|\cdot$ if $y = 0$; $B_2|\cdot \prec A_2|\cdot$ if $y < 0$; thus determined over the full range.
3) $B_2|\cdot \succ A_2|\cdot$ if $\varepsilon_1 \leq y$; $B_2|\cdot \prec A_2|\cdot$ if $y \leq -\varepsilon_1$; generally ambiguous if $-\varepsilon_1 < y < \varepsilon_1$.

comparison of these gambles reduces to, namely $[B_n|\cdot\langle A_n|\cdot\rangle]|[0]$ vs. $[A_n|\cdot\langle B_n|\cdot\rangle]|[0]$. If and only if the former gamble is preferred to the latter, $B_n|\cdot \succ A_n|\cdot$. In general, our findings for exogenous reference points depend on the functional form of the evaluation function $u(\cdot)$. As in the previous sections, we again distinguish between common assumptions about $u(\cdot)$ and present their corresponding results in columns three to seven.

Table 2.2: Third-Order Effects with Exogenous References

Reference Point	Equivalent to Evaluating	Globally Linear	Linear w/ Loss Aversion	Globally Standard	Symmetric S-shape	S-shape (w/ Loss Aversion)
$\cdot\|[0]$	$[y-k; y+\tilde{\varepsilon}_1]\|[0]$ vs. $[y; y-k+\tilde{\varepsilon}_1]\|[0]$	$B_3\|\cdot \sim A_3\|\cdot$	$B_3\|\cdot \overset{\succ}{\underset{\prec}{\sim}}^{1} A_3\|\cdot$	$B_3\|\cdot \succ A_3\|\cdot$	$B_3\|\cdot \overset{\succ}{\underset{\prec}{\sim}}^{2} A_3\|\cdot$	$B_3\|\cdot \overset{\succ}{\underset{\prec}{\sim}}^{2} A_3\|\cdot$
$\cdot\|[y]$	$[-k; \tilde{\varepsilon}_1]\|[0]$ vs. $[0; -k+\tilde{\varepsilon}_1]\|[0]$	$B_3\|\cdot \sim A_3\|\cdot$	$B_3\|\cdot \prec A_3\|\cdot$	$B_3\|\cdot \succ A_3\|\cdot$	$B_3\|\cdot \prec A_3\|\cdot$	$B_3\|\cdot \prec A_3\|\cdot$
$\cdot\|[y; y+\tilde{\varepsilon}_1]$	$[-k; 2\tilde{\varepsilon}_1]\|[0]$ vs. $[0; -k+2\tilde{\varepsilon}_1]\|[0]$	$B_3\|\cdot \sim A_3\|\cdot$	$B_3\|\cdot \prec A_3\|\cdot$	$B_3\|\cdot \succ A_3\|\cdot$	$B_3\|\cdot \prec A_3\|\cdot$	$B_3\|\cdot \prec A_3\|\cdot$
$\cdot\|[y; y-k]$	$[-k; k+\tilde{\varepsilon}_1]\|[0]$ vs. $[k; -k+\tilde{\varepsilon}_1]\|[0]$	$B_3\|\cdot \sim A_3\|\cdot$	$B_3\|\cdot \precsim^3 A_3\|\cdot$	$B_3\|\cdot \succ A_3\|\cdot$	$B_3\|\cdot \prec A_3\|\cdot$	$B_3\|\cdot \prec A_3\|\cdot$

Notes:
1) $B_3|\cdot \sim A_3|\cdot$ if $y \leq -\varepsilon_1$, or $\varepsilon_1 + k \leq y$; $B_3|\cdot \prec A_3|\cdot$ if $y = 0$; generally ambiguous if $-\varepsilon_1 < y < \varepsilon_1 + k \wedge y \neq 0$, but $B_3|\cdot \succ A_3|\cdot$ if $k < y < \varepsilon_1$, $B_3|\cdot \sim A_3|\cdot$ if $\varepsilon_1 < y < k - \varepsilon_1$, and $B_3|\cdot \prec A_3|\cdot$ if $k - \varepsilon_1 < y < 0$.
2) $B_3|\cdot \succ (\sim, \prec)\ A_3|\cdot$ if either $y \leq -\varepsilon_1$ or $\varepsilon_1 + k \leq y$, and additionally $u'''(\cdot) > (=, <)\ 0$; generally ambiguous if $-\varepsilon_1 < y < \varepsilon_1 + k$, or the sign of $u'''(\cdot)$ is not constant over the full range.
3) $B_3|\cdot \sim A_3|\cdot$ if $\varepsilon_1 \leq k$; $B_3|\cdot \prec A_3|\cdot$ if $k < \varepsilon_1$; thus determined over the full range.

In column three $u(\cdot)$ is globally linear. In the fourth column we consider piecewise linear gain-loss utility with a unique kink at 0 inducing loss aversion. In the fifth column we analyze a standard u function that has derivatives that alternate in sign with $u'(\cdot)$ being positive. Column six considers a function that exhibits identical diminishing sensitivity for losses and gains. In column seven we report results for S-shaped u functions that are asymmetric, either due to loss aversion or faster diminishing sensitivity in gains than in losses, or both. This functional form is the traditional assumption made in prospect theory. All results in Tables 2.1 to 2.3 are formally derived for two-outcome $\tilde{\varepsilon}_i$, hence $\tilde{\varepsilon}_i = [-\varepsilon_i, \varepsilon_i]$.

As can be seen from Tables 2.1 to 2.3 results are rather ambiguous in case the reference point is the status quo. This is due to the fact that individuals can end up making gains or losses depending on the specific parameter combinations of y, k, ε_1, and

Table 2.3: Fourth-Order Effects with Exogenous References

Reference Point	Equivalent to Evaluating	Globally Linear	Linear w/ Loss Aversion	Globally Standard	Symmetric S-shape	S-shape (w/ Loss Aversion)
$\cdot\|[0]$	$[y+\tilde{\varepsilon}_1; y+\tilde{\varepsilon}_2]\|[0]$ vs. $[y; y+\tilde{\varepsilon}_1+\tilde{\varepsilon}_2]\|[0]$	$B_4\|\cdot \sim A_4\|\cdot$	$B_4\|\cdot \overset{\succ}{\underset{\prec}{\sim}}{}^{1} A_4\|\cdot$	$B_4\|\cdot \succ A_4\|\cdot$	$B_4\|\cdot \overset{\succ}{\underset{\prec}{\sim}}{}^{2} A_4\|\cdot$	$B_4\|\cdot \overset{\succ}{\underset{\prec}{\sim}}{}^{2} A_4\|\cdot$
$\cdot\|[y]$	$[\tilde{\varepsilon}_1; \tilde{\varepsilon}_2]\|[0]$ vs. $[0; \tilde{\varepsilon}_1+\tilde{\varepsilon}_2]\|[0]$	$B_4\|\cdot \sim A_4\|\cdot$	$B_4\|\cdot \prec A_4\|\cdot$	$B_4\|\cdot \succ A_4\|\cdot$	$B_4\|\cdot \sim A_4\|\cdot$	$B_4\|\cdot \prec A_4\|\cdot$
$\cdot\|[y; y+\tilde{\varepsilon}_2]$	$[2\tilde{\varepsilon}_2; \tilde{\varepsilon}_1]\|[0]$ vs. $[0; 2\tilde{\varepsilon}_2+\tilde{\varepsilon}_1]\|[0]$	$B_4\|\cdot \sim A_4\|\cdot$	$B_4\|\cdot \prec A_4\|\cdot$	$B_4\|\cdot \succ A_4\|\cdot$	$B_4\|\cdot \sim A_4\|\cdot$	$B_4\|\cdot \prec A_4\|\cdot$

Notes:
1) $B_4|\cdot \sim A_4|\cdot$ if $\varepsilon_1+\varepsilon_2 \leq y$, $y=0$, or $y \leq -\varepsilon_1-\varepsilon_2$; $B_4|\cdot \succ A_4|\cdot$ if $-\varepsilon_1-\varepsilon_2 < y \leq -\max\{\varepsilon_1,\varepsilon_2\}$, or $\max\{\varepsilon_1,\varepsilon_2\} < y < \varepsilon_1+\varepsilon_2$; generally ambiguous if $-\max\{\varepsilon_1,\varepsilon_2\} < y < \max\{\varepsilon_1,\varepsilon_2\}$.
2) $B_4|\cdot \succ (\sim, \prec)\ A_4|\cdot$ if either $y \leq -\varepsilon_1-\varepsilon_2$ or $\varepsilon_1+\varepsilon_2 \leq y$, and additionally $u''''(\cdot) < (=, >)\ 0$; generally ambiguous if $-\varepsilon_1-\varepsilon_2 < y < \varepsilon_1+\varepsilon_2$, or the sign of $u''''(\cdot)$ is not constant over the full range.

ε_2. In the other reference scenarios initial wealth y cancels out since it is part of both the reference point and the gambles evaluated. Results for these reference scenarios are therefore more straightforward.[29] Generally, they vary widely with the functional form considered. Nevertheless, a few tendencies seem prevalent. While not stated explicitly, exogenous reference points also yield monotone preferences. As already extensively discussed in the existing literature, we further find that loss aversion induces risk aversion. But, more surprisingly, loss aversion also induces imprudence and intemperance in many cases. Assuming a S-shaped u function does only generate risk aversion in the asymmetric case (column seven). Such a requirement is not needed for the third order. Here, an S-shaped u function always induces imprudence in cases where y is an element of the reference point. These results clearly show that exogenous reference points do not yield the same predictions on third-order risk preferences as expectational references. Again, the implications of expectational references we derived for precautionary saving and insurance demand would therefore not obtain with exogenous reference points.

2.6 Conclusion

Models of reference-dependent preferences have been proposed to resolve empirical puzzles concerning second-order risk preferences. There has been no attempt to consider higher orders, like the third-order effect of prudence or the fourth-order effect of temperance, in these models in order to see whether they can resolve empirical puzzles concerning higher orders as well. This chapter is a first attempt in this direction. We first analyzed higher-order risk preferences in Kőszegi and Rabin's (2006, 2007) model of expectation-based reference dependence. Risk preferences of odd orders (except or-

[29] Note that in case of a fixed, real-numbered reference point, as in the first two rows of Tables 2.1 to 2.3, our reference-dependent model reduces to a general EUT model if the functional form of $u(\cdot)$ does not depend on the reference point (see Wakker, 2005).

der one) were generally found to be absent while even-order risk attitudes still exist in their model. These results show that higher-order risk preferences under reference dependence follow a completely different pattern than in classical EUT models as they do not depend on the previous order. For example, if a certain order delivers indifference in EUT, the next higher order has to deliver indifference as well. This is not the case under reference dependence. Therefore, our results have strong implications in situations where characteristics of one order are used to derive statements on a succeeding order. We considered precautionary saving and optimal demand for insurance as illustrating examples for the implications of our results. In a simple model we showed that seemingly sub-optimal amounts of precautionary saving under EUT may well be optimal under reference dependence. Also, the importance of insurance is less mitigated by buffer-stock saving with such preferences. These implications were formally derived in a two-period model. In order to verify whether the observed empirical patterns also match the optimal behavior of individuals under reference dependence quantitatively, a dynamic analysis would be necessary. However, this would additionally require a dynamic model of belief formation concerning own behavior in all future periods. Such a task is left for future research.

In a second step, we further showed that alternative behavioral models cannot resolve these empirical puzzles as they predict different patterns of risk attitudes toward higher orders. In this robustness analysis we considered other endogenous as well as exogenous reference points. While exogenous reference points often seem to be of an ad-hoc nature, endogenous reference points generally account for the fact that individuals anticipate how their behavior will have a feedback on their preferences. With reference points being endogenously formed by the alternative choice, as in models of regret, risk preferences toward even orders are absent while those of odd orders are present. With reference points being endogenously formed by the mean of expectations, as in models of disappointment, results can be considered as in between expectational and regret references. Note that all these models predict monotone preferences. However, only concerning higher *even* orders expectation-based reference dependence and disappointment models yield the same predicted behavior and only with respect to higher *odd* orders models of regret and disappointment yield the same behavior. Regret models and expectation-based reference dependence never induce the same behavior as they yield opposite patterns of higher-order risk attitudes.

Together with our results on exogenous reference points, which generally differ from expectation-based reference dependence, our robustness analysis clearly showed that higher-order risk preferences under expectation-based reference dependence follow a different pattern than under alternative assumptions on the reference point. Alternative behavioral models can therefore in general not resolve the same empirical puzzles that can be resolved with expectation-based reference dependence. Besides this observation, comparing our results on endogenous reference points offers further interesting insights. On the one hand, they emphasize a new way, for instance via third-order risk preferences, to discriminate between expectation-based reference dependence and models of disappointment in laboratory experiments. This has not been possible in the past and the gamble definitions of Eeckhoudt and Schlesinger (2006) we used in our analysis are especially advantageous in this respect. On the other hand, they show that models of regret and expectation-based reference dependence cannot influence risk preferences of a certain order simultaneously. Thus, risk preferences of a particular order can only be influenced by either expectational or regret references. The notion that expectations and regret influence human decision making under uncertainty in distinctively different ways is also suggested by neurological research. For instance, Camille et al. (2004) and Coricelli et al. (2005) find that counterfactual thinking on the alternative choice is activated in a different brain region than counterfactual thinking on one's choice taken.

Although we analyzed the implications of our results for two specific examples only, i.e. precautionary saving and insurance demand, they may well be of significance in other areas as well where higher orders have shown to play an important role. For instance, they could lead to further insights for asset demand under background risk or for precautionary labor supply. Future research should therefore analyze further applications where our results on reference-dependent risk preferences of higher orders may be helpful to yield better predictions than those of classical EUT models.

2.7 Appendix: Proofs

Proof. **[Proposition 2.1]** Since $U(B_1|B_1) = u(m(y) - m(y)) = u(0)$ and $U(A_1|A_1) = u(m(y-k) - m(y-k)) = u(0)$, it follows that $U(B_1|B_1) = U(A_1|A_1)$. Then, $V(B_1|B_1) = M(B_1) + U(B_1|B_1) = m(y) + u(0)$ and $V(A_1|A_1) = M(A_1) + U(A_1|A_1) = m(y-k) + u(0)$. Since $m(\cdot)$ is strictly increasing and $k > 0$, it follows that $V(B_1|B_1) > V(A_1|A_1)$ and thus $B_1|B_1 \succ A_1|A_1$. ■

Lemma 2.16 *$M(B_n) = M(A_n)$ holds for all $n \geq 2$. It therefore holds for all $n \geq 2$ that $B_n|C \succ (\prec, \sim)\ A_n|C \Leftrightarrow V(B_n|C) > (<, =)\ V(A_n|C) \Leftrightarrow U(B_n|C) > (<, =)\ U(A_n|C)$ where C is any arbitrary gamble.*

Proof. **[Lemma 2.16]** From $\mathbb{E}[\tilde{\varepsilon}_1] = 0$, the linearity of $m(\cdot)$, and the definitions of B_2, A_2, B_3, A_3, it follows that $M(B_2) = M(A_2)$ and $M(B_3) = M(A_3)$. Using these results, $\mathbb{E}[\tilde{\varepsilon}_i] = 0\ \forall i$, and the linearity of $m(\cdot)$, it follows from the definitions of B_n and A_n that $M(B_n) = M(A_n)$ also holds for $n \geq 4$. Thus, for all $n \geq 2$ and C being any arbitrary gamble it holds that $B_n|C \succ (\prec, \sim)\ A_n|C \Leftrightarrow V(B_n|C) > (<, =)\ V(A_n|C) \Leftrightarrow U(B_n|C) > (<, =)\ U(A_n|C)$. ■

Proof. **[Proposition 2.2]** From Lemma 2.16 it follows that $B_2|B_2 \succ (\prec, \sim)\ A_2|A_2 \Leftrightarrow U(B_2|B_2) > (<, =)\ U(A_2|A_2)$. Evaluating $B_2|B_2$ by $U(\cdot)$ means getting $B_2 = y$ and also expecting $B_2 = y$. Therefore,

$$U(B_2|B_2) = u(m(y) - m(y)) = u(0).$$

Evaluating $A_2|A_2$ by $U(\cdot)$ means getting $A_2 = y + \tilde{\varepsilon}_1$ and also expecting $A_2 = y + \tilde{\varepsilon}'_1$. Therefore,

$$\begin{aligned} U(A_2|A_2) &= \iint u(m(y + \tilde{\varepsilon}_1) - m(y + \tilde{\varepsilon}'_1)) dF(\tilde{\varepsilon}_1) dF(\tilde{\varepsilon}'_1) \\ &= \iint u(\tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1) dF(\tilde{\varepsilon}_1) dF(\tilde{\varepsilon}'_1), \end{aligned}$$

where the last equality holds because of the symmetry of all $\tilde{\varepsilon}_i$ and the linearity of $m(\cdot)$. From $\mathbb{E}[\tilde{\varepsilon}_1] = 0$, $\mathbb{E}[\tilde{\varepsilon}'_1] = 0$ and the stochastic independence of $\tilde{\varepsilon}_1$ and $\tilde{\varepsilon}'_1$ it follows that $\mathbb{E}[\tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1] = 0$. From $\mathbb{E}[\tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1] = 0$ and the linearity of $m(\cdot)$ it follows that $M([\tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1]) = M([0])$. Thus, we can state that

$$V(B_2|B_2) > (<, =)\ V(A_2|A_2) \Leftrightarrow V([0]|[0]) > (<, =)\ V([\tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1]|[0]),$$

and therefore

$$B_2|B_2 \succ (\prec, \sim)\ A_2|A_2 \Leftrightarrow [0]|[0] \succ (\prec, \sim)\ [\tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1]|[0].$$

Alternatively, we can also state that both gambles reduce to

$$[B_2|B_{2\langle A_2|A_2\rangle}]\,|[0] = [0]|[0] \quad \text{and} \quad [A_2|A_{2\langle B_2|B_2\rangle}]\,|[0] = [\tilde{\varepsilon}_1 + \tilde{\varepsilon}_1']|[0]$$

when compared with each other. ■

Proof. [Proposition 2.4] Evaluating $B_3|B_3$ by $U(\cdot)$ delivers

$$\begin{aligned} U(B_3|B_3) &= \iint \big[\tfrac{1}{4}u(m(y-k)-m(y-k)) + \tfrac{1}{4}u(m(y-k)-m(y+\tilde{\varepsilon}_1')) \\ &\quad + \tfrac{1}{4}u(m(y+\tilde{\varepsilon}_1)-m(y-k)) + \tfrac{1}{4}u(m(y+\tilde{\varepsilon}_1)-m(y+\tilde{\varepsilon}_1'))\big]\, dF(\tilde{\varepsilon}_1)dF(\tilde{\varepsilon}_1') \\ &= \iint \big[\tfrac{1}{4}u(0) + \tfrac{1}{4}u(-k+\tilde{\varepsilon}_1') + \tfrac{1}{4}u(k+\tilde{\varepsilon}_1) + \tfrac{1}{4}u(\tilde{\varepsilon}_1+\tilde{\varepsilon}_1')\big]\, dF(\tilde{\varepsilon}_1)dF(\tilde{\varepsilon}_1'), \end{aligned}$$

where the last equality holds because of the symmetry of all $\tilde{\varepsilon}_i$ and the linearity of $m(\cdot)$. Evaluating $A_3|A_3$ by $U(\cdot)$ delivers

$$\begin{aligned} U(A_3|A_3) &= \iint \big[\tfrac{1}{4}u(m(y)-m(y)) + \tfrac{1}{4}u(m(y)-m(y-k+\tilde{\varepsilon}_1')) + \tfrac{1}{4}u(m(y-k+\tilde{\varepsilon}_1)-m(y)) \\ &\quad + \tfrac{1}{4}u(m(y-k+\tilde{\varepsilon}_1)-m(y-k+\tilde{\varepsilon}_1'))\big]\, dF(\tilde{\varepsilon}_1)dF(\tilde{\varepsilon}_1') \\ &= \iint \big[\tfrac{1}{4}u(0) + \tfrac{1}{4}u(k+\tilde{\varepsilon}_1') + \tfrac{1}{4}u(-k+\tilde{\varepsilon}_1) + \tfrac{1}{4}u(\tilde{\varepsilon}_1+\tilde{\varepsilon}_1')\big]\, dF(\tilde{\varepsilon}_1)dF(\tilde{\varepsilon}_1'), \end{aligned}$$

where the last equality holds because of the symmetry of all $\tilde{\varepsilon}_i$. Since $\tilde{\varepsilon}_1$ and $\tilde{\varepsilon}_1'$ are identically distributed we can conclude that $U(B_3|B_3) = U(A_3|A_3)$ regardless of the shape of $u(\cdot)$. Lemma 2.16 showed that $M(B_3) = M(A_3)$. From $M(B_3) = M(A_3)$ and $U(B_3|B_3) = U(A_3|A_3)$ it follows that $V(B_3|B_3) = V(A_3|A_3)$ and thus $B_3|B_3 \sim A_3|A_3$. Alternatively, we can also state that both gambles reduce to

$$[B_3|B_{3\langle A_3|A_3\rangle}]\,|[0] = [C]|[0] \quad \text{and} \quad [A_3|A_{3\langle B_3|B_3\rangle}]\,|[0] = [C]|[0] \tag{2.16}$$

when compared with each other, where C denotes any arbitrary gamble. ■

Proof. [Proposition 2.5] Follows directly from Theorem 2.7. ■

Proof. **[Theorem 2.7]** To prove both part (i) and part (ii) of Theorem 2.7 (where we use Lemma 2.16, $\tilde{\varepsilon}_i$ being symmetric with $\mathbb{E}[\tilde{\varepsilon}_i] = 0\ \forall i$, and the linearity of $m(\cdot)$) it helps to show first that for all $n \geq 3$

$$\begin{aligned}
& [B_n | B_{n\langle A_n|A_n\rangle}]\,|[0] \\
= \; & \left[B_{n-2}|B_{n-2\langle A_{n-2}|A_{n-2}\rangle}; A_{n-2}|A_{n-2\langle B_{n-2}|B_{n-2}\rangle} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\right]|[0], \quad (2.17)\\
& [A_n | A_{n\langle B_n|B_n\rangle}]\,|[0] \\
= \; & \left[A_{n-2}|A_{n-2\langle B_{n-2}|B_{n-2}\rangle}; B_{n-2}|B_{n-2\langle A_{n-2}|A_{n-2}\rangle} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\right]|[0]. \quad (2.18)
\end{aligned}$$

To show this, note that $[B_n|B_{n\langle A_n|A_n\rangle}]|[0]$ and $[A_n|A_{n\langle B_n|B_n\rangle}]|[0]$ can be expressed as gambles containing only B_n and A_n.

$$\begin{aligned}
B_n|B_n = [B_n - B_n]\,|[0] \quad &\text{and} \quad A_n|A_n = [A_n - A_n]\,|[0] \\
\Rightarrow [B_n|B_{n\langle A_n|A_n\rangle}]|[0] = [B_n - B_n]\,|[0] \quad &\text{and} \quad [A_n|A_{n\langle B_n|B_n\rangle}]|[0] = [A_n - A_n]\,|[0],
\end{aligned}$$

and therefore

$$\begin{aligned}
[B_{n-2}|B_{n-2\langle A_{n-2}|A_{n-2}\rangle}]|[0] \;&=\; [B_{n-2} - B_{n-2}]\,|[0] \quad (2.19)\\
\text{and} \quad [A_{n-2}|A_{n-2\langle B_{n-2}|B_{n-2}\rangle}]|[0] \;&=\; [A_{n-2} - A_{n-2}]\,|[0]. \quad (2.20)
\end{aligned}$$

However, $[B_n|B_{n\langle A_n|A_n\rangle}]|[0]$ and $[A_n|A_{n\langle B_n|B_n\rangle}]|[0]$ can also be expressed as gambles containing only B_{n-2} and A_{n-2} since $B_n = [A_{n-2}; B_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]$ and $A_n = [B_{n-2}; A_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]$.

$$\begin{aligned}
B_n|B_n \;&=\; \Big[A_{n-2} - A_{n-2}; A_{n-2} - B_{n-2} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}; B_{n-2} - A_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})};\\
&\qquad B_{n-2} - B_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\Big]\,|[0]\\
\text{and} \quad A_n|A_n \;&=\; \Big[B_{n-2} - B_{n-2}; B_{n-2} - A_{n-2} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}; A_{n-2} - B_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})};\\
&\qquad A_{n-2} - A_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\Big]\,|[0]\\
\Rightarrow \quad [B_n|B_{n\langle A_n|A_n\rangle}]|[0] \;&=\; \Big[A_{n-2} - A_{n-2}; B_{n-2} - B_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\Big]\,|[0] \quad (2.21)\\
\text{and} \quad [A_n|A_{n\langle B_n|B_n\rangle}]|[0] \;&=\; \Big[B_{n-2} - B_{n-2}; A_{n-2} - A_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\Big]\,|[0], \quad (2.22)
\end{aligned}$$

because two of the elements cancel out when the two gambles are compared with each other. Finally, inserting (2.19) and (2.20) into (2.21) and (2.22) delivers (2.17) and (2.18). Part (i) of Theorem 2.7 follows directly from (2.17) and (2.18) when we note that for $\frac{n}{2} \in \mathbb{N} \Rightarrow \text{int}(\frac{n}{2}) = \frac{n}{2}$.

In order to prove part (ii) of Theorem 2.7 ($\frac{n}{2} \notin \mathbb{N}$) we again use (2.17) and (2.18). Suppose $n = 5$, then

$$\begin{aligned} [B_5|B_{5\langle A_5|A_5\rangle}]|[0] &= \left[A_3|A_{3\langle B_3|B_3\rangle}; B_3|B_{3\langle A_3|A_3\rangle} + \tilde{\varepsilon}_2 + \tilde{\varepsilon}_2'\right]|[0] & (2.23) \\ \text{and} \quad [A_5|A_{5\langle B_5|B_5\rangle}]|[0] &= \left[B_3|B_{3\langle A_3|A_3\rangle}; A_3|A_{3\langle B_3|B_3\rangle} + \tilde{\varepsilon}_2 + \tilde{\varepsilon}_2'\right]|[0]. & (2.24) \end{aligned}$$

By inserting (2.16) into (2.23) and (2.24) and canceling $\tilde{\varepsilon}_2$ and $\tilde{\varepsilon}_2'$ in both gambles we then get

$$\begin{aligned} [B_5|B_{5\langle A_5|A_5\rangle}]|[0] = [C]|[0] \quad &\text{and} \quad [A_5|A_{5\langle B_5|B_5\rangle}]|[0] = [C]|[0]. \qquad (2.25) \\ \Rightarrow \qquad B_5|B_5 \quad &\sim \quad A_5|A_5. \end{aligned}$$

Since (2.25), by the same logic it must also hold that $[B_7|B_{7\langle A_7|A_7\rangle}]|[0] = [C]|[0]$ and $[A_7|A_{7\langle B_7|B_7\rangle}]|[0] = [C]|[0]$. Continuing this reasoning yields

$$\begin{aligned} [B_n|B_{n\langle A_n|A_n\rangle}]|[0] = [C]|[0] \quad &\text{and} \quad [A_n|A_{n\langle B_n|B_n\rangle}]|[0] = [C]|[0] \\ \Rightarrow \qquad B_n|B_n \quad &\sim \quad A_n|A_n \end{aligned}$$

for $\frac{n}{2} \notin \mathbb{N}$ and C being any arbitrary gamble. ■

Proof. [Proposition 2.8] First, we show that the second-order conditions are met for the optimal values in the benchmark model. The second partial derivatives of $V^b(\cdot)$ are

$$\begin{aligned} \frac{\partial^2 V^b}{(\partial k^i)^2} &= (1+\lambda)^2 c''(y - k^s - (1+\lambda)k^i) + 2c''(y - \varepsilon_1 + k^s + 2k^i) < 0, \\ \frac{\partial^2 V^b}{(\partial k^s)^2} &= c''(y - k^s - (1+\lambda)k^i) + \tfrac{1}{2}c''(y - \varepsilon_1 + k^s + 2k^i) + \tfrac{1}{2}c''(y + \varepsilon_1 + k^s) < 0, \\ \frac{\partial^2 V^b}{\partial k^s \partial k^i} &= (1+\lambda)c''(y - k^s - (1+\lambda)k^i) + c''(y - \varepsilon_1 + k^s + 2k^i) < 0. \end{aligned}$$

Then, the determinant of the Hessian in the benchmark case reduces to

$$\begin{aligned} |\mathcal{H}^b| &= \tfrac{1}{2}\left(1+\lambda^2\right) c''(y - \varepsilon_1 + k^s + 2k^i) c''(y - k^s - (1+\lambda)k^i) \\ &\quad + \tfrac{1}{2}c''(y + \varepsilon_1 + k^s)\left[(1+\lambda)^2 c''(y - k^s - (1+\lambda)k^i) + 2c''(y - \varepsilon_1 + k^s + 2k^i)\right] > 0. \end{aligned}$$

Now, we show that the second-order conditions are met for the values solving the first-order conditions in the hybrid model. Then conditions (2.13) and (2.15) in fact have to hold in the optimum. The second partial derivatives of $V^h(\cdot)$ are

$$\begin{aligned}
\frac{\partial^2 V^h}{(\partial k^i)^2} &= (1+\lambda)^2 c''(y-k^s-(1+\lambda)k^i) + 2c''(y-\varepsilon_1+k^s+2k^i) \\
&\quad + u''(2\varepsilon-2k^i) + u''(-2\varepsilon+2k^i) < 0, && (2.26)\\
\frac{\partial^2 V^h}{(\partial k^s)^2} &= c''(y-k^s-(1+\lambda)k^i) + \tfrac{1}{2}c''(y-\varepsilon_1+k^s+2k^i) + \tfrac{1}{2}c''(y+\varepsilon_1+k^s) < 0, && (2.27)\\
\frac{\partial^2 V^h}{\partial k^s \partial k^i} &= (1+\lambda)c''(y-k^s-(1+\lambda)k^i) + c''(y-\varepsilon_1+k^s+2k^i) < 0, && (2.28)
\end{aligned}$$

where we used the assumption of $u''(\tau) \leq -u''(-\tau)\ \forall \tau > 0$ to derive (2.26). The determinant of the Hessian in the hybrid case reduces to

$$\begin{aligned}
|\mathcal{H}^h| &= \tfrac{1}{2}\left(1+\lambda^2\right) c''(y-\varepsilon_1+k^s+2k^i)c''(y-k^s-(1+\lambda)k^i) \\
&\quad + \tfrac{1}{2}c''(y+\varepsilon_1+k^s)\left[(1+\lambda)^2 c''(y-k^s-(1+\lambda)k^i) + 2c''(y-\varepsilon_1+k^s+2k^i)\right] \\
&\quad + \frac{\partial^2 V^h}{(\partial k^s)^2}\left[u''(2\varepsilon-2k^i) + u''(-2\varepsilon+2k^i)\right] > 0, && (2.29)
\end{aligned}$$

where we again used the assumption $u''(\tau) \leq -u''(-\tau)\ \forall \tau > 0$ as well as (2.27). We can now conclude from (2.15) that $dk^i > 0$ has to hold if we use (2.27) and (2.29). Using this result it follows from (2.13), (2.27), and (2.28) that $dk^s < 0$ which completes the proof. ■

Proof. [Proposition 2.9] Parts (i)-(iv) of Proposition 2.9 will be proved subsequently. Part (i) follows directly from the proof of Proposition 2.1 since $B_1 = \mathbb{E}[B_1]$ and $A_1 = \mathbb{E}[A_1]$.

For proving part (ii) first note that from Lemma 2.16 it follows that $B_2|\mathbb{E}[B_2] \succ (\prec, \sim)\ A_2|\mathbb{E}[A_2] \Leftrightarrow U(B_2|\mathbb{E}[B_2]) > (<, =)\ U(A_2|\mathbb{E}[A_2])$. Evaluating $B_2|\mathbb{E}[B_2]$ by $U(\cdot)$ means getting $B_2 = y$ and expecting $\mathbb{E}[B_2] = y$. Thus,

$$U(B_2|\mathbb{E}[B_2]) = u(m(y) - m(y)) = u(0).$$

Evaluating $A_2|\mathbb{E}[A_2]$ by $U(\cdot)$ means getting $A_2 = y + \tilde{\varepsilon}_1$ and expecting $\mathbb{E}[A_2] = y$ since $\mathbb{E}[\tilde{\varepsilon}'_1] = 0$. Therefore,

$$\begin{aligned}
U(A_2|\mathbb{E}[A_2]) &= \int u(m(y+\tilde{\varepsilon}_1) - m(\int (y+\tilde{\varepsilon}'_1)dF(\tilde{\varepsilon}'_1)))dF(\tilde{\varepsilon}_1) \\
&= \int u(m(y+\tilde{\varepsilon}_1) - m(y))dF(\tilde{\varepsilon}_1) = \int u(\tilde{\varepsilon}_1)dF(\tilde{\varepsilon}_1),
\end{aligned}$$

where the last equality holds because of the linearity of $m(\cdot)$. From $\mathbb{E}[\tilde{\varepsilon}_1] = 0$ and $m(\cdot)$ being linear it follows that $M([\tilde{\varepsilon}_1]) = M([0])$. We can therefore state that

$$V(B_2|\mathbb{E}[B_2]) > (<, =)\ V(A_2|\mathbb{E}[A_2]) \Leftrightarrow V([0]|[0]) > (<, =)\ V([\tilde{\varepsilon}_1]|[0]),$$

and hence

$$B_2|\mathbb{E}[B_2] \succ (\prec, \sim)\ A_2|\mathbb{E}[A_2] \Leftrightarrow [0]|[0] \succ (\prec, \sim)\ [\tilde{\varepsilon}_1]|[0].$$

For proving part (iii) we first note that evaluating $B_3|\mathbb{E}[B_3]$ by $U(\cdot)$ delivers

$$\begin{aligned} U(B_3|\mathbb{E}[B_3]) &= \int \Big[\tfrac{1}{2}u(m(y-k) - m(\int (\tfrac{1}{2}(y-k) + \tfrac{1}{2}(y+\tilde{\varepsilon}'_1))dF(\tilde{\varepsilon}'_1))) \\ &\quad + \tfrac{1}{2}u(m(y+\tilde{\varepsilon}_1) - m(\int (\tfrac{1}{2}(y-k) + \tfrac{1}{2}(y+\tilde{\varepsilon}'_1))dF(\tilde{\varepsilon}'_1)))\Big]\, dF(\tilde{\varepsilon}_1) \\ &= \int \left[\tfrac{1}{2}u(m(y-k) - m(y-\tfrac{k}{2})) + \tfrac{1}{2}u(m(y+\tilde{\varepsilon}_1) - m(y-\tfrac{k}{2}))\right] dF(\tilde{\varepsilon}_1) \\ &= \int \left[\tfrac{1}{2}u(-\tfrac{k}{2}) + \tfrac{1}{2}u(\tfrac{k}{2}+\tilde{\varepsilon}_1)\right] dF(\tilde{\varepsilon}_1), \end{aligned}$$

where the second equality holds because $\mathbb{E}[\tilde{\varepsilon}'_1] = 0$ and the last equality holds since $m(\cdot)$ is linear. Evaluating $A_3|\mathbb{E}[A_3]$ by $U(\cdot)$ yields

$$\begin{aligned} U(A_3|\mathbb{E}[A_3]) &= \int \Big[\tfrac{1}{2}u(m(y-k+\tilde{\varepsilon}_1) - m(\int (\tfrac{1}{2}(y-k+\tilde{\varepsilon}'_1) + \tfrac{1}{2}y)dF(\tilde{\varepsilon}'_1))) \\ &\quad + \tfrac{1}{2}u(m(y) - m(\int (\tfrac{1}{2}(y-k+\tilde{\varepsilon}'_1) + \tfrac{1}{2}y)dF(\tilde{\varepsilon}'_1)))\Big]\, dF(\tilde{\varepsilon}_1) \\ &= \int \left[\tfrac{1}{2}u(m(y-k+\tilde{\varepsilon}_1) - m(y-\tfrac{k}{2})) + \tfrac{1}{2}u(m(y) - m(y-\tfrac{k}{2}))\right] dF(\tilde{\varepsilon}_1) \\ &= \int \left[\tfrac{1}{2}u(-\tfrac{k}{2}+\tilde{\varepsilon}_1) + \tfrac{1}{2}u(\tfrac{1}{2})\right] dF(\tilde{\varepsilon}_1), \end{aligned}$$

where again the second equality holds because $\mathbb{E}[\tilde{\varepsilon}'_1] = 0$ and the last equality holds since $m(\cdot)$ is linear. Lemma 2.16 showed that $M(B_3) = M(A_3)$. From $\mathbb{E}[\tilde{\varepsilon}_1] = 0$ and the linearity of $m(\cdot)$ it follows that $M([-\frac{k}{2}; \frac{k}{2} + \tilde{\varepsilon}_1]) = M([-\frac{k}{2} + \tilde{\varepsilon}_1; \frac{k}{2}])$, and we can therefore state that

$$V(B_3|\mathbb{E}[B_3]) > (<, =)\ V(A_3|\mathbb{E}[A_3]) \Leftrightarrow V([-\tfrac{k}{2}; \tfrac{k}{2} + \tilde{\varepsilon}_1]|[0]) > (<, =)\ V([-\tfrac{k}{2} + \tilde{\varepsilon}_1; \tfrac{k}{2}]|[0]),$$

and hence

$$B_3|\mathbb{E}[B_3] \succ (\prec, \sim)\ A_3|\mathbb{E}[A_3] \Leftrightarrow [-\tfrac{k}{2}; \tfrac{k}{2} + \tilde{\varepsilon}_1]|[0] \succ (\prec, \sim)\ [-\tfrac{k}{2} + \tilde{\varepsilon}_1; \tfrac{k}{2}]|[0].$$

In order to prove part (iv) note that from the definitions of B_n and A_n it follows that $\mathbb{E}[B_n] = \mathbb{E}[A_n] = y$ if $\frac{n}{2} \in \mathbb{N} \wedge n \geq 2$ and that $\mathbb{E}[B_n] = \mathbb{E}[A_n] = y - \frac{k}{2}$ if $\frac{n}{2} \notin \mathbb{N} \wedge n \geq 3$. Consider first the case where $\frac{n}{2} \in \mathbb{N} \wedge n \geq 4$. Then, $B_n|\mathbb{E}[B_n] = [B_n - y]$ and $A_n|\mathbb{E}[A_n] = [A_n - y]$ holds for all $n \geq 2$ with $\frac{n}{2} \in \mathbb{N}$. Since $B_n = [A_{n-2}; B_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]$ and $A_n = [B_{n-2}; A_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]$ we get for any arbitrary $n \geq 4$ with $\frac{n}{2} \in \mathbb{N}$ that

$$\begin{aligned} B_n|\mathbb{E}[B_n] &= [A_{n-2} - y; B_{n-2} - y + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]|[0] \\ &= [A_{n-2}|\mathbb{E}[A_{n-2}]; B_{n-2}|\mathbb{E}[B_{n-2}] + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]|[0], \\ \text{and} \quad A_n|\mathbb{E}[A_n] &= [B_{n-2} - y; A_{n-2} - y + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]|[0] \\ &= [B_{n-2}|\mathbb{E}[B_{n-2}]; A_{n-2}|\mathbb{E}[A_{n-2}] + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]|[0]. \end{aligned}$$

Now consider the other case where $\frac{n}{2} \notin \mathbb{N} \wedge n \geq 5$. Then, $B_n|\mathbb{E}[B_n] = [B_n - (y - \frac{k}{2})]$ and $A_n|\mathbb{E}[A_n] = [A_n - (y - \frac{k}{2})]$ holds for all $n \geq 3$ with $\frac{n}{2} \notin \mathbb{N}$. Since $B_n = [A_{n-2}; B_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]$ and $A_n = [B_{n-2}; A_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]$ we get for any arbitrary $n \geq 5$ with $\frac{n}{2} \notin \mathbb{N}$ that

$$\begin{aligned} B_n|\mathbb{E}[B_n] &= [A_{n-2} - (y - \tfrac{k}{2}); B_{n-2} - (y - \tfrac{k}{2}) + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]|[0] \\ &= [A_{n-2}|\mathbb{E}[A_{n-2}]; B_{n-2}|\mathbb{E}[B_{n-2}] + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]|[0], \\ \text{and} \quad A_n|\mathbb{E}[A_n] &= [B_{n-2} - (y - \tfrac{k}{2}); A_{n-2} - (y - \tfrac{k}{2}) + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]|[0] \\ &= [B_{n-2}|\mathbb{E}[B_{n-2}]; A_{n-2}|\mathbb{E}[A_{n-2}] + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]|[0]. \end{aligned}$$

Lemma 2.16 showed that $M(B_n) = M(A_n)\ \forall n \geq 2$. From the fact that $\mathbb{E}[\tilde{\varepsilon}_i] = 0$ always holds and the linearity of $m(\cdot)$ it follows that $M([A_{n-2}|\mathbb{E}[A_{n-2}]; B_{n-2}|\mathbb{E}[B_{n-2}] + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]) = M([B_{n-2}|\mathbb{E}[B_{n-2}]; A_{n-2}|\mathbb{E}[A_{n-2}] + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}])$, and we can therefore state that

$$\begin{aligned} & V(B_n|\mathbb{E}[B_n]) > (<,=)\ V(A_n|\mathbb{E}[A_n]) \\ \Leftrightarrow \quad & V([A_{n-2}|\mathbb{E}[A_{n-2}]; B_{n-2}|\mathbb{E}[B_{n-2}] + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]|[0]) \\ > (<,=) \quad & V([B_{n-2}|\mathbb{E}[B_{n-2}]; A_{n-2}|\mathbb{E}[A_{n-2}] + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]|[0]), \end{aligned}$$

and hence

$$\begin{aligned} B_n|\mathbb{E}[B_n] \succ (\prec,\sim)\ A_n|\mathbb{E}[A_n] \quad \Leftrightarrow \quad & [A_{n-2}|\mathbb{E}[A_{n-2}]; B_{n-2}|\mathbb{E}[B_{n-2}] + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]|[0] \\ & \succ (\prec,\sim)\ [B_{n-2}|\mathbb{E}[B_{n-2}]; A_{n-2}|\mathbb{E}[A_{n-2}] + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]|[0]. \end{aligned}$$

■

Proof. [Proposition 2.11] Since $U(B_1|A_1) = u(m(y) - m(y-k)) = u(k)$ and $U(A_1|B_1) = u(m(y-k) - m(y)) = u(-k)$ (because of the linearity of $m(\cdot)$), it follows that $V(B_1|A_1) = M(B_1) + U(B_1|A_1) = m(y) + u(k)$ and $V(A_1|A_1) = M(A_1) + U(A_1|B_1) = m(y-k) + u(-k)$. Since $m(\cdot)$ and $u(\cdot)$ are strictly increasing and $k > 0$ it follows that $V(B_1|A_1) > V(A_1|B_1)$ and therefore $B_1|A_1 \succ A_1|B_1$. ■

Proof. [Proposition 2.12] Evaluating $B_2|A_2$ by $U(\cdot)$ delivers

$$U(B_2|A_2) = \int u(m(y) - m(y + \tilde{\varepsilon}_1'))dF(\tilde{\varepsilon}_1') = \int u(\tilde{\varepsilon}_1')dF(\tilde{\varepsilon}_1').$$

Evaluating $A_2|B_2$ by $U(\cdot)$ delivers

$$U(A_2|B_2) = \int u(m(y + \tilde{\varepsilon}_1) - m(y))dF(\tilde{\varepsilon}_1) = \int u(\tilde{\varepsilon}_1)dF(\tilde{\varepsilon}_1).$$

The last equalities in both cases hold due to the symmetry of all $\tilde{\varepsilon}_i$ and the linearity of $m(\cdot)$. Since $\tilde{\varepsilon}_1$ and $\tilde{\varepsilon}_1'$ are identically distributed we can conclude that $U(B_2|A_2) = U(A_2|B_2)$ regardless of the shape of $u(\cdot)$. Using $M(B_2) = M(A_2)$ from Lemma 2.16 and $U(B_2|A_2) = U(A_2|B_2)$, it follows that $V(B_2|A_2) = V(A_2|B_2)$ and thus $B_2|A_2 \sim A_2|B_2$. Alternatively, we can also state that both gambles reduce to

$$[B_2|A_{2\langle A_2|B_2\rangle}]\,|[0] = [C]|[0] \quad \text{and} \quad [A_2|B_{2\langle B_2|A_2\rangle}]\,|[0] = [C]|[0] \tag{2.30}$$

when compared with each other, where C denotes any arbitrary gamble. ■

Proof. [Proposition 2.13] From Lemma 2.16 we know that $B_3|A_3 \succ (\prec, \sim)\ A_3|B_3 \Leftrightarrow U(B_3|A_3) > (<, =)\ U(A_3|B_3)$. Evaluating $B_3|A_3$ by $U(\cdot)$ yields

$$\begin{aligned} U(B_3|A_3) &= \iint \left[\tfrac{1}{4}u(m(y-k) - m(y)) + \tfrac{1}{4}u(m(y-k) - m(y-k+\tilde{\varepsilon}_1'))\right. \\ &\quad \left. + \tfrac{1}{4}u(m(y+\tilde{\varepsilon}_1) - m(y)) + \tfrac{1}{4}u(m(y+\tilde{\varepsilon}_1) - m(y-k+\tilde{\varepsilon}_1'))\right] dF(\tilde{\varepsilon}_1)dF(\tilde{\varepsilon}_1') \\ &= \iint \left[\tfrac{1}{4}u(-k) + \tfrac{1}{4}u(\tilde{\varepsilon}_1') + \tfrac{1}{4}u(\tilde{\varepsilon}_1) + \tfrac{1}{4}u(k+\tilde{\varepsilon}_1+\tilde{\varepsilon}_1')\right] dF(\tilde{\varepsilon}_1)dF(\tilde{\varepsilon}_1'), \end{aligned}$$

where the last equality holds because of the symmetry of all $\tilde{\varepsilon}_i$ and the linearity of $m(\cdot)$. Evaluating $A_3|B_3$ by $U(\cdot)$ delivers

$$\begin{aligned} U(A_3|B_3) &= \iint \left[\tfrac{1}{4}u(m(y) - m(y-k)) + \tfrac{1}{4}u(m(y) - m(y+\tilde{\varepsilon}_1'))+\right. \\ &\quad \left. \tfrac{1}{4}u(m(y-k+\tilde{\varepsilon}_1) - m(y-k)) + \tfrac{1}{4}u(m(y-k+\tilde{\varepsilon}_1) - m(y+\tilde{\varepsilon}_1'))\right] dF(\tilde{\varepsilon}_1)dF(\tilde{\varepsilon}_1') \\ &= \iint \left[\tfrac{1}{4}u(k) + \tfrac{1}{4}u(\tilde{\varepsilon}_1') + \tfrac{1}{4}u(\tilde{\varepsilon}_1) + \tfrac{1}{4}u(-k+\tilde{\varepsilon}_1+\tilde{\varepsilon}_1')\right] dF(\tilde{\varepsilon}_1)dF(\tilde{\varepsilon}_1'), \end{aligned}$$

where again the last equality holds because of the symmetry of all $\tilde{\varepsilon}_i$ and the linearity of $m(\cdot)$. Comparing $U(B_3|A_3)$ to $U(A_3|B_3)$ yields

$$\begin{aligned} U(B_3|A_3) \quad &> (<, =) \quad U(A_3|B_3) \Leftrightarrow \\ &\qquad \iint \left[\tfrac{1}{2}u(-k) + \tfrac{1}{2}u(k+\tilde{\varepsilon}_1+\tilde{\varepsilon}_1')\right] dF(\tilde{\varepsilon}_1)dF(\tilde{\varepsilon}_1') \\ &> (<, =) \quad \iint \left[\tfrac{1}{2}u(k) + \tfrac{1}{2}u(-k+\tilde{\varepsilon}_1+\tilde{\varepsilon}_1')\right] dF(\tilde{\varepsilon}_1)dF(\tilde{\varepsilon}_1'). \end{aligned}$$

From $\mathbb{E}[\tilde{\varepsilon}_1] = 0$ and $\mathbb{E}[\tilde{\varepsilon}'_1] = 0$ it follows that $\mathbb{E}[k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1] = k$ and $\mathbb{E}[-k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1] = -k$. With $m(\cdot)$ being linear it follows that $M([-k; k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1]) = M([k; -k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1])$. Thus, we can state that

$$V(B_3|A_3) > (<,=)\ V(A_3|B_3) \Leftrightarrow V([-k; k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1]|[0]) > (<,=)\ V([k; -k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1]|[0]),$$

and thus

$$B_3|A_3 \succ (\prec,\sim)\ A_3|B_3 \Leftrightarrow [-k; k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1]|[0] \succ (\prec,\sim)\ [k; -k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1]|[0].$$

Alternatively, we can also state that both gambles reduce to

$$[B_3|A_{3\langle A_3|B_3\rangle}]\,|[0] = [-k; k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1]|[0] \quad \text{and} \quad [A_3|B_{3\langle B_3|A_3\rangle}]\,|[0] = [k; -k + \tilde{\varepsilon}_1 + \tilde{\varepsilon}'_1]|[0]$$

when compared with each other. ■

Proof. [Proposition 2.14] Follows directly from Theorem 2.15. ■

Proof. [Theorem 2.15] In the following proof of Theorem 2.15 we again first show that a generalization of the first part holds and then that this generalization implies both parts of the theorem (where we use Lemma 2.16, $\tilde{\varepsilon}_i$ being symmetric with $\mathbb{E}[\tilde{\varepsilon}_i] = 0\ \forall i$, and the linearity of $m(\cdot)$). The general part is

$$\begin{aligned} & [B_n|A_{n\langle A_n|B_n\rangle}]\,|[0] \\ = & \left[B_{n-2}|A_{n-2\langle A_{n-2}|B_{n-2}\rangle}; A_{n-2}|B_{n-2\langle B_{n-2}|A_{n-2}\rangle} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\right]|[0], \quad (2.31) \\ & [A_n|B_{n\langle B_n|A_n\rangle}]\,|[0] \\ = & \left[A_{n-2}|B_{n-2\langle B_{n-2}|A_{n-2}\rangle}; B_{n-2}|A_{n-2\langle A_{n-2}|B_{n-2}\rangle} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\right]|[0]. \quad (2.32) \end{aligned}$$

To show this, note that $[B_n|A_{n\langle A_n|B_n\rangle}]\,|[0]$ and $[A_n|B_{n\langle B_n|A_n\rangle}]\,|[0]$ can be expressed as gambles containing only B_n and A_n.

$$\begin{aligned} B_n|A_n = [B_n - A_n]\,|[0] \quad &\text{and} \quad A_n|B_n = [A_n - B_n]\,|[0] \\ \Rightarrow [B_n|A_{n\langle A_n|B_n\rangle}]\,|[0] = [B_n - A_n]\,|[0] \quad &\text{and} \quad [A_n|B_{n\langle B_n|A_n\rangle}]\,|[0] = [A_n - B_n]\,|[0], \end{aligned}$$

and therefore

$$\begin{aligned} [B_{n-2}|A_{n-2\langle A_{n-2}|B_{n-2}\rangle}]\,|[0] &= [B_{n-2} - A_{n-2}]\,|[0] \quad (2.33) \\ \text{and} \quad [A_{n-2}|B_{n-2\langle B_{n-2}|A_{n-2}\rangle}]\,|[0] &= [A_{n-2} - B_{n-2}]\,|[0]. \quad (2.34) \end{aligned}$$

But $[B_n|A_{n\langle A_n|B_n\rangle}]\,|[0]$ and $[A_n|B_{n\langle B_n|A_n\rangle}]\,|[0]$ can also be expressed as gambles containing only B_{n-2} and A_{n-2} as $B_n = [A_{n-2}; B_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]$ and $A_n = [B_{n-2}; A_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}]$.

$$\begin{aligned}
B_n|A_n &= \Big[A_{n-2} - B_{n-2}; A_{n-2} - A_{n-2} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}; B_{n-2} - B_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}; \\
&\quad B_{n-2} - A_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\Big]\,|[0] \\
\text{and} \quad A_n|B_n &= \Big[B_{n-2} - A_{n-2}; B_{n-2} - B_{n-2} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}; A_{n-2} - A_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})}; \\
&\quad A_{n-2} - B_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\Big]\,|[0] \\
\Rightarrow \quad [B_n|A_{n\langle A_n|B_n\rangle}]\,|[0] &= \Big[A_{n-2} - B_{n-2}; B_{n-2} - A_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\Big]\,|[0] \qquad (2.35) \\
\text{and} \quad [A_n|B_{n\langle B_n|A_n\rangle}]\,|[0] &= \Big[B_{n-2} - A_{n-2}; A_{n-2} - B_{n-2} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\Big]\,|[0], \qquad (2.36)
\end{aligned}$$

since $\tilde{\varepsilon}_{\text{int}(\frac{n}{2})}$ and $\tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}$ are valued equally and therefore two of the elements cancel out when the two gambles are compared. Finally, inserting (2.33) and (2.34) into (2.35) and (2.36) delivers (2.31) and (2.32). Part (i) of Theorem 2.15 follows directly from (2.31) and (2.32) when we note that for $\frac{n}{2} \notin \mathbb{N} \Rightarrow \text{int}(\frac{n}{2}) = \frac{n-1}{2}$.

To prove part (ii) of Theorem 2.15 ($\frac{n}{2} \in \mathbb{N}$) we again use (2.31) and (2.32). Suppose $n = 4$, then

$$\begin{aligned}
[B_4|A_{4\langle A_4|B_4\rangle}]\,|[0] &= \Big[A_2|B_{2\langle B_2|A_2\rangle}; B_2|A_{2\langle A_2|B_2\rangle} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\Big]\,|[0] \quad (2.37) \\
\text{and} \quad [A_4|B_{4\langle B_4|A_4\rangle}]\,|[0] &= \Big[B_2|A_{2\langle A_2|B_2\rangle}; A_2|B_{2\langle B_2|A_2\rangle} + \tilde{\varepsilon}_{\text{int}(\frac{n}{2})} + \tilde{\varepsilon}'_{\text{int}(\frac{n}{2})}\Big]\,|[0]. \quad (2.38)
\end{aligned}$$

By inserting (2.30) into (2.37) and (2.38) and canceling $\tilde{\varepsilon}_2$ and $\tilde{\varepsilon}'_2$ in both gambles we then get

$$\begin{aligned}
&[B_4|A_{4\langle A_4|B_4\rangle}]|[0] = [C]|[0] \quad \text{and} \quad [A_4|B_{4\langle B_4|A_4\rangle}]|[0] = [C]|[0]. \qquad (2.39) \\
&\Rightarrow \qquad B_4|A_4 \quad \sim \quad A_4|B_4.
\end{aligned}$$

Since (2.39), by the same logic it must also hold that $[B_6|A_{6\langle A_6|B_6\rangle}]|[0] = [C]|[0]$ and $[A_6|B_{6\langle B_6|A_6\rangle}]|[0] = [C]|[0]$. Continuing this reasoning yields

$$\begin{aligned}
&[B_n|A_{n\langle A_n|B_n\rangle}]|[0] = [C]|[0] \quad \text{and} \quad [A_n|B_{n\langle B_n|A_n\rangle}]|[0] = [C]|[0] \\
&\Rightarrow \qquad B_n|A_n \quad \sim \quad A_n|B_n
\end{aligned}$$

for $\frac{n}{2} \in \mathbb{N}$ and C being any arbitrary gamble. ■

Chapter 3

Experimental Evidence on Higher-Order Risk Preferences[0]

3.1 Introduction

Higher-order risk preferences are crucial for economically important behavior under risk which is not solely characterizable with the second-order preference of risk aversion. Most prominently, the third-order preference of prudence (or downside risk aversion) has been identified to drive precautionary savings (see e.g. Leland, 1968; Sandmo, 1970; Kimball, 1990; or Eeckhoudt and Schlesinger, 2008), but has also been found to determine behavior in other domains, like insurance demand, auction bidding, bargaining, prevention, rent-seeking, or monitoring.[1] The fourth-order risk attitude of temperance (or outer risk aversion) has mostly been linked to behavior under background risk (see e.g. Kimball, 1992; Gollier and Pratt, 1996; or Eeckhoudt, Gollier, and Schlesinger, 1996), and since background risks are eminently present in everyday life, its importance is undisputed. While there is a lot of empirical field evidence on precautionary saving (for an overview see e.g. Carroll and Kimball, 2008; or Browning and Lusardi,

[0]This chapter is based on joint work with Maximilian Rüger.

[1]Although being not necessary *and* sufficient as for the case of precautionary savings, prudence has also been identified to play an important role for behavior in rent-seeking games (Treich, 2010), auctions (Eső and White, 2004), or bargaining interactions (White, 2008). It has been further applied to problems of principal-agent monitoring (Fagart and Sinclair-Desagné, 2007), global commons (Bramoullé and Treich, 2009), inventory management (Eeckhoudt, Gollier and Schlesinger, 1995), insurance demand (Fei and Schlesinger, 2008) or optimal prevention (Eeckhoudt and Gollier, 2005; or Courbage and Rey, 2006). It has also been used in a more normative context to derive an optimal ecological discount rate (Gollier, 2010) or to justify the so-called precautionary principle (Gollier, Jullien and Treich, 2000; and Gollier and Treich, 2003).

1996), experimental evidence on both derived behavior as well as direct measurement of risk attitudes beyond risk aversion is scarce.

In this chapter we present experimental evidence on higher-order risk preferences. More specifically, we investigate second- to fourth-order risk attitudes, and how they relate to each other, not only when decisions can yield gains but also when they lead to *real* losses. Many theories that have been proposed as alternatives to expected utility theory (EUT) predict a substantial difference in risk preferences concerning gains and losses. Our elicitation of higher-order risk preferences is based on the model-independent gamble definitions of Eeckhoudt and Schlesinger (2006) and our results are therefore not confined to EUT.

In contrast to the extensively studied concept of risk aversion, higher-order risk preferences have only recently found their way to the experimental literature. First *indirect* evidence on prudence has been provided by Ballinger, Palumbo, and Wilcox (2003). They showed in an experiment simulating life-cycle consumption and saving that individuals make precautionary savings, for which prudence is necessary and sufficient in EUT. Kocher, Pahlke, and Trautmann (2010) observe precautionary bidding in first-price auctions, again behavior for which prudence is necessary under the assumption of EUT. Harrison, List, and Towe (2007) found in a field experiment that the risk aversion of coin traders is significantly higher when they faced an additional background risk. This supports the notion that temperance is present. However, experiments are well suited not only to measure behavior *associated* with higher-order risk preferences, but also to analyze whether risk preferences exhibit the hypothesized patterns.

In doing so, another strand of literature has emerged providing rather *direct* experimental evidence on higher-order risk attitudes. Tarazona-Gomez (2004) elicits prudence based on certainty equivalents. While her elicitation method is only valid under strong assumptions within an EUT framework, she finds a modest amount of prudent subjects. All other existing approaches apply the model-independent gamble definitions of Eeckhoudt and Schlesinger (2006) and do therefore not require EUT. Intuitively, these definitions can be interpreted as apportioning risks properly, and thereby provide a compound representation of lottery choices.[2] Deck and Schlesinger (2010) test whether subjects apportion risks consistent with prudence and temperance. They find evidence

[2] For instance, an individual is said to be prudent if she prefers to add risk to the better rather than worse wealth state. Likewise, an individual is said to be temperate if she prefers to add risk to the less risky rather than more risky state. This interpretation of higher-order risk preferences stimulated the notion of proper risk apportionment. In addition, odd-order risk attitudes can be understood not only as a question on where to add risk, but also on where to add a sure reduction in wealth. If both risk and sure reductions in wealth are 'harms', the intuitive interpretation of higher-order risk preferences becomes an issue of where to add such harms.

for prudence but also against temperance (i.e. intemperance). Ebert and Wiesen (2011) test only for risk apportionment of the third order, but also find supportive evidence for prudence. Noussair, Trautmann, and van de Kuilen (2011) examine second-to fourth-order risk apportionment not only among student subjects, but also in a demographically representative sample. They find evidence for prudence as well as temperance. Assuming the status quo as reference point, all these studies have in common that they elicit preferences only in the gain domain.

Deck and Schlesinger (2010) conclude that common specifications of EUT, such as the classes of constant relative risk-averse (CRRA) or constant absolute risk-averse functions (CARA), are not consistent with their found combination of prudence and intemperance. However, popular specifications of cumulative prospect theory (Tversky and Kahneman, 1992) are consistent with both prudence and intemperance. Hence, they interpret their findings as support of cumulative prospect theory (CPT) rather than EUT.

Compared to EUT, one crucial difference of reference-dependent models, such as prospect theory, is the distinct evaluation of gains and losses. In experiments, it is usually difficult to impose losses on subjects. Most experiments endow subjects with enough money right before they impose 'losses' on them. Since subjects in these experiments cannot make real losses, it is hard to believe that they behave as if they would. The problem of 'house money' effects in economic experiments of risky choice is well-established since Thaler and Johnson (1990). The house money effect describes the behavior of subjects to treat all money earned in the experiment as money to play with. Consequently, they are willing to take risks they normally would not accept. This problem arises in any design where subjects are endowed with a participation fee or previous earnings and can subsequently only lose this amount of money. Therefore, the focus of our experimental design was to inflict *real* monetary losses on subjects.

Bosch-Domènech and Silvestre (2006) introduced a design where subjects have to attend two dates of the experiment. This design enabled them to impose real losses on the second date. However, they only considered risk aversion. To our knowledge, no experiment has been conducted so far that tests higher-order risk preferences in the domain of real monetary losses. In our experiment, we investigate second- to fourth-order risk preferences over gains *and* losses. Subjects attended two dates of the experiment. The purpose of the first date was to let subjects earn enough money which they possibly could lose at the second date, several weeks later. At the second date subjects made 84 binary choices that are also based on the gamble definitions of Eeckhoudt and Schlesinger (2006). However, while Deck and Schlesinger (2010), Ebert and Wiesen

(2011), and Noussair, Trautmann, and van de Kuilen (2011) asked subjects where they wanted to add 'harms', we rather presented the gambles in final outcomes. We chose this design, as it is well established in the literature that in experiments subjects have major difficulties in solving compound lotteries (see e.g. Bar-Hillel, 1973; Kahneman and Tversky, 1979 in their discussion of the isolation effect; Bernasconi, 1994; or Friedman, 2005).

Our results indicate that the prevalent preference pattern exhibits risk aversion, prudence, and temperance both at an aggregate level and on the level of the individual subject. Third- and fourth-order risk preferences seem to be equally present as second-order risk preferences. While our result on prudence closely resembles the results of previous experiments, our results on temperance are in stark contrast to those of Deck and Schlesinger (2010). Since we elicit temperance using 28 choices by each subject while Deck and Schlesinger (2010) used only four choices by each subject, our result provides some confidence that temperance rather than intemperance is the prevailing preference pattern. As a consequence, we do not reject the possibility that common specifications of EUT can be consistent with the data. This conclusion is additionally supported by two further results of our experiment. Probably surprisingly, we find no variation of elicited risk preferences between the domain of gains and losses. Although this result does not speak against the concept of a reference point or loss aversion, it indicates that the shape of the value function does not change radically at the reference point. Furthermore, we find that risk preferences of different orders are empirically highly correlated, another feature of common EUT specifications. In other words, subjects who show a higher level of consistency with risk-averse behavior are also more consistent with prudent and temperate behavior.

This chapter is structured as follows. Section 3.2 provides some theoretical prerequisites which are needed in order to understand what we exactly measure in our experiment. The experimental design and the choices subjects had to make in our experiment are described in Section 3.3. Subsequently, we present and discuss our results in Section 3.4 and conclude in Section 3.5. In Appendix A we give a detailed description of all decisions in our experiment, and our experimental instructions can be found in Appendix B.

3.2 Theoretical Prerequisites

In this section we outline some basic concepts and definitions that clarify what is precisely measured in our experiment. This illustrates what is meant by stating that individuals exhibit certain risk preferences.

Let $y \in \mathbb{R}$ be the level of wealth. Let $k \in \mathbb{R}$ (with $k > 0$) be a sure reduction in wealth. $\tilde{\varepsilon}_1$ and $\tilde{\varepsilon}_2$ are random variables which are non-degenerate, independent of all other random variables that affect wealth, and have $\mathbb{E}[\tilde{\varepsilon}_1] = \mathbb{E}[\tilde{\varepsilon}_2] = 0$. Following Eeckhoudt and Schlesinger (2006), we can now define the following standard gambles, where each element denotes an outcome and each outcome of a specific gamble is realized with equal probability.[3] Such defined outcomes can be either deterministic or stochastic themselves.

$$\begin{array}{ll} B_2 \equiv [y] & A_2 \equiv [y + \tilde{\varepsilon}_1] \\ B_3 \equiv [y - k, y + \tilde{\varepsilon}_1] & A_3 \equiv [y, y - k + \tilde{\varepsilon}_1] \\ B_4 \equiv [y + \tilde{\varepsilon}_1, y + \tilde{\varepsilon}_2] & A_4 \equiv [y, y + \tilde{\varepsilon}_1 + \tilde{\varepsilon}_2] \end{array}$$

Intuitively, B_2 vs. A_2 is the comparison between a certain level of wealth and a risky alternative with an identical mean. B_3 vs. A_3 is equivalent to the question whether one prefers to add an unavoidable random variable to the higher or lower outcome. It is also equivalent to the question whether one prefers to accept a sure reduction in wealth in the certain or uncertain state. B_4 vs. A_4 can be seen as the question whether one prefers to aggregate or disaggregate two independent random variables. Although such a compound interpretation has the virtue of capturing the intuition of higher-order risk apportionment, the following definition is not based on these compound characteristics.

Definition 3.1 [Eeckhoudt and Schlesinger, 2006] *Individual i is*

$$\begin{array}{llllll} \textit{risk-averse} & \Leftrightarrow & B_2 \succsim^i A_2 & \forall\, y, \tilde{\varepsilon}_1 & (\Leftrightarrow u_i''(x) & \leq 0 \textit{ in EUT models}), \\ \textit{prudent} & \Leftrightarrow & B_3 \succsim^i A_3 & \forall\, y, \tilde{\varepsilon}_1, k & (\Leftrightarrow u_i'''(x) & \geq 0 \textit{ in EUT models}), \\ \textit{temperate} & \Leftrightarrow & B_4 \succsim^i A_4 & \forall\, y, \tilde{\varepsilon}_1, \tilde{\varepsilon}_2 & (\Leftrightarrow u_i''''(x) & \leq 0 \textit{ in EUT models}). \end{array}$$

Note that these definitions are deterministic. They define preferences, which are underlying characteristics of the subjects. There are several approaches to relate these preferences to observations. One way to proceed would be to state that an individual exhibits a certain risk preference only if *all* observed actions are in accordance with its definition. However, then a single observation contradicting the definition would

[3] We defined all gambles having a mean of y if $k = 0$ instead of 0 as in Eeckhoudt and Schlesinger (2006). We do so because only then gambles can completely lie in the domain of gains or losses.

lead to a rejection of the concept. For instance, somebody who is choosing B_3 over A_3 in 27 out of 28 decisions, but chooses A_3 once, would be termed neither prudent nor imprudent. It becomes clear, both from our experimental results as well as from those of Deck and Schlesinger (2010), Ebert and Wiesen (2011), and Noussair, Trautmann, and van de Kuilen (2011) that with such a strict application of Definition 3.1, almost no subject could be classified as having a certain risk preference. However, we think that the concepts of risk aversion, prudence and temperance still are valuable to characterize individuals who often but not always act according to the definitions above.

This leads us to a stochastic definition of higher-order risk attitudes, replacing the deterministic definition. Deck and Schlesinger (2010), Ebert and Wiesen (2011), and Noussair, Trautmann, and van de Kuilen (2011) do so implicitly when making inferences from their results. We state these stochastic definitions explicitly in order to clarify the concepts used to analyze the data. The notion that deterministic theories only represent the structural part of preferences whereas the entire system of preferences also incorporates a random or stochastic part is widespread among researchers who try to closely match theories and empirical observations. The fact that subjects of an experiment sometimes make different choices on the very same decision problem and under the same conditions is only reconcilable with stochastic theories.[4] In this regard, Wilcox (2007) provides a comprehensive survey on the different variants of stochastic utility models.

Let P_n^i be the probability that individual i prefers B_n over A_n for a given set of parameters. Then it follows that P_2^i is the probability that individual i chooses B_2 over A_2 for given y and $\tilde{\varepsilon}_1$. Similarly, P_3^i is the probability that individual i chooses B_3 over A_3 for given y, $\tilde{\varepsilon}_1$, and k. And finally, P_4^i is the probability that individual i chooses B_4 over A_4 for given y, $\tilde{\varepsilon}_1$, and $\tilde{\varepsilon}_2$. Stochastic risk preferences are then defined in the following way.

Definition 3.2 *Individual i is*

stochastically risk-averse (-seeking, -neutral)	$\Leftrightarrow$	P_2^i	$> (<, =)$	$1/2$,
stochastically prudent (imprudent, neither nor)	$\Leftrightarrow$	P_3^i	$> (<, =)$	$1/2$,
stochastically temperate (intemperate, neither nor)	$\Leftrightarrow$	P_4^i	$> (<, =)$	$1/2$.

Definition 3.2 will allow us to make statements on the risk attitudes of a single individual. Sometimes, however, we will make interpersonal comparisons. For such

[4] On experimental evidence on this issue see e.g. Camerer (1989), Starmer and Sugden (1989), Ballinger and Wilcox (1997), or Loomes and Sugden (1998).

comparisons we need a definition that allows us to order subjects in terms of the level of consistency with a certain risk attitude.

Definition 3.3 *Individual i is*

$$\begin{array}{llcll} \textit{stochastically more risk-averse} & \textit{than} \; j & \Leftrightarrow & P_2^i > P_2^j & \forall\, y, \tilde{\varepsilon}_1, \\ \textit{stochastically more prudent} & \textit{than} \; j & \Leftrightarrow & P_3^i > P_3^j & \forall\, y, \tilde{\varepsilon}_1, k, \\ \textit{stochastically more temperate} & \textit{than} \; j & \Leftrightarrow & P_4^i > P_4^j & \forall\, y, \tilde{\varepsilon}_1, \tilde{\varepsilon}_2. \end{array}$$

The definition of stochastically more risk-averse as stated in Definition 3.3 was formally first stated by Hilton (1989). Recently, Wilcox (2011) extended the definition for stochastic y. The definitions of stochastically more prudent and stochastically more temperate were not stated explicitly before but are a straightforward extension of the stochastically more risk-averse definition to higher orders.

It should be emphasized that the concepts of Definition 3.3 are not to be confounded with the definitions of higher intensities of risk aversion (or parallel concepts of higher downside and outer risk aversion). Here, we define the level of consistency with a qualitative concept whereas in the literature on the intensities of risk preferences one is concerned with the quantitative strength of a risk preference. Only under very specific conditions these concepts relate to each other in a one-to-one fashion. If we allow the non-systematic part of preferences to differ among subjects (e.g. some subjects make more errors than others) the two measures may yield different orderings of subjects. For instance, it is well conceivable that a subject is very confident about rejecting a mean-preserving spread but is not willing to give up a lot in order to avoid one. Another subject might be ready to pay a lot to avoid such a risk in principle but has not thought about such decisions much and is therefore prone to make many mistakes.

Definitions 3.2 and 3.3 are still in terms of underlying characteristics of an individual. However, using our observations, we can make statistical inferences on these unknown characteristics of the subjects. When labeling a subject as risk averse, prudent or temperate in the subsequent analysis we imply these stochastic definitions.

3.3 The Experiment

The experiment was divided into two distinct dates for each subject. This was necessary in order to induce real monetary losses. Institutional restraints ruled out that in a single date-experiment subjects could make real monetary losses not compensated by any fixed payment such as the participation fee. However, by dividing the experiment into two dates we were able to induce real monetary losses at the second date.

At the first date everybody made real monetary gains. These gains were the lower bound of losses we could inflict at the second date. By letting the second date be separated from the first date by several weeks, it seems likely that the gains of the first date were internalized by the subjects. Furthermore, the first date demanded the completion of real tasks by the subjects. The real gains of the first task thus were likely perceived as rightfully earned and not as house money to gamble with at the second date. It is therefore reasonable to assume that the status quo prior to the second date is the relevant reference point for decisions made during the second date. A similar approach to inflict real monetary losses was employed by Bosch-Domènech and Silvestre (2006). They restricted their experiment to the elicitation of risk aversion. Our results for risk aversion resemble their results closely, but higher-order risk preferences in the domain of real monetary losses have, to our knowledge, not been investigated before.

Both dates of the computer-based experiment were conducted at the MELESSA laboratory of the University of Munich. Subjects (graduate students were excluded) were recruited using the software ORSEE by Greiner (2004). At both dates subjects first received written instructions that were read privately by them. Then, they answered control questions ensuring comprehension of the instructions. There was no time limit for the instructions and subjects had the opportunity to ask questions in private. The experiment started on the computer screen only after everybody had answered the control questions correctly and there were no further questions. All questions in the experiment were displayed and answered through a computer interface. The interfaces were programmed using the z-Tree software of Fischbacher (2007). All random processes that determined the monetary outcomes were realized through the drawings of numbered balls from urns by the subjects themselves.[5] The two dates of our experiment are further explained below.

3.3.1 First Date

72 subjects participated at the first date of the experiment, which took place on November 25, 2008. In each of the three sessions 24 subjects participated. They were randomly assigned to seats. The first task subjects conducted was to fill out tables measuring intensities of risk preferences. These tables are not analyzed in this chapter. However, it is of importance that the subjects carried out a meaningful task, so that the payment at the end of the first date are interpreted as rightfully earned and are not

[5] We chose not to let the randomization be done by a computer because we agree with Harrison and Rutsröm (2008, p133): "In our experience subjects are suspicious of randomization generated by computers. Given the propensity of many experimenters in other disciplines to engage in deception, we avoid computer randomization whenever feasible."

perceived as 'windfall gains' by the subjects. After filling out the tables, subjects further answered a detailed questionnaire on personal characteristics. The most important feature of the first date for the present analysis is that subjects received payments between 23.30 and 29.20 euros. These were payed out in cash, in private, and not by the experimenters. The payments at the first date include both the participation fee for the first and for the second date of the experiment. Answering the tasks and the questionnaire, drawing balls from urns, and paying participants took around 60 minutes.

3.3.2 Second Date

The second date took place exactly three weeks after the first date, on December 16, 2008, in the same room as the first date, and with an equal time duration. Again, subjects were randomly assigned to seats. 67 participants (93%) showed up at the second date. The five subjects who did not show up at the second date could in principle have decided not to show up because of fundamentally different risk preferences compared with the other participants. However, we have detailed information on the subjects to check whether this is the case. First, we have the detailed answers to the questionnaires and second, we have the responses to the tables eliciting the intensities of risk preferences. We found no systematic differences in the answers of those subjects participating only at the first date.

Each subject made 84 choices at the second date of the experiment. Each choice was a binary decision between two options, called option A and option B. The framing of these choices was always in final outcomes. That is, each option was represented only by the final outcomes that could result if the option was played out. For a given decision, all stated outcomes of a decision were realized with equal probability. Thus, a gamble can be fully represented by the outcomes without stating probabilities separately. In the instructions, the options were described in a way that even subjects who are not familiar with the basic notions of probabilities could understand how the outcome was determined. Only the comprehension that each ball was equally likely to be drawn from an urn was needed. To assist those subjects who are more familiar with representations in terms of probabilities we provided additionally an alternative explanation. The order of appearance of the 84 decisions was randomly chosen for each of the three sessions, hence we can test whether the order of appearance influences results.

A key feature of the decisions was that their outcomes lie either in the domain of gains, in the domain of losses or in a mixed domain including both. We classify a decision as being a decision in gains if all final outcomes that can potentially occur regardless of the actual choice have positive values. Likewise, we define decisions as being in losses if all outcomes have negative values. We define a decision as being over a mixed domain if both positive and negative outcome values can occur.

28 of the 84 decisions were designed to elicit risk aversion. With $\tilde{\varepsilon}_1 = [-\varepsilon_1, \varepsilon_1]$ the definition of risk aversion becomes

$$B_2 = [y] = [y, y] \succsim [y - \varepsilon_1, y + \varepsilon_1] = [y + \tilde{\varepsilon}_1] = A_2$$

In order to make both B_2 and A_2 more comparable in terms of complexity we displayed B_2 also as a two-outcome of equal probability gamble, and therefore chose the representation $B_2 = [y, y]$. Table 3.6 in Appendix A describes the 28 gambles to elicit risk aversion both in terms of chosen parameters (y and ε_1) and in terms of final outcomes displayed (A1, A2, B1 and B2). The values of y were drawn from the range of -10 to 62 euros, and y also equals the expected value of both B_2 and A_2. The values of ε_1 were drawn from the range of 2 to 32 euros, and ε_1 is also the standard deviation of A_2. This leads to final outcomes in the range of -12 to 79 euros. Column 8 of Table 3.6 in Appendix A states whether the decision lies in the domain of gains, losses, or in mixed domains (indicated by a G, L, or M, respectively). A decision eliciting risk aversion is in gains if $y - \varepsilon_1 > 0$, in losses if $y + \varepsilon_1 < 0$, and in mixed domains if neither is the case. Of the 28 decisions to elicit risk aversion, 13 were in gains, seven in losses, and eight in mixed domains. Table 3.6 in Appendix A further displays the order of appearance of the respective decision in the three sessions (in columns 9, 10, and 11). The last column of Table 3.6 in Appendix A shows the percentage of risk-averse choices for each of the 28 decisions.

Another 28 of the 84 decisions were designed to elicit prudence. With $\tilde{\varepsilon}_1 = [-\varepsilon_1, \varepsilon_1]$ the definition of prudence can be stated as

$$\begin{aligned} & B_3 = [y - k, y + \tilde{\varepsilon}_1] = [y - k, y - k, y - \varepsilon_1, y + \varepsilon_1] \\ \succsim \; & [y, y, y - k - \varepsilon_1, y - k + \varepsilon_1] = [y, y - k + \tilde{\varepsilon}_1] = A_3 \end{aligned}$$

Table 3.7 in Appendix A describes the 28 gambles to elicit prudence in terms of parameters (y, k, and ε_1) and in terms of final outcomes (A1 - A4, and B1 - B4) which were actually displayed to the subjects. y was drawn from the range of -8 to 67 euros, k from the range of 2 to 15 euros, and ε_1 from the range of 2 to 21 euros. The expected value of both B_3 and A_3 is $y - k/2$ and ranges from -10 to 61.50 euros. The standard deviation of both B_3 and A_3 is $\varepsilon_1/2$ and ranges from 1 to 10.50 euros. This leads to final outcomes in the range of -12 to 72 euros. Column 13 of Table 3.7 in Appendix

A states whether the decision lies in the domain of gains, losses, or in mixed domains. Decisions eliciting prudence are in gains if $y - k - \varepsilon_1 > 0$, in losses if $y + \varepsilon_1 < 0$, and in mixed domains if neither is the case. Of the 28 decisions to elicit prudence ten were in gains, seven in losses, and eleven were in mixed domains. Columns 14, 15, and 16 of Table 3.7 in Appendix A show the orders of appearance of the respective decision in the three sessions, and the last column shows the percentage of prudent choices for each decision.

Finally, 28 of the decisions were used to elicit temperance. With $\tilde{\varepsilon}_1 = [-\varepsilon_1, \varepsilon_1]$ and $\tilde{\varepsilon}_2 = [-\varepsilon_2, \varepsilon_2]$ the definition of temperance can be stated as

$$\begin{aligned} B_4 &= [y + \tilde{\varepsilon}_1, y + \tilde{\varepsilon}_2] = [y - \varepsilon_1, y + \varepsilon_1, y - \varepsilon_2, y + \varepsilon_2] \\ &= [y - \varepsilon_1, y - \varepsilon_1, y + \varepsilon_1, y + \varepsilon_1, y - \varepsilon_2, y - \varepsilon_2, y + \varepsilon_2, y + \varepsilon_2] \\ \succsim\ & [y, y, y, y, y - \varepsilon_1 - \varepsilon_2, y - \varepsilon_1 + \varepsilon_2, y + \varepsilon_1 - \varepsilon_2, y + \varepsilon_1 + \varepsilon_2] \\ &= [y, y + \tilde{\varepsilon}_1 + \tilde{\varepsilon}_2] = A_4 \end{aligned}$$

We again used the representation of $[y-\varepsilon_1, y-\varepsilon_1, y+\varepsilon_1, y+\varepsilon_1, y-\varepsilon_2, y-\varepsilon_2, y+\varepsilon_2, y+\varepsilon_2]$ for B_4 in order to make B_4 and A_4 gambles with an equal number of outcomes and thus to make them similarly complex. Table 3.8 in Appendix A describes the 28 gambles to elicit temperance in terms of parameters (y, ε_1, and ε_2) and in terms of displayed final outcomes (A1 - A8, and B1 - B8). y was drawn from the range of -8 to 62 euros and it also equals the expected value of B_4 and A_4. The values for ε_1 were drawn from the range of 2 to 22 euros, and those of ε_2 from the range of 2 to 12 euros. Without loss of generality we chose $\varepsilon_1 \geq \varepsilon_2$. The standard deviation of B_4 and A_4 is $\sqrt{(\varepsilon_1^2 + \varepsilon_2^2)/2}$ and ranges from 2 to $\sqrt{246.5} \approx 15.70$ euros. As a consequence, the final outcomes range from -12 to 84 euros. Again, the subsequent column states whether the decision lies in the domain of gains, losses, or in mixed domains. A decision eliciting temperance is in gains if $y - \varepsilon_1 - \varepsilon_2 > 0$, in losses if $y + \varepsilon_1 + \varepsilon_2 < 0$, and in mixed domains if neither is the case. Of the 28 decisions to elicit temperance ten were in gains, seven in losses, and eleven in mixed domains. Also displayed are the orders of appearance of the respective decision in the three sessions and the percentage of temperate choices for each decision.

3.4 Results

In total, we observed 5628 binary choices that can be used in our empirical analysis, 84 choices for each of the 67 subjects. In the following, we first analyze aggregate choices which treats all choices as independent events. This gives us a first general view of the results. However, it neglects the fact that subsets of 84 choices each are the actions of a single individual. We therefore focus in the remaining part of this section on a within-subject analysis. The potential of an actual within-subject analysis is the benefit we receive (compared to other related experiments) in exchange for having subjects exposed to answer 84 decisions. Our within-subject analysis proceeds in three steps. We first investigate whether risk preferences are influenced by the domain where we distinguish between the gain, loss, and mixed domain (includes potential gains *and* losses). Then, we look at the sign of subjects' risk preferences for each order separately. And finally, we shed light on the relations of the qualitative strengths or consistency of risk preferences across orders. A robustness analysis and a discussion on the predictive power of CPT and EUT is provided at the end of this section.

3.4.1 Pooling Subjects

Table 3.1 outlines our results if we pool all subjects. The second column of Table 3.1 states in how many decisions B_n was chosen over A_n for each $n = 2, 3, 4$. In this column no distinction between the decisions were made other than whether they were designed to elicit risk aversion, prudence, or temperance. The hypothesis that the probability of making a risk-averse choice in a second-order decision, a prudent choice in a third-order decision, or a temperate choice in a fourth-order decision is equal or lower than 1/2 is rejected by binomial tests at the 1%-level for all three tests ($p = 0.0000$; $N = 1876$).

Table 3.1: Descriptive Results when Pooling Subjects

Choices Eliciting	Across Domains	Within Gains	Within Losses	In Mixed Gambles
Risk Aversion	56 %	55 %	57 %	57 %
(N)	(1876)	(871)	(469)	(536)
Prudence	56 %	60 %	55 %	52 %
(N)	(1876)	(670)	(469)	(737)
Temperance	56 %	58 %	54 %	56 %
(N)	(1876)	(670)	(469)	(737)

In decisions eliciting third-order risk preferences, 56% of the choices are in favor of the prudent choice, that is B_3 was preferred over A_3. Deck and Schlesinger (2010), Ebert and Wiesen (2011), and Noussair, Trautmann, and van de Kuilen (2011) also observed a majority of prudent choices. At first glance their results indicate a higher consistency with the concept of prudence since they found 61% (in Deck and Schlesinger, 2010), 65% (in Ebert and Wiesen, 2011), and 69% (in Noussair, Trautmann, and van de Kuilen, 2011) of choices in favor of the prudent alternative. Note, however, that in their experiments subjects could only make real monetary gains. The result which is best comparable to their analysis is therefore reported in the third column of Table 3.1. Here, we find that 60% of choices in gains are in favor of the prudent choice, which is closer to the results of the other experiments.

In decisions eliciting fourth-order risk attitudes, 56% of the choices are in favor of the temperate choice. This stands in contrast to the results of Deck and Schlesinger (2010). They report that only 38% of the choices favor the temperate choice and conclude that individuals are on average intemperate. In their experiment, only four decisions for each subject elicited temperance. Noussair, Trautmann, and van de Kuilen (2011) used five decisions for each subject and found that 60% of the choices were temperate. Our result on temperance, based on 28 decisions for each subject, adds confidence that the prevalent choice pattern indeed exhibits temperance. There are two reasons why we expect this result to be robust. First, our design was in final outcomes only and thus could be very easily understood by subjects and second, our analysis of temperance rests on 1876 observations while Deck and Schlesinger (2010) only used 396 observations to elicit temperance. This result could have strong consequences since the results of Deck and Schlesinger (2010) on intemperance led them to conclude that CPT better explains data on higher-order risk attitudes than EUT with CRRA or CARA functions. When choosing only temperate choices in gains as the appropriate comparison to the results of Deck and Schlesinger (2010) the gap between our and their result even broadens by two percentage points.

A striking feature of the results in Table 3.1 is the fact that the percentage of risk-averse over risk-seeking, of prudent over imprudent, and of temperate over intemperate choices are so close that they all are rounded to 56% (the more exact percentages are 56.24% risk-averse, 55.70% prudent, and 56.02% temperate choices).[6] Later, we will explore whether this stems from the fact that the same individuals who are risk averse, are also prudent and temperate. But it already indicates that third- and fourth-order risk preferences are equally present as second-order risk preferences among subjects.

Based on these first results we can summarize the following. If we would accept the assumptions that individuals are homogeneous and observations are independent, we could state with a high degree of confidence that the behavior of the representative subject exhibits stochastic risk aversion, stochastic prudence, and stochastic temperance.

[6]If we compute the 90% confidence intervals of the underlying population percentages we receive 54.14%-58.14%, 53.78%-57.61%, and 54.11%-57.93%, respectively. The fact that these intervals widely overlap again highlights the possibility that the underlying percentage of choosing the 'averse' option may well be equal or very close for different orders.

This first picture, however, is based on risk preferences across domains. Columns 3 to 5 of Table 3.1 state for each domain separately in how many decisions B_n was chosen over A_n (for each $n = 2, 3, 4$). Choices eliciting risk aversion are very homogeneous across domains. Also, choices eliciting temperance do not show much variation. Only choices eliciting prudence seem to *slightly* depend on the domain they lie in. We will make more precise statements in the following section which explores the relations of higher-order risk preferences within different domains in more detail. For each of the nine cells distinguishing domains in Table 3.1, we conduct a binomial test to determine whether our observed pattern across domains can also be observed within different domains. For all these nine cells, i.e. for all domains and orders, we can reject the hypothesis that the underlying probability of choosing B_n is equal or below 1/2 (with one marginally insignificant exception for prudence in mixed domains).[7] Based on all twelve binomial tests, we can now formulate our first result.

Result 3.1 *Aggregated choices over all subjects represent stochastic risk aversion, prudence, and temperance not only across domains but also within the gain, loss, and mixed domain.*

While Result 3.1 is based on pooling subjects, the remaining results are based on a within-subject analysis.

3.4.2 Comparing Domains

We now examine the question whether risk preferences differ according to the domain the gambles are defined over: gains, losses, or mixed domains. This is of interest because domain-specific risk preferences are a key feature of many non-EUT models, including CPT. For each subject, we compare how often B_n was chosen across different domains, separately for each n.

We start our analysis with risk aversion. Since every subject answered all 28 questions eliciting risk aversion we can use two-sided Wilcoxon signed-rank tests in order to test the null hypothesis that the underlying distribution of risk aversion is the same in the domain to the left of a cell than in the domain above the cell in Table 3.2. We find no pairwise comparison of domains where we can reject the null at a 10%-significance

[7] The corresponding p-values and number of independent observations N of the one-sided binomial tests are the following: $p = 0.0014$, $N = 871$ for risk aversion/gains; $p = 0.0000$, $N = 670$ for prudence/gains; $p = 0.0000$, $N = 670$ for temperance/gains; $p = 0.0008$, $N = 469$ for risk aversion/losses; $p = 0.0104$, $N = 469$ for prudence/losses; $p = 0.0697$, $N = 871$ for temperance/losses; $p = 0.0006$, $N = 536$ for risk aversion/mixed; $p = 0.1052$, $N = 737$ for prudence/mixed; and $p = 0.0010$, $N = 737$ for temperance/mixed.

Table 3.2: Wilcoxon Signed-Rank Tests: Does Risk Aversion Depend on Domains?

p-value / z-statistic / (N)	Losses	Mixed
Losses		0.6482 / 0.456 / (67)
Gains	0.1822 / -1.334 / (67)	0.4095 / -0.825 / (67)

level. The literature on prospect theory found in numerous experiments that individuals are risk-seeking in losses. However, these conclusions were predominantly based on experiments with hypothetical choices or subject to house-money effects. In contrast, the experiment of Bosch-Domènech and Silvestre (2006) finds a similar pattern as we do, subjects are risk-averse for (sufficiently large) losses. Since they also inflict real monetary losses on subjects, this feature seems to be driving our probably surprising result.

Table 3.3: Wilcoxon Signed-Rank Tests: Does Prudence Depend on Domains?

p-value / z-statistic / (N)	Losses	Mixed
Losses		0.4860 / 0.697 / (67)
Gains	0.1981 / 1.287 / (67)	0.0176 / 2.374 / (67)

Table 3.3 reports the results of two-sided Wilcoxon signed-rank tests for the distribution of prudence in different domains. The only statistically significant effect of domains on any measure of risk preferences is found for prudence, between the distributions of prudence in the gain and in the mixed domain. This confirms our conjecture from the descriptive results of Table 3.1, which already indicated less prudence in the mixed than in the gain domain. In contrast, it cannot be rejected that the distributions of prudence in gains and in losses are identical. Neither can the hypothesis be rejected that the distributions in the loss and in the mixed domain are identical.

Last, Table 3.4 displays two-sided Wilcoxon signed-rank test results for the distribution of temperance across domains. Similar to the analysis of risk aversion, we cannot reject the hypothesis that the underlying distributions of fourth-order risk preferences in different domains are identical. To summarize, we do not find that higher-order

Table 3.4: Wilcoxon Signed-Rank Tests: Does Temperance Depend on Domains?

p-value z-statistic (N)		Mixed
	Losses	0.3834 -0.872 (67)
Gains	0.1507 1.437 (67)	0.6260 0.487 (67)

risk preferences in the loss domain are any different than in the gain domain. Also, risk preferences in the mixed domain do not differ from those in one of the two pure domains, with the exception prudence in the gain vs. mixed domain.

Result 3.2 *Risk preferences of any order seem not to depend on the domain. Subjects have the same stochastic risk preference in the loss and in the gain domain. Those who are stochastically more risk-averse (prudent, temperate) in the gain domain are also stochastically more risk-averse (prudent, temperate) in the loss domain.*

Although being consistent with EUT, Result 3.2 should by no means be interpreted as a rejection of the general concept of a reference point. The concept of loss aversion is not systematically testable in our experimental design.[8] Also, we are not in a position to test whether the strength of risk preferences depends on the distance of outcomes to a reference point. What we can conclude is that the direction of the qualitative concepts of risk aversion, prudence and temperance seems not to depend on the domain of the decision. Therefore, the functional form of the value function seems not to change radically at the reference point.

Note that Result 3.2 makes no statement about the sign of risk preferences. Being stochastically more risk-averse (prudent, temperate) is equivalent to being stochastically less risk-seeking (imprudent, intemperate). Next, we focus on the sign of risk preferences.

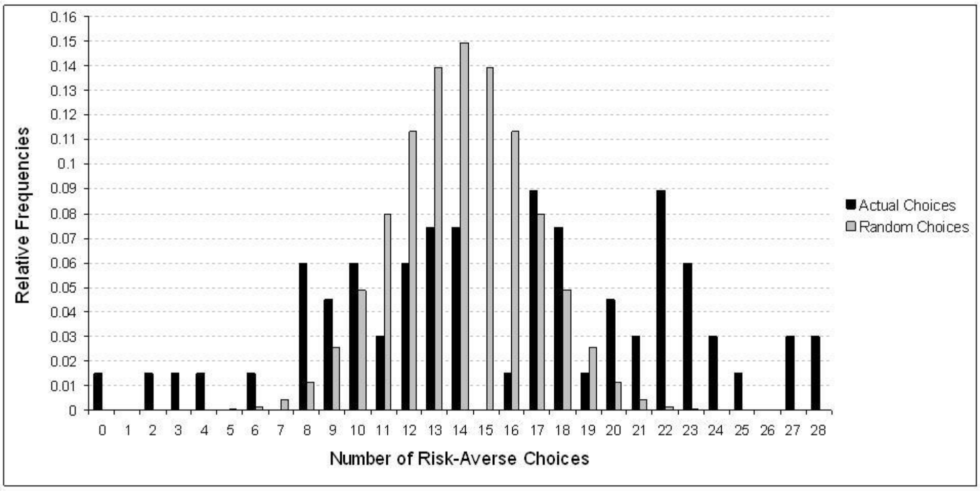
3.4.3 Comparing Subjects

While Table 3.1 already showed how aggregated choices represent risk aversion, prudence, and temperance, Figures 3.1, 3.2, and 3.3 rather present these results on the

[8] If we examine the second-order decisions over the mixed domain, we find some indications that loss aversion is present. Subjects choose B_2 noticeably more often (65 % versus 50%) if this choice provides a sure gain (decisions 21 and 26) than if this choice leads to a sure loss (decisions 2 and 22). However, the number of decisions is too small to allow for a detailed formal analysis.

level of the individual subject. Based on Result 3.2, we do not distinguish between the gain, loss, and mixed domain but only between orders of risk preferences. We count how often each individual chooses B_n over A_n for each n and show the relative frequencies of subjects with a certain number of risk-averse, prudent, and temperate choices. The dark distributions in Figures 3.1, 3.2, and 3.3 show the empirical frequency distributions of the respective number of risk-averse, prudent, and temperate choices made by the subjects. The light distributions show the frequency distributions that would have occurred if every subject had chosen randomly between B_n and A_n. This would also have been the distribution if every subject had been stochastically risk neutral, since this is equivalent to the behavior of choosing B_n with a probability of 1/2.

Figure 3.1: Characterizing Subjects: Risk Aversion



Concerning second-order risk preferences, the empirical distribution in Figure 3.1 does clearly not follow random behavior.[9] The median of the empirical distribution (17 risk-averse choices) is above the median of the random distribution (14 risk-averse choices). More subjects are located in the right tail than in the left tail of the empirical distribution. This indicates risk aversion. In order to see how many subjects are in fact risk-averse, we can classify them using binomial tests. For each subject, it is tested whether her probability of choosing B_2 is equal or lower than 1/2. Likewise, in order to classify risk-seeking subjects, it is tested whether the probability of choosing A_2 is equal or lower than 1/2. Then, subjects choosing B_2 in 0-10 out of 28 decisions are risk-seeking and those choosing B_2 in 18-28 out of 28 decisions are risk-averse. The

[9]We can reject that the two distributions are equal with a two-sided Kolmogorov-Smirnov test ($p = 0.017$; $D = 0.3793$; $N = 58$). We can further reject that the empirical distribution follows a normal distribution with a two-sided Shapiro-Wilk test ($p = 0.0617$; $W = 0.9319$; $N = 29$).

remaining subjects, those who choose B_2 in 11-17 out of 28 decisions, stay unclassified and cannot be distinguished from choosing randomly.[10] We find that 42% of the subjects can be classified as risk-averse and only 24% as risk-seeking. The remaining 34% cannot be distinguished from behaving randomly. Therefore, the number of subjects who are risk-averse is almost twice as high as the number of risk-seeking subjects.

Figure 3.2: Characterizing Subjects: Prudence

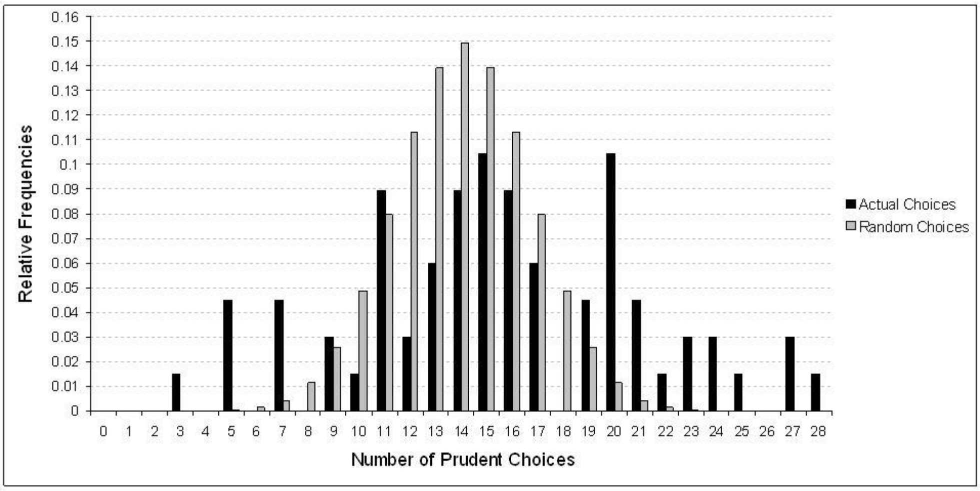


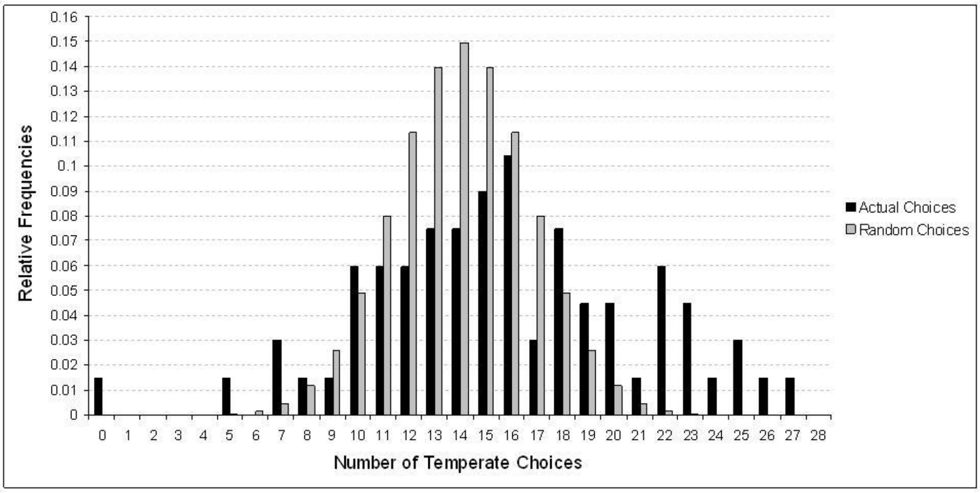
Figure 3.2 shows our results for third-order risk preferences. As before, the two distributions are distinct.[11] The median of the empirical distribution is again higher in the actual distribution (15 prudent choices) and the right tail is fatter than the left tail. This means subjects tend to be prudent. When performing the same classification as for second-order risk preferences, we find that 33% of all subjects are prudent, 15% are imprudent, and 52% cannot be distinguished from random behavior. Similar as with risk aversion, more than twice as many subjects can be classified as being prudent rather than imprudent. Ebert and Wiesen (2011) use a comparable classification and find that 47% are prudent, 8% are imprudent, and 45% remain unclassified for the 16 decisions subjects made in their experiment. Deck and Schlesinger (2010) analyze only six decisions and a comparable classification would yield 14% prudent, 2% imprudent, and 84% unclassified subjects. Moreover, a comparable classification based on five prudent decisions in Noussair, Trautmann, and van de Kuilen (2011) would show that approximately 45% are prudent, 13% are imprudent, and 42% cannot be distinguished

[10]This classification implements a one-sided binomial test at the 10%-significance level.

[11]We can reject the null hypothesis that subjects were choosing randomly in favor of the one-sided alternative that subjects are prudent with a Kolmogorov-Smirnov test ($p = 0.077$; $D = 0.2759$; $N = 58$). Also, we can again reject that the empirical distribution follows a normal distribution with a two-sided Shapiro-Wilk test ($p = 0.0091$; $W = 0.8987$; $N = 29$).

from being neither prudent nor imprudent.[12] So, our within-subject result on prudence seems to be in line with existing evidence.[13]

Figure 3.3: Characterizing Subjects: Temperance



Finally, Figure 3.3 shows the results of temperate choices against the distribution that would occur under random behavior. It shows that individuals tend to be temperate, the median individual chooses 15 times the temperate option. The left tail has only slightly more mass than what would be predicted under random behavior, indicating that the share of subjects who are clearly intemperate is very low. In contrast, the right tail of the distribution has significantly more mass than under the random distribution.[14] A classification of subjects shows that 36% are temperate, 15% are intemperate, and 49% remain unclassified. More than twice as many subjects are therefore temperate rather than intemperate. A comparable classification of fourth-order risk preferences with the data of Deck and Schlesinger (2010) yields the opposite pattern. Here, only 6% are temperate while 21% of the subjects are intemperate. The remaining 73% can-

[12]Noussair, Trautmann, and van de Kuilen (2011) do not provide exact data on this and these approximations are therefore inferred from the graphs displaying the empirical distribution over the five choices subjects made.

[13]When comparing our empirical third-order distribution with the ones of Deck and Schlesinger (2010), Ebert and Wiesen (2011), and Noussair, Trautmann, and van de Kuilen (2011), it can be observed that they share central features. The empirical medians are above the medians of the random distributions (4 vs. 3 prudent choices in Deck and Schlesinger, 2010; 11 vs. 8 prudent choices in Ebert and Wiesen, 2011; and 4 vs. 2 or 3 prudent choices in Noussair, Trautmann, and van de Kuilen, 2011) and more mass is in the right tail of their empirical distributions than in a random counterpart.

[14]A two-sided Kolmogorov-Smirnov test for equality of the two distributions yields a clear rejection ($p = 0.038$; $D = 0.3448$; $N = 58$). Furthermore, the empirical distribution does not follow a normal distribution when tested with a two-sided Shapiro-Wilk test ($p = 0.0028$; $W = 0.9185$; $N = 29$).

not be distinguished from random behavior in a classification based on four temperate choices subjects made in their experiment.

While the empirical distribution in Figure 3.3 might not look surprising when compared with the empirical distributions of risk aversion (Figure 3.1) and prudence (Figure 3.2), it is very different from the results of Deck and Schlesinger (2010). The empirical distribution of Deck and Schlesinger (2010) is skewed to the right, while ours is skewed to the left. Deck and Schlesinger (2010) conclude that subjects must be intemperate.[15] They deduce that conventional functional forms of EUT are not reconcilable with individuals being intemperate. Based on our data on temperance, we come to a different conclusion. Since we elicited 28 decisions for each subject whereas Deck and Schlesinger (2010) elicited only four, we are confident that our result on temperance is robust.

Although being also based on only five temperate choices, the results of Noussair, Trautmann, and van de Kuilen (2011) provide further confidence for our result on fourth-order risk preferences. A comparable classification with their data would show that approximately 34% are temperate, 15% are intemperate, and 51% cannot be distinguished from being neither temperate nor intemperate.[16] Also, the median of their empirical distribution is not lower than under random behavior (3 vs. 2 or 3 temperate choices) and it has more mass in the right tail.

Result 3.3 *Of those subjects that can be distinguished from random behavior, approximately twice as many are stochastically risk-averse, prudent, and temperate rather than stochastically risk-seeking, imprudent, and intemperate.*

Result 3.3 complements Result 3.1 and highlights that the prevalent pattern of risk preferences satisfies risk aversion, prudence, and temperance also on a within-subject level.

3.4.4 Relating Risk Preferences of Different Order

Having identified the sign of risk preferences, we now turn to the question what the relations of these higher-order risk preferences are. Already our observation that aggregate accounts of different higher-order risk preferences closely resemble each other (see Table 3.1) posed the question of how tightly risk preferences of different orders are

[15] Whether the exact median of their observed data equals one or two temperate choices (versus a median of two choices under the random distribution) is not inferable from their reported data. They report that 50% of subjects made either no or a single temperate choice. Since they analyze 99 subjects the median has to be either one or two choices.

[16] These classifications are again approximations based the graphs displaying the empirical distribution over the five choices subjects made.

in fact related. Based on Result 3.2 we do again not distinguish between decisions in gains, in losses, and in the mixed domain. We can conduct a within subject analysis of this question, since we can link the decisions of a single subject on different orders. The relation of risk preferences of different orders is of interest because it indicates whether the elicited risk preferences of one order can be a good predictor for risk preferences of a different order. This has been a standard procedure in many empirical studies on precautionary saving. In this literature, measures of risk aversion are used to predict within common EUT models the amount of prudence, and subsequently the presumable optimal levels of precautionary savings.[17]

Figure 3.4: Relation of Risk Aversion and Prudence

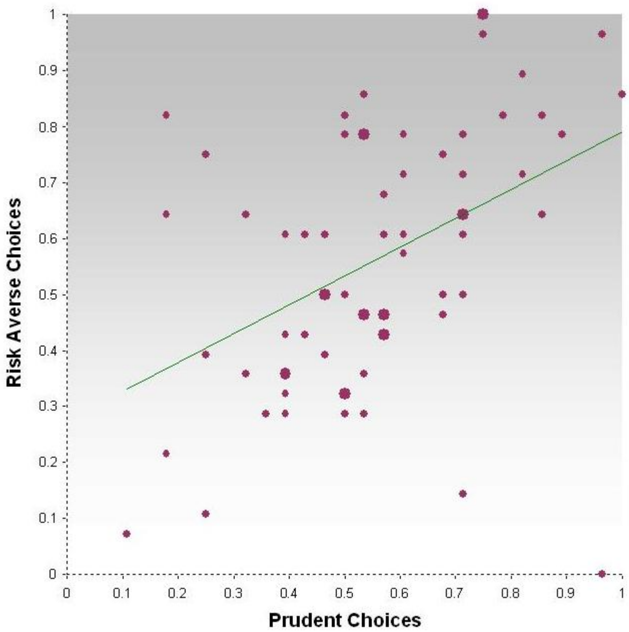


Figure 3.4 illustrates whether the same subjects that are risk-averse also are prudent. We map each subjects' relative frequency of risk-averse and prudent choices. Each small point represents a single individual and each larger point represents two individuals who share exactly the same pattern. There is a clear positive relation between the frequencies of risk-averse and prudent choices. The straight line represents a least squares regression line. It has the functional form $Y = 0.2758 + 0.5144 \cdot X$, where X represents the relative frequency of prudent choices and Y the relative frequency of risk averse choices. The coefficient of determination of the regression line is $R^2 = 0.193$ and the Spearman rank order correlation coefficient is $\rho = 0.4948$ ($p = 0.000$; $N = 67$).

[17]For a survey article discussing this approach and its results, see Carroll and Kimball (2008).

It shows that there exists a considerable positive and significant relation between risk aversion and prudence.

Figure 3.5: Relation of Risk Aversion and Temperance

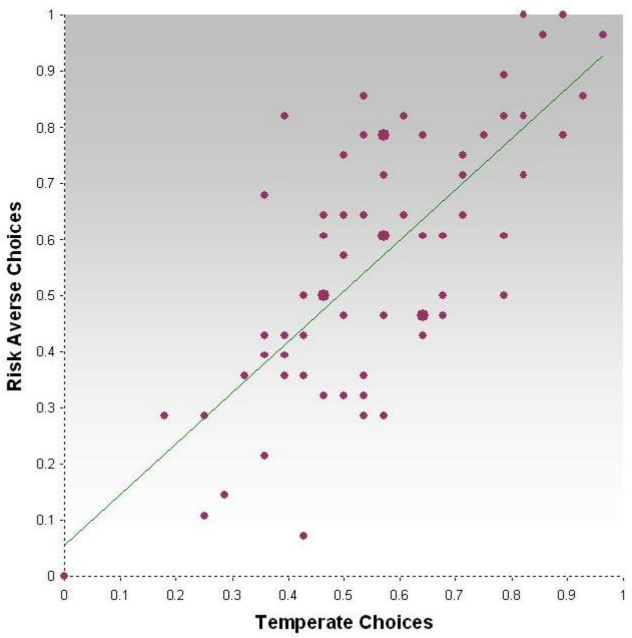
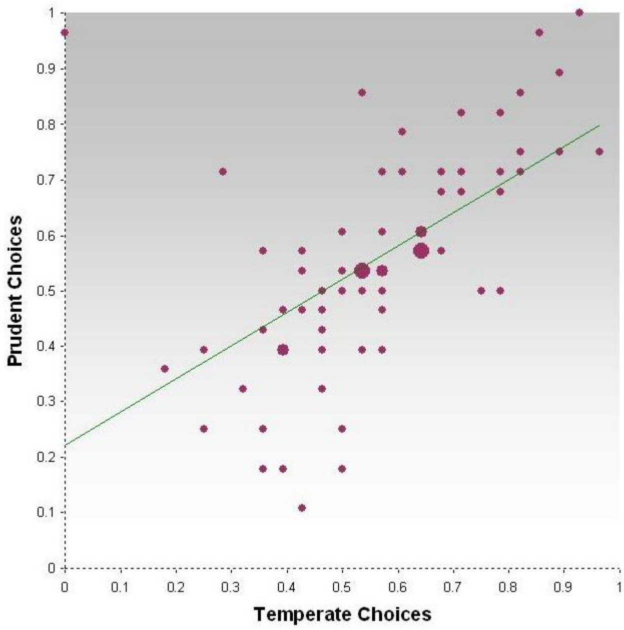


Figure 3.5 illustrates a similar relation for risk aversion and temperance. Again, each small point represents a single individual and each larger point represents two individuals. The positive relation between the frequency of risk aversion and prudence is even more eminent. The least squares regression line is $Y = 0.0547 + 0.9062 \cdot X$, where X represents the relative frequency of temperate choices and Y the relative frequency of risk-averse choices (with $R^2 = 0.5391$). The Spearman rank order correlation coefficient is $\rho = 0.6853$, and if we test whether ρ is not different from 0 we can again reject this at the 1%-significance level ($p = 0.0000$; $N = 67$). This shows that there is a very strong relation between the degree of consistency of risk aversion and temperance for an individual.

Finally, Figure 3.6 relates prudence and temperance. In this diagram there are not two but three sizes of data points. Each small point is the representation of a single individual, each medium sized point that of two, and each largest sized point that of three individuals. The least squares regression line is $Y = 0.2221 + 0.5978 \cdot X$, where X represents the relative frequency of temperate choices and Y the relative frequency of prudent choices (with $R^2 = 0.3216$). The Spearman rank order correlation coefficient is $\rho = 0.6795$ ($p = 0.0000$; $N = 67$). The relationship between prudence and temperance

Figure 3.6: Relation of Prudence and Temperance



is almost as strong as the one between risk aversion and temperance, and stronger than the one between risk aversion and prudence.

Observing such a high degree of correlation between risk preferences of different orders[18] raises the question whether risk preferences of different orders might be drawn from the same distribution. Therefore, we now analyze whether they are statistically distinct. Since we have a frequency of risk-averse, prudent, and temperate choices for every single individual we can employ two-sided Wilcoxon signed-rank tests to compare a pair of two orders of risk preferences. We report the results in Table 3.5. The null hypothesis is that the underlying distribution of the risk preference to the left of a cell and the underlying distribution of the risk preference above the cell are identical.

All p-values in Table 3.5 are very far from any customary significance level. Hence, we clearly cannot reject the null hypothesis that the distributions of risk preferences of different orders are identical. This means that subjects are equally stochastically risk-averse as prudent, equally stochastically risk-averse as temperate, and equally stochastically prudent as temperate. In other words, subjects who are stochastically more risk-averse (or less risk-seeking) are also stochastically more prudent (or less imprudent), with the same relation holding for the other two comparisons.

[18]Noussair, Trautmann, and van de Kuilen (2011) also investigate correlations between second- to fourth-order risk preferences and find significant Spearman rank order correlation coefficients of about half the size we find.

Table 3.5: Wilcoxon Signed-Rank Tests: Relations of Risk Preferences Across Orders

p-value z-statistic (N)	Prudence	Temperance
Prudence		0.3600 -0.915 (67)
Risk Aversion	0.9253 0.094 (67)	0.9054 0.119 (67)

Result 3.4 *There is a high correlation between risk preferences of different orders. Subjects who are stochastically more risk-averse are also stochastically more prudent and temperate. And subjects who are stochastically more prudent are also stochastically more temperate.*

Note that our definition of stochastically more risk-averse, prudent, or temperate (see Definition 3.3) does not distinguish whether subjects are stochastically risk-averse, prudent, or temperate or whether they are stochastically risk-seeking, imprudent, or intemperate (see Definition 3.2). Similar to Result 3.2, Result 3.4 therefore describes the qualitative strength or consistency of risk preferences. Here, stochastically more risk-averse, prudent, and temperate is equivalent to being stochastically less risk-seeking, imprudent, and intemperate.

3.4.5 Robustness

As a robustness analysis of our previous results, we now investigate whether the sequence in which the decision problems were displayed and answered by the subjects had any effect. In the three sessions we conducted the decisions were displayed in different sequences, but all determined by a random process. Therefore, we approach this question by analyzing whether the pattern of answers differs between the three sessions. In doing so, we perform two-sided Kolmogorov-Smirnov tests for equality of distribution functions. For each order of risk preferences we compare answers of sessions 1, 2, and 3 to determine whether the probability of choosing the ‘averse’ option differs across sessions.

With respect to risk aversion, a majority of choices in all sessions was in favor of the risk-averse option (61% in session 1, 52% in session 2, and 55% in session 3). We cannot reject the null hypothesis that the underlying probability of choosing the risk-averse choice is identical in any of the session comparisons at the 5%-significance level.[19] When looking at prudent decisions, we again observe that a majority of choices was in favor of the prudent option in all sessions (54% in session 1, 60% in session 2, and 54% in session 3). None of the session comparisons yields a rejection at any customary significance level of the null hypothesis that probabilities of choosing the prudent option are identical across sessions.[20] Finally, we test the robustness of temperance elicitation with respect to sequence effects. In all sessions, a majority of choices was in favor of the temperate option (56% in session 1, 58% in session 2, and 54% in session 3) Again, neither of the three two-sided Kolmogorov-Smirnov tests indicates that the hypothesis of identical distributions can be rejected.[21]

3.4.6 Cumulative Prospect Theory

We now examine whether CPT can be a better predictor of higher-order risk preferences than EUT with a commonly used utility function.[22] This is a key conclusion, Deck and Schlesinger (2010) draw from their experiment. They use the parametrization of Tversky and Kahneman (1992) and assume that the reference point is equal to the mean of the alternatives in a specific decision. With these assumptions, the predictions of CPT over the decisions of their experiment is that subjects will always prefer the prudent choice and will always prefer the intemperate choice. EUT with a commonly used utility function predicts that subjects choose always the prudent and the temperate choice. In all six decisions eliciting third-order risk preferences in the experiment of Deck and Schlesinger (2010), more subjects preferred the prudent choice and in all four decisions eliciting fourth-order risk preferences more subjects preferred the intemperate choice. Therefore, while both CPT and EUT predict prudence correctly, the finding of intemperance is only predicted by CPT.

[19]The corresponding results are $p = 0.387$, $D = 0.2727$, $N = 44$ for session 1 vs. session 3; $p = 0.251$, $D = 0.2866$, $N = 45$ for session 2 vs. session 3; and $p = 0.061$, $D = 0.3755$, $N = 45$ for session 1 vs. session 2. Thus, at the 10%-significance level, the only significant result is obtained for the comparison of session 1 vs. session 2. This is, however, the only significant comparison at the 10%-significance level out of nine tests, and we therefore do not derive any conclusions from this.

[20]The corresponding results are $p = 0.842$, $D = 0.1818$, $N = 44$ for session 1 vs. session 3; $p = 0.954$, $D = 0.1364$, $N = 45$ for session 2 vs. session 3; and $p = 0.313$, $D = 0.2727$, $N = 45$ for session 1 vs. session 2.

[21]The corresponding results are $p = 0.632$, $D = 0.1818$, $N = 44$ for session 1 vs. session 3; $p = 0.194$, $D = 0.2984$, $N = 45$ for session 2 vs. session 3; and $p = 0.854$, $D = 0.1621$, $N = 45$ for session 1 vs. session 2. One may be puzzled by the test results between session 1 vs. session 3 for third- and fourth-order risk preferences. They yield the same D-statistic and are based on the same sample size, yet they exhibit different p-values. The reason is that we report the exact p-values of the tests while the p-values directly based on the D-statistics are approximations in case of ties (and are likewise both 0.860 for the two tests under consideration).

[22]As commonly used utility functions we denote all functional forms that have derivatives with alternating signs and positive marginal utility, such as CRRA or CARA utility functions.

We carefully designed our experiment to implement the status quo prior to the second date as the reference point. If we adopt the parametric assumptions of Tversky and Kahneman (1992), we find that there is a variety of predictions for each order. CPT predicts in seven of our decisions eliciting second-order risk preferences that the risk-seeking choice will be taken and in 21 decisions it is predicted that the risk-averse choice will be preferred. Out of the 28 decisions eliciting third-order risk preferences, CPT predicts only one imprudent choice. In the decisions eliciting fourth-order risk preferences, it predicts twelve temperate choices versus 14 intemperate choices.[23] We can compute in how many decisions CPT correctly predicted the majority of choices. In 59 out of 84 decisions the prediction was correct. To receive an intuition on the performance of CPT as a predictive theory, we compare it to the result under EUT with standard utility functions. Here, 69 out of the 84 decisions were predicted correctly.[24] To summarize, based on our experiment, we conclude that the parametric version of CPT most commonly used is performing worse as a predictive theory than EUT with any standard utility function.

If we adopted our definition of the reference point to the experiment of Deck and Schlesinger (2010), the two compared theories would predict exactly the same behavior. We defined the reference point to be the status quo prior to the experiment. According to this definition, all decisions in Deck and Schlesinger (2010) lie in the domain of gains. In gains, the value function of CPT is identical to a specific CRRA function. In the decision problems of Deck and Schlesinger (2010) the nonlinear influence of probabilities does not change these predictions. Hence, both theories yield identical predictions. And if we adopted the definition of the reference point from Deck and Schlesinger (2010) to our experiment we find that in this case CPT predicts that subjects always choose the risk-averse, the prudent and the intemperate alternative. Only in 53 out of 84 decisions the predictions of CPT would be correct.

3.5 Conclusion

We analyzed higher-order risk preferences, namely risk aversion, prudence, and temperance in a laboratory experiment. Each subject participated at two distinct

[23]EUT is independent of domains. If we additionally assume a commonly used utility function, it therefore predicts that subjects always make the risk-averse, the prudent and the temperate choice.

[24]Choices in decisions 10, 11, 16, 19, 20, 25, 57, 61, 63, 64, 67, 68, 69, 71, 72, 73, and 83 were correctly predicted by EUT but not by CPT. Choices in decisions 24, 43, 65, 76, 77, 80, and 84 were correctly predicted by CPT but not by EUT. Choices in decisions 12, 22, 30, 31, 34, 40, 55, and 74 were wrongly predicted by both theories. Choices in the 52 remaining decisions were correctly predicted by both theories.

dates of the experiment, separated from each other by three weeks. This allowed us to inflict real monetary losses on subjects at the second date (in case decisions and chance determined that the pay-offs were negative). Thus, we designed our experiment in order to clearly implement the status quo prior to the second date experiment as the reference point.

Our findings reveal that the prevailing preference pattern exhibits risk aversion, prudence, and temperance both at an aggregate level but also at the level of the individual subject. While our results on risk aversion and prudence are in line with existing evidence, our finding on temperance contrasts the result of Deck and Schlesinger (2010) which led them to conclude that CPT but not EUT is consistent with the data. Since a crucial feature of reference-dependent models, such as prospect theory (Kahneman and Tversky, 1979), is the distinct evaluation of gains and losses, we could shed further light on risk preferences in the gain, loss, and mixed domain. Similar to Bosch-Domènech and Silvestre (2006) who also investigate second-order risk preferences in the (real) loss domain, we find no evidence for risk-seeking in losses, a behavior that has been found in the early experimental literature relying on hypothetical questions only, and in some studies subject to 'house-money' problems. More generally, we find that risk preferences of any order are not influenced by the domain decisions are taken. Subjects have the same second- to fourth-order risk preference in the gain and in the loss domain. These results are well reconcilable with standard assumptions in EUT, such as CRRA or CARA, but not with the favored variant of cumulative prospect theory. Another feature of common EUT models is the tight connection of risk preferences across orders. Consistent with this prediction, we find very strong relations between risk preferences of all orders. Subjects showing a higher level of consistency with a risk preference in one order also show a higher level of consistency with a risk preference in another order.

The shape of gain-loss utility does not seem to be of a fundamentally different nature for gains and for losses. As a qualification of our results, we have to state that the stakes in our experiment are small compared to some decisions individuals face outside the laboratory. If individuals are threatened by huge losses, they may be tempted to speculate on a series of favorable outcomes of events and take risks they otherwise would shun. They might be willing to aggregate risks in order to leave at least one possible positive outcome. This outcome could then function as a glimmer of hope. However, behavior in reaction to monetary losses that do not stir such existential fears seems more accessible to economic analysis. Although we cannot infer the functional shape of the value function for all levels of stakes from our results, we can dismiss the notion that at the reference point the function radically changes its characteristics.

3.6 Appendix A: Decisions in the Experiment

Table 3.6: Decisions Eliciting Risk Aversion

Dec-ision	Para-meters		Displayed Outcomes				In Do-main	Appearance in Session			% of
			Option A		Option B						
#	y	ε_1	A1	A2	B1	B2		1	2	3	B_2
01	45	30	15	75	45	45	G	29	12	75	58
02	-5	6	-11	1	-5	-5	M	1	39	76	52
03	10	2	8	12	10	10	G	2	40	77	51
04	5	3	2	8	5	5	G	3	64	19	69
05	0	12	-12	12	0	0	M	57	41	49	55
06	57	5	52	62	57	57	G	30	65	50	51
07	28	20	8	48	28	28	G	4	66	20	51
08	0	9	-9	9	0	0	M	58	67	78	55
09	62	17	45	79	62	62	G	59	68	21	57
10	-9	3	-12	-6	-9	-9	L	31	13	79	51
11	-5	3	-8	-2	-5	-5	L	32	42	22	67
12	6	5	1	11	6	6	G	5	43	80	45
13	46	32	14	78	46	46	G	33	69	51	57
14	3	2	1	5	3	3	G	60	14	81	58
15	15	11	4	26	15	15	G	34	44	52	57
16	-6	5	-11	-1	-6	-6	L	61	45	23	51
17	0	2	-2	2	0	0	M	62	70	82	52
18	0	5	-5	5	0	0	M	6	15	53	64
19	-3	2	-5	-1	-3	-3	L	7	46	54	66
20	-7	4	-11	-3	-7	-7	L	35	47	24	66
21	5	6	-1	11	5	5	M	8	71	25	60
22	-1	10	-11	9	-1	-1	M	63	72	26	48
23	9	3	6	12	9	9	G	9	73	83	55
24	-10	2	-12	-8	-10	-10	L	36	16	55	49
25	-7	5	-12	-2	-7	-7	L	10	17	84	52
26	1	10	-9	11	1	1	M	11	18	27	70
27	7	4	3	11	7	7	G	37	74	56	55
28	7	5	2	12	7	7	G	38	48	28	54

Table 3.7: Decisions Eliciting Prudence

Dec-	Para-			Displayed Outcomes								In	Appearance			%
ision	meters			Option A				Option B				Do-	in Session			of
#	y	k	ε_1	A1	A2	A3	A4	B1	B2	B3	B4	main	1	2	3	B_3
29	-4	3	3	-10	-4	-4	-4	-7	-7	-7	-1	L	12	75	29	58
30	-5	2	3	-10	-5	-5	-4	-8	-7	-7	-2	L	39	76	1	49
31	10	2	2	6	10	10	10	8	8	8	12	G	40	77	2	49
32	-6	4	2	-12	-6	-6	-8	-8	-10	-10	-4	L	64	19	3	58
33	9	6	2	1	9	9	5	7	3	3	11	G	41	49	57	60
34	-6	2	4	-12	-6	-6	-4	-10	-8	-8	-2	L	65	50	30	49
35	57	5	20	32	57	57	72	37	52	52	77	G	66	20	4	54
36	0	6	4	-10	0	0	-2	-4	-6	-6	4	M	67	78	58	54
37	10	4	2	4	10	10	8	8	6	6	12	G	68	21	59	57
38	62	15	21	26	62	62	68	41	47	47	83	G	13	79	31	70
39	0	2	2	-4	0	0	0	-2	-2	-2	2	M	42	22	32	58
40	0	4	6	-10	0	0	2	-6	-4	-4	6	M	43	80	5	49
41	8	2	4	2	8	8	10	4	6	6	12	G	69	51	33	57
42	51	12	5	34	51	51	44	46	39	39	56	G	14	81	60	61
43	0	10	2	-12	0	0	-8	-2	-10	-10	2	M	44	52	34	55
44	67	11	11	45	67	67	67	56	56	56	78	G	45	23	61	51
45	-3	2	2	-7	-3	-3	-3	-5	-5	-5	-1	L	70	82	62	63
46	-3	6	2	-11	-3	-3	-7	-5	-9	-9	-1	L	15	53	6	52
47	-3	2	6	-11	-3	-3	1	-9	-5	-5	3	M	46	54	7	54
48	5	2	2	1	5	5	5	3	3	3	7	G	47	24	35	67
49	-4	2	5	-11	-4	-4	-1	-9	-6	-6	1	M	71	25	8	55
50	4	2	5	-3	4	4	7	-1	2	2	9	M	72	26	63	51
51	42	8	16	18	42	42	50	26	34	34	58	G	73	83	9	70
52	10	7	5	-2	10	10	8	5	3	3	15	M	16	55	36	52
53	0	6	6	-12	0	0	0	-6	-6	-6	6	M	17	84	10	58
54	-8	2	2	-12	-8	-8	-8	-10	-10	-10	-6	L	18	27	11	58
55	0	7	3	-10	0	0	-4	-3	-7	-7	3	M	74	56	37	48
56	0	2	10	-12	0	0	8	-10	-2	-2	10	M	48	28	38	42

Table 3.8: Decisions Eliciting Temperance

Dec-ision	Para-meters			Displayed Outcomes																In Do-main	Appearance in Session			% of
				Option A								Option B												
#	y	ε_1	ε_2	A1	A2	A3	A4	A5	A6	A7	A8	B1	B2	B3	B4	B1	B2	B3	B4		1	2	3	B_4
57	0	4	3	-7	-1	0	0	0	0	1	7	-4	-4	-3	-3	3	3	4	4	M	75	29	12	55
58	52	18	5	29	39	52	52	52	52	65	75	34	34	47	47	57	57	70	70	G	76	1	39	63
59	-4	3	2	-9	-5	-4	-4	-4	-4	-3	1	-7	-7	-6	-6	-2	-2	-1	-1	M	77	2	40	69
60	6	4	2	0	4	6	-6	-6	-6	8	12	2	2	4	4	8	8	10	10	G	19	3	64	55
61	-8	2	2	-12	-8	-8	-8	-8	-8	-8	-4	-10	-10	-10	-10	-6	-6	-6	-6	L	49	57	41	60
62	48	12	12	24	48	48	48	48	48	48	72	36	36	36	36	60	60	60	60	G	50	30	65	57
63	-4	2	1	-7	-5	-4	-4	-4	-4	-3	-1	-6	-6	-5	-5	-3	-3	-2	-2	L	20	4	66	52
64	0	8	4	-12	-4	0	0	0	0	4	12	-8	-8	-4	-4	4	4	8	8	M	78	58	67	57
65	0	6	6	-12	0	0	0	0	0	0	12	-6	-6	-6	-6	6	6	6	6	M	21	59	68	45
66	-1	6	2	-9	3	-1	-1	-1	-1	-5	7	-3	-3	-7	-7	5	5	1	1	M	79	31	13	76
67	-6	3	2	-11	-7	-6	-6	-6	-6	-5	-1	-9	-9	-8	-8	-4	-4	-3	-3	L	22	32	42	69
68	35	22	4	9	17	35	35	35	35	53	61	13	13	31	31	39	39	57	57	G	80	5	43	55
69	-5	2	2	-9	-5	-5	-5	-5	-5	-5	-1	-7	-7	-7	-7	-3	-3	-3	-3	L	51	33	69	60
70	7	4	1	2	4	7	7	7	7	10	12	3	3	6	6	8	8	11	11	G	81	60	14	57
71	0	7	5	-12	-2	0	0	0	0	2	12	-7	-7	-5	-5	5	5	7	7	M	52	34	44	51
72	0	9	2	-11	-7	0	0	0	0	7	11	-9	-9	-2	-2	2	2	9	9	M	23	61	45	60
73	0	10	2	-12	-8	0	0	0	0	8	12	-10	-10	-2	-2	2	2	10	10	M	82	62	70	58
74	62	13	9	40	58	62	62	62	62	66	84	49	49	53	53	71	71	75	75	G	53	6	15	43
75	10	7	5	-2	12	10	10	10	10	8	22	5	5	3	3	17	17	15	15	M	54	7	46	61
76	-7	4	1	-12	-10	-7	-7	-7	-7	-4	-2	-11	-11	-8	-8	-6	-6	-3	-3	L	24	35	47	43
77	1	6	2	-7	5	1	1	1	1	-3	9	-1	-1	-5	-5	7	7	3	3	M	25	8	71	37
78	18	9	6	3	15	18	18	18	18	21	33	9	9	12	12	24	24	27	27	G	26	63	72	73
79	45	6	6	33	45	45	45	45	45	45	57	39	39	39	39	51	51	51	51	G	83	9	73	57
80	0	2	2	-4	0	0	0	0	0	0	4	-2	-2	-2	-2	2	2	2	2	M	55	36	16	45
81	8	2	2	4	8	8	8	8	8	8	12	6	6	6	6	10	10	10	10	G	84	10	17	69
82	7	3	2	2	6	7	7	7	7	8	12	4	4	5	5	9	9	10	10	G	27	11	18	52
83	-6	4	2	-12	-8	-6	-6	-6	-6	-4	0	-10	-10	-8	-8	-4	-4	-2	-2	L	56	37	74	51
84	-7	3	2	-12	-8	-7	-7	-7	-7	-6	-2	-10	-10	-9	-9	-5	-5	-4	-4	L	28	38	48	40

3.7 Appendix B: Instructions (translated from German)

Dear participant,

thank you very much for your attendance. You are now participating in an experiment at the MELESSA laboratory. The experiment consists of two parts. As it was described in the invitation E-mail, the first part will take place today and the second part in exactly three weeks.

The experiment is organized as follows. You participate today and earn money. At the second date in three weeks you will also get the opportunity to earn money, but it is also possible that you make losses. Over the two dates you cannot lose money. This means that the minimum amount you earn today is assured to be greater than the amount you can lose in three weeks. Thus, the total amount over both experimental parts will be positive with certainty (larger than zero).

It is necessary for the experimental results that you participate at both dates. If you cannot participate at the second date because you did not read the notice in the mail or do not want to participate for other reasons, please raise your hand now. You will get your show-up fee and do not have to participate in the experiment. There are no further consequences for you if you decide not to participate. If you do not raise your hand now, you commit to take part at both parts. This means, if you do not raise your hand now and do not show up at the second date, you will be excluded from the experimental recruiting system. You will not be allowed to take part in further experiments of this laboratory.

Instructions for Today's First Part of the Experiment:

From now on you are not allowed to communicate with the other participants. If you have any questions, please alert one of the experimenters by raising your hand. The experimenter will then come to your seat to answer your question. It is prohibited to use cell phones or to open other programs on your computer. If you violate any of these rules, we will have to exclude you from the experiment without compensation.

In today's part of the experiment you work on two tasks. For each task you will be paid separately. Therefore, your decisions in task 1 do not influence the payment in task 2. Your decisions in task 2 also do not influence your income in task 1.

The **procedure** of today's first part is as follows:

1. At first, please read the instructions for task 1 at hand. Control questions will show, if you understood the instructions.
2. Afterwards, you will work on task 1 on the computer.
3. After working on task 1 you will get the instruction for task 2 on the computer screen. Please also read them carefully.
4. Subsequently, you work on task 2 on the computer.
5. After all participants finished task 2, chance and your decisions will determine the amount of your payoff.
6. Your payoffs appear on the screen and you wait at your seat until your number is announced.
7. When your number is announced, leave the room and you will receive your payment.
8. Then the first part of the experiment is over and you can leave. We expect you back for the second part of the experiment in three weeks.

Instructions Task 1:

In task 1 you have to make decisions in three tables. The tables appear one after another on the screen. When you finished working on a table (by pressing the OK button on the screen) you cannot revise them afterwards.

Instructions for Table 1:

In each row you have to decide between two alternatives (alternative A and alternative B). Alternative A as well as alternative B may have two possible realizations.
The possible realizations of alternative A are called A1 and A2. The possible realizations of alternative B are called B1 and B2. If you decided for alternative A, chance will determine if you will get realization A1 or realization A2. Similarly, if you decide for alternative B, you will get either B1 or B2 by chance. Both realizations of an alternative are equally likely. This means that it can be expected that the first realization occurs as often as the second realization.

Here, you see an example table (this is not the table you will work on later):

Tabelle 1 (von 3)

	Alternative A			Alternative B		
Zeile	Ausprägung A 1	Ausprägung A 2	Ihre Wahl	Ausprägung B 1	Ausprägung B 2	Zeile
1.	4,20	12,90	A ○ ○ B	1,00	12,70	1.
2.	0,30	6,50	A ○ ○ B	5,90	11,00	2.
3.	0,80	1,80	A ○ ○ B	7,70	12,70	3.
4.	0,60	10,80	A ○ ○ B	0,10	7,70	4.
5.	5,20	15,90	A ○ ○ B	1,00	6,50	5.
6.	1,70	9,90	A ○ ○ B	0,60	4,20	6.
7.	10,20	15,40	A ○ ○ B	6,10	8,80	7.
8.	4,80	7,10	A ○ ○ B	5,70	10,60	8.
9.	1,60	8,00	A ○ ○ B	2,60	8,30	9.
10.	14,10	15,20	A ○ ○ B	8,30	12,80	10.
11.	2,60	14,80	A ○ ○ B	10,60	15,30	11.
12.	5,60	8,20	A ○ ○ B	1,10	3,90	12.

OK

In each row you have to decide between alternative A and alternative B. There are no objectively right or wrong decisions. But you should consider carefully in each row if you rather want to have the potential realizations of alternative A (A1 and A2) or the potential realizations of alternative B (B1 and B2). Depending on your decision, it is decided by pure chance which of the realizations (A1 or A2 if A was chosen; B1 or B2 if B was chosen) determines your income.

Instructions for Tables 2 and 3:

Again, you have to decide in each row between two alternatives (alternative A and alternative B). The only difference to table 1 is that both alternatives (A and B) do have not only two but four different realizations.
Again, all realizations are equally likely. This means that it can be expected that each of the four realizations will occur as often as the other ones. So, you have to consider if you rather want to have the possible realizations A1, A2, A3, and A4 (alternative A) or the possible realizations B1, B2, B3, and B4 (alternative B).

Here you have two example tables (again, these are not the tables you will work on later):

[SIMILAR SCREENS AS THE PREVIOUSLY, FOUR OUTCOMES EACH OPTION]

Your **income** from task 1 will be calculated at the end of today's first part as follows:

The person who by chance sits on seat 1 draws one ball from a box (called table box) with 3 numbered balls. The drawn number (1, 2, or 3) determines, which of the 3 tables is relevant for your income.

The person rolls a twelve sided die afterwards. The result (1 - 12) determines which row in the chart is relevant for your income.

Finally, the person draws from two boxes (called realization box A and realization box B) one ball each. Both realization boxes contain 2 numbered balls, if the ball with number 1 was drawn from the table box. Both realization boxes contain 4 numbered balls each, if the ball with number 2 or 3 was drawn from the table box.

If you decided for alternative A in the diced row of the drawn table, you get the amount of the drawn realization of A. If you decided for alternative B in the diced row of the drawn table, you get the amount of the drawn realization of B.

An Example:

Suppose the person with seat number 1 draws the ball with number 2 from the table box at first. Afterwards, this person dices a 10. Finally the ball with number 2 from realization box A and the ball with number 4 from the realization box B is drawn. If you have chosen alternative A for row 10 in table 2, you get paid according to realization A2 in this row. This means you receive an income of 8.30 euros. If you have chosen alternative B for row 10 in table 2, you get paid according to realization B4 in this row. In this case you get 15.10 euros.

The following three control questions will help you to see whether you have understood the instructions for task 1. Please fill in the solutions and give a signal by raising your hand after you answered all questions.

Control Question 1:

Suppose the person with seat number 1 draws from the table box the ball number 3. Afterwards, the person dices 11. Finally, ball 4 from realizations box A and ball 1 from realizations box B is drawn.

What is the amount you receive if you decided in row 11 of table 3 for alternative A? ________

What is the amount you receive if you decided in row 11 of table 3 for alternative B? ________

Control question 2:

Suppose the person with seat number 1 draws from the table box the ball number 1. Afterwards, the person dices 9. Finally, ball 1 from realizations box A and ball 2 from realizations box B is drawn.

What is the amount you receive if you decided in row 9 of table 1 for alternative A? ________

What is the amount you receive if you decided in row 9 of table 1 for alternative B? ________

Control question 3:

Suppose the person with seat number 1 draws from the table box the ball number 2. Afterwards the person dices 4. Finally, ball 3 from realizations box A and ball 3 from realizations box B is drawn.

What is the amount you receive if you decided in row 4 of table 2 for alternative A? ________

What is the amount you receive if you decided in row 4 of table 2 for alternative B? ________

After you answered all control questions, please raise your hand and an experimenter will come to your seat to assure that all answers are correct. If you have further questions they can then be answered as well.

After all participants have answered the control questions and if there are no further questions, the experiment will start on the computer with task 1.

Reminder:

This note is a **reminder**.

As explained, this experiment consists of two parts.

Your **payoff** for the first part of the experiment today was:
________.________ euros.

The **second part** of the experiment takes place in exactly three weeks, therefore on December 16th, 2008 at ________ : ________ o'clock in the same rooms of MELESSA.

In the second part of the experiment you will have the opportunity to obtain income. But this time, depending on chance and your decisions, losses can occur in the second part of the experiment.

Because of that we ask you to bring the today obtained amount in cash with you.

The profits or losses you can obtain in the second part of the experiment do not depend on today's obtained payoffs.

Dear participants,
thank you very much for coming to the second and last part of this experiment.
You earned money at the first part of the experiment exactly three weeks ago. Today you have to work on **one task** only. Depending on chance and your decisions you have the opportunity to make **gains** as well as **losses**.

The **procedure** of today's second and last part of the experiment is as follows:

1. At first, carefully read the instructions for the task.

2. Afterwards, you will work on the task on the computer.

3. After finishing the task, a drawing conducted by you determines the paid amount for today's second part of the experiment.

4. Your payment appears on the screen and you wait at your seat until your number is announced.

5. When your number is announced, please leave the room. If the amount of your payment is positive, you get paid this amount. If the amount of your payment is negative, you have to pay this amount. With this, the experiment ends and you can leave.

Instructions for Today's Second and Last Part of the Experiment:

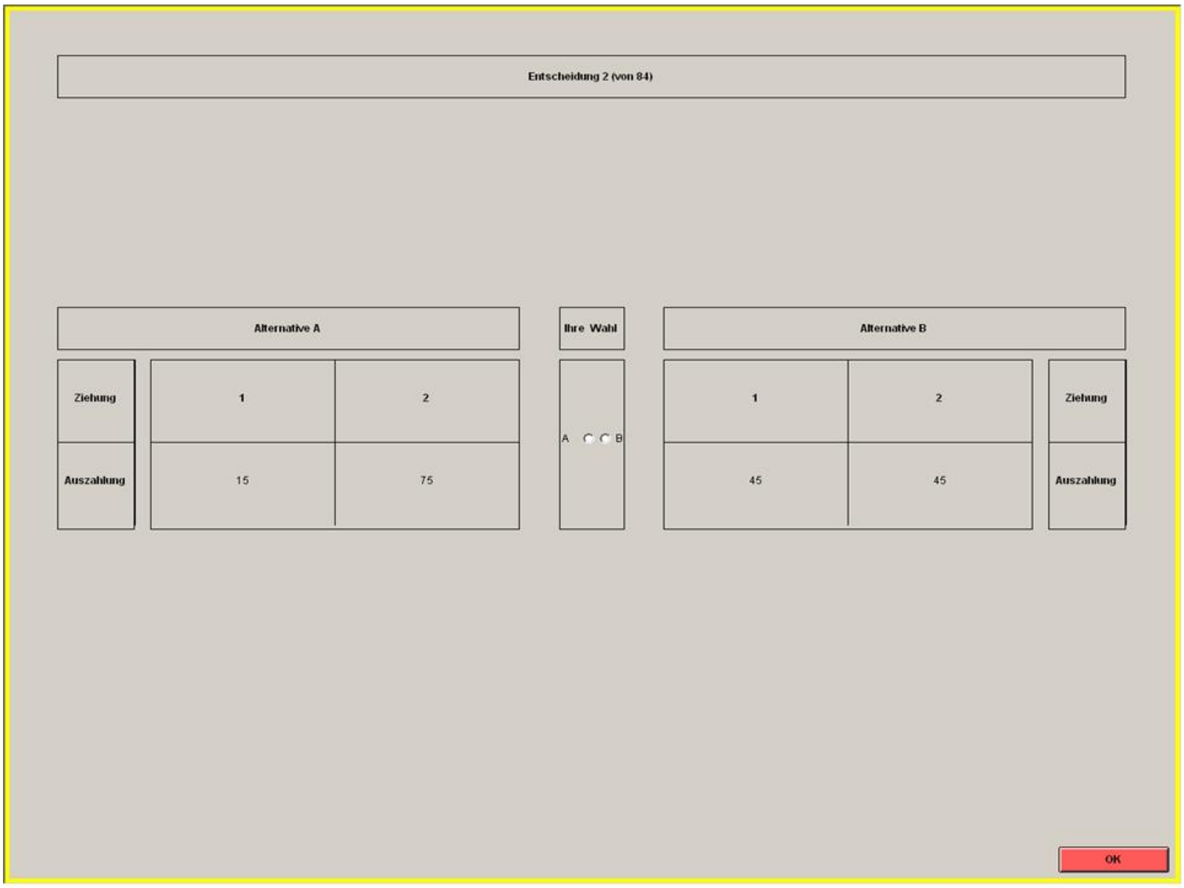
From now on you are not allowed to communicate with the other participants. If you have any questions, please alert one of the experimenters by raising your hand. The experimenter will then come to your seat to answer your question. It is prohibited to use cell phones or to open other programs on your computer. If you violate any of these rules, we will have to exclude you from the experiment without compensation.

In today's second part of the experiment you will work on **one tastk** only. You will get paid only for this task. This means, that chance and your decisions in this task determine the complete amount of payment for today's part of the experiment. Because the show-up fee for today was already paid three weeks ago, the mentioned amounts fully determine your payment for today's part of the experiment. All mentioned amounts are in **euros**.

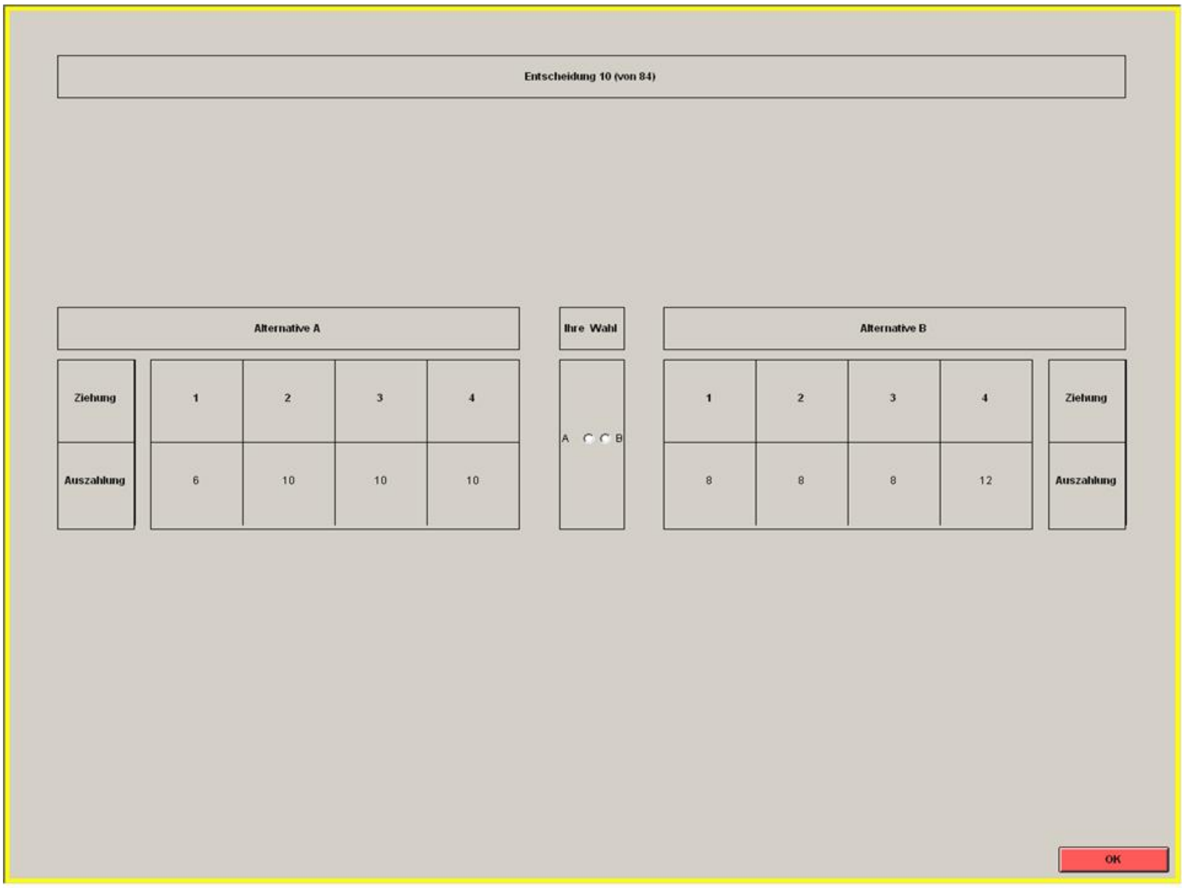
Task Instructions:

Your task for today's second part of the experiment is to make 84 **decisions** on **two alternatives, A** and **B**. There are three different types of alternatives A and B:

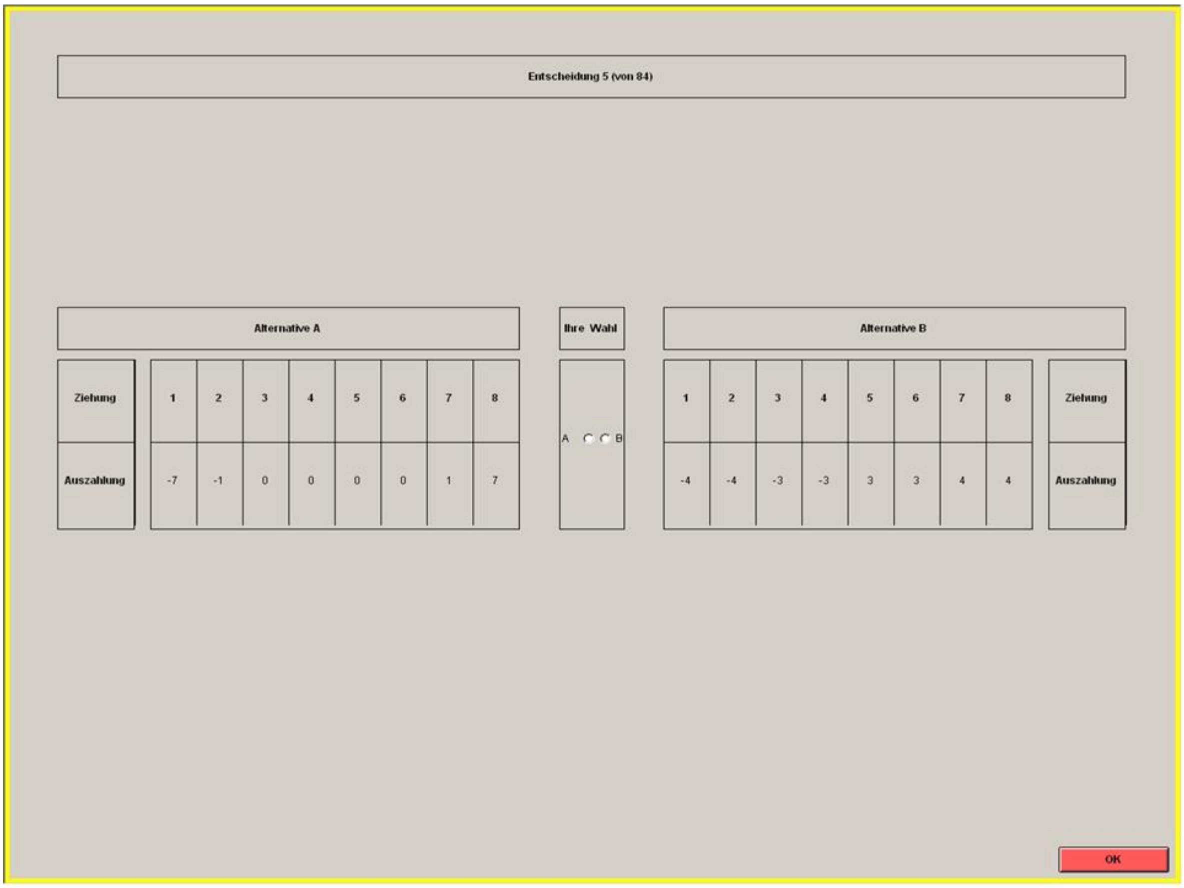
1. Alternative A as well as alternative B can have **2 potential realizations**. Both realizations (drawing 1 and 2) are equally likely. This means it can be expected that the first realization occurs as often as the second realization. You have to consider if you rather want to have the potential realizations of alternative A (drawing 1 and 2) or the potential realizations of alternative B (drawing 1 and 2). This is an example:



2. Alternative A as well as alternative B can have **4 potential realizations**. All realizations (drawing 1, 2, 3, and 4) are equally likely. This means it can be expected that each of the four realizations occurs as often as the other ones. You have to consider if you rather want to have the potential realizations of alternative A (drawing 1, 2, 3, and 4) or the potential realizations of alternative B (drawing 1, 2, 3, and 4). This is an example:



3. Alternative A as well as alternative B can have **8 potential realizations**. All realizations (drawing 1-8) are equally likely. This means it can be expected that each of the eight realizations occurs as often as the other ones. You have to consider if you rather want to have the potential realizations of alternative A (drawing 1-8) or the potential realizations of alternative B (drawing 1-8). This is an example:



For each of the 84 decisions, you choose between alternative A and alternative B by pressing the OK button. Remember, that you cannot revise your decision after confirming by pressing OK. When you made all 84 decisions, please raise your hand. An experimenter will come to your seat to make the drawing which determines your payoff. After your draw has been made, you have to enter the drawing result **while being monitored by an experimenter** on the provided screen. You are not allowed to enter something until the experimenter has arrived at your seat. If you violate this rule, you will be excluded from the experiment, including all payments.

Your **payment** for this task, and for the entire second part of this experiment, is calculated as follows:

After you made your 84 decisions, an experimenter will come to your seat to make the drawing which determines your payoff.
For this you have to first draw a ball from a box (called decision box) with 84 numbered balls. The drawn number (1-84) determines which of the decisions is relevant for your payment.
Afterwards, you have to draw a ball each out of two boxes (called realization box A and realization box B). Both realization boxes contain either 2, 4 or 8 numbered balls, according to which of the 84 decisions is relevant for your payoff.
If you have chosen alternative A for the drawn decision, you get the payoff for the drawn realization of A. If this amount is negative, you have to pay this amount. If you have chosen alternative B in the relevant decision, you get the amount of the drawn realization of B. Again, you have to pay the amount if it is negative.

An example:
Suppose that the last example screen (see above) was chosen as the relevant decision by your first draw from the decision box. Therefore, there are 8 balls in each of the realization boxes. Further suppose you now draw ball number 8 from realization box A and ball number 2 from realization box B. If you have chosen alternative A for this decision, your total payoff for today's second part of the experiment is 7 euros. In this case you are paid 7 euros at the end of the experiment. If you have chosen alternative B for this decision, your total payoff for today's second part of the experiment is -4 euros. In this case, you have to pay 4 euros at the end of the experiment.

If you have further questions, please raise your hand. An experimenter will come to your seat and will answer your questions. If you do not have any questions, please put the instructions upside down on the table to signal us that you understood everything.

Chapter 4

Conditional Cooperation in Repeated Public Goods Games: Theory and Experimental Evidence[0]

4.1 Introduction

Conditional cooperation describes a preference where people are willing to cooperate given that others cooperate as well.[1] However, cooperative behavior usually exhibits a self-serving bias which deters conditional cooperation from being perfect. Such imperfect conditional cooperation has been identified as the median as well as average cooperation preference at work in public goods experiments.[2] Recently, Fischbacher and Gächter (2010) argued *empirically* that the imperfectness of conditional cooperation can explain the dynamics of contribution behavior in repeated public goods games and that the interaction of different types of players, where some are purely selfish, is

[0]The empirical part of this chapter (Sections 4.2, 4.3, and 4.5) is based on joint work with Jan Schikora.

[1]The importance of conditional cooperation was early recognized by social psychologists (e.g. Kelley and Stahelski, 1970). Its prevalence has been found in lab as well as field environments (see Gächter, 2007 for an overview).

[2]In order to directly elicit various preference types, it has become common to use the elicitation method of Fischbacher, Gächter, and Fehr (2001) (which builds upon Selten's, 1967 strategy method). Classifications based on this method usually distinguish between conditional cooperators, free-riders, triangle contributors, and unclassifiable subjects, with most weight clearly on the class of conditional cooperators. For instance, Fischbacher, Gächter, and Fehr (2001) find 50% are conditional cooperators and 30% free-riders, Kocher et al. (2008) find between 42% and 81% are conditional cooperators (depending on the location) and, respectively, between 36% and 8% are free-riders. Herrmann and Thöni (2009) find 56% are conditional cooperators while 6% are free-riders. Fischbacher and Gächter (2010) find 55% and 23% are conditional cooperators and free-riders, respectively. And Aurélie and Riedl (2010) report 67% are conditional cooperators while 22% are free-riders.

not needed for contributions to decline over time.[3] Theoretically, however, contributing nothing is the only equilibrium if everybody exhibits a preference for imperfect conditional cooperation.[4]

In this chapter we show that imperfect conditional cooperation can explain the dynamics of contribution behavior also *theoretically* within a non-equilibrium level-k model of strategic thinking. Since previous studies examining conditional cooperation invariably used homogeneous endowments, they have remained silent about the issue whether people condition their contributions on the absolute or relative contributions of others. Under homogeneous endowments absolute and relative conditional cooperation are not distinguishable. In a first step, we implement heterogeneous endowments in a one-shot public goods experiment based on the strategy method, and show that people are on average imperfect *relative* conditional cooperators. It is this preference that becomes important when explaining behavior under the quite realistic and important scenarios of heterogeneous and uncertain endowments.[5] In a second step, we then not only investigate theoretically the role of conditional cooperation for the dynamics of public goods contributions but we additionally use the specific preference of relative conditional cooperation to derive predictions for the effects of heterogeneous and uncertain endowments. In a third step, we finally test the model's predictions in a repeated public goods experiment and find that all predictions of the model are confirmed. More specifically, given their identified preference subjects strategically over-contribute and these over-contributions as well as total contributions are declining over time. Heterogeneously endowed groups contribute less than homogeneous groups under certain endowments and are therefore less efficient in providing the public good. This is not the case under uncertain endowments. Here, poor groups contribute less and rich groups contribute more compared to a situation where endowment levels of players are common

[3] Or as Fischbacher and Gächter (2010, p542) put it: "Our main result [...] is that contributions decline because, on average, people are "imperfect conditional cooperators" who match others' only partly. The presence of free-rider types is not necessary for this result; contributions also decline if everyone is an imperfect conditional cooperator." By contrast, in a recent theoretical model of Ambrus and Pathak (2011) decaying contributions in equilibrium are explained by purely selfish types who induce cooperation among reciprocal types.

[4] This is because in each period the imperfectness of conditional cooperation means that players contribute less than what they expect others to contribute. Hence, the only fixed point is where contributions are zero.

[5] While the literature on the effects of heterogeneous endowments is quite large, it is rather sparse when it comes to uncertain endowments. This is surprising as one can think of many situations where cooperation is important and where individuals usually do not know the endowment of others. Charitable giving, for example, usually occurs without any knowledge about the income of those who donated, but donations become known (Gächter, 2007 lists campaigns where it was practice to reveal the donated amounts of all donators). Team work often involves the observability of contributions by individual team members while they do not know how much disposable time each team member actually has.

knowledge. Thus, uncertainty causes the inequality of the overall income distribution to increase.

Our experimental design, described in Section 4.2, builds upon the design of Fischbacher and Gächter (2010) who also used an experiment combining two parts, the P- and the C-experiment.[6] Subjects take part in both parts of our experiment and keep their endowment level (high (H) or low (L)) throughout. Each subject is paired with two other subjects leading to four possible group structures (LLL, LLH, LHH, or HHH). In our P-experiment we ask subjects to indicate their own contribution for all combinations of others' contributions and endowments. This specific design of the P-experiment allows us to refine the preference of conditional cooperation in two dimensions. The first dimension concerns the question whether subjects condition their contribution on the *average* of others' contributions. This is an important robustness question for previous studies on conditional cooperation that used the elicitation method of Fischbacher, Gächter, and Fehr (2001). All of these studies only asked for averages and thereby implicitly assumed that this is the case, despite the fact that most theoretical models of outcome-based social preferences predict that the inequality of others' contributions matters for own contributions (see e.g. Fehr and Schmidt, 1999; Charness and Rabin, 2002; or Cox and Sadiraj, 2007). The second dimension where our design allows us to refine conditional cooperation preferences concerns the question whether subjects condition their own contribution on the absolute level of others' contributions or whether they take endowment heterogeneity into account and condition on relative amounts. With homogeneous endowments as previously implemented, these two forms of conditional cooperation coincide. Our C-Experiment consists of two treatments, the certainty and the uncertainty treatment. In both treatments subjects play a repeated linear public goods game over ten periods. While subjects know others' endowments in the certainty treatment they know neither others' endowments nor the total group endowment in the uncertainty treatment. The purpose of the C-experiment is to test predictions of our theoretical model.

For this model, we use the refined preference of conditional cooperation as found via the P-experiment. Section 4.3 presents the results of the P-experiment. Here, we indeed find that subjects condition their contributions on the *average* of others' contributions so that the inequality of contributions holding the mean constant does not have an effect on own contributions. Further, the average subject is an imperfect *relative* conditional cooperator who conditions her own contribution on the relative average contribution of

[6] Fischbacher and Gächter (2010) were the first who used the elicitation method of Fischbacher, Gächter, and Fehr (2001), the P-Experiment, to derive predictions for a (simultaneous) repeated public goods game, the C-Experiment (see also Aurélie and Riedl, 2010).

others. Importantly, the slope of the contribution schedule decreases, thereby making relative conditional cooperation more imperfect, in case a subject is in the minority according to her endowment level, i.e. subjects exhibit a minority bias.

This specific preference is then used in Section 4.4 in order to theoretically analyze contribution behavior in a repeated public goods game with $N \in \mathbb{N}$ players and $T \in \mathbb{N}$ periods. The model we present is a non-equilibrium level-k model, where level-k players best respond to level-$k-1$ players (Crawford, Costa-Gomes, and Iriberri, 2010 provide a recent overview of the literature on strategic thinking). Any equilibrium model where all players have preferences that are imperfectly conditional cooperative would predict zero contributions in every period, which is clearly at odds with empirical findings. By contrast, within our non-equilibrium model we show that level-2 players, but not level-1 players, with such preferences strategically over-contribute and that (over-) contributions of all players are declining over time.[7] Intuitively, level-2 players over-contribute because they anticipate that higher own contributions induce higher contributions by other players in the future. This leads them to contribute more than what would have been myopically optimal, despite the fact that these higher contributions by other players have to be matched in the future (since players are partly conditionally cooperative) and this matching is costly (since players are partly selfish). Contributions decrease because level-2 players best respond to level-1 players who in turn best respond to observed contributions in the previous period, because they best respond to level-0 players who just contribute the same as in the previous period. So, in contrast to level-2 players, level-1 players cannot anticipate that their own contribution influences others' future contributions. Hence, level-1 players do not over-contribute and therefore contribute less than what was contributed in the previous period (since players are partly selfish). Level-2 players best respond to that and their contributions are also lower than in the previous period since over-contribution is not too large.

While this first set of results would also obtain under a preference for absolute conditional cooperation, the second set of results is specifically concerned with the effects of heterogeneous and uncertain endowments on contribution behavior, and therefore only obtains under relative conditional cooperation. We show that the minority bias, either acting on the player herself or being anticipated to act on other players, decreases contributions in heterogeneous groups. This is another well-known pattern in public goods experiments and our results show that the specific preference of relative

[7](Strategic) over-contribution is the amount that players contribute more in a repeated public goods game compared to its one-shot version for given beliefs of others' contributions.

conditional cooperation, and especially the minority bias, is responsible for this result.[8] Under uncertain endowments all subjects have similar symmetric beliefs about others' endowments. As a result, subjects in rich groups wrongly believe to be in a poorer group and subjects in poor groups wrongly believe to be in a richer group. Theoretically, these deviating beliefs cause rich groups to contribute more and poor groups to contribute less. Thus, uncertainty causes the inequality of the income distribution over all players to increase. The efficiency loss of heterogeneous (rather than homogeneous) endowments in the provision of the public good no longer remains under uncertainty. In contrast, under uncertainty an efficiency loss (gain) obtains in poor (rich) groups compared to certainty.

In Section 4.5 we present the results of the C-Experiment and test whether the predictions of our model are verified by the data. We find that all predictions of the model are confirmed. Interestingly, the adverse effects of uncertain endowments on poor and rich groups are neutralized when looking at aggregate contributions. That uncertain endowments have no overall effect on contribution behavior is consistent with the sparse literature touching endowment uncertainty.[9] However, adverse effects on poor vs. rich groups have not been investigated. Our result shows that only considering aggregate contributions is not sufficient with regard to policy implications. Furthermore, the results show that the specific preference of relative conditional cooperation drives contribution behavior, not only under certainty but also under uncertainty.

Section 4.6 concludes. In Appendix A we derive and discuss further results of the model that are not testable with the specific design of our C-Experiment, but which have been tested in previous studies. We consider the effect of increasing the efficiency of the public good (or marginal per capita return), the group size (or number of players), the time horizon (or number of periods), and the endowments of all players. We also look at the one-shot prediction of a public goods game and on the effect of increased selfishness on contributions behavior. Our model confirms empirical patterns that have been found with respect to these variations of the game. All proofs of the theoretical model appear in Appendix B and the experimental instructions are devoted to Appendix C.

[8]Studies investigating the effects of heterogeneous endowments include Marwell and Ames (1979), Aquino, Steisel, and Kay (1992), Rapoport and Suleiman (1993), van Dijk and Wilke (1994), Ledyard (1995), Cardenas (2003), Zelmer (2003), Cherry, Kroll, and Shogren (2005), Buckley and Croson (2006), and Anderson, Mellor, and Milyo (2008). Most of these studies indicate that making endowments heterogeneous decreases cooperation. A more detailed discussion is provided in Section 4.5.

[9]The only studies we are aware of that consider the effect of uncertain endowments (although with a quite different focus from ours) are Isaac and Walker (1988), van Dijk and Grodzka (1992), Chan et al. (1999), and Levati, Sutter, and van der Heijden (2007). We discuss how these studies relate to ours in Section 4.5.

4.2 The Experiment

The experiment was computer-based and was conducted at the experimental laboratory MELESSA of the University of Munich. It used the experimental software z-Tree (Fischbacher, 2007) and the organizational software Orsee (Greiner, 2004). 180 subjects (graduate students were excluded) participated in 8 sessions and earned 14.76 Euros (including 4 Euros show-up fee) on average (with a minimum of 10.14 Euros and a maximum of 22.56 Euros) for a duration of approximately 90 minutes. At the beginning of the experiment subjects received written instructions that were read privately by them. At the end of these instructions they had to answer test questions that showed whether everything was understood. There was no time limit for the instructions, and subjects had the opportunity to ask questions in private. The experiment started on the computer screen only after everybody had answered the test questions correctly and there were no further questions.

Contributions in the public goods game were modeled through a standard linear voluntary contribution mechanism (VCM). In each period, the payoff of player i was given by

$$\Pi_i = E_i^k - x_i + \tfrac{A}{N} \sum_{j=1}^{N} x_j \tag{4.1}$$

where E_i^k is player i's endowment, $x_i \in [0, E_i^k]$ her individual contribution, N is the number of group members (with $i, j = \{1, 2, 3\}$), and $\frac{A}{N} = 0.6$ (hence $A = 1.8$) is the efficiency factor that determines how valuable the public good is. Under standard selfish preferences it is predicted that all subjects choose $x_i = 0$, whereas $x_i = E_i^k$ is the social optimum.

In each session of the experiment, half of the subjects had a low endowment of 4 EMU (Experimental Monetary Units, where the exchange rate was such that 1 EMU corresponded to 0.10 Euros) and the other half of the subjects was highly endowed with 8 EMU. Hence, $k = \{L, H\}$ with $E_i^L = 4$ and $E_i^H = 8$. All groups in the experiment consisted of three subjects and thus $N = 3$. Endowments were randomly assigned to subjects (via the instructions[10]) and everybody knew that each subject had the same

[10]The instructions were given in three sets. The first set of instructions at the beginning of the experiment gave information about the general setup and how the VCM worked. The second set of instructions was handed out after everybody had correctly answered the test questions from the first set of instructions. In the second set subjects received information about their type (H or L) and about the first part of the experiment (i.e. the P-Experiment). At this point they only knew that there would be a second part (i.e. the C-Experiment), but did not know how this second part would look like. After the first part of the experiment was completed, subjects finally received the third set of instructions, which gave them information about their treatment and the second part that followed.

chance to be lowly or highly endowed, since they knew that half of the subjects in their session had low and the other half had high endowments.

Subjects kept their endowment type in both parts of the experiment. The first part of the experiment, the P-Experiment, adopts a design first proposed by Fischbacher, Gächter, and Fehr (2001). Here, subjects stated their contributions conditional on others' contributions separately for all possible group structures. Since they only knew their own type but neither the type of the other two group members nor the group structure and thus the total group endowment, we used strategy tables where they stated their contribution given every combination of others' contributions and others' endowments (LL, LH, and HH). Table 4.1 illustrates our elicitation method for the case where one other group member was highly and the other lowly endowed.

Table 4.1: Conditional Contribution Tables

Example for group structure LH		
Contribution of Type L	Contribution of Type H	Own Contribution
...	...	...
0	4	?
1	3	?
2	2	?
2	3	?
...	...	...

Notes: For a complete version of the table refer to the instructions in Appendix C.

In order to make these conditional contributions incentive compatible we also asked for their unconditional contribution in each of the three possible group structures.[11] The stated contribution preferences from the P-Experiment are used in order to explain contribution behavior in the second part of the experiment.[12] While the seminal P-Experiment of Fischbacher, Gächter, and Fehr (2001) and other studies that used this method elicit cooperation preferences conditional on the *average* of others' contributions in *homogeneous* groups, our P-Experiment elicits cooperation preferences conditional on every possible combination of others' contributions in homogeneous *and*

[11]For two randomly selected subjects in each group the unconditional contribution was payoff relevant and for one randomly selected subject the conditional contribution was payoff relevant. Payoffs were then realized according to (4.1) but feedback about the P-Experiment was only given at the end of the session.

[12]The P-Experiment is an 'as if' sequential one-shot public goods game with conditional contributions being contributions of the second mover and unconditional contributions being first-mover contributions. We only use the conditional contributions as stated contribution preferences for the second part of the experiment.

heterogeneous groups. This modification allows us to investigate two open questions concerning conditional cooperation preferences. First, whether subjects condition their contributions on the averages of others' contributions or whether inequality in contributions among others matters. And second, whether subjects condition on relative or absolute contributions.

In the second part, the C-Experiment, subjects played the standard (simultaneous) linear public goods game as described above repeatedly for ten rounds. We used a partner design where groups stayed together for all rounds.[13] Subjects were randomly allocated to groups with the structures LLL, LLH, LHH, and HHH.[14] These group structures appeared equally often in each of the following two treatments.

In the *certainty* treatment, subjects learned before the first period, whether the other two group members were lowly or highly endowed. They further received information after each period how much each of the other group members contributed (together with the information about their endowments). In each period, subjects first stated their belief about others' contributions (knowing others' endowments) before we asked them to state their own contribution (which could be at most as high as their own endowment).

In the *uncertainty* treatment, subjects neither received information about the endowments of the other two group members nor about the total group endowment.[15] They only knew their own endowment. After each period they also received information about others' contribution (but without endowment information). Within each period, we first asked each subject to state her probabilistic belief about the group

[13]Using two parts and connecting stated preferences from the P-Experiment to observed behavior in the C-Experiment, we adopt a design similar to Fischbacher and Gächter (2010). The difference between their and our P-Experiment is outlined above since they used the original design of Fischbacher, Gächter, and Fehr (2001). Moreover, their C-Experiment consists of ten-period linear public goods games with random stranger matching and a group size of four. The design of our C-Experiment is further explained below.

[14]Although these group structures were also used to determine the payoffs from the P-Experiment, feedback about the P-Experiment was only given at the end and after the C-Experiment.

[15]Note that all subjects knew that half of the them were highly and the other half was lowly endowed and therefore knew the endowment distribution in the experiment. They did, however, not know the frequency of each group structure. Conditional on that they believed in an equal frequency of each group structure, the rational expectation of an L (H) type is to be matched with other group members being LL, LH, and HH with probability $\frac{1}{2}$ ($\frac{1}{6}$), $\frac{1}{3}$ ($\frac{1}{3}$), and $\frac{1}{6}$ ($\frac{1}{2}$), respectively. Conditional on that they believed in a random assignment of group structures, the rational expectation of an L or H type converges to probability $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ for being matched with LL, LH, and HH, respectively, if the number of subjects goes to infinity. With 24 subjects in a session, L (H) types would rationally expect to be grouped with LL, LH, and HH with probability $\frac{5}{23}$ ($\frac{6}{23}$), $\frac{12}{23}$ ($\frac{12}{23}$), and $\frac{6}{23}$ ($\frac{5}{23}$), respectively. However, note that we elicit subjects' true expectations to be in each possible group.

structure, then her belief about others' contributions, and finally asked for her own contribution (with the upper limit being her own endowment).[16]

The experiment consisted of eight sessions in total. In seven of those sessions, 24 subjects participated (12 in the certainty and 12 in the uncertainty treatment) and in one session only 12 subjects participated (in the uncertainty treatment). So, we had each possible group structure (LLL, LLH, LHH, and HHH) eight times in the uncertainty treatment and seven times in the certainty treatment. Having finished both parts of the experiment, first the P-Experiment and then the C-Experiment,[17] subjects answered a short questionnaire about socio-economic characteristics and then received feedback about their total earnings from the experiment. Before they left, each subject was paid in private and by a person that was *not* the experimenter.

4.3 Results P-Experiment

The preference for conditional cooperation has been extensively used to empirically explain why people contribute in public goods games. While some studies used more indirect approaches (see e.g. Sonnemans, Schram, and Offerman, 1999; Keser and van Winden, 2000; Brandts and Schram, 2001; Croson, Fatas, and Neugebauer, 2005; Croson, 2007; or Neugebauer et al., 2009), Fischbacher, Gächter, and Fehr (2001) proposed a direct mechanism (based on the strategy method of Selten, 1967) to elicit individuals' willingness to cooperate given that others cooperate as well (see Ockenfels, 1999 for a similar approach). This elicitation method asks subjects to choose their own contribution for every possible average of others' contributions and thereby elicits an individual contribution function. It has been used in other experimental studies in order to investigate the robustness (see Kocher, 2008) or cultural differences (see Kocher et al., 2008; or Herrmann and Thöni, 2009) of conditional cooperation. Fischbacher and Gächter (2010) used it in order to explain the decline of cooperation in public goods games and Aurélie and Riedl (2010) used it to investigate how different degrees of the (im)perfectness of conditional cooperation manifest in contribution behavior.

[16]In both treatments, certainty and uncertainty, beliefs about others' contributions were incentivized in the following way. If their belief was correct, subjects received 2 EMU, if their belief deviated from the actual contribution by only 1 EMU they received 1 EMU, and if their belief deviated by 2 EMU or more, they received nothing. Again, feedback was only given at the end of the experiment. We did not incentivize the stated probabilistic beliefs about others' endowment in the uncertainty treatment, since the only existing method we are aware of (i.e. quadratic scoring rule) requires the assumption of risk neutrality.

[17]We did not reverse the order of conducting the P- and C-Experiment since Fischbacher and Gächter (2010) already showed that it has no effect on outcomes.

In our P-experiment subjects stated their contributions conditional on every combination of others' contributions for each possible group structure. The main purpose of the P-Experiment was to elicit individual contribution preferences that can be used to explain cooperation behavior in the C-Experiment. Previous experiments already showed that the preference of conditional cooperation, i.e. imperfectly matching others contributions, is the average as well as the median preference among subjects. Nevertheless, there are two open questions regarding conditional cooperation that have not been captured by previous experiments. First, whether subjects condition their contribution on the average of others' contributions or whether they take individual contribution differences into account. And second, whether subjects condition their contribution on the relative amount of others' contributions or rather on the absolute amount, which is not distinguishable under homogeneous endowments as previously implemented. We start with the first question.[18]

4.3.1 Conditioning on Averages

The seminal work by Fischbacher, Gächter, and Fehr (2001) that proposed the method we adopted in our P-Experiment in order to elicit conditional cooperation preferences, as well as other papers that used it, asks subjects to state their own contributions conditional on the *averages* of others' contributions. The specific design of our P-Experiment allows us to test whether the implied assumption that subjects condition their own contributions only on the averages of others' contributions can indeed be justified. This is an important question as most theoretical models of social preferences (as discussed below) would predict that the inequality of others' contributions matters for own contributions. Asking only for averages may therefore impose uncontrolled noise on the elicited preference since the inequality of others' contributions may be incorporated in the beliefs of subjects. Moreover, for future research on the theoretical foundations of conditional cooperation it seems important to know not only whether conditional cooperation is relative or absolute (which will be answered in the next section), but also whether inequality increases in others' contributions matter for own contributions.

Suppose imperfect conditional cooperation behavior was driven by inequity aversion as in Fehr and Schmidt (1999). This would imply that own contributions are sensitive to the inequality of others' contributions if one's own group is the relevant reference group (which is assumed). That this is the case, can already be seen in the linear specification of Fehr and Schmidt (1999). For simplicity, suppose groups have

[18] Note that in the entire empirical analysis of the P- and the C-Experiment, all reported p-values are those of two-sided tests. N always denotes the number of independent observations.

homogeneous endowments. Then, one's own predicted contribution is either zero or equal to the lowest of the others' individual contributions. This could reason *imperfect* conditional cooperation behavior in cases where others contributed unequal amounts, since then one's own contribution would match the lowest of the other contributions, which is lower than the average of others' contributions by definition. In these cases the model of Fehr and Schmidt (1999) would predict a change in own contributions for increases in the inequality of others' contributions (holding their average contribution constant) since increases in inequality imply that the lowest of the other contributions changes.[19] A similar prediction could be made in the outcome-based version of the model by Charness and Rabin (2002) for situations where their model can generate (perfect or imperfect) conditional cooperation. The reason is that increasing inequality of others' contributions always decreases the income of *another worst-off* person. Increasing the inequality of others' contributions would also change own contributions in the egocentric-altruism model of Cox and Sadiraj (2007) in case subjects with such convex preferences behaved imperfectly conditional cooperative. The ERC model of Bolton and Ockenfels (2000) can also generate conditional cooperation. However, in their model inequality of others' contributions has no effect on own contributions, since it is only important how one's own payoff compares to the *average* payoff. Note that it is not our intention to test models of social preferences in the context of conditional cooperation. We only use these models as inspiration for the tests we perform in this section and to illustrate that not controlling for the inequality of others' contributions (which may be incorporated in the beliefs when only asking for averages) may have an effect on contribution behavior.

Since all previous studies on conditional cooperation endowed all subjects with the same income, we focus on homogeneous groups but additionally report the results for heterogeneous groups. From the data of our P-Experiment we know how much each subject contributes for any given combination of others' individual contributions and all combinations of others' endowments. We can therefore make multiple pairwise comparisons (i.e. matched pairs) of own contributions where we increase the inequality of others' individual contributions and, at the same time, hold their average contribution constant. As an example, consider a subject that contributes a certain amount given the other group members contribute 2 and 3 units. We then compare this amount to her contribution when the other group members contribute 1 and 4 units. We can make

[19]Note that own contributions would also change for increases in the inequality of others' contributions in cases where their model would predict *perfect* conditional cooperation, i.e. when others give the same amounts.

7 such pairwise comparisons when the other group members are LL and 50 comparisons when they are HH.

In homogeneous poor groups (LLL) we find that none of the pairwise comparisons is significant.[20] Likewise, in homogeneous rich groups (HHH) none of the pairwise comparisons is significant.[21] Moreover, there is no tendency of subjects contributing more or less for increases in inequality since the z-values of the Wilcoxon signed-rank tests are negative *and* positive.[22] This suggests that the implicitly made assumption in the elicitation method of Fischbacher, Gächter, and Fehr (2001) can be justified. Making these multiple pairwise comparisons, we do not find any evidence that subjects change their contributions for increases in the inequality of others' contributions when holding the average contribution of others constant. As stated above, a model of inequity aversion with a linear specification would rather predict changes in own contributions. One can, however, construct a model of inequity aversion with a non-linear specification (where the function of own payoffs is strictly concave rather than linear) that would predict imperfect conditional cooperation where certain inequality increases would not have an effect. In such a model, there would be no effect of inequality increases as long as they do not lead to rank changes in the income distribution. For inequality increases leading to a rank change, the model would still predict changes in own contributions.

We therefore separate the pairwise comparisons into three classes. The first class consists of those comparisons that do not lead to a rank change. For instance, suppose a subject contributes 3 units, when others contribute 2 and 6 units. Then, this subject's rank does not change in the comparison of others contributing 1 and 7 units.[23] The second class consists of comparisons leading to negative rank changes. As an example, consider a subject contributing 1 unit given others contribute both 2 units. Then, there is a negative rank change for the comparison where others contribute 0 and 4 units. If this subject had contributed 3 instead of 1 units, there would have been a positive rank change. This is the third class of pairwise comparisons that we consider.

In homogeneous poor groups (LLL) there is only one potential situation (out of 7) which can generate a rank change. It is the comparison of $x(2,2)$ vs. $x(0,4)$. This comparison generates a negative rank change if $x(2,2) = 1$ and a positive rank change if

[20] Using Wilcoxon signed-rank tests we find $p = 0.182$ for $x(1,1)$ vs. $x(0,2)$, $p = 0.738$ for $x(1,2)$ vs. $x(0,3)$, $p = 0.658$ for $x(1,3)$ vs. $x(0,4)$, $p = 0.248$ for $x(2,2)$ vs. $x(0,4)$, $p = 0.680$ for $x(2,2)$ vs. $x(1,3)$, $p = 0.808$ for $x(2,3)$ vs. $x(1,4)$, and $p = 0.860$ for $x(3,3)$ vs. $x(2,4)$ with $N = 90$ and $x(y_1, y_2)$ being own contribution given the other group members contribute y_1 and y_2.

[21] Using Wilcoxon signed-rank tests we find $0.111 \leq p \leq 0.988$ with $N = 90$ for the 50 comparisons.

[22] In LLL two are positive and five are negative. In HHH 12 are positive and 38 are negative.

[23] Note that we defined rank changes to be strict. For instance, there is no rank change in the comparison of $x(2,2) = 2$ vs. $x(1,3)$, but there is a rank change when comparing $x(2,2) = 1$ vs. $x(0,4)$.

$x(2,2) = 3$. Out of 90 subjects, 67 subjects contributed amounts that did not generate a rank change in this comparison, 9 subjects had a positive rank change, and 14 subjects had a negative rank change. Those with a positive rank change contribute significantly more and those with a negative rank change contribute significantly less (as predicted by the above mentioned model of inequity aversion).[24] However, since there is only one potential situation generating a rank change in homogeneous poor groups and only few observations that in fact experience rank changes in this situation, we should not derive any conclusions from this result before having analyzed homogeneous rich groups.

In homogeneous rich groups (HHH) there are 22 pairwise comparisons (out of 50) that can possibly generate rank changes. Of those 22 situations that induce positive rank changes only one is significant.[25] Of the 22 situations inducing negative rank changes only two are significant.[26] Although the direction of change in the three significant situations is consistent with the predictions of the model, in the remaining 39 comparisons rank changes do not have a significant effect.[27] Again, even the z-values of the Wilcoxon signed-rank tests are both negative and positive suggesting that there is not even a tendency that can be observed in these 39 non-significant cases.[28]

Moreover, in none of the 57 comparisons inducing no rank change, neither in poor (7 comparisons) nor in rich groups (50 comparisons), do we observe any significant effect of inequality increases. Overall, distinguishing between rank changes and no rank changes does not lead to any results different from those for the pooled data. Rank changes do not have an effect in virtually all cases. In those few cases where they have an effect, the direction of change is such that own contributions increase for higher ranks and decrease for lower ranks in the income distribution.

So far, our analysis used data which was aggregated over subjects and tested whether there is a prevalent pattern that can be observed. Since each subject makes

[24]Using Wilcoxon signed-rank tests we find for the comparison of $x(2,2)$ vs. $x(0,4)$ $p = 0.467$ with $N = 67$ if there is no rank change, $p = 0.083$ with $N = 9$ if there is a positive rank change, and $p = 0.046$ with $N = 14$ if there is a negative rank change.

[25]Using Wilcoxon signed-rank tests we find for the comparison of $x(2,2)$ vs. $x(0,4)$ $p = 0.189$ with $N = 68$ if there is no rank change, $p = 0.083$ with $N = 7$ if there is a positive rank change, and $p = 0.317$ with $N = 15$ if there is a negative rank change.

[26]Using Wilcoxon signed-rank tests we find for the comparison of $x(3,4)$ vs. $x(0,7)$ $p = 0.240$ with $N = 60$ if there is no rank change, $p = 0.934$ with $N = 9$ if there is a positive rank change, and $p = 0.046$ with $N = 21$ if there is a negative rank change. For the comparison of $x(4,4)$ vs. $x(1,7)$ we find $p = 0.290$ with $N = 67$ if there is no rank change, $p = 1.000$ with $N = 4$ if there is a positive rank change, and $p = 0.080$ with $N = 19$ if there is a negative rank change

[27]Using Wilcoxon signed-rank tests we find $0.317 \leq p \leq 1.000$ with $0 \leq N \leq 13$ in 21 situations inducing positive rank changes and $0.159 \leq p \leq 0.934$ with $6 \leq N \leq 27$ in 20 situation inducing negative rank changes.

[28]In situations inducing positive rank changes five are negative, nine are positive, and seven have $p = 1.000$. Concerning negative rank changes 15 are negative and five are positive.

multiple decisions where the average but not the inequality of others' contributions remains the same, we can further use Binomial tests to classify subjects on the individual level. Subjects are classified if their behavior can be significantly distinguished from random behavior based on the 5%- (1%-) significance level. Otherwise, they remain unclassified.

Our within-subject analysis yields a result similar to our across-subject analysis. In homogeneous poor groups (LLL), 77 (62) out of 90 subjects can be classified as contributing the same for inequality increases in others' contributions. No subject can be classified as contributing differently and 13 (28) subjects remain unclassified. In homogeneous rich groups (HHH), 79 (78) out of 90 subjects are classified as contributing the same, none is classified as changing her contribution, and 11 (12) remain unclassified.

Our results do not change when we consider heterogeneous instead of homogeneous groups. Out of 90 H types 69 (56) can be classified as contributing the same when the other group members are LL, none can be classified as contributing differently, and 21 (34) remain unclassified. With the other group members being LH 65 (53) H types contribute the same, none contributes differently, and 25 (37) remain unclassified. Out of 90 L types 87 (87) can be classified as contributing the same when the other group members are HH, none can be classified as contributing differently, and only 3 (3) remain unclassified. If the other group members are LH instead, 73 (68) L types contribute the same, none differently, and 17 (22) remain unclassified.

Neither in homogeneous nor in heterogeneous groups can we classify a single subject as changing her contribution for increases in the inequality of others' contributions. Our within-subject analysis therefore supports the results from our across-subject analysis.

Result 4.1 *The hypothesis that subjects condition their contribution on the average of others' contributions cannot be rejected. Increasing the inequality of others' contributions while holding the mean constant does not have an effect on own contributions.*

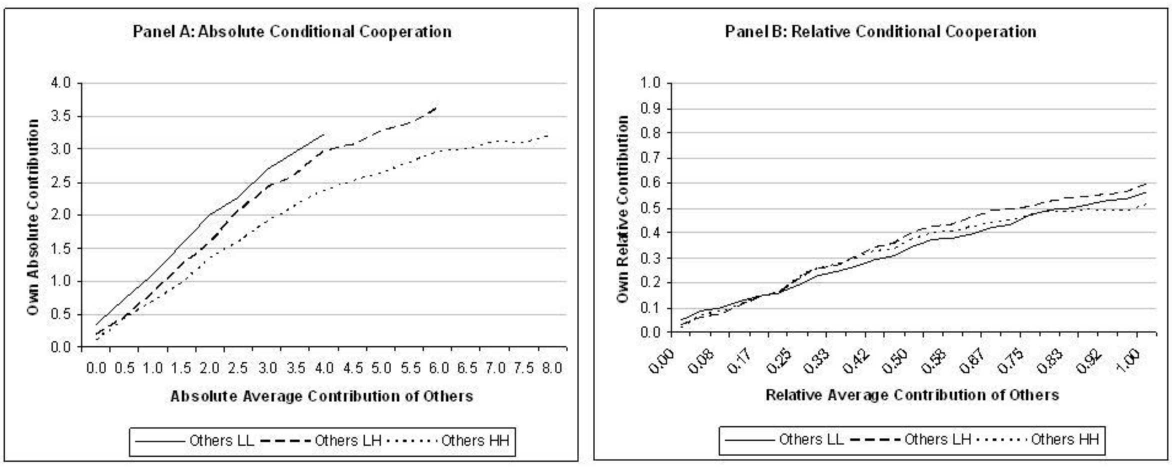
Result 4.1 is important for two reasons. First, for empirical purposes it shows that the elicitation method of Fischbacher, Gächter, and Fehr (2001) is robust towards our design variation and conditioning on averages is in fact an assumption that can be justified. And second, for theoretical foundations of conditional cooperation it shows that a conditioning on averages rather than on individual contributions is important for contribution behavior.

4.3.2 Relative Conditional Cooperation

In all previous studies on conditional cooperation individuals had the same endowment. Under homogeneous endowments, there is no difference whether subjects condition on absolute or rather on relative contribution levels. Under heterogeneous endowments it becomes, however, important whether subjects (imperfectly) choose to match the absolute amount of the average contribution of other group members or whether they choose contributions depending on others' average contribution relative to their endowments. In the latter but not in the former case would subjects contribute more the poorer their group members are.[29]

Figure 4.1 shows the average contribution schedule separately for each combination of others' endowments.[30] Figure 4.1A shows own absolute contributions (on the vertical axis) given the absolute average level of others' contributions (on the horizontal axis) separately for the cases where the other group members are LL, LH, and HH. Subjects contribute more in poorer groups, holding the absolute level of others' average contribution constant. The slopes of the average contribution schedules in Figure 4.1A are less than one, and the larger they are the poorer are the group members. While positive slopes of less than one just confirm the stylized fact of *imperfect* conditional cooperation, the relation between slopes and others' endowments, i.e. greater slopes in poorer groups, remains to be explained.

Figure 4.1: Conditional Cooperation Preferences: Absolute and Relative



[29]Note that it is not our intention to test any foundations of *why* people may want to condition their contributions on relative rather than on absolute contributions of others. We are rather interested in disclosing the systematic behavioral patterns that can be empirically observed.

[30]Figure 4.1 as well as the regression analysis is based on all subjects and does not separate into specific *preference types* like free-riders, conditional or triangle cooperators. We therefore investigate the average preference of all individuals.

That subjects contribute more in poorer groups may be due to the fact that relative rather than absolute contributions matter. Indeed, that this is the case can be seen in Figure 4.1B. Figure 4.1B plots own relative contributions (on the vertical axis) given others' relative average contributions (on the horizontal axis) separately for the cases when others are LL, LH, and HH. Since subjects condition their contributions on others' relative rather than absolute contributions, we do not observe a difference in contribution schedules between the combinations of others' endowments in Figure 4.1B. The differences in the slopes between the situations in which group members are LL, LH, or HH in Figure 4.1A disappear almost completely when we control for their endowment level in Figure 4.1B. We interpret this observation as subjects matching relative rather than absolute average contributions of others. Still, the slopes are smaller than one, so relative conditional cooperation is imperfect as well.

There are two ways subjects can conditionally cooperate on relative levels. They can condition their own contribution either on others' relative average contributions (as in Figure 4.1B) or on the average of others' relative contributions. In the first specification they would take the average of others' individual absolute contributions and then divide it by the average of others' absolute endowment. In the second specification subjects would take others' individual absolute contributions, divide them by others' individual absolute endowments, and then take the average of these relative individual contributions. Both ways of relative conditional cooperation coincide when the endowments of other group members are the same. But they do not coincide when others are differently endowed. Consider an example where one group member is lowly and the other is highly endowed. Suppose the L type contributes 4 EMU and the H type contributes 0 EMU. In the first case a subject would (imperfectly) match a contribution of 33.3% ($= \frac{4+0}{4+8} * 100$) of her endowment, and in the second case she would (imperfectly) match a contribution of 50% ($= [\frac{4}{4} + \frac{0}{8}] * 100$) of her endowment. So we need to find out whether the first specification, which we used in Figure 4.1B, can explain our data better than the second specification. In the following quantitative analysis we will detect which of these specifications yields more reasonable results.

To quantify conditional cooperation in absolute as well as in relative terms we estimate OLS regressions[31] of the following specification.

$$OwnCon_k = \beta_0 + \beta_1 \; OthCon_g + \beta_2 \; OthCon_g * D_{LH} + \beta_3 \; OthCon_g * D_{LL} + \epsilon. \quad (4.2)$$

[31] We additionally ran a Tobit regression, which disregards the clustering of the data but takes the censoring of the dependent variable into account. Since we received similar results we only report those of the OLS regression.

The dependent variable $OwnCon_k$ is an individual's own contribution where $k = \{A, R\}$. $OwnCon_A$ stands for own absolute contribution and $OwnCon_R$ for own relative contribution (i.e. own absolute contribution relative to own endowment). $OthCon_g$ is the average of others' contributions where $g = \{A, R1, R2\}$. $OthCon_A$ represents others' absolute average contribution, $OthCon_{R1}$ is others' relative average contribution (first specification from above), and $OthCon_{R2}$ is the average of others' relative contributions (second specification from above). D_{LL} is a dummy variable that takes the value one for being in a group where the others are LL, D_{LH} is a dummy variable for others being LH, and ϵ is the individual error term. For relative conditional cooperation under the first specification (with $k = R$ and $g = R1$) we ran (4.2) with all subjects (M2; Table 4.2), only with H types (M4; Table 4.2), and only with L types (M5; Table 4.2). For relative conditional cooperation under the second specification (with $k = R$ and $g = R2$) and absolute conditional cooperation (with $k = A$ and $g = A$) we ran (4.2) with all subjects in (M3) in Table 4.2 and in (M1) in Table 4.2, respectively.

Table 4.2 reports regression results for all five specifications.[32] Let us first look at (M1) in Table 4.2. The marginal effect of an additional unit of others' absolute average contribution is 0.4177 ($= \beta_1$) when others are HH, 0.5831 ($= \beta_1 + \beta_2$) when others are LH, and 0.7234 ($= \beta_1 + \beta_3$) when others are LL. So, individuals contribute more the more others contribute. The marginal effect is strictly lower than one. This indicates a slope of absolute conditional cooperation which is smaller than one. Moreover, holding others' absolute average contribution constant, individuals contribute significantly more the poorer their group members are. The slope of the contribution schedule significantly increases by 0.1654 when going from HH to LH and by another 0.1403 when going from LH to LL, since the difference between β_2 and β_3 is also significant at the 1%-level (F-test; $p = 0.0000$).

These differences (almost) disappear when we consider own relative contribution given the relative average contribution of others in (M2) in Table 4.2. The corresponding marginal effects are 0.5258 ($= \beta_1$) when others are HH, 0.5764 ($= \beta_1 + \beta_2$) when others are LH, and 0.5088 ($= \beta_1 + \beta_3$) with others being LL. Now, individuals increase their relative contribution by roughly half of the increase of others' relative average contribution. Interestingly, they make almost no difference between groups anymore. The difference in the slopes between HH and LL is not significant. However, despite being very small there is a significant difference both between HH and LH and between LL and LH, since the difference between β_2 and β_3 is significant at the 5%-level (F-test; $p = 0.0122$). Before investigating the reasons for the small but significant differences in

[32] The results of Table 4.2 are qualitatively similar when we include dummy variables for the different group structures (LL and LH) in our regression.

these slopes, we first consider the second specification of relative conditional cooperation below.

Table 4.2: Regression Results of Absolute and Relative Conditional Cooperation

	Dependent Variables				
	(M1) Abs.: $OwnCon_A$ and $g = A$	(M2) Rel. 1: $OwnCon_R$ and $g = R1$	(M3) Rel. 2: $OwnCon_R$ and $g = R2$	(M4) Rel. H: $OwnCon_R$ and $g = R1$	(M5) Rel. L: $OwnCon_R$ and $g = R1$
$OthCon_g$	0.4177***	0.5258***	0.4600***	0.6192***	0.4324***
	(0.0329)	(0.0380)	(0.0335)	(0.0520)	(0.0538)
$OthCon_g * D_{LH}$	0.1654***	0.0506**	0.1443***	-0.0227	0.1238***
	(0.0197)	(0.0231)	(0.0225)	(0.0298)	(0.0338)
$OthCon_g * D_{LL}$	0.3057***	-0.0170	0.0698**	-0.1900***	0.1560***
	(0.0299)	(0.0322)	(0.0305)	(0.0359)	(0.0471)
Constant	0.4646***	0.0773***	0.0878***	0.0778***	0.0767***
	(0.0778)	(0.0122)	(0.0119)	(0.0172)	(0.0175)
Observations	18900	18900	18900	9450	9450
Clusters	180	180	180	90	90
R^2	0.1433	0.1650	0.1441	0.2183	0.1417

Notes: OLS Regression with data from the P-Experiment. Robust standard errors (clustered in subjects) are in parentheses. *** denotes significance at the 1%-level, ** denotes significance at the 5%-level.

In (M3) in Table 4.2 we consider own relative contribution given the average of others' relative contributions. We find that the differences in the slopes of the absolute contribution schedules across different endowments of others (LL, LH, or HH) do not disappear when we consider our second alternative relative specification. The marginal effect is 0.4600 ($= \beta_1$) when others are HH, 0.6043 ($= \beta_1 + \beta_2$) when others are LH, and 0.5298 ($= \beta_1 + \beta_3$) when others are LL. Therefore, all differences in the slopes between group structures are larger and remain significant when the second specification of relative conditional cooperation is used. Note that the difference between LH (β_2) and LL (β_3) is also significant at the 1%-level (F-test; $p = 0.0092$). In addition, the coefficient of determination (i.e. R^2) of (M2) is larger than that of both (M3) and (M1). We therefore choose (M2), the first specification of relative conditional cooperation, as our preferred model.

As mentioned above, the preference of absolute conditional cooperation and relative conditional cooperation coincide with homogeneous endowments. Our results from Figure 4.1 and Table 4.2 therefore extend existing results of conditional cooperation preferences to heterogeneous endowments. On average, individuals condition their own contribution on others' relative average contribution but not on the average of others' relative contribution. So far, our analysis used the data of all subjects and we did not investigate whether there are differences between H and L types. The last two columns (M4 and M5) of Table 4.2 estimate (4.2) separately for H and L types, using the first specification of relative conditional cooperation.

In (M4) in Table 4.2 we see that for H types the marginal effect of others' relative average contribution on own relative contribution is 0.6192 ($= \beta_1$) when others are HH, 0.5965 ($= \beta_1 + \beta_2$) when others are LH, and 0.4292 ($= \beta_1 + \beta_3$) when others are LL. The difference between HH and LH is very small and not significant, but the differences between HH and LL and also between LL and LH are significant, since the difference between β_2 and β_3 is significant at the 1%-level as well (F-test; $p = 0.0000$). This shows that H types contribute the same in relative terms when others are HH and LH, but they contribute significantly less when others are LL.

In (M5) in Table 4.2 we consider own relative contribution for L types given the relative average contribution of others. The slopes of their contribution schedule are 0.4324 ($= \beta_1$) when others are HH, 0.5562 ($= \beta_1 + \beta_2$) when others are LH, and 0.5884 ($= \beta_1 + \beta_3$) when others are LL. The differences between HH and LH and between HH and LL are significant, while the difference between LH and LL is not (F-test; $p = 0.4166$ for the difference between β_2 and β_3). L types contribute the same in relative terms when others are LL and LH, but they contribute significantly less when others are HH. This is the reverse pattern of H types. Both types seem to have a preference for relative conditional cooperation, but with a negative bias if their type is the minority in the group. Note, however, that both types still contribute more in absolute terms for a given absolute average contribution of others the poorer their group members are.

In order to investigate whether both L and H types show the same relative conditional cooperation behavior in situations without negative minority bias, we ran additional OLS regressions of the following form.

$$OwnCon_R = \beta_0 + \beta_1\ OthCon_{R1} + \beta_2\ OthCon_{R1} * D_H + \epsilon. \tag{4.3}$$

Our dependent variable remains $OwnCon_R$. D_H is a dummy variable that is one for being an H type (and zero for being an L type). We ran (4.3) when others are LL (M6; Table 4.3), LH (M7; Table 4.3), and HH (M8; Table 4.3).

Table 4.3: Regression Results of Type-Specific Relative Conditional Cooperation

	Dependent Variables		
	(M6) $OwnCon_R$ when others LL	(M7) $OwnCon_R$ when others LH	(M8) $OwnCon_R$ when others HH
$OthCon_{R1}$	0.6041***	0.5707***	0.4085***
	(0.0471)	(0.0528)	(0.0524)
$OthCon_{R1} * D_H$	-0.1576***	0.0419	0.1885**
	(0.0551)	(0.0674)	(0.0738)
Constant	0.0666***	0.0677***	0.0914***
	(0.0150)	(0.0143)	(0.0142)
Observations	2700	8100	8100
Clusters	180	180	180
R^2	0.2414	0.1903	0.1488

Notes: OLS Regression with data from the P-Experiment. Robust standard errors (clustered in subjects) are in parentheses. *** denotes significance at the 1%-level, ** denotes significance at the 5%-level.

First, consider (M7) in Table 4.3. The slopes of the relative contribution schedule are not statistically different between L types ($\beta_1 = 0.5707$) and H types ($\beta_1 + \beta_2 = 0.6126$) when others are LH. This picture changes, however, when we consider (M6) and (M8) in Table 4.3. In (M6) we see that H types ($\beta_1 + \beta_2 = 0.4465$) contribute significantly less than L types ($\beta_1 = 0.6041$) when others are LL. L types still contribute the relative amount they contributed when others were LH (this can be seen from (M5) in Table 4.2), but H types contribute less now. When looking at (M8) in Table 4.3 we see the reverse picture. Now, L types ($\beta_1 = 0.4085$) contribute significantly less than H types ($\beta_1 + \beta_2 = 0.5970$) when others are HH. H types still contribute at the relative level when others were LH (again this can be seen from (M4) in Table 4.2), but L types contribute less.

Combining the analysis of both tables yields a complete picture of conditional cooperation preferences. We first showed that on average subjects are imperfect relative conditional cooperators who condition their own contribution on others' relative average contributions rather than on the average of others' relative contributions (this can be seen from (M1) to (M3) in Table 4.2). Instead of looking at the average subject we then considered L and H types separately in a second step. Here, we showed that both types, despite having a relative conditional cooperation preference, have a negative bias for being the minority type. H types contribute less in relative terms when others are LL compared to when others are LH or HH. L types contribute less in relative terms when others are HH than what they contribute when others are LH or LL (this can be seen from (M4) and (M5) in Table 4.2 and (M6) to (M8) in Table 4.3).

The observed minority bias may capture the effect of group identity on social preferences which has recently been analyzed by Chen and Li (2009). They show that outgroup (as opposed to ingroup) identity decreases charity concerns, reciprocity, and the likelihood of choosing the social-welfare-maximizing action. In our experiment, subjects may identify themselves as outside the group if they are the only one's with a certain endowment. When considering the average subject, the minority bias causes relative contributions to shift downward in cases where others are LL and HH, but not in cases where others are LH. This explains why we found (in (M2) in Table 4.2) that on average subjects contribute significantly more when others are LH than when others are LL or HH. Nevertheless, despite having a negative minority bias making the imperfectness in the preference of relative conditional cooperation larger, the clear pattern of giving more in absolute terms the poorer the group members are, still holds.

Result 4.2 *On average subjects are imperfect relative (rather than absolute) conditional cooperators who condition their own contribution on the relative average contribution of others (rather than on the average of others' relative contribution). In addition, the slope of their contribution schedule decreases (leading to increased imperfectness) in case they are the minority type with respect to their endowment, i.e. they exhibit a minority bias.*

As a robustness check for the preceding regression analysis, we can additionally perform non-parametric tests. For any given combination of others' individual contributions (not averages) we can test whether subjects' own contributions are higher or lower for changes in the endowment of other group members. For instance, assume that one group member contributes 2 EMU and the other group member contributes 4 EMU. We can then compare a subject's own contribution for these given contributions of others in the situations where the other group members are LL, LH, and HH.

In doing so we use Wilcoxon signed-rank tests to make pairwise comparisons for every possible combination of others' contributions between the situations where the other group members are LL vs. LH and LH vs. HH.[33] Our results strongly support the regression analysis. For the comparison of LL vs. LH we find that out of 15 combinations of others' contributions 14 are significant ($0.0000 \leq p \leq 0.0151$; $N = 180$). The only one that is not significant in the comparison of LL vs. LH is when others both contribute nothing ($p = 0.3210$; $N = 180$), where we indeed should not expect any difference. If others contribute a given individual amount that is positive for at least one of them, subjects contribute significantly more when others are LL than what they do when

[33]We could also test LL vs. HH. However, if we can show a significant difference in a consistent direction for the comparisons of LL vs. LH *and* LH vs. HH, testing LL vs. HH will be redundant.

others are LH.[34] For the comparison of LH vs. HH we get similar results. Out of 25 combinations of others' contributions 23 are significant at the 1%-level ($0.0000 \leq p \leq 0.0017$; $N = 180$) in the comparison of LH vs. HH. Again, as expected, when both others contribute nothing we do not observe a difference ($p = 0.4624$; $N = 180$). We also do not observe a significant difference between LH vs. HH for the combination where one of the others contributes nothing and the other contributes 1 EMU ($p = 0.5146$; $N = 180$).[35] However, for all other combinations subjects contribute significantly more when others are LH instead of HH.

In a recent study Reuben and Riedl (2011) investigate contribution norms. In a public goods setting comparable to ours (heterogeneous endowments and unrestricted contribution possibilities), they find (using a questionnaire study of uninvolved subjects) contributions proportional to endowments as the modal normative contribution rule. Furthermore, in their 10-period public goods experiment, relative contributions of the minority type (the high type) are lower than those of the majority types (the low types).

4.4 Deriving Hypotheses

Having empirically determined the preference of subjects as measured in the P-Experiment, we will now ask what this preference theoretically implies for the standard repeated public goods game as implemented in the C-Experiment. We will approach this question in three steps. In a first step, we will present a theoretical framework that accounts for Results 4.1 and 4.2 of our (one-shot) P-Experiment. In a second step, we will then use this framework to analyze play in the repeated public goods game. The model we propose is a non-equilibrium model of level-k thinking. Our first set of results are general and do not rely on the specific preference of *relative* conditional cooperation, but would also obtain under *absolute* conditional cooperation. In a third step, we will thereafter present the second set of results that rely on the assumption that players are relative and not absolute conditional cooperators. More specifically, these are results that are important in the scenarios of heterogeneous and uncertain endowments. For all theoretical results discussed in this section we will derive testable empirical hypotheses. Additional results of the model which are not testable with the specific design of our C-Experiment can be found in Appendix A.

[34] In the comparisons with LH, the lower contribution is attached to the L type and the higher contribution to the H type.

[35] The non-significance of this combination disappears for the comparison of LL vs. HH. Here, subjects contribute significantly more at the 1%-level if one of the others contributes nothing and the other contributes 1 EMU ($p = 0.0000$; $N = 180$). Obviously, for all other combinations the comparison of LL vs. HH also yields significant results.

4.4.1 Theoretical Framework

The theoretical framework we present in this section presumes a preference of conditional cooperation and is not a foundation of why players have such a preference. In that sense our approach is agnostic towards where conditional cooperation comes from. Any model that is able to generate relative conditional cooperation could be embedded in our framework. We choose such an agnostic approach because our focus is on what imperfect conditional cooperation theoretically implies in a finitely repeated public goods game.

Each player i's utility is given by

$$u_i = \delta_i \ m(\Pi_i) + (1 - \delta_i) \ v(x_i, g(\cdot)) \tag{4.4}$$

where $\Pi_i = E_i + (\frac{A}{N} - 1)x_i + \frac{A}{N} \sum_{j \neq i} x_j$ is the payoff from private and public investment of player i (as already introduced in Section 4.2), where $\frac{A}{N} < 1$, $A > 1$, and $E_i > 0$ is player i's endowment which is common knowledge to all players. $m(\cdot)$ is the utility of this payoff. $v(\cdot)$ is a function that generates disutility depending how much i's contribution $x_i \in [0, E_i]$ deviates from the target contribution generated by the function $g(\cdot)$. For simplicity, we further assume that $m(y) = y$ and $v(\cdot) = -(x_i - g(\cdot))^2$.[36] Relative conditional cooperation is captured by the function $g(\cdot)$ for which it is assumed that

$$g(\cdot) = \alpha_i \frac{\sum_{j \neq i} x_j}{\sum_{j \neq i} E_j} E_i. \tag{4.5}$$

So, a player i's target contribution in $v(\cdot)$ is determined by the contributions of other players $j \neq i$ relative to their endowment and her own endowment. According to (4.5) the target is to contribute a relative amount that is an α-match of others' relative contributions. Hence, α_i measures how (im)perfect a player aims to match others' contributions. In order to capture the minority bias of our P-Experiment we further assume that

$$\alpha_i = \begin{cases} 0 < \overline{\alpha}_i \leq 1 & \text{if } |\{j \neq i \text{ with } E_j = E_i\}| \geq |\{j \text{ with } E_j \neq E_i\}| \\ 0 < \hat{\alpha}_i < \overline{\alpha}_i & \text{if } |\{j \neq i \text{ with } E_j = E_i\}| < |\{j \text{ with } E_j \neq E_i\}|. \end{cases} \tag{4.6}$$

[36] With these functional form assumptions on $m(\cdot)$ and $v(\cdot)$ we follow Ambrus and Pathak (2011). However, since Ambrus and Pathak (2011) focus in their equilibrium model on type interactions, they only consider players with either $\delta_i = 1$ (so that $u_i = \Pi_i$) or $\delta_i = 0$ (so that $u_i = -(x_i - g(\cdot))^2$). Moreover, since our focus is on conditional cooperation, utility in our model also departs from its specification in Ambrus and Pathak (2011) by the assumption on the g function. We discuss further deviations of the two models at the end of Section 4.4.

If there is no minority bias the target is to match others' relative average contributions, either perfectly ($\overline{\alpha}_i = 1$) or imperfectly ($\overline{\alpha}_i < 1$). And if there is a minority bias, this matching becomes more imperfect ($\hat{\alpha}_i < \overline{\alpha}_i$). In (4.6), the minority bias is specified weakly and will not occur if and only if an individual is part of the majority group. Thus, even if the number of players with equal endowments is the same as the number of players with different but equal endowments, the minority bias occurs. The exact specification of when the minority bias occurs as specified in (4.6) is, however, not crucial for our results. Any other specification where some other minority of players has this bias would generate the same results.[37]

If only social considerations played a role, a player would exactly match the target contribution generated by $g(\cdot)$. However, matching the target comes at the cost of a reduced material payoff from public and private investment. δ_i is a measure of selfishness with $0 \leq \delta_i \leq 1$, where a larger value indicates that player i is more concerned with increasing her material payoff and less concerned with behaving socially, i.e. conditionally cooperative.

In our P-Experiment subjects acted as if they knew how much others contributed and then decided on their own contribution. So, each player i maximizes (4.4) given others' contributions x_j and endowments E_j. The first-order condition specifying the optimal contribution x_i^* is

$$\delta_i(\tfrac{A}{N} - 1) = 2(1 - \delta_i)(x_i - g(\cdot)). \tag{4.7}$$

Since $x_i \geq 0$, we need to restrict δ_i to be sufficiently low in order for (4.7) to hold with equality, i.e. to have an interior solution. So, for given contributions of others, we get a corner solution in which $x_i^* = 0$ if δ_i is too large. For all δ_i smaller than this critical value, (4.7) specifies the optimal contribution with $x_i^* = \frac{\delta_i}{2(1-\delta_i)}(\frac{A}{N} - 1) + \alpha_i \frac{\sum_{j \neq i} x_j}{\sum_{j \neq i} E_j} E_i$. Note that for $\delta_i > 0$ we have $x_i^* < g(\cdot)$. This represents the self-serving bias of conditional cooperation. If $\delta_i = 0$, only social concerns are relevant, hence there will be no self-serving bias and $x_i^* = g(\cdot)$.

[37]Moreover, in the three-player setup of our experiment we can actually not distinguish whether α_i is specified as in (4.6) or whether it is specified by another minority definition, for instance one in which the minority bias only occurs if there is a single subject in the minority (which is in a way the other 'extreme' of (4.6)), i.e.

$$\alpha_i = \begin{cases} 0 < \overline{\alpha}_i \leq 1 & \text{if } \exists\, j \neq i \text{ with } E_i = E_j \\ 0 < \hat{\alpha}_i < \overline{\alpha}_i & \text{if } \nexists\, j \neq i \text{ with } E_i = E_j. \end{cases}$$

In order to find the appropriate general specification of a minority bias, more future research is needed.

Within this framework, the imperfection of relative conditional cooperation may be caused by two different effects. First, a smaller α_i decreases the target that is socially aimed to be matched. And second, a larger δ_i decreases the weight attached to execute this social concern. Nevertheless, pure selfishness as defined via contributing zero no matter what others contribute occurs if and only if $\delta_i = 1$ (since $\alpha_i > 0$). Despite simultaneously contributing to the imperfectness of conditional cooperation, α_i and δ_i measure different causes. α_i measures how imperfect own contributions increase in response to increases of others' contributions, and thus measures the slope of the contribution schedule. δ_i, on the other hand, measures how much own contributions negatively deviate from the target contribution specified by $g(\cdot)$, and thus measures the downward shift of the contribution schedule.[38]

From (4.7) several comparative statics results can be derived that are in line with the observed behavior from our P-Experiment. If others give more, i.e. $\sum_{j\neq i} x_j$ increases, x_i^* will increase by the amount $\alpha_i \frac{E_i}{\sum_{j\neq i} E_j}$. This means that if others increase their relative contributions $\sum_{j\neq i} x_j / \sum_{j\neq i} E_j$ by one percentage point, player i increases her relative contribution also by one percentage point if $\alpha_i = 1$, and by less than one percentage point if $\alpha_i < 1$. This behavior is represented by Figure 4.1B and the regression results of (M2) in Table 4.2. If others are poorer, i.e. $\sum_{j\neq i} E_j$ is smaller, the slope of i's best response function will also be smaller. This is represented by Figure 4.1A and corresponds to the regression results of (M1) in Table 4.2. The minority bias causes x_i^* to be smaller if player i is in the minority. This again alters the slope of i's best response function as can be seen from regression results of (M4) and (M5) in Table 4.2 and from the regression results in Table 4.3. Further, note that if player i becomes more selfish, i.e. δ_i increases, x_i^* will decrease.

Having outlined how the presented theoretical framework accounts for our P-Experiment results, we can now incorporate this framework into the analysis of a repeated game. This allows us to derive hypotheses for the C-Experiment.

4.4.2 A Level-k Model of Strategic Thinking

In this section we analyze a repeated public good game with a finite horizon of $T \in \mathbb{N}$ periods and $N \geq 2$ players. The model we present is a non-equilibrium model of level-k strategic thinking. Level-k thinking was introduced by Nagel (1995) and Stahl and Wilson (1994, 1995) and has been further developed and applied to guessing games by

[38] More specifically, the parameter δ_i is needed for allowing subjects to have corner solutions, i.e. $x_i^* = 0$, for some positive contributions of others, i.e. $\sum_{j\neq i} x_j > 0$. We do, however, not exploit this additional degree of freedom in our results. In fact, results similar to ours could also be derived with a fixed δ_i.

Ho, Camerer, and Weigelt (1998), Bosch-Domènech et al. (2002), and Costa-Gomes and Crawford (2006), to auctions by Crawford and Iriberri (2007a) and related mechanism design questions by Crawford et al. (2009), to hide-and-seek games by Crawford and Iriberri (2007b), to coordination games by Costa-Gomes, Crawford, and Iriberri (2009), and to other normal-form complete-information games by Costa-Gomes, Crawford, and Broseta (2001), and Costa-Gomes and Weizsäcker (2008).[39]

In a level-k model, level-1 (L1) players best respond to the behavior of level-0 (L0) players, level-2 (L2) players best respond to L1 players, and so on. Generally, level-k (Lk) players best respond to level-$k-1$ players (L$k-1$).[40] In these models it is crucial how L0 players, who only exist as hypothetical constructs of higher-level players' beliefs, are specified. Often, it is assumed that L0 players act randomly according to a uniform distribution, which may be a reasonable assumption in some one-shot environments like the Beauty-Contest game. Salience has also been considered to be of importance for L0 players' responses in coordination games. In a repeated game, however, other assumptions on L0 players seem more reasonable. We assume that in each (except the first) period L0 players contribute the same as in the last period. In the first period, they contribute a certain percentage of their endowment, which can be positive or zero.[41] That L0 players are expected to contribute in period t the same as in period $t-1$ captures, in a simple way, higher-level players' adjustment of beliefs about what others will contribute in period t, which only depends on what happened in the previous period $t-1$. It is similar to belief-based learning under 'Cournot best-response'.[42]

It is further assumed that all players are either of type L1 or of type L2. This is consistent with the evidence on the level type distribution in previous studies. "The estimated frequency of L0 is normally zero or small; and the type distribution is fairly stable across games, with most weight on L1 and L2 [...]." (Costa-Gomes, Crawford,

---

[39] Crawford, Costa-Gomes, and Iriberri (2010) provide a great overview of the literature on strategic thinking. In addition to the presented laboratory evidence on level-k reasoning, they also report field evidence supporting such limited strategic thinking (one recent example is Brown, Camerer, and Lovallo, 2011).

[40] This is the difference to a cognitive hierarchy model (see Camerer, Ho, and Chong, 2004), where Lk players best respond to a distribution of lower-level types.

[41] Alternatively and without affecting our results, we could also assume that L0 players act randomly in the first period, where their contribution is drawn according to some common distribution function.

[42] One could think of more sophisticated learning, in which L0 players' contributions in period t depend on more of the history than only the last period. We conjecture that results similar to ours could also be derived under more complicated belief learning. Fischbacher and Gächter (2010) provide empirical evidence that, although a sophisticated belief formation model in which beliefs are a weighted average of previous-period contributions and previous-period beliefs performs best, a naïve version in which beliefs entirely depend on the previous period, also captures the important contribution dynamics. Furthermore, they provide evidence that the way people form their beliefs does not change over time (which is also assumed by our L0 specification).

and Iriberri, 2009, p369) We do not restrict the frequencies of L1 and L2 players and the results of the model will be derived for any possible distribution of types being either L1 or L2 players. The derived empirical hypotheses thereof assume some positive fraction of L2 players.

Both L1 and L2 players are assumed to be homogeneous *only* according to their belief about the level-type of the other players. That is, all L1 players are homogeneous in that they believe all others are L0 players, but they can well be heterogeneous according to their own preferences, i.e. their α_i and δ_i. Similarly, L2 players are homogeneous in that they believe all others are L1 players (who in turn best respond to L0 players), but they can be heterogeneous concerning their belief about the preferences of L1 players. Hence, the believed δ_j and α_j can be heterogeneous among L2 players and need not be correct.[43] Furthermore, L2 players can be heterogeneous according to their own preferences, namely δ_i and α_i. Our results are derived for others' believed preferences satisfying $0 \leq \delta_j \leq 1$ and $0 < \alpha_j \leq 1$, and for own preferences satisfying $0 < \delta_i < 1$ and $0 < \alpha_i \leq 1$. We will, however, additionally discuss our results in the special cases of $\delta_i = 0$ and $\delta_i = 1$.

We also assume that neither the level-type, nor believed and own preferences change over time. So, we exclude any updating of believed preferences. All players anchor their belief about others on the same L0 players, who simply contribute in periods $t \neq 1$ the same as in the previous period and in period $t = 1$ some percentage of their endowment. This 'home-grown' belief for L0 players' first period contribution is the same for all L1 and L2 players.

Since L1 players best respond to L0 players, their utility in period t is

$$u_{iL1}^t = \delta_i \; \Pi_i^t - (1 - \delta_i)(x_i^t - \alpha_i \frac{\sum_{j \neq i} x_j^{t-1}}{\sum_{j \neq i} E_j} E_i)^2 \tag{4.8}$$

where $\Pi_i^t = E_i + (\frac{A}{N} - 1)x_i^t + \frac{A}{N} \sum_{j \neq i} x_j^{t-1}$ for $t = 1, 2, \ldots T$. Since they believe that others contribute some 'home-grown' amount in the first period and the same as in the previous in all other periods, L1 players do not anticipate that their own behavior affects the behavior of others. Their behavior is therefore myopic and does not take future periods into account when they choose the optimal contribution in a given period via a maximization of $U_{iL1}^t = \sum_{\tau=t}^{T} u_{iL1}^{\tau}$. In each period t L1 players face a trade-off where contributing decreases the material payoff Π_i^t but also decreases the disutility from not matching the target contribution specified by what has already happened the

[43] This contains the case where these believed preferences are correct. In this special case we are in a world of complete information with respect to others' preferences. Note that all our results would also hold (and could be derived more easily) in this case.

period before. This is because L1 players believe others contribute in the current period t the same as in the previous period $t-1$.

L2 players best respond to L1 players and, in contrast to L1 players, they therefore anticipate that their own behavior influences others' behavior. Thus, in period t L2 players' utility is

$$u_{iL2}^t = \delta_i \; \Pi_i^t - (1-\delta_i)(x_i^t - \alpha_i \frac{\mathbb{E}[\sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})]}{\sum_{j\neq i} E_j} E_i)^2 \tag{4.9}$$

where $\Pi_i^t = E_i + (\frac{A}{N}-1)x_i^t + \frac{A}{N}\mathbb{E}[\sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})]$. $\sum_{j\neq i} x_j^t(\cdot)$ is the sum of L1 players' best response functions in period t which is drawn according to $F_i(\cdot)$, so that $\mathbb{E}[\sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})] = \int \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1}) \; dF_i(\sum_{j\neq i} x_j^t(\cdot, x_i^{t-1}))$. More specifically, $F_i(\cdot)$ is the distribution over the sum of best response functions of all other players and it is determined by the joint distributions over the believed $\tilde{\delta}_j$ and $\tilde{\alpha}_j$ for each of the other L1 players, where $\tilde{\delta}_j$ and $\tilde{\alpha}_j$ are independent random variables such that the density $f_{ij}(x_j^t(\cdot, x_i^{t-1}, \alpha_j, \delta_j)) = f_{i\tilde{\alpha}_j}(\tilde{\alpha}_j = \alpha_j) f_{i\tilde{\delta}_j}(\tilde{\delta}_j = \delta_j)$.[44] This simply captures the notion that beliefs about others' preferences may be represented by a distribution and are not necessarily fixed to a single point. It results in a distribution over the believed best response function for each other player. Summing these believed best responses over all players yields $F_i(\cdot)$. To simplify notation, we will call $\int \sum_{j\neq i} x_j^t(\cdot) \; dF_i(\sum_{j\neq i} x_j^t(\cdot))$ the believed $\sum_{j\neq i} x_j^t(\cdot)$. Again, in any period t the optimal contribution for L2 player i is derived via a maximization of $U_{iL2}^t = \sum_{\tau=t}^{T} u_{iL2}^{\tau}$.

According to (4.9), contributing does not only have the myopic effect as described above for L1 players and what would be solely derived via a maximization of u_{iL2}^t instead of U_{iL2}^t, but additional effects for future utility. Contributions increase Π_i^{t+1} as other L1 players will increase their contributions in the next period in response. These increased contributions by others, however, also lead to more disutility if they are not matched accordingly in the next period. This raises the question whether contributions of L2 players are higher (or lower) than what is myopically optimal. The difference between optimal and myopic contributions is strategic over-contribution (or under-contribution). So, contributions can be disentangled into two parts, myopic contributions and strategic over-contributions. The latter can only exist for L2 players. Intuitively, if L2 player i anticipates that a higher current contribution yields higher contributions by other players $j \neq i$ in later periods, it may well make sense for her to over-contribute although

[44] That $\tilde{\delta}_j$ and $\tilde{\alpha}_j$ are independent random variables is only used for simplicity. All our results would hold and would even be stronger (in case of Proposition 4.8) if $\tilde{\delta}_j$ and $\tilde{\alpha}_j$ would be negatively correlated.

these higher contributions of others have to be matched by her own contribution in later periods.

Proposition 4.1 *L1 players do neither over- nor under-contribute and their contributions are equal to their myopic contributions in all periods. Contributions of L2 players are not below myopic contributions, i.e. there is weak strategic over-contribution, in period $t \neq T$ if*

$$\tfrac{A}{N} \geq (1 - \tfrac{A}{N})\alpha_i \frac{E_i}{\sum_{j \neq i} E_j}. \tag{4.10}$$

In the last period T, L2 players' contributions are equal to their myopic contributions, i.e. there is no over-contribution.

Proposition 4.1 holds generally and does not require an interior solution, so there is no restriction imposed on (δ_i, α_i) or the believed $(\tilde{\delta}_j, \tilde{\alpha}_j)$ for a given value of the other variables. L1 players always contribute according to their myopic contributions, but L2 players do so only in the last period as future beliefs of others cannot be influenced by own contributions anymore. In earlier periods (4.10) is sufficient for under-contribution not to occur. A special case arises for $\delta_i = 1$, where over-contribution is the only motivation to contribute. Then, contributions of L2 players in the last period are clearly zero, but such a player would contribute everything (nothing) in earlier periods if the believed sum of the slopes of others' best response functions is larger (smaller) than $(1 - \frac{A}{N})/\frac{A}{N}$.[45] By contrast, L1 players would not contribute anything. On the other hand, if $\delta_i = 0$, i.e. only social concerns are relevant, there will be no over-contribution since matching others' contributions is the only concern.

Since Proposition 4.1 also holds in the case where the first-order approach is not valid in a certain period, imposing an assumption by which we can rely on an interior solution in all periods allows us to obtain a clearer statement regarding over-contribution.

Assumption 4.1 *In period $t \neq T$, players plan to have positive contributions in all remaining periods, i.e. $x_i^{*t}, x_i^{*t+1}, x_i^{*t+2}, \ldots, x_i^{*T} > 0$.*

For the interior solutions of the maximization problem it is important what players *plan* to do.[46] However, all theoretical results describe what players *actually* will do as this is of empirical interest. The next result refines Proposition 4.1 for L2 players.

[45] The reason for these boundary solutions in the case of $\delta_i = 1$ lies in the fact that $g(\cdot)$ is linear in x_j.

[46] We omit notation that separates planned actions from actions, since it should be contextually clear which one is used.

Corollary 4.2 *Suppose Assumption 4.1 holds. Then, L2 players over-contribute (neither over- nor under-contribute, under-contribute) in period $t \neq T$ if and only if (4.10) holds strictly (holds with equality, does not hold).*

Contributing more now increases the marginal utility of L2 players in the next period by $\frac{A}{N}$, weighted by the believed sum of the slopes of others' best response function with respect to own contributions and the degree of selfishness. Given optimal contributions in the next period, contributing more now also decreases marginal utility in the next period by $(1 - \frac{A}{N})\alpha_i \frac{E_i}{\sum_{j \neq i} E_j}$, since this is what it costs to contribute more in the next period to match others' higher contributions. Again, it is weighted by the believed sum of the slopes of others' best response function and the degree of selfishness, but neither determines the decision whether to over-contribute or not. They do, however, determine the quantity of over-contribution. Both, a larger believed sum of the slopes of others' best response functions and more selfishness increases over-contribution if (4.10) holds strictly. Note that over-contributions occur solely in the attempt to manipulate future beliefs of others. So, it is quite intuitive that more selfishness and a larger reaction by others increases over-contribution.

The assumption used in Corollary 4.2 *directly* assumes an interior solution in all remaining periods for the maximization problem in period t. Thus, Assumption 4.1 is arguably strong. We can, however, guarantee interior solutions with another set of assumptions.

Assumption 4.2 *In period $t \neq T$, players plan to have positive myopic contributions in the last period T, i.e. $x_i^{*T} > 0$.*

Assumption 4.3 *Condition (4.10) holds, i.e. in period $t \neq T$ players do not plan to under-contribute in all remaining periods $t, t+1, t+2, \ldots, T$.*

Assumption 4.4 *In period $t \neq T$, players do not plan to contribute more in all remaining periods $k = \{t, t+1, t+2, \ldots, T\}$ than the amount of perfect relative conditional cooperation, i.e. $x_{iL1}^{*k} \leq \frac{E_i}{\sum_{j \neq i} E_j} \sum_{j \neq i} x_j^{k-1} \leq E_i$ for L1 players, and $x_{iL2}^{*k} \leq \frac{E_i}{\sum_{j \neq i} E_j} \int \sum_{j \neq i} x_j^k(\cdot)\, dF_i(\sum_{j \neq i} x_j^k(\cdot)) \leq E_i$ for L2 players.*

Assumption 4.4 restricts over-contribution not to be overly large. It cannot be binding in the last period as players do not over-contribute. For the same reason, it cannot be binding for L1 players. Assumption 4.3 says that (4.10) is satisfied such that players do not under-contribute. Again, this is not important in the last period. Assumption 4.2 presumes positive myopic contributions in the last period and since there are no strategic considerations in the last period, this is equivalent to assuming positive contributions in the last period. Jointly, these assumptions imply Assumption 4.1.

Lemma 4.3 *Assumptions 4.2, 4.3 and 4.4 imply Assumption 4.1.*

That all our assumptions are formulated as plans has technical reasons. The drawback is that plans cannot be empirically observed but only actions. There is, however, nothing which would prevent planned actions to be actually executed.[47] For instance, planned as well as actual under-contribution cannot occur if (4.10) holds. Assumption 4.3 is justified by our experimental design, so (4.10) holds strictly since $\frac{A}{N} = 0.6 > 0.4 = 1 - \frac{A}{N}$, and $\frac{E_i}{\sum_{j \neq i} E_j} \leq 1$. In Section 4.5 we will empirically assess whether Assumptions 4.1, 4.2, and 4.4 are also reasonable.[48] With the data from the C-Experiment we can unambiguously support all of them. For expository reasons we will, however, present the empirical analysis of these assumptions later in Section 4.5 before we test our derived hypotheses.

Because our empirical analysis supports these theoretical assumptions and since (4.10) holds strictly in our experiment, we can base our first hypothesis not only on Proposition 4.1 but additionally on Corollary 4.2. Given that there is a positive fraction of L2 players, we should expect the following.

Hypothesis 4.1 *There is over-contribution in all but the last period.*

We now turn to the dynamics of contribution behavior and start with over-contribution. Since L1 players do not over-contribute, the following dynamics of over-contribution only apply to L2 players.

Proposition 4.4 *Suppose Assumptions 4.2 to 4.4 hold. Then, over-contributions in period t are lower than in the previous period $t-1$.*

Note that in Proposition 4.4, we presumed that there is over-contribution. If (4.10) holds with equality, there is no over-contribution which can change over time. As it stands, Proposition 4.4 implicitly assumes that (4.10) holds strictly. Indeed, as we

[47] If we were in a world of complete information, i.e. L2 players would know the preferences of others, it would even be the case that actual contributions being positive in period t (and Assumptions 4.4 and 4.3 being fulfilled) would imply that planned contributions for that period must have been positive as well. The reason is that L2 players believe that others are L1 players. However, some others may in fact also be L2 players who contribute at most equally as if they were L1 players. Hence, actual contributions are at most as high as planned contributions.

[48] Note that empirically supporting Assumptions 4.2 and 4.4 means that Assumption 4.1 should be empirically supported as well (by Lemma 4.3). The empirical support of all three assumptions is therefore suggestive evidence that Lemma 4.3 holds.

already showed in our experiment (4.10) holds strictly and we therefore expect over-contributions to decrease over time. Since Proposition 4.4 implicitly assumes an interior solution in all periods (by Lemma 4.3), it is assumed that δ_i is small enough (for a given value of the other variables). For $\delta_i = 1$, we get that over-contribution is either zero in all periods or constant but dropping to zero in some period (at the latest in the last period), however not increasing. For $\delta_i = 0$, we already saw that there are no incentives to over-contribute, so over-contributions will not change over time.

Intuitively, over-contribution is zero in the last period. In the second to last period, there is over-contribution if L2 player i believes that these higher contributions will be reciprocated in the last period, despite that these higher contributions in the last period have to be matched. In the third to last period, i knows that over-contributing is optimal in the second to last period, and hence over-contributing slightly more than one period later is optimal if over-contribution in the second to last period was already optimal. This reasoning continues until the first period, so clearly planned over-contributions are decreasing. However, also actual over-contributions are decreasing as there is no updating over time for the believed preferences of the other players and Assumption 4.4 assures that the believed sum of the slopes of others' best response functions are weakly decreasing.

Hypothesis 4.2 *Over-contributions decline over time.*

One of the most well-known patterns in the public goods literature is that contributions are declining over time. Contributions are composed of myopic contributions and strategic over-contributions. We already considered (the dynamics of) over-contribution and will now turn to the dynamics of myopic and total contributions.

Proposition 4.5 *Suppose Assumptions 4.2 to 4.4 hold. Then, (myopic) contributions of L1 and L2 players in period t are lower than in the previous period $t-1$.*

Myopic contributions in period t are determined by $\sum_{j \neq i} x_j^{t-1}$ for L1 players and by the believed $\sum_{j \neq i} x_j^t(\cdot, x_i^{t-1})$ for L2 players. A necessary and sufficient condition for myopic contributions to be weakly lower in period t than in $t-1$ is that $\sum_{j \neq i} x_j^{t-1} \leq \sum_{j \neq i} x_j^{t-2}$ for L1 players and the believed $\sum_{j \neq i} x_j^t(\cdot, x_i^{t-1}) \leq \sum_{j \neq i} x_j^{t-1}(\cdot, x_i^{t-2})$ for L2 players. In turn, a sufficient condition for this to hold is that actual contributions of all players are weakly decreasing. In fact, this is the case in all periods if subjects do not cooperate more than perfectly (Assumption 4.4) and if over-contributions are not increasing (Proposition 4.4), since then players contribute in period t an amount at most equal in relative terms to observed relative contributions from the previous period $t-1$ and the conditions assure that what players observe can never be higher than what they observed in the previous period. With these two conditions fulfilled, Proposition 4.5 holds *weakly* for all contribution paths, even if Assumption 4.2 does not hold (because zero contributions trigger zero contributions). However, since we could show the pattern of decreasing over-contributions only under Assumptions 4.2 to 4.4, we impose these assumptions in Proposition 4.5 in which case the proposition even holds *strictly* for $0 < \delta_i < 1$. If $\delta_i = 0$ or $\delta_i = 1$, Proposition 4.5 again holds weakly. Also note that (myopic) contributions will always *ultimately* decrease even if over-contributions are increasing since the latter drop to zero in the last period.

Because the empirical analysis of Section 4.5 supports our assumptions, we can state the following hypothesis.

Hypothesis 4.3 *(Myopic) contributions decline over time.*

In Appendix A we derive and discuss further theoretical results of our model that are not testable with the specific design of our experiment, but which have been found to hold in previous experiments. Among them are well-known and robust patterns of the public goods literature. In Proposition 4.10 we show that increasing the efficiency of the public good (i.e. marginal per capita return) leads myopic as well as over-contributions to increase and hence increases contributions of L1 and L2 players. Increasing the group size (Proposition 4.11), increases over-contributions so that L2 but not L1 players increase their contributions. Over-contributions are also increased if the number of periods increases (Proposition 4.12) so that again L2 but not L1 players increase their contributions. In one-shot public goods games (Proposition 4.13), both L1 and L2 players contribute according to their myopic contributions, which can well be positive. This is in stark contrast to an equilibrium model, where imperfectly conditional cooperative players would also contribute zero in the one-shot version of the game. In Propositions 4.14 and Corollary 4.15 we show that, perhaps surprisingly, larger contributions are not necessarily made by L2 players who are less selfish. In contrast, less selfish L1 players always contribute more. Last, increasing the endowments of all players (Proposition 4.16) increases myopic contributions, thus L1 and L2 players' contributions are increased.

It is interesting to note that all results we discussed so far, do not rely on the specific preference of *relative* conditional cooperation. That subjects condition their own contribution on others' relative rather than absolute contributions implies that they somehow take differences in their endowments into account. An important fact for such behavior may be that these endowments were *randomly* assigned in our experiment. Other scenarios seem plausible where it may be the case that subjects only condition on contributions independent of endowments, for instance when endowments were really earned so that subjects are made responsible for their endowments.[49] Nevertheless, all previous results still hold in this case.

Remark 4.6 *Propositions 4.1 to 4.5 (including Corollary 4.2 and Lemma 4.3) and Propositions 4.10 to 4.16 in Appendix A (including Corollary 4.15) would also hold if players conditioned their own contribution not on relative but on absolute contributions of others, i.e.* $\hat{g}(\cdot) = \alpha_i \frac{\sum_{j \neq i} x_j^t}{N-1}$.

In case subjects rather conditioned on absolute contributions of others we could simply set $E_i = 1$ and $\sum_{j \neq i} E_j = N - 1$, and all results we derived or discussed so far would still hold. The remaining results, however, only obtain in case subjects are *relative* conditional cooperators. So, in the next two sections we focus on what a preference for relative rather than absolute conditional cooperation implies for contribution behavior in the repeated public goods game. Behavior is affected by relative but not by absolute conditional cooperation in the environments of heterogeneous and uncertain endowments.

4.4.3 Heterogeneous Endowments

Under heterogeneous endowments it is the minority bias that has an effect on contribution behavior. This minority bias can either act on player i (*own minority bias*) and/or be anticipated by player i to act on another player $j \neq i$ (*anticipated minority bias*).

Own Minority Bias: Our P-Experiment showed that if player i is the minority type, her matching of others relative contributions becomes (more) imperfect in the sense that α_i decreases. This has two effects in period $t \neq T$. Decreasing α_i reduces myopic contributions since it decreases the target amount aimed to be matched in the current period. However, since it also decreases the target amount in the next period, over-contributing now may become less harmful. Hence, the total effect of player i's minority bias on her contributions is unambiguous for L1 players, but is ambiguous for L2 players. More specifically, the effect on over-contribution is ambiguous and depends on the values of $\frac{A}{N}$ (efficiency of the public good), $\alpha_i \frac{E_i}{\sum_{j \neq i} E_j}$ (marginal matching target), $\sum_{j \neq i} \frac{\partial x_j^{t+l}(\cdot)}{\partial x_i^{t+l-1}}$ (believed sum of slopes of others' best response functions), and $T - t$ (number of periods left). With only one period left, one's own minority bias has a clearly positive effect on over-contribution in that period $(T - 1)$ since there are

[49] In contrast to this, Cherry, Kroll, and Shogren (2005) show that contributions to the public good are *not* any different when endowments were earned instead of randomly allocated.

only myopic contributions in the last period T and these reduced myopic contributions decrease the anticipated matching costs in period $T-1$. However, as the number of periods left increases, (4.10) holds strictly, and $\alpha_i \frac{E_i}{\sum_{j \neq i} E_j} \leq 1$ it becomes more and more likely that the effect on over-contribution is rather negative. The reason is that over-contributions in period t are determined by an optimal reaction to contributions in period $t+1$. With only one period left, one's own minority bias clearly decreases these contributions, but with more periods left one's own minority bias may increase these contributions since not only myopic but also over-contributions are affected. Therefore, the effect of one's own minority bias on over-contribution is ambiguous in $t < T-1$ and positive in $t = T-1$.

Proposition 4.7 *Suppose Assumptions 4.2 to 4.4 hold. Then, ceteris paribus, one's own minority bias decreases L1 players' contributions in period t but has an ambiguous effect on L2 players' contributions in period $t \neq T$. In the last period T, one's own minority bias also decreases L2 players' contributions.*

In order to see how many left periods are needed such that one's own minority bias has a negative effect on over-contributions for L2 players, we calibrate the model with reasonable parameter values of our experiment. From this calibration we are then able to derive hypotheses for our experiment. The calibration shows that the effect on over-contribution should be negative, and is thus in line with the effect on myopic contributions, most of the time. Table 4.4 summarizes the results of this calibration. The efficiency of the public good is not varied in our experiment and is $\frac{A}{N} = 0.6$ throughout. The believed $\sum_{j \neq i} \frac{\partial x_j^{t+l}(\cdot)}{\partial x_i^{t+l-1}}$ is varied and depends on the believed $(\tilde{\delta}_j, \tilde{\alpha}_j)$.[50] The specific numbers implement reasonable lower and upper bounds as found in the P-experiment (see Section 4.3). For instance, suppose the believed $\sum_{j \neq i} \frac{\partial x_j^{t+l}(\cdot)}{\partial x_i^{t+l-1}} = 1$, then one's own minority bias decreases over-contribution with two periods left if and only if $\alpha_i \frac{E_i}{\sum_{j \neq i} E_j} < \frac{1}{4}$, and with three periods left if and only if $\alpha_i \frac{E_i}{\sum_{j \neq i} E_j} < 0.61$. Since $\frac{E_i}{\sum_{j \neq i} E_j} = \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1\}$ in our experiment, the effect on over-contribution is negative in most periods of the game.

[50] For simplicity, the calibration in Table 4.4 assumes that the believed sum of slopes of others' best response functions is constant over time, hence $\sum_{j \neq i}(\partial x_j^{t+l}(\cdot)/\partial x_i^{t+l-1})\ \forall l$. This holds in the model if and only if the believed contribution path does not shift some best responses of others in the support of $F_{ij}(\cdot)$ into the corner solution. Otherwise, the believed sum of slopes of others' best response functions decreases over time, but is still always positive.

Table 4.4: Own Minority Bias Leading to Negative Over-contribution for L2 Players

Believed $\sum_{j\neq i}\frac{\partial x_j^{t+l}(\cdot)}{\partial x_i^{t+l-1}}$	$T-t$		
	2	3	4
0.7	0.04	0.21	0.40
1	0.25	0.61	0.85
1.4	0.39	0.77	0.98
2	0.50	0.86	1.05

For given values of $\sum_{j\neq i}\frac{\partial x_j^{t+l}(\cdot)}{\partial x_i^{t+l-1}}$ and $T-t$, each cell of Table 4.4 gives the critical value of $\alpha_i\frac{E_i}{\sum_{j\neq i}E_j}$ such that over-contribution is negative for all values smaller than this critical value of $\alpha_i\frac{E_i}{\sum_{j\neq i}E_j}$.

For L2 players, one's own minority bias has a negative effect on myopic contributions in all periods, so the total effect on contributions is clearly negative in the last period T and very likely to be negative early in the game. Using reasonable parameter values of our experiment, one's own minority bias should decrease over-contributions for almost the entire game, so we expect average contributions over all periods to decline with one's own minority bias.

Proposition 4.7 requires a ceteris paribus assumption. This is a strong assumption with respect to the same given believed sum of contributions of other players in period t, i.e. $\sum_{j\neq i}x_j^{t-1}$ for L1 players and $\sum_{j\neq i}x_j^t(\cdot,x_i^{t-1})$ for L2 players, because the contribution path may be different with one's own minority bias being present. In fact, if the minority bias decreases contributions early in the game, which is clearly the case since both L1 players and L2 players contribute less (see calibration), it would even reinforce the fact that contributions are lower over all periods with the minority bias being present. This is because the minority bias causes lower contributions in early periods and this leads to lower (myopic) contributions in later periods even if there was no minority bias present in these later periods.

Hypothesis 4.4 *Contributions (over-, myopic, and total) of L and H types in groups with, respectively, two H and two L types are lower than contributions of L and H types in groups with, respectively, two L and two H types over all periods.*

Hypothesis 4.4 isolates the effect of one's own minority bias from an anticipated minority bias as contributions are considered where only one's own minority bias can be present (L types in HH and H types in LL) or no bias at all (L types in LL and H types in HH).

Anticipated Minority Bias: Anticipating that others have a minority bias has an unambiguous effect on contributions. L1 players cannot anticipate others' minority bias as they only react to observed contributions from the previous period. So, the anticipated minority bias can only *indirectly* affect L1 players' contributions since observed behavior may be affected if there are L2 players. L2 players, on the other hand, can anticipate the minority bias of others. Other L1 players' minority bias decreases their contributions (since for them only myopic contributions are relevant) as well as the slope of their best response functions. Thus, L2 players decrease their myopic as well as over-contributions in response.

Proposition 4.8 *Suppose Assumptions 4.2 to 4.4 hold. Then, ceteris paribus, the anticipated minority bias weakly decreases L2 players' contributions and has no effect on L1 players' contributions in period t.*

The ceteris paribus assumption in Proposition 4.8 is again innocuous. Since the anticipated minority bias decreases L2 players contributions early, reacting to these decreased contributions in later periods means that L1 players also decrease their contributions simply because they are on a different contribution path and even though they cannot anticipate the minority bias of others. This reinforces the described effect in Proposition 4.8 since lower contributions of L2 players trigger lower (myopic) contributions of all other players.

As stated in Proposition 4.8, the anticipated minority bias only *weakly* decreases L2 players' contributions, because it may happen (depending on $(\tilde{\delta}_j, \tilde{\alpha}_j)$) that for another player for whom the minority bias is anticipated L2 player i believes that the other player's best response is in the corner solution for all best responses in the support of $F_{ij}(\cdot)$. As a result neither the believed sum of others' best responses nor the believed sum of the slopes of these best response functions is affected. Hence, neither myopic nor over-contributions are decreased. However, as soon as one of the best responses in the support of $F_{ij}(\cdot)$ is an interior solution, Proposition 4.8 holds strictly. Moreover, if L2 players beliefs about others' preferences are not any different across other players, Proposition 4.8 will also hold strictly, since then Assumption 4.1 guarantees that at least some of the best responses in the support of $F_{ij}(\cdot)$ are in the interior. Since this seems very likely and with a positive fraction of L2 players, we can formulate the next hypothesis strictly. It should hold for subjects who anticipate the minority bias of other players.

Hypothesis 4.5 *Contributions (over-, myopic, and total) of L and H types in groups with one L and one H type are lower than contributions of L and H types in groups with, respectively, two L and two H types in all periods.*

While Hypotheses 4.4 and 4.5 only capture the isolated respective effects of one's own and the anticipated minority bias, the next hypothesis is concerned with the overall effect of heterogeneous endowments. Over all periods both types of minority biases go in the same direction. We can therefore expect an effect of heterogeneous endowments on contribution behavior. With homogeneous endowments neither of the minority biases exists.

Hypothesis 4.6 *Heterogeneous groups (LLH and LHH) contribute less than homogeneous groups (LLL and HHH), so they are less efficient in providing the public good over all periods.*

Hypothesis 4.6 has interesting policy implications as it suggests that unequal income distributions are not only per se lowering aggregate welfare, but also that they cause less efficiency in increasing this welfare based on voluntary public goods provisions. Most likely, however, precise income distributions are not common knowledge in groups interacting in the provision of public goods. Such uncertainty is considered next.

4.4.4 Uncertain Endowments

So far, our analysis assumed certain endowments. We relax this assumption in this section and instead assume that players only have beliefs about others' endowments that do not necessarily coincide with their true endowments. Players still know their own endowments. Since we are interested in how systematic and stable deviations of these beliefs from reality affect behavior, we abstract from issues of updating and assume that believed endowments of others are fixed over time. In order for the absence of endowments belief updating to be a reasonable assumption, we restrict beliefs in a way such that the believed sum of others' contributions never exceeds the believed sum of others' endowments.[51] It is assumed that this holds for L1 and L2 players, and also for L2 players' beliefs about L1 players. Then, Propositions 4.1 and 4.4 and Corollary 4.2 still hold if endowments are uncertain, but not necessarily Proposition 4.5 as players are relative conditional cooperators.[52] In this section, we will derive an effect of uncertain endowments that arises if believed endowments are higher or lower than true endowments. We assume that players' second-order belief, namely their belief

[51] This is implicitly assumed in Assumption 4.4 if we exchange $\sum_{j\neq i} E_j$ with $\int \sum_{j\neq i} E_j \; dG_i(\sum_{j\neq i} E_j)$, where others' endowments $\sum_{j\neq i} E_j$ are drawn according to the distribution $G_i(\cdot)$.

[52] The imposed restriction on the believed sum of others' endowments is arguably strong when looking at contribution dynamics. Therefore, our focus in this section rather lies in a specific comparative statics effect that arises compared to certain endowments.

about others' beliefs about the endowments of others is the same as their own belief. This is a reasonable assumption when players have the same information and know that others have the same information as the one they have themselves.

Similar as with the effect of one's own minority bias, opposing effects arise when L2 player i thinks others are richer than they actually are. Such a deviating belief decreases myopic contributions since for a given contribution of others, their relative contribution which is aimed to be matched decreases. At the same time over-contributions may increase (and they definitely do so with only one period left) because the pain of over-contributing due to matching higher contributions in the next period may decrease. Thus, the effect on total contributions is ambiguous in a similar way as in Proposition 4.7 for L2 players, but is unambiguous for L1 players.

Proposition 4.9 *Suppose Assumptions 4.2 to 4.4 hold. Then, ceteris paribus, thinking others are richer (poorer) decreases (increases) L1 players' contributions in period t and L2 players' contributions in the last period T. It decreases (increases) L2 players' (over-) contributions in period $t \neq T$ if one's own minority bias decreases their (over-) contributions.*

By the calibration analysis of Proposition 4.7, we expect clearcut behavior under uncertainty based on Proposition 4.9. Thinking others are richer decreases myopic and should also decrease over-contributions for most periods and especially early in the game. We would expect an even stronger effect on (myopic) contributions without the ceteris paribus assumption. By contrast, thinking others are poorer increases myopic and over-contributions and can even lead to a contribution path that is not decreasing over time (as opposed to Proposition 4.5). Again, the effect on (myopic) contributions would be even stronger without the ceteris paribus assumption.

Before formulating the next empirical hypothesis, it is important to investigate whether there is a systematic bias in the beliefs about others' endowments. Do subjects in fact believe that others are poorer or richer than they actually are? This constitutes an empirical question which we can answer with our C-Experiment data. In the experiment, subjects knew that half of them had low and the other half had high endowments. So, they may have had the belief that one of the group members is highly and one lowly endowed. Then, subjects in LL believe to be in LH, so they believe to be in a richer group, and subjects in HH believe to be in LH, thus believe to be in a poorer group. In LLL and HHH, all subjects have wrong beliefs, and in LLH and LHH only one subject has wrong beliefs. Nevertheless, subjects in poor groups (LLL and LLH) believe that others are weakly richer than they are and subjects in rich groups (HHH and LHH) believe that others are weakly poorer than they are. The same effects would arise if subjects believed to be in LL, LH, and HH with equal probability or with any other symmetric probability distribution over these possibilities of others' endowments. In Section 4.5 we test whether this is indeed the case and find that subjects in poor groups believe to be in a richer group and subjects in rich groups believe to be in a poorer group than they actually are. Subjects' beliefs are symmetrically distributed around a group structure with one other being an L and one other being an H type. Interestingly, such beliefs are (approximately) rational if subjects supposed a random group assignment in the experiment.

Given this empirical observation, Proposition 4.9 implies that poor groups decrease their over-, myopic, and total contributions while rich groups should increase their over-, myopic, and total contributions due to deviating beliefs. We can test whether this is indeed the case since we know subjects' probabilistic beliefs about their group structure in the C-Experiment.

Hypothesis 4.7 *With uncertain rather than certain endowments, poor groups (LLL and LLH) decrease their contributions (over-, myopic, and total), while rich groups (LHH and HHH) increase their contributions (over-, myopic, and total) over all periods. The effects are stronger in homogeneous groups (LLL and HHH) since the deviation of beliefs is larger than in heterogeneous groups (LLH and LHH).*

An interesting implication of Hypothesis 4.7 is that under uncertainty, the inequality of the overall income distribution increases compared to certain endowments. Hypothesis 4.7 is derived for group behavior. Under uncertainty and given our observation that subjects' beliefs are symmetrically distributed around the LH group structure, one's own and the anticipated minority bias act on all group structures in the same way since there is neither a reason nor an indication for subjects having not the same belief distribution under uncertainty. In our analysis of the deviating beliefs in Section 4.5 one can actually see that there is not much updating going on in the belief distribution for the different group structures, so that this pattern does no change much over time. As an example, suppose that subjects believe to be paired with one L and one H player. Then, L1 players can neither have their own nor an anticipated minority bias affecting their behavior. L2 players believe that others have the same belief than the one they have themselves so that they also do not execute their own or an anticipated minority bias. More generally, symmetric beliefs guarantee that any type of minority bias does not cause a distortion for the effect of uncertainty as described in Hypothesis 4.7 because such a bias would act on all groups in the same way. So, the uncertainty effect is solely driven by the deviating beliefs as postulated in Proposition 4.9. Moreover, the systematic difference between homogeneous and heterogeneous groups under certainty should break down under uncertainty.

Hypothesis 4.8 *Under uncertainty, heterogeneous groups (LLH and LHH) do not contribute less than homogeneous groups (LLL and HHH), so they are equally efficient in providing the public good over all periods.*

4.4.5 Discussion

Before testing the derived hypotheses of this section, we will discuss two related approaches in more detail, one theoretical by Ambrus and Pathak (2011) and one empirical by Fischbacher and Gächter (2010), that share some of the presented results.

Ambrus and Pathak (2011) propose an *equilibrium* model where players are either completely selfish or completely reciprocal. Selfish players in their model solely maximize their material payoff Π_i. Reciprocal players solely maximize $v(\cdot)$, however with a quite different target function $g(\cdot)$. Our *non-equilibrium* model builds upon the model of Ambrus and Pathak (2011) in that we adopt their functional forms, namely $m(y) = y$ and $v(\cdot) = -(x_i - g(\cdot))^2$. However, besides obtaining non-equilibrium rather than equilibrium predictions, our model differs from the model of Ambrus and Pathak (2011) in several respects. First of all, they nicely focus on how the interaction of selfish and reciprocal types explains contribution dynamics in a repeated public goods game. Our focus is rather on how the preference of imperfect conditional cooperation affects behavior. We therefore allow types to be in between completely selfish and conditionally cooperative and thereby account for the self-serving bias of conditional cooperation. Second, they assume complete information about the types or preferences of other players. This assumption is not necessary in our setup. Third, they need at least one selfish type to generate contributions in the game while there are almost no restrictions on the type distribution in our model.[53] Fourth, the target contribution in their model is a function of past and current contributions while we impose more structure on $g(\cdot)$ in that we rather model conditional cooperation where the target contribution depends only on the belief what others will contribute in the current period.[54] Fifth, for their results it is crucial that players stay together for the entire game since contributions are driven by reciprocity over the history of the game. By contrast, our results would also obtain if partners change each period as long as they would get contribution information about

[53]In our model, positive contributions cannot occur if all players are either purely selfish L1 players or purely selfish L2 players with sufficiently selfish beliefs about other players.

[54]While it may well be reasonable that players retaliate or reward past behavior, we abstract from such issues and solely focus on conditional cooperation for which it is only important in a given period what the belief is about others' contributions in that period.

the new partners for the last period. So, while in their model strategic concerns are driven by the known reciprocity of others, in our model strategic concerns rather arise as a signal for future contribution behavior.

Despite these differences, some results are similar between Ambrus and Pathak (2011) and our model. They show a weak decay of contributions and the positive effect of increasing the efficiency of the public good. Further, they show that a larger time horizon results in higher contributions (see Isaac, Walker, and Williams, 1994) and they explain the 'restart effect' first stated by Andreoni (1988). If subjects treat the restarted game as a new game with new initial beliefs in the first period, our model would also be consistent with the 'restart effect'. Besides that our model offers various additional results, there is also one result where our model crucially differs from the model of Ambrus and Pathak (2011). This is the one-shot prediction of the game. In their setup, players contribute nothing while they contribute according to their myopic contributions in our model. That players also contribute in one-shot public goods games is a common finding in the literature (see e.g. Marwell and Ames, 1981). "Cooperation never falls to zero, even in one-trial games or in the last period of multi-trial games when it can never be selfishly rational to cooperate." (Dawes and Thaler, 1988, p191) Nevertheless, our level-k model could be seen as a short-run complement to the equilibrium model of Ambrus and Pathak (2011), also because they need that subjects learn the exact type of others which may require quite some time.

Our theoretical approach may also be seen as a complement to the empirical investigation on the dynamics of contribution behavior of Fischbacher and Gächter (2010). Not only the design of our experiment builds upon their two-part design, but also their empirical results are predicted by our theoretical model. They find that "[...] subjects behave according to a contribution pattern that is in between their elicited contribution schedule and perfect conditional cooperation. Since most people's elicited contribution preferences lie below the diagonal, that is, below the schedule of perfect conditional cooperators, this intermediate contribution pattern lies above the predicted cooperation. This means that subjects are more conditionally cooperative in the C-Experiment than predicted from their decisions in the P-Experiment." (Fischbacher and Gächter, 2010, p550) So, Fischbacher and Gächter (2010) also find over-contribution and that this over-contribution is bounded above such that contributions never exceed the amount of perfect conditional cooperation (our Assumption 4.4).[55] Note that according to our Proposition 4.1, over-contribution would also be optimal in the setup of Fischbacher

[55] Ambrus and Pathak (2011) assume the similar condition of 'no over-reciprocation' (Assumption A5 in their model).

and Gächter (2010) since $\frac{E_i}{\sum_{j\neq i} E_j} = \frac{1}{3}$ and $\frac{A}{N} = 0.4$ in their experiment. Moreover, they also find that over-contributions are somehow decreasing because 'beliefs' are found to be more important in the first half of the experiment than in the second half, and in later periods 'predicted contributions' become more important than 'beliefs'. In their setup with homogeneous endowments, perfect relative conditional cooperation directly translates into contributions being equal to 'beliefs' about others' contributions, and what they call 'predicted contributions' is simply myopic contributions in our model. Interestingly, with the help of simulation models, Fischbacher and Gächter (2010) further show that beliefs decline because contributions decline and not vice versa, that "[...] preference heterogeneity is surprisingly unimportant in explaining the decay of cooperation [...]", and that a belief formation process in which the belief in period t is equal to others' contributions in period $t-1$ explains the data well.[56] (Fischbacher and Gächter, 2010, p553) All these findings are captured in our model.

However, there is one crucial difference between our setup and theirs. They use a random stranger design while our experiment as well as our model relies on a partner setup. Botelho et al. (2009) compare a random stranger to a perfect stranger design. The former is usually intended to be a representation of the latter since perfect stranger designs produce too little data. They find a significant difference between these two designs with more free-riding under perfect stranger. "This is evidence that a Random Strangers environment does not elicit the same behavior as a comparable Perfect Strangers environment, and that the direction of the change in behavior is consistent with the Perfect Strangers environment being more conducive to subjects viewing the stage game as one-shot." (Botelho et al., 2009, p261) They further review the literature on comparing random stranger to partner designs and stress that some of these studies find no significant difference[57], while other studies find contributions in partner treatments are higher than in random stranger treatments (see also Gächter, 2007). Nevertheless, contribution patterns are similar and differ mainly with respect to the contribution level.

Sonnemans, Schram, and Offerman (1999) use a design where subjects in a group gradually change and can never meet again afterwards, so their design changes from partner to perfect stranger. The authors find that contributions (which is a binary

[56]Note that the preferred model of belief formation in Fischbacher and Gächter (2010) is, however, one in which beliefs in period t are a weighted average of others' contributions in period $t-1$ and own beliefs in period $t-1$.

[57]More specifically, Botelho et al. (2009) classify the studies of Andreoni (1988, with new appropriate econometric specification), Weimann (1994), Palfrey and Prisbrey (1996), Burlando and Hey (1997, overall data), and Brandts and Schram (2001) as having found no difference, while Croson (1996) and Keser and van Winden (2000) find that partners contribute more. See Andreoni and Croson (2008) for another review.

decision) decrease when a change is approaching, that subjects leaving contribute less than subjects staying, that subjects temporarily decrease their contributions in the period they leave, that contributions are not influenced by the history but by the future suggesting strategic forward looking behavior, and that beliefs about contributions are highly correlated with contributions in the previous period. All these results are fully consistent with our model.

The probabilities of being rematched in a random stranger design are quite low (and thus probabilities of not being rematched are high), nevertheless subjects seem to rather act as if being rematched with some positive probability than as if not being rematched. Interestingly, the usual distortion of probabilities where low probabilities are over-weighted and high probabilities are under-weighted further acts in this direction. A probabilistic partner (or random stranger) setup would also induce over-contributions in our model, although quantitatively less than our real partner setup. This could well explain the level effect of studies which found a difference. By referring to the study of Botelho et al. (2009), Fischbacher and Gächter (2010, p550) conclude that "one potential explanation for why beliefs matter in addition to "predicted contribution" is subjects' willingness to invest in cooperation in order to induce high beliefs and contributions in the population, even in our random matching design [...]." This idea is what we tried to capture in the model.

4.5 Results C-Experiment

We organize the empirical results of the repeated public goods game in the following way. Before testing our derived hypotheses, we first present the analysis of our stated assumptions from the previous section and investigate how beliefs deviate under uncertain endowments.

4.5.1 Assumptions

All our assumptions in Section 4.4 were formulated as plans. However, plans are not observable but only actions. We can therefore only back our assumptions with observed behavior. Since Assumption 4.3 is satisfied by our experimental design, we only need to consider Assumptions 4.1, 4.2, and 4.4.

Assumption 4.1: We first turn to Assumption 4.1, for which we look at *actual* contributions (X^a_{tr}) as observed from the C-Experiment, where $tr = \{c, u\}$ stands for the treatment, being either the uncertainty (u) or the certainty treatment (c). In all of the periods from period 1 to 9, the mean and median actual contribution is positive, both under certainty and uncertainty. Most subjects have $X^a_{tr} > 0$ under certainty and uncertainty in all nine periods. Over all nine periods, an average of 66% under certainty and 62% under uncertainty have $X^a_{tr} > 0$ (with a maximum/minimum of 80%/52% and 72%/51% under certainty and uncertainty, respectively). The average value of actual (or total) contributions over all subjects and all nine periods is $X^a_c = 2.26$ EMU and $X^a_u = 2.09$ EMU.

Observation 4.3 *Median contributions are positive in periods 1-9.*

While Assumption 4.1 requires positive contributions in all periods, Observation 4.3 considers only the first nine periods. This is because the last period is considered in the following analysis of Assumption 4.2.

Assumption 4.2: For Assumption 4.2 we look at myopic contributions in the last period.[58] In the C-Experiment *predicted* contributions (X^p_{tr}) represent myopic contributions. X^p_{tr} are calculated from the individual contribution preferences based on the P-Experiment, the beliefs about others' contributions, and the actual probabilistic beliefs about the group structure (only in the uncertainty treatment) as stated in every period in the C-Experiment. Note that predicted contributions are the point predictions that do not impose any assumptions like conditioning on averages or behaving relative conditionally cooperative. The calculation is based on the exact stated preference of each individual given others' individual contributions (and not the average of their contributions) in each specific group structure. Under certainty and uncertainty, the mean and median predicted contribution is positive in the last period. Most subjects, namely 58% under certainty and 67% under uncertainty, have $X^p_{tr} > 0$. The average value of predicted (or myopic) contributions over all subjects under certainty is $X^p_c = 1.77$ EMU and it is $X^p_u = 1.81$ EMU under uncertainty.

Observation 4.4 *Median myopic contributions are positive in period 10.*

Assumption 4.4: Since we know subjects' beliefs about others' contributions in every period, we can calculate the amount of perfect relative conditional cooperation for every subject in every period and compare it to their actual contribution X^a_{tr}. Under certainty, the calculation is such that the sum of these beliefs is divided by the sum of others' endowments, and multiplied by one's own endowment. Under uncertainty, the amount of perfect relative conditional cooperation is the sum of believed contributions of others, divided by the linear combination of the believed sum of others' endowments, and multiplied by one's own endowment. The question we approach is

[58]In the model positive myopic contributions in the last period are necessary and sufficient for positive contributions in the last period (see Observation 4.3).

then whether subjects ever contribute more than what would be optimal under perfect relative conditional cooperation. Fischbacher and Gächter (2010) and Neugebauer et al. (2009) provide evidence under certain and homogeneous endowments that contributions lie below beliefs about others' contributions. In their setup this finding is equal to contributions being below the amount of perfect relative conditional cooperation.

Under certainty, we find that the z-values of a Wilcoxon signed-rank test are negative in all periods indicating that, indeed, contributions never exceed the amount of perfect conditional cooperation. Further, over periods 1-9 contributions are significantly below this amount ($p = 0.0333$; $N = 28$). Also in the last period 10, contributions are significantly below the amount of perfect conditional cooperation ($p = 0.0003$; $N = 28$) which is, however, not surprising since over-contribution should be absent in the last period.[59]

Under uncertainty, we again observe that the z-values of a Wilcoxon signed-rank test are negative in all periods. Although contributions over periods 1-9 are not significantly below the amount of perfect relative conditional cooperation ($p = 0.2034$; $N = 32$), they are significantly below this amount in period 10 ($p = 0.0023$; $N = 32$) and also over all periods ($p = 0.0963$; $N = 32$).

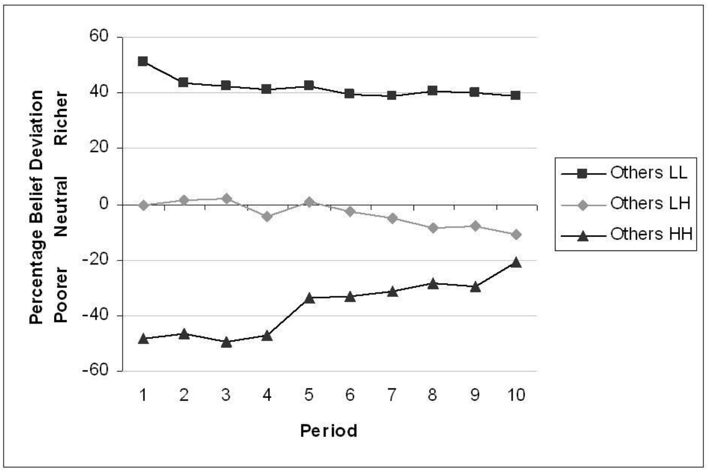
Observation 4.5 *Contributions are (weakly) below the amount that represents perfect relative conditional cooperation in periods 1-10.*

Endowment Beliefs: Under uncertainty it is important what subjects believe about their group members' endowments, since any deviation of this belief from reality affects contribution behavior compared to certainty. Figure 4.2 shows how correct subjects' beliefs are in the C-Experiment. The variable 'deviation of beliefs' measures how realistic subjects' beliefs about the endowments of their group members are. For instance, consider a subject who has group members that are both L types. If this subject believed to be in HH with probability 1, her belief would deviate by 100% from reality. If she believed to be in LH with probability 1, her belief would deviate by 50%, and if she believed to be in LL with probability 1, her belief would not deviate (i.e. deviate by 0%).[60] Since subjects know that half of the subjects in the experiment are L types and the other half are H types and that there are three possible group

[59] Of course, over all periods contributions are also significantly below the amount of perfect conditional cooperation ($p = 0.0074$; $N = 28$).

[60] We restricted beliefs to be certain in this example for illustration only. Since subjects in the C-Experiment stated their *probabilistic* beliefs about their group members' endowments, the variable 'deviation of beliefs' is calculated as a linear combination of the different group structures that give 100%, 50%, or 0% deviation. For instance, if this subject believed to be in LL with 33% probability, in LH with 33% probability, and in HH with 34% probability, her belief would also deviate by 50%.

Figure 4.2: Deviation of Beliefs about Group Structure from Reality



structures each subject could possibly be in, they mostly start out with a symmetric belief around LH. Hence, subjects who actually are in LH have realistic beliefs since positive and negative deviations cancel out. Subjects who actually are in LL or HH start out with beliefs that deviate by about 50% from reality. Over the ten periods, there is not much updating. Subjects in LH more or less keep their realistic beliefs, and subjects in LL and HH mostly keep their deviating beliefs. However, subjects in HH update better than subjects in LL, since they sometimes get informative signals of contributions exceeding 4 EMU.

Figure 4.2 also displays whether the deviation of beliefs is positive or negative, i.e. whether subjects incorrectly believe to be in a poorer (negative) or richer (positive) group than they actually are. Subjects in LL think they are in a richer group than they actually are and subjects in HH think they are in a poorer group than they actually are. The effect of deviating beliefs is strongest in LLL and HHH groups since individual effects of all group members go in the same direction. Moreover, the fact that subjects with other group members being HH exhibit more realistic beliefs than subjects who have LL group members suggests that the impact of deviating beliefs is larger for poor groups. Note that although updating is rather small, this leads to a slight asymmetry of the effect of uncertainty as predicted by Hypothesis 4.7, namely that poor groups lose more than rich groups gain through the uncertainty.

In order to see whether the observed patterns in Figure 4.2 are statistically significant we can perform t-tests. Subjects in LL have beliefs that significantly deviate from reality in all periods ($0.0001 \leq p \leq 0.0003$; $N = 16$). The same holds for subjects in HH ($0.0001 \leq p \leq 0.0021$; $N = 16$). In contrast, subjects in LH have beliefs that do not significantly deviate from reality in almost all periods (($0.1546 \leq p \leq 0.8214$ in periods

1-8; $p = 0.0740$ in period 9; $p = 0.0153$ in period 10; $N = 16$). Belief deviations over all periods are significantly different from zero when others are LL or HH ($p = 0.0001$; $N = 16$), but not when others are LH ($p = 0.2091$; $N = 16$).

Furthermore, belief distributions are symmetric around LH in all periods. Since each subject stated her probabilistic belief about others being LL, LH, and HH in each period, we can perform Wilcoxon signed-rank tests in order to see whether the stated probability to be in LL is any different from the stated probability to be in HH. There is no period where this difference is significant ($0.2420 \leq p \leq 0.9386$; $N = 32$) and the z-values are positive in five periods and negative in five periods. Over all periods, subjects believe to be in LL with 30%, in HH with 31%, and in LH with 39% probability. Again, over all periods no significant difference between LL and HH is observed when performing Wilcoxon signed-rank tests ($p = 0.8895$; $N = 32$).

Observation 4.6 *Under uncertainty, subjects with poor (LL) group members believe to be in a richer group and subjects with rich (HH) group members believe to be in a poorer group than they actually are in periods 1-10. Subjects' beliefs are symmetrically distributed around a group structure with one other being an L and one other being an H type in periods 1-10. Thus, subjects with mixed (LH) group members do not have deviating beliefs.*

While Observations 4.3, 4.4, and 4.5 are needed to justify the respective Assumptions 4.1, 4.2, and 4.4, which are used for some of the theoretical results, Observation 4.6 is solely needed for deriving empirical hypotheses and not as a justification of theoretical results.

4.5.2 Hypotheses

Having shown that the assumptions we presumed in the previous section to derive theoretical results and hypotheses thereof are empirically justified, we can now turn to the empirical analysis of these hypotheses. As Hypotheses 4.1 to 4.6 were derived under the assumption of certain endowments, we will present the corresponding results based on the data from the certainty treatment and only use the uncertainty treatment for comparison. By contrast, Hypotheses 4.7 and 4.8 are especially concerned with the effects of uncertain endowments, so respective results are based on the data of both treatments.

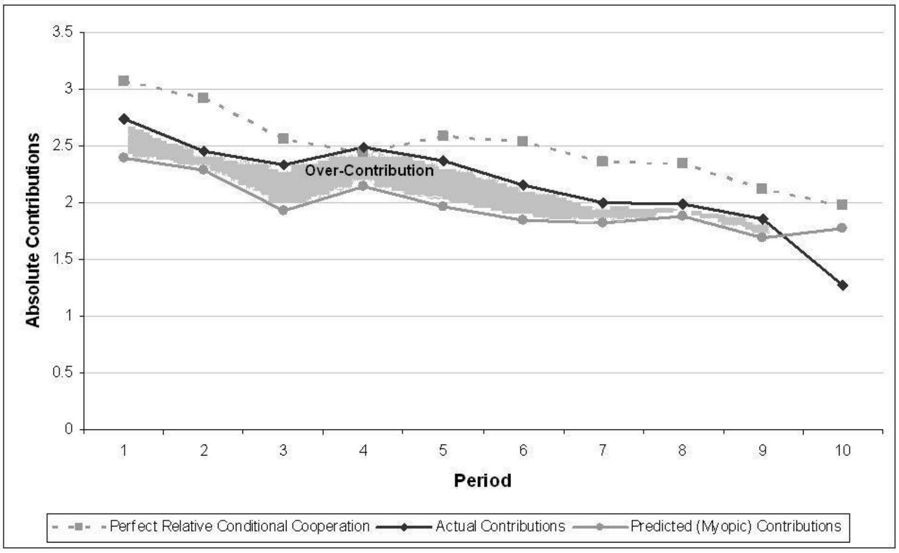
Strategic Over-Contribution

In the C-Experiment strategic over-contribution is calculated as the difference between *predicted* contributions (X_{tr}^p) and *actual* contributions (X_{tr}^a). Strategic over-contribution is then defined as $SOC_{tr} = X_{tr}^a - X_{tr}^p$. Under certainty, we observe that the z-values of a Wilcoxon signed-rank test are positive in all but the last period, so under-contribution never occurs in period $t \neq T$. Over the first nine periods SOC_c is significantly positive ($p = 0.0157$; $N = 28$) but not in period 10 as predicted by Hypothesis 4.1.[61]

Result 4.7 [Hypothesis 4.1] *Strategic over-contribution is positive in periods 1-9 but not in period 10.*

Result 4.7 is further visualized in Figure 4.3. The actual contribution schedule in periods 1-9 lies between predicted (or myopic) contributions and the contributions that would be optimal under perfect relative conditional cooperation (see Observation 4.5).[62]

Figure 4.3: Contributions under Certainty: Myopic, Total, and Perfectly Cooperative



[61]Note that there is even significant under-contribution in period 10 ($p = 0.0002$; $N = 28$) showing a specific end-game effect not captured by our model which would have predicted zero over-contribution in the last period.

[62]Under uncertainty and also consistent with our model, the z-value of a Wilcoxon signed-rank test is positive over periods 1-9, though not significantly ($p = 0.6298$; $N = 32$). As under certainty, there is no over-contribution in period 10, but significant under-contribution ($p = 0.0001$; $N = 32$).

Decreasing Contributions

Figure 4.3 further shows that actual (or total) contributions are decreasing over time. The Spearman's rank order correlation coefficient is $\rho = -0.3075$ ($p = 0.0000$; $N = 280$). Over-contribution, represented by the difference of actual and predicted contributions, is also decreasing in Figure 4.3 and its Spearman rank order correlation coefficient is $\rho = -0.1575$ ($p = 0.0083$; $N = 280$).[63] Moreover, it can be seen that predicted (or myopic) contributions are decreasing as well, with a Spearman rank order correlation coefficient of $\rho = -0.1256$ ($p = 0.0030$; $N = 280$).[64]

Result 4.8 [Hypotheses 4.2 and 4.3] *Contributions (over-, myopic, and total) are decreasing over time.*

While Results 4.7 and 4.8 could have obtained under absolute conditional cooperation as well (as pointed out by Remark 4.6), the remaining results only obtain under relative conditional cooperation since they are concerned with the effects of making endowments uncertain and heterogeneous. Before considering the general efficiency of the public goods provision within these two treatment variations, we first focus on the specific hypotheses that test isolated theoretical results of the underlying effects at work.

Minority Bias Effects

With heterogeneous endowments, the minority bias can either directly affect subjects or be anticipated to affect other subjects and thereby causing an indirect effect. Under certainty, only subjects in specific groups can experience a minority bias. We start with Hypothesis 4.4 which focused on the effect of one's own minority bias. We hypothesized that contributions over all periods should be lower with one's own minority bias at work. Indeed, subjects for whom only one's own minority can affect cooperation have lower contributions in eight out of ten periods than players with no possible bias. Using a Mann Whitney U test, we find that over all periods subjects with an own minority bias (L types in HH and H types in LL) contribute significantly less

[63]Note that the reported correlation coefficients use individual contributions in each period and not the average contributions as displayed in Figure 4.3. Using average contributions instead yields much higher correlations.

[64]While Result 4.8 is derived for the certainty treatment, we obtain similar results (which are again consistent with our model) for the uncertainty treatment, where $\rho = -0.2228$ ($p = 0.0001$; $N = 320$) for total contributions, $\rho = -0.1517$ ($p = 0.0065$; $N = 320$) for over-contributions, and $\rho = -0.0712$ ($p = 0.0273$; $N = 320$) for myopic contributions.

in relative terms than subjects without any bias (L types in LL and H types in HH) ($p = 0.0605$; $N = 28$).[65]

Result 4.9 [Hypothesis 4.4] *Relative Contributions of L and H types in groups with, respectively, two H and two L types are lower than relative contributions of L and H types in groups with, respectively, two L and two H types over periods 1-10.*

We now turn to the anticipated minority bias. Hypothesis 4.5 suggested that players who can only anticipate the minority bias of other players should contribute less than players without any bias. Indeed, this is what we find. In nine out of ten periods anticipating the minority bias of others results in lower contributions. A Mann Whitney U test shows that over all periods subjects who anticipate the minority bias of other subjects (L types in LH and H types in LH) have significant lower relative contributions than subjects without any bias (L types in LL and H types in HH) ($p = 0.0649$; $N = 28$).[66]

Result 4.10 [Hypothesis 4.5] *Relative Contributions of L and H types in groups with one L and one H type H are lower than relative contributions of L and H types in groups with, respectively, two L and two H types over periods 1-10.*

Effects of Deviating Beliefs

Next, we consider the model's uncertainty prediction as claimed in Hypothesis 4.7. In doing so, we first consider myopic contributions, then over-contributions, and finally total contributions.

Myopic Contributions: In discussing strategic over-contribution we used the variable predicted contributions (X^p_{tr}) which corresponds to myopic contributions. Under uncertainty, this variable is calculated as under certainty except that now the actual probabilistic beliefs about the group structure are taken into account. These beliefs were shown to be deviating from the true group structure in a systematic fashion (see Observation 4.6). Since such predicted contributions measure the myopic contributions under these deviating beliefs, we want to know how the myopic contributions would look like had subjects correct beliefs. So, we have to compare the *predicted* contributions that are based on actual and often deviating beliefs to predicted contributions that would

[65] We do, however, not observe significant differences in relative contributions between these two groups of subjects with regard to over-contributions and myopic contributions.

[66] Here, we do not observe a significant difference in relative contributions between these two groups of subjects with regard to myopic contributions, but we do with respect to over-contributions. Subjects anticipating the minority bias over-contribute significantly less over all periods than subjects without a bias ($p = 0.0294$; $N = 28$).

have occurred if subjects had held the correct beliefs about their group structure. We will call the latter type of predicted contributions *hypothetical* contributions (X^h_{tr}).[67]

The difference between predicted (X^p_u) and hypothetical (X^h_u) contributions gives the net amount that subjects myopically contributed more or less under uncertainty because of their deviating beliefs. We refer to this difference as $DOB_u = X^p_u - X^h_u$. For subjects in LL we expect $DOB_u < 0$ since they would give more if they knew others' true endowments. For subjects in HH it should hold that $DOB_u > 0$, since they would give less if they knew others' true endowments. We expect subjects in LH to have $DOB_u = 0$ since their belief is (more or less) correct. Under certainty, beliefs cannot be wrong, $X^p_c = X^h_c$, and therefore $DOB_c = 0$. So, the difference in myopic contributions between uncertainty and certainty, which is caused by deviating beliefs, is captured by $\Delta DOB = DOB_u - DOB_c = DOB_u$.

Over all periods, we find that poor groups myopically contribute less under uncertainty and rich groups myopically contribute more because of their deviating beliefs. The ΔDOB is -0.24 EMU for LLL, -0.03 EMU for LLH, 0.11 EMU for LHH, and 0.27 EMU for HHH. When performing Mann Whitney U tests, we find that the differences for homogeneous groups are significant, but they are not for heterogeneous groups ($p = 0.0240$ for LLL, $p = 0.6432$ for LLH, $p = 0.2196$ for LHH, and $p = 0.0176$ for HHH; $N = 8$). Pooling poor and rich groups, ΔDOB is significantly different from zero for LLL/LLH ($p = 0.0507$; $N = 16$) but not for LHH/HHH ($p = 0.6412$; $N = 16$). The absolute value of ΔDOB is larger in homogeneous groups (LLL/HHH), namely $|\Delta DOB| = 0.51$ EMU, than heterogeneous groups (LLH/LHH) where it is $|\Delta DOB| = 0.14$ EMU. These findings of the effect of deviating beliefs on myopic contributions are in line with Hypothesis 4.7. In LLL, all subjects have $DOB_u < 0$. Likewise, all subjects in HHH have $DOB_u > 0$. So, all individual effects go in the same direction in homogeneous groups. By contrast, in heterogeneous groups individual effects do not go in the same direction. To see this, consider first an LLH group. L types in this group are in LH leading to $DOB_u = 0$, and the H type in this group is in LL leading to $DOB_u < 0$. So the total effect for the LLH group should be slightly negative. Second, consider the LHH group. Here, the L type is in HH leading to $DOB_u > 0$, and

[67] Hypothetical contributions are calculated from the individual contribution preferences (elicited in the P-Experiment), the beliefs about others' contributions in each period (elicited in the C-Experiment), and under the assumption of correct beliefs about the group structure. Both predicted and hypothetical contributions are the point predictions that do not impose any specific assumption on cooperation preferences (like conditioning on averages or relative conditional cooperation). Again, the calculation is based on the exact stated preference of each individual given others' individual contributions in each specific group structure.

both H types are in LH leading to $DOB_u = 0$. Thus, the total effect in LHH should be slightly positive.

Over-Contributions: The same patterns should be observed when considering strategic over-contribution. We denote the difference in over-contributions between the uncertainty and certainty treatment as $\Delta SOC = SOC_u - SOC_c$. ΔSOC is -0.56 EMU ($p = 0.0178$; $N = 15$) for LLL, -0.11 EMU ($p = 0.0526$; $N = 15$) for LLH, 0.04 EMU ($p = 0.2640$; $N = 15$) for LHH, and 0.26 EMU ($p = 0.0874$; $N = 15$) for HHH. Note that all differences are significant using a Mann Whitney U test, except for LHH. Pooling poor and rich groups, ΔSOC is significantly different from zero for LLL/LLH ($p = 0.0387$; $N = 30$) but not for LHH/HHH ($p = 0.8680$; $N = 30$). The absolute value of ΔSOC is larger in homogeneous groups (LLL/HHH), namely $|\Delta SOC| = 0.82$ EMU, than in heterogeneous groups (LLH/LHH) where it is $|\Delta DOB| = 0.15$ EMU. Therefore, the results on over-contributions seem also to be in line with Hypothesis 4.7.

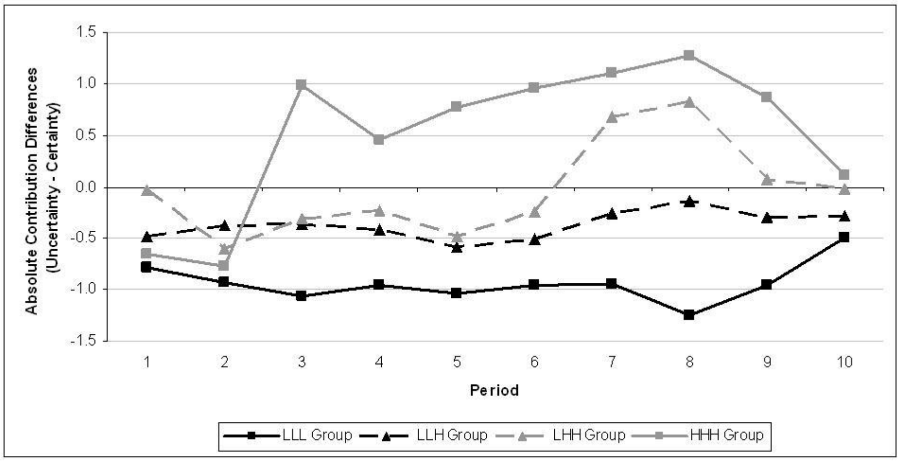
Strategic over-contribution captures the amount subjects contribute more in a repeated set-up given their one-shot preferences and actual beliefs. The previous analysis on ΔDOB showed that poor (rich) groups would myopically contribute more (less) under uncertainty if they had correct beliefs. Thus, the SOC_u of poor groups would be higher and therefore ΔSOC would be less negative if subjects had correct beliefs. So, ΔSOC includes the myopic effect caused by deviating beliefs, which is ΔDOB. For rich groups we showed that they would myopically contribute less if they had correct beliefs. Hence, the SOC_u of rich groups would be lower and the ΔSOC therefore less positive. Again, ΔSOC includes the myopic effect caused by deviating beliefs.

In order to isolate the pure effect of deviating beliefs on strategic over-contribution, we can use the variable net strategic over-contribution, denoted $NSOC_{tr}$. Net strategic over-contribution ($NSOC_{tr}$) is the amount subjects would over-contribute if they had correct beliefs, namely $NSOC_{tr} = X^{a'}_{tr} - X^{p}_{tr}$, where $X^{a'}_{tr} = X^{a}_{tr} + (X^{h}_{tr} - X^{p}_{tr})$. The term in brackets is the amount subjects would myopically contribute more (poor groups) or less (rich groups) if they had correct beliefs.[68] Under certainty, $NSOC_c = SOC_c$ since $DOB_c = 0$. We denote the difference in net strategic over-contribution between the uncertainty and certainty treatment as $\Delta NSOC = NSOC_u - NSOC_c$. Using Mann Whitney U tests, we find that $\Delta NSOC$ is significantly negative for LLL, but we do not observe a significant difference for the other groups.[69] Thus, when completely disentangling the effect of deviating beliefs on over-contributions from the effect on

[68]Note that $SOC_{tr} = NSOC_{tr} + DOB_{tr} \Leftrightarrow DOB_{tr} = SOC_{tr} - NSOC_{tr} = [X^{a}_{tr} - X^{p}_{tr}] - [X^{a}_{tr} + (X^{h}_{tr} - X^{p}_{tr}) - X^{p}_{tr}] = X^{p}_{tr} - X^{h}_{tr}$. Under certainty, $X^{p}_{c} - X^{h}_{c} \Leftrightarrow X^{a}_{c} = X^{a'}_{c}$ and hence $SOC_c = NSOC_c$.

[69]The $\Delta NSOC$ is -0.31 EMU ($p = 0.0732$; $N = 15$) for LLL, -0.07 EMU ($p = 0.3092$; $N = 15$) for LLH, -0.08 EMU ($p = 0.4514$; $N = 15$) for LHH, and -0.01 EMU ($p = 0.7894$; $N = 15$) for HHH.

Figure 4.4: Group-Specific Contribution Differences between Certainty and Uncertainty

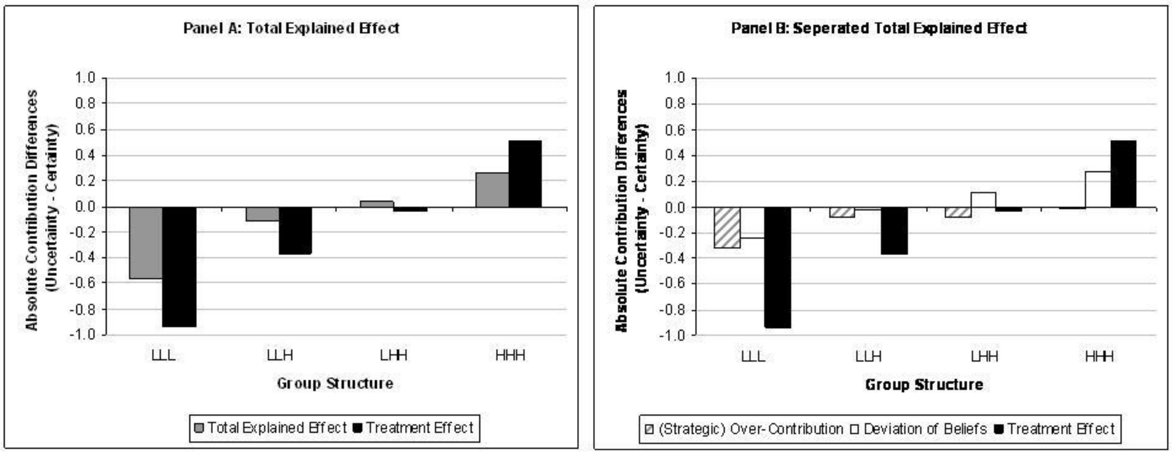


myopic contributions, we only observe a significant effect in line with Hypothesis 4.7 for poor homogeneous groups. For the other groups, the effect is not significant.

Total Contributions: Last and most importantly, we consider actual contributions and turn to the question whether rich and poor groups totally contribute more or less under uncertainty than under certainty. Figure 4.4 shows the treatment differences (uncertainty minus certainty) of group-specific absolute contributions over the ten periods. Roughly speaking, rich groups gain from the uncertainty and contribute more, whereas poor groups lose from making endowments uncertain and contribute less. The average treatment difference over the ten periods is -0.94 EMU for LLL, -0.37 EMU for LLH, -0.03 EMU for LHH, and 0.51 EMU for HHH. With a Mann Whitney U test, these treatment differences deviate significantly from zero only for homogeneous groups (p = 0.0000 for LLL, p = 0.1536 for LLH, p = 0.1204 for LHH, and p = 0.0622 for HHH; N = 15). When pooling rich and poor groups, we find that poor groups contribute significantly less under uncertainty while rich groups' larger contributions under uncertainty are not significant (p = 0.0021 for LLL/LLH, p = 0.4515 for LHH/HHH; N = 30). Treatment differences for poor groups are significantly larger than for rich groups (p = 0.0047 for LLL/LLH vs. LHH/HHH; N = 60).[70] This is due to the fact that there is slight updating of endowment beliefs in rich groups since here subjects sometimes get informative signal (see Observation 4.6). Hence, over all periods the beliefs of subjects in rich groups do not deviate as much as the beliefs of subjects in

[70]We can further test whether these treatment differences differ between the various group structures. Here we find that treatment differences are significantly different between all group structures except between LLH vs. LHH and LHH vs. HHH ($p = 0.0782$ for LLL vs. LLH, $p = 0.0176$ for LLL vs. LHH, $p = 0.0000$ for LLL vs. HHH, $p = 0.3378$ for LLH vs. LHH, $p = 0.0646$ for LLH vs. HHH, and $p = 0.1740$ for LHH vs. HHH; $N = 30$).

Figure 4.5: Explaining Treatment Effects



poor groups. The absolute value of the treatment differences with respect to total contributions is larger in homogeneous groups (LLL/HHH), namely 1.45 EMU, than in heterogeneous groups (LLH/LHH), where it is 0.40 EMU.

Since the analysis above shows that contribution behavior with respect to over-, myopic, and total contributions is mainly in line with Hypothesis 4.7, we can formulate our result on the effects of deviating beliefs.

Result 4.11 [Hypothesis 4.7] *With uncertain rather than certain endowments, poor groups decrease their contributions, while rich groups increase their contributions over periods 1-10. The effects are stronger in homogeneous than in heterogeneous groups.*

The next question we will approach is whether our explanation, namely that subjects have deviating beliefs, also *quantitatively* captures the treatment differences in actual contributions reasonably well. We outlined that ΔSOC captures two effects, since $\Delta SOC = \Delta NSOC + \Delta DOB$. Deviating beliefs have a pure effect on over-contributions, i.e. $\Delta NSOC$, as well as a pure effect on myopic contributions, i.e. ΔDOB. Since our objective is to explain the results from the C-Experiment, we now have to compare ΔSOC to the treatment effects. The gray bars in Figure 4.5A illustrate ΔSOC graphically for all group structures. The black bars in Figures 4.5A and 4.5B depict our treatment differences from the C-Experiment, namely $X_u^a - X_c^a$. Roughly speaking, they show that poor groups contribute less under uncertainty and rich groups contribute more than under certainty. Figure 4.5A shows that the ΔSOC captures the observed treatment effects reasonably well. In Figure 4.5B, ΔSOC is split graphically into its components, $\Delta NSOC$ and ΔDOB. The shaded bars capture the amount explained by net (strategic) over-contribution and the white bars capture the myopic effect of deviating beliefs.

Having identified the specific driving forces at work under heterogeneous and uncertain endowments, we now turn to the total effects of these driving forces on general contribution behavior.

Efficiency under Heterogeneous and Uncertain Endowments

Figure 4.6 shows group-specific mean contributions over the ten periods relative to the total group endowment. This determines how efficient a group is in providing the public good in the two treatments. Under certainty and turning to Hypothesis 4.6, there is a U-shaped relationship between the number of H types in a group and the efficiency of the group. LLL groups contribute on average 40% of their endowment, LLH groups 32%, LHH groups 31%, and HHH groups contribute 43% of their endowment. Thus, heterogeneous groups are less efficient than homogeneous groups. When testing these differences with Mann Whitney U tests, we find that they are significant between homogeneous and heterogeneous groups but not within these two categories.[71] Pooling homogeneous and heterogeneous groups, LLL and HHH have significantly higher relative contributions than LLH and LHH ($p = 0.0098$; $N = 28$).

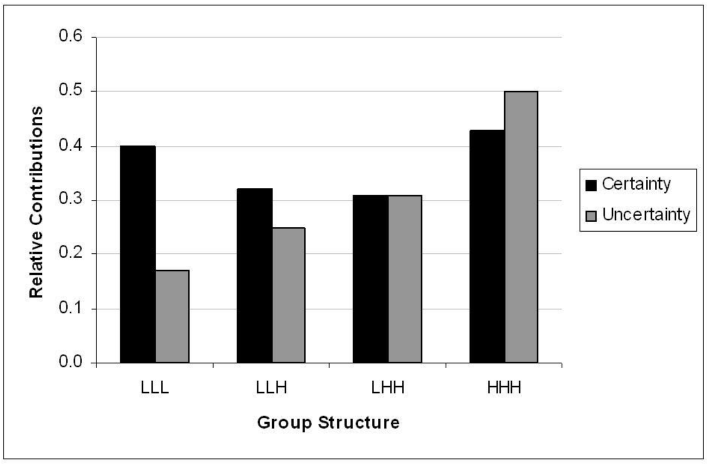
Result 4.12 [Hypothesis 4.6] *Under certainty, heterogeneous groups contribute relatively less than homogeneous groups over periods 1-10, so they are less efficient in providing the public good.*

Under certainty, Result 4.12 is consistent with the existing literature. Most studies show that endowment heterogeneity decreases cooperation (for an overview see Ledyard, 1995; or Zelmer, 2003). For instance, Cherry, Kroll, and Shogren (2005) also use a linear public goods game and find that contributions are significantly lower in heterogeneous (33.1% of endowment) than in homogeneous groups (42.1% of endowment). Note that we find 31.4% and 41.8%, respectively. Anderson, Mellor, and Milyo (2008) vary the distribution of a fixed payment for participating in the public goods experiment and also find that inequality reduces contributions. In a field experiment, Cardenas (2003) examines the effect of real wealth on cooperation. Villagers in rural Columbia knew each other and others' wealth that they brought into the public goods experiment. Again, increased inequality decreased contributions.

Under uncertainty, it can be seen in Figure 4.6 that the relationship between the number of H types and group efficiency is not U-shaped anymore, but becomes increas-

[71] Note that this holds with the exception of LLL vs. LLH where the difference is marginally insignificant. For the comparisons of homogeneous vs. heterogeneous groups we find $p = 0.1182$ for LLL vs. LLH, $p = 0.0395$ for LLL vs. LHH, $p = 0.0822$ for HHH vs. LLH, and $p = 0.0379$ for HHH vs. LHH; $N = 14$. Within these two categories we get $p = 0.9750$ for LLL vs. HHH and $p = 0.9483$ for LLH vs. LHH; $N = 14$.

Figure 4.6: Group Efficiency under Certainty and Uncertainty



ing in the number of H types. The effect of heterogeneous endowments under certainty (less efficiency than with homogeneous endowments) breaks down under uncertainty. Instead, it becomes important how rich a group is. LLL groups now contribute on average only 17% of their endowment, LLH groups 25%, LHH groups 31%, and HHH groups contribute 50% of their endowment. The efficiency differences between groups under uncertainty are all significant using Mann Whitney U tests, with two marginally insignificant exceptions.[72] Clearly, the richer a group the more efficient it is in providing the public good. When pooling homogeneous and heterogeneous groups, we now observe under uncertainty that LLL and HHH do not have significantly higher relative contributions than LLH and LHH anymore ($p = 0.1724$; $N = 32$).

Result 4.13 [Hypothesis 4.8] *Under uncertainty, heterogeneous groups do not contribute relatively less than homogeneous groups over periods 1-10, so they are equally efficient in providing the public good.*

Figure 4.6 further shows that under uncertainty poorer groups are much less efficient and richer groups are slightly more efficient than under certainty. Figure 4.6 therefore complements Figure 4.4 in that average relative contributions rather than per-period absolute contributions are considered.[73] Furthermore, it suggests that certain endowments in homogeneous groups yield the highest overall efficiency.

[72] $p = 0.1181$ for LLL vs. LLH, $p = 0.0001$ for LLL vs. LHH, $p = 0.0004$ for LLL vs. HHH, $p = 0.1429$ for LLH vs. LHH, $p = 0.0003$ for LLH vs. HHH, and $p = 0.0004$ for LHH vs. HHH; $N = 16$.

[73] Again, the difference between certainty and uncertainty is however significant only for homogeneous groups but not for heterogeneous groups. Testing the differences in efficiency between certainty and uncertainty with a Mann Whitney U test, we find $p = 0.0000$ for LLL, $p = 0.2406$ for LLH, $p = 0.3561$ for LHH, and $p = 0.0563$ for HHH; $N = 15$.

By Result 4.11 we know that uncertainty has opposing effects on poor and rich groups. The analysis on deviating beliefs showed that these adverse effects are somewhat asymmetric because beliefs deviate more in poor groups. In our experiment, we had an equal number of groups for each group structure in each treatment. We therefore expect that the total effect of uncertain endowments over all groups is slightly negative. Figure 4.7 illustrates contribution behavior under the certainty and uncertainty treatment in all ten periods. The mean contribution over all ten periods is 2.00 EMU in the uncertainty treatment and 2.21 EMU in the certainty treatment. This difference of 9.5% less contribution in the uncertainty treatment is, however, not significant when performing a Mann Whitney U test ($p = 0.1891$; $N = 60$).

Figure 4.7: Contributions under Certainty and Uncertainty

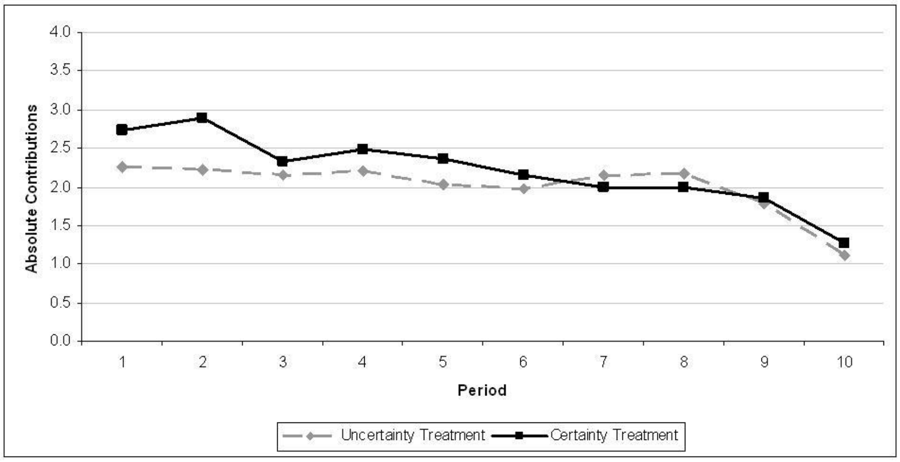


Figure 4.7 suggests that uncertainty about others' endowment levels does not have much of an effect on cooperation. In fact, this is what the sparse literature touching endowment uncertainty concludes. Isaac and Walker (1988) use a linear public goods game and find that mean contributions under certainty are 8.2% higher (though not significantly) than under uncertainty. Their interest in the effects of communication, however, may have confounded these results.[74] Chan et al. (1999) are also interested in the effects of communication. They use a non-linear public goods setting and find that uncertainty has a small but significant negative effect (7.4% less) on mean contributions.[75] Levati, Sutter, and van der Heijden (2007) use a linear public goods game to study the effects of leadership. They do not find any effect of uncertainty on contributions without a leader. However, in their uncertainty setting subjects knew the total

[74]Isaac and Walker (1988) find 18% higher levels of contributions with homogeneous than with heterogeneous endowments. These findings are also similar to our results under certainty.

[75]In their setup without communication and without preference heterogeneity, they also find 16% higher mean contributions with homogeneous than with heterogeneous endowments.

group endowment, and since we find that subjects condition their own contribution on others' relative average contribution, they indeed should not see any difference in contributions between certainty and uncertainty based on our results. Van Dijk and Grodzka (1992) use a one-shot step-level public goods setting where they are interested in contribution rules. All groups in their experiment consisted of two H and two L types. They find no effect of uncertainty on actual contributions. However, in their uncertainty treatment subjects were not informed about any endowment asymmetry. So, subjects probably believed that endowments were symmetrically distributed. With subjects being relative conditional cooperators, L types should myopically contribute more and H types less compared to certainty. Strategic over-contribution should be absent in a one-shot public goods game, so it is not surprising that van Dijk and Grodzka (1992) do not find an overall effect of uncertainty.

None of these studies has attempted to examine the specific effects of uncertainty on groups and types. Although our result is consistent with these studies in that there is only a slight negative effect of uncertainty on the overall level, we showed that this is due to two opposing effects canceling each other. Moreover, despite the overall effect is almost neutralized, uncertainty causes inequality to increase compared to certainty. When measuring inequality in the income distribution over all subjects via the Gini coefficient[76], we find that cooperation causes inequality to decrease in both treatments, but much less so under uncertainty. Without any cooperation, the Gini coefficient is 0.167. It decreases in both treatments, to 0.154 in the certainty treatment and to 0.161 in the uncertainty treatment.

Note that the two opposing effects arising for uncertain endowments should also be observable when considering type behavior, since there are more L (H) types in poor (rich) groups. As a robustness analysis of the previously considered group-specific contribution behavior, we will now analyze type-specific contributions in order to see whether the behavior of L and H types is in line with the previous results.

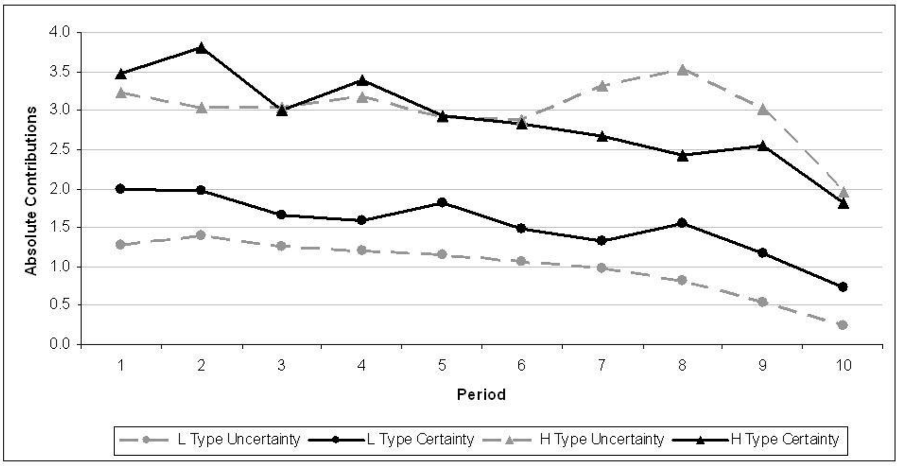
Robustness: Type-Specific Contributions

We will first consider type-specific behavior with respect to the effects of deviating beliefs. When considering types instead of groups, we find that the ΔDOB for L types is -0.28 EMU and for H types it is 0.34 EMU. For both types these differences are

[76]The Gini coefficient is based on the Lorenz curve which plots the proportion of the total income (on the vertical axis) that is cumulatively earned by the bottom $x\%$ of the population (where x goes from 0 to 100 on the horizontal axis). The Gini coefficient is then the ratio of the area that lies between the line of equality (i.e. the 45 degree line) and the Lorenz curve over the total area under the line of equality. The Gini coefficient lies always between 0 and 1 and the closer its value is to 0 the more equal is the income distribution.

significant using a Mann Whitney U test ($p = 0.0032$ for L types, and $p = 0.0012$ for H types; $N = 16$). Hence, L types myopically contribute less and H types more under uncertainty compared to certainty. For ΔSOC we observe that $\Delta SOC = -0.48$ EMU ($p = 0.0134$; $N = 45$) for L types and $\Delta SOC = 0.30$ EMU ($p = 0.1924$; $N = 45$) for H types. Thus, ΔSOC is significantly different from zero only for L types. Considering $\Delta NSOC$ we find, again consistent with our previous results, that $\Delta NSOC = -0.20$ EMU for L types ($p = 0.0936$; $N = 45$), and $\Delta NSOC = -0.04$ EMU for H types ($p = 0.8666$; $N = 45$). The difference of net (strategic) over-contribution between uncertainty and certainty is therefore only significant for L types but not for H types.

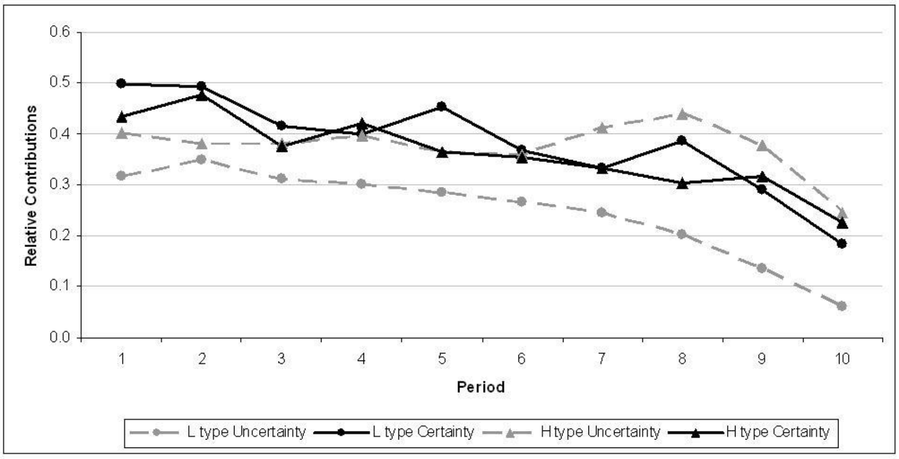
Figure 4.8: Type-Specific Absolute Contributions under Certainty and Uncertainty



Treatment differences in actual absolute contributions are illustrated in Figure 4.8. We find that L types contribute substantially less under uncertainty. The mean absolute contribution over all ten periods is 0.99 EMU in the uncertainty treatment and 1.53 EMU in the certainty treatment. This difference of 35.3% less contributions in the uncertainty treatment is significant using a Mann Whitney U test ($p = 0.0002$; $N = 45$). By contrast, H types contribute slightly more under uncertainty. Here, the mean absolute contribution over all ten periods is 3.01 EMU in the uncertainty treatment and 2.89 EMU in the certainty treatment. The difference of 4.2% more contribution in uncertainty treatment is, however, not significant ($p = 0.6186$; $N = 45$). Nevertheless, it suggests that L types contribute substantially less and H types slightly more under uncertainty. Interestingly, it seems that H types do not try to mimic L types in order to increase their payoff. This suggests that subjects follow their elicited preference of relative conditional cooperation.

Under certainty and according to the predicted effect of the minority bias, we should not expect any difference between relative contributions of L and H types. The reason is that we defined the minority bias independent of absolute endowment levels. Hence, the minority bias acts on both L and H types in a symmetric fashion and they are equally often in groups where the minority bias (own and anticipated) affects players' contributions. Figure 4.9 shows type-specific relative contributions in the ten periods for both treatments. We find that under certainty L and H types contribute similar amounts relative to their endowment. Over the ten periods L types contribute on average 38% and H types 36% of their endowment. Thus, using a Mann Whitney U test, there is no significant difference in relative contributions between types under certainty (p = 0.3212; N = 28). This picture changes however under uncertainty. Average relative contributions of L types over the ten periods drop to 25%, whereas average relative contributions of H types slightly increase to 38%. When testing for differences in these relative contributions, we now find a significant difference in relative contributions between types under uncertainty. L types contribute significantly less relative to their endowment than H types (p = 0.0232; N = 32). Moreover, between treatments there is no significant change in relative contributions among H types (testing the difference between certainty and uncertainty yields p = 0.6184; N = 45), but L types contribute significantly less under uncertainty (p = 0.0000; N = 45).

Figure 4.9: Type-Specific Relative Contributions under Certainty and Uncertainty



To conclude, type-specific contribution behavior under certainty and uncertainty is consistent with our model and supports our previous results on group-specific behavior since more L (H) types are in poor (rich) groups.

Again, the results on type-specific behavior under certainty support existing findings on the effect of heterogeneous endowments. There are several studies (see e.g. Marwell and Ames, 1979; or Wit, Wilke, and Oppewal, 1992) that show that high endowed individuals contribute more in absolute terms than low endowed individuals, and many studies even indicate that subjects contribute approximately the same in relative terms (see e.g. Rapoport, 1988; Rapoport and Suleiman, 1993; van Dijk and Grodzka, 1992; van Dijk and Wilke, 1994). A notable exception is the study by Buckley and Croson (2006), which finds that L types contribute relatively more than H types. However, when looking at field evidence, Buckley and Croson (2006, p937) note the following: "Data from the Social Welfare Research Institute at Boston College shows that in 2000, U.S. families with incomes under $125,000 (91.39% of families in the U.S.) gave an average of 2.34% of their income to charity. There was little variation of giving across incomes with the poorest families, those with incomes under $10,000 giving 2.25% (Giving USA, 2003). Thus low-income families give the same percentage of their income as high-income families."

4.6 Conclusion

This chapter establishes the preference of imperfect relative conditional cooperation as the driving force for contribution behavior in repeated public goods games. In accomplishing this task, we provided theoretical as well as experimental results supporting this claim and proceeded in three consecutive steps.

In the first part of the chapter, we presented new evidence on a refinement of conditional cooperation preferences in two dimensions. With respect to the first dimension, we showed that the conditioning of cooperation is based on average contributions of others and is not sensitive to inequalities in others' contributions. For previous and future studies eliciting the preference of conditional cooperation with the well-known method of Fischbacher, Gächter, and Fehr (2001), this is an important robustness result as this method only elicits own contributions for given average contributions of others. The second dimension concerns the question whether subjects take endowment heterogeneity into account and condition on relative or on absolute contributions of others. Here, we found that subjects are on average relative rather than absolute conditional cooperators. Furthermore, they exhibit a minority bias which causes the imperfectness of relative conditional cooperation to increase in case they are not in the majority with respect to their endowment. Future research should be conducted on other aspects where a minority bias may exist.

The second part of the chapter formalized this refined preference and analyzed contribution behavior in a repeated public goods game. In contrast to an equilibrium model where the imperfectness of conditional cooperation would trigger zero contributions, we applied a non-equilibrium model of level-k strategic thinking and showed that such a model predicts well-known patterns of contribution behavior. For instance, it was shown that players contribute more in a repeated set-up than in the one-shot version of the game (called strategic over-contribution) when their objective is to induce higher contributions by others in the future. This is likely to be a fruitful attempt in the model because others' beliefs about others' contributions are determined by their past behavior. Strategic over-contributions are decreasing because they are more effective early in the game. Moreover, total contributions are decreasing because best responses of players still represent the imperfectness of conditional cooperation. Our theoretical predictions therefore complement the empirical investigation on contribution dynamics of Fischbacher and Gächter (2010). While these theoretical results do not rely on the relativeness of conditional cooperation but would also obtain under absolute conditional cooperation, we also analyzed the effects of making endowments heterogeneous and uncertain. In these environments effects for contribution behavior only arise under relative conditional cooperation preferences. It is the minority bias that causes heterogeneous groups to contribute less than homogeneous groups under certainty. Deviating beliefs about others' endowments under uncertainty causes poor groups to contribute less and rich groups to contribute more compared to certainty. In contrast to certainty, heterogeneous groups are equally efficient in providing the public good as homogeneously endowed groups under uncertainty.

In the third part of the chapter, we tested our theoretical predictions in a 10-period public goods experiment. All predictions of the model were found to hold. With respect to uncertain endowments, our results suggest that despite having no effect on the aggregate level, the adverse effects of uncertain endowments on poor and rich groups cause the inequality of the income distribution to increase. As regards further possible policy implications, our results shed light on what drives lower contributions in heterogeneous groups under certainty, namely the minority bias. Depending on the belief distribution about others' endowments, this minority bias may, however, not be present under uncertainty. Hence, heterogeneous groups are equally efficient than homogeneous groups. But, in contrast to certainty, poor groups are less efficient in providing the public good than rich groups. Certain endowments in homogeneous groups provide the highest efficiency in the public goods provision.

It is important to note that the presented model is agnostic towards the foundations of conditional cooperation since its focus is on explaining contribution behavior and corresponding contribution dynamics in repeated games. While such a task is clearly beyond the scope of this chapter, future research should incorporate our findings in an attempt to provide a foundation for the preference of relative conditional cooperation.

Only then are we able to answer the question of *why* people behave that way. This is warranted in order to fully understand conditionally cooperative behavior and not only its consequences. The present chapter stops at showing that this preference is able to explain contribution behavior in repeated public goods games. Nevertheless, it shows that level-k thinking can provide fruitful insights not only in static environments but also in repeated games.

4.7 Appendix A: Further Theoretical Results

In this appendix we offer further theoretical results, which are not experimental variations in our experiment but have been tested in the previous literature. Cox and Sadiraj (2007) report four stylized facts from linear public goods games. First, contributions amount to a significant fraction of total endowment and that even in the last round, up to 57% of subjects (see Andreoni, 1995b) make positive contributions (our Assumptions 4.1 and 4.2). Second, a larger marginal per capita return, i.e. $\frac{A}{N}$, results in higher contributions (see e.g. Marwell and Ames, 1979; Isaac, Walker, and Thomas, 1984; Kim and Walker, 1984; Isaac and Walker, 1988; Saijo and Nakamura, 1995; Ledyard, 1995; or Zelmer, 2003). Third, a larger group size, while holding the marginal per capita return constant, results in larger contributions (see e.g. Isaac and Walker, 1988; Isaac, Walker, and Williams, 1994; or Zelmer, 2003). And fourth, larger endowments (under endowment homogeneity) lead to larger absolute contributions (see e.g. Cherry, Kroll, and Shogren, 2005; or Andreoni, 1988, 1995a).

Another well-known result from the public goods literature is that a larger time horizon increases contributions in a given period. Or as Isaac, Walker, and Williams (1994, p30) put it: "Clearly, the rate of decay of allocations to the group account is inversely related to the number of decision rounds." While all our results are based on a repeated game, it is also a stylized fact from previous experiments that even in simultaneous one-shot public goods games, subjects make positive contributions (see e.g. Marwell and Ames, 1981; Dawes and Thaler, 1988; or Ledyard, 1995). We therefore consider such a setup in this section. Finally, there is recent puzzling evidence that more selfish players do not contribute less to the public good in a repeated game (see Aurélie and Riedl, 2010). In the remaining of this section we consider all these variations. To be upfront with it, our model predicts all of the mentioned patterns.

Marginal Per Capita Return

We start with the effect of increasing the efficiency of the public good. From (4.10) it directly follows that over-contributions increase if $\frac{A}{N}$ rises (either because the return from the public investment A increases or the number of players benefiting from the public good N decreases). This is because the gain from over-contributing (LHS of (4.10)) increases whereas the loss of over-contribution (RHS of (4.10)) decreases. Myopic contributions also clearly increase with $\frac{A}{N}$ as contributing to the public good becomes less harmful in the trade-off to match others' contributions.

Proposition 4.10 *Suppose players contribute positively in period t, i.e. $x_i^{*t} > 0$. Then, ceteris paribus, increasing the efficiency of the public good (marginal per capita return) $\frac{A}{N}$ increases contributions of L1 and L2 players in period t.*

The effect of increasing $\frac{A}{N}$ in Proposition 4.10 is even stronger without the ceteris paribus assumption as the effect on myopic contributions is reinforced.

If players contribute zero in period t, i.e. $x_i^t = 0$, increasing the efficiency only weakly increases contributions for both L1 and L2 players. With $\delta_i = 0$, there is no effect of increasing $\frac{A}{N}$ for L1 players because they do not over-contribute and their only concern is then to (imperfectly) match others' contributions irrespective of how costly it is. The same holds true for L2 players with $\delta_i = 0$, but since they best respond to L1 players, increasing $\frac{A}{N}$ increases the believed contribution of other L1 players with $0 < \delta_j < 1$. In order to match these higher contributions, L2 players increase their contributions even if $\delta_i = 0$. With $\delta_i = 1$, L1 players would not contribute anything and increasing $\frac{A}{N}$ would not have an effect. By contrast, L2 players with $\delta_i = 1$ contribute everything (nothing) if the believed sum of slopes of others' best response function is larger (smaller) than $1 - \frac{A}{N} / \frac{A}{N}$. Increasing $\frac{A}{N}$ decreases the threshold for which they would contribute everything, so increasing $\frac{A}{N}$ weakly increases contributions of L2 players with $\delta_i = 1$.

Group Size

Now, we turn to the effect of increasing the group size N. In order to isolate the pure effect of increasing the number of players we have to hold various things constant. First, the marginal per capita return $\frac{A}{N}$ has to remain the same as this has an effect as shown in Proposition 4.10. Second, the believed relative contribution of others, namely $\frac{\sum_{j \neq i} x_j^{t-1}}{\sum_{j \neq i} E_j}$ for L1 players and $\frac{\sum_{j \neq i} x_j^t(\cdot, x_i^{t-1})}{\sum_{j \neq i} E_j}$ for L2 players, has to remain constant as this affects behavior simply because players are relative conditional cooperators. Third, for L2 players it is important that the sum of the slopes of others' believed best response functions stays the same as this, for instance, would simply be affected by changing the average believed preferences. If these conditions are fulfilled, myopic contributions are not affected, but over-contributions increase with N.

Proposition 4.11 *Suppose Assumption 4.1 holds. Then, ceteris paribus, increasing the number of players N while holding $\frac{A}{N}$, $\frac{\sum_{j \neq i} x_j^{t-1}}{\sum_{j \neq i} E_j}$ (for L1 players), believed $\frac{\sum_{j \neq i} x_j^t(\cdot, x_i^{t-1})}{\sum_{j \neq i} E_j}$ (for L2 players), and believed $\sum_{j \neq i} \partial x_j^{t+l}(\cdot) / \partial x_i^{t+l-1}$ (for L2 players) constant, has no effect on L1 players' contributions in period t and on L2 players' contributions in period T, but increases L2 players' contributions in period $t \neq T$.*

Under homogeneous endowments, as implemented in previous experiments showing the effect postulated in Proposition 4.11, the condition of holding the believed $\sum_{j \neq i} \partial x_j^{t+l}(\cdot) / \partial x_i^{t+l-1}$ constant directly translates into the condition that the average believed preferences of other players remain constant, which seems quite reasonable as there is no reason why additional players should be any different than others. Under heterogeneous endowments, it additionally depends on the endowments of additional players and it may indeed be the case that the believed $\sum_{j \neq i} \partial x_j^{t+l}(\cdot)/\partial x_i^{t+l-1}$ changes when increasing N, even if additional players have the same preferences. In this case a sufficient condition for Proposition 4.11 still to hold is that this change has a positive sign. Moreover, if this sign is positive, then L2 players with $\delta_i = 1$ weakly increase contributions as the believed sum of slopes of others best response functions is larger while $1 - \frac{A}{N}/\frac{A}{N}$ is not affected. However, if the believed $\sum_{j \neq i} \partial x_j^{t+l}(\cdot)/\partial x_i^{t+l-1}$ stays constant (which is the case under homogeneous endowments with unchanged average preferences), then L2 players with $\delta_i = 1$ are not affected, and neither L1 players in any case. Also, with $\delta_i = 0$, there would be no effect observed for both L1 and L2 players.

Without the ceteris paribus assumption in Proposition 4.11, myopic contributions of both L1 and L2 players would be positively affected since L2 players' higher over-contributions shift the contribution path upward.

Number of Periods

Increasing the number of periods T can only have an effect for L2 players, since L1 players are myopic. L2 players increase their over-contributions as the amount of over-contribution in period $t \neq T$ depends on the number of periods left $T - t$.

Proposition 4.12 *Suppose Assumptions 4.2 to 4.4 hold and (4.10) holds strictly. Then, ceteris paribus, increasing the number of periods T has no effect for L1 players' contributions, but increases L2 players contributions in period t.*

In the last period T, a larger time horizon also increases contributions of L2 players as over-contribution is only present when T is increased. If $\delta_i = 1$, L1 players do never contribute anything and L2 players contribute nothing or everything but in neither case does the number of periods T change this behavior as it is solely determined by how the believed $\sum_{j \neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t}$ compares to $1 - \frac{A}{N}/\frac{A}{N}$. If $\delta_i = 0$, there is again no effect observed for L1 and L2 players as over-contribution is absent.

Proposition 4.12 considers the effect of increasing T for contributions in period t. It implies that the effect on aggregate contributions over the entire horizon is positive as well. This is, however, a rather trivial result and would be obtained under much weaker assumptions. If contributions in the last round of a shorter game are sufficiently positive, then aggregate contributions in a longer game have to be larger as contributions occur for a longer time horizon. This can happen even if contributions in all periods up to the very end of the shorter game would not be any different than in the same periods of the longer game. By contrast, in Proposition 4.12 contributions are considered in a given period but for different time horizons. "Clearly, the rate of decay is faster the shorter the time horizon of the experiment. This result is inconsistent with backward induction models, and purely adaptive or learning models based on the number of rounds completed. This aspect of the VCM data is consistent with a forward-looking modelling approach based on the potential gains from cooperation."

(Isaac, Walker, and Williams, 1994, p30) Without the ceteris paribus assumption in Proposition 4.12, also myopic contributions would be positively affected as higher over-contributions of L2 players trigger higher contributions of all players in later periods. The rate of decay should therefore clearly be slower in a longer game.

One-Shot Game

The one-shot prediction of an equilibrium model with our specified preferences would be to contribute zero to the public good. This is because players *imperfectly* match others contributions so that the only fixed point is where all players do not contribute. By contrast, in our non-equilibrium model both L1 and L2 players simply contribute according to their myopic contributions in the one-shot version of the game. Over-contribution does not arise as there are no future periods where higher current contributions could induce anything from others.

Proposition 4.13 *Suppose there is only one period in the game, i.e. $T = 1$. Then, L1 and L2 players contribute according to their myopic contributions. For any given positive belief of others contributions, i.e. $\sum_{j \neq i} x_j^0 > 0$ for L1 players and $\sum_{j \neq i} x_j^1(\cdot, x_i^0) > 0$ for L2 players, contributions are positive if δ_i is sufficiently small (for given values of α_i, E_i, $\sum_{j \neq i} E_j$, and $\frac{A}{N}$).*

Interestingly, for any positive common believed relative 'home-grown' contribution of L0 players, it may be the case that L2 players contribute more than L1 players, despite the fact that they best respond and imperfectly match others. This is because, L2 players react on the believed $\sum_{j \neq i} x_j^1(\cdot, x_i^0)$, so their contribution additionally depends on the believed $\tilde{\delta}_j$ and $\tilde{\alpha}_j$ of the other L1 players. These believed preferences of others are not necessarily correct, so they may end up contributing more than L1 players. If $\delta_i = 1$, no player would ever contribute. And if $\delta_i = 0$, Proposition 4.13 still holds.

Also, in the sequential version of the one-shot public goods game, both L1 and L2 players contribute according to their myopic contributions. Thus, as second movers they would simply react to first-mover contributions. As first movers, they would contribute the same as when they were second movers if the believed sum of others' contributions is identical to the actual sum of others' contributions in the situation where they are second movers. With such consistent beliefs, L1 and L2 players' contributions would be perfectly correlated in the role of first and second movers (for recent evidence in this direction see Gächter et al., forthcoming for the case of a public goods game; Blanco, Engelmann, and Normann, 2011 for a prisoners dilemma game; or Altmann, Dohmen, and Wibral, 2008 for a trust game).

Selfishness

While on the one hand a higher degree of selfishness leads to higher over-contributions since strategic concerns increase, more selfishness, on the other hand, decreases myopic contributions as it becomes less important to match others' contributions. So, the question arises whether more selfish players can also have higher total contributions than less selfish players. This is of course only relevant in periods $t \neq T$ as there is no over-contribution which counteracts the myopic effect in the last period T. For the same reason, the question only arises if (4.10) holds strictly. Otherwise, total contributions clearly decrease with selfishness.

Proposition 4.14 *Suppose Assumptions 4.2 to 4.4 hold and (4.10) holds strictly. Then, ceteris paribus, more selfishness (increasing δ_i) decreases L1 players' contributions in period t and L2 players' contributions in period T. In period $t \neq T$ more selfishness not leading to higher contributions for L2 players is sufficient but not necessary for the assumption to hold that contributions in period t are not higher than under perfect relative conditional cooperation, i.e. $x^{*t}_{iL2} \leq \frac{E_i}{\sum_{j\neq i} E_j} \int \sum_{j\neq i} x^t_j(\cdot)\ dF_i(\sum_{j\neq i} x^t_j(\cdot))$.*

So, more selfishness may increase L2 players' contributions in period $t \neq T$ even if the assumption holds that contributions never exceed the amount of perfect relative conditional cooperation (see Assumption 4.4). Higher contributions are therefore not necessarily made by players who are less selfish. Intuitively, for small α_i the pain of over-contributing now caused by the need to match these higher contributions in the future is small and can be so small compared to its benefit that total contributions increase with selfishness.

Using the same two-part experiment as Fischbacher and Gächter (2010), Aurélie and Riedl (2010) find that more selfish players as measured in the P-Experiment do not contribute lower amounts than less selfish players in the C-Experiment, despite that beliefs of different preference types do not differ. This is consistent with Proposition 4.14.

An interesting question arising for L2 players in this context is which assumption implementing an upper bound on contributions would be necessary and sufficient for the, probably at first sight, more intuitive notion that higher selfishness leads to lower contributions.

Corollary 4.15 *Suppose Assumptions 4.2 to 4.4 hold and (4.10) holds strictly. Then, ceteris paribus, contributing not more than under full relative conditional cooperation, i.e. $x^{*t}_{iL2} \leq \alpha_i \frac{E_i}{\sum_{j\neq i} E_j} \int \sum_{j\neq i} x^t_j(\cdot)\ dF_i(\sum_{j\neq i} x^t_j(\cdot))$, is a necessary and sufficient condition for more selfishness not leading to higher contributions of L2 players in period $t \neq T$.*

Endowments

If endowments of all subjects increase by the same factor, both L1 and L2 players contributions are positively affected. Over-contributions are not affected as the sum of the believed slopes of others best response functions stays constant. But myopic contributions (weakly) increase if and only if believed contributions of others (weakly) increase. This is indeed the case if players' 'home-grown' belief in the first period is that L0 players give some *positive* percentage of their endowment (which is implied by Assumption 4.1).[77]

Proposition 4.16 *Suppose Assumption 4.1 holds. Then, increasing the endowment of all players by the same factor $z > 1$ increases L1 and L2 players' contributions in period t.*

If players 'home-grown' belief in the first period is that L0 players contribute zero percent of their endowment. Then, increasing all endowments by some factor would not change anything because L1 and L2 players would still believe that others do not contribute anything in the first period. If $\delta_i = 1$, L1 players would not contribute and this does not change when endowments increase. L2 players either contribute nothing or everything but this decision is not affected by an increase of endowments as it depends on the believed sum of slopes of others best response functions (which stays constant if all endowments increase by the same factor). If $\delta_i = 0$, the same behavior would be observed as described in Proposition 4.16.

[77] This would even hold if players were absolute conditional cooperators since such a 'home-grown' belief would shift them to a higher contribution path.

4.8 Appendix B: Proofs

Proof. [Proposition 4.1 and Corollary 4.2] We will first prove the proposition for L2 players and then for L1 players. To simplify notation we will denote

$$\int \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})\ dF_i(\sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})) = \mathbb{E}[\sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})] \equiv \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1}),$$

so that $\sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})$ is the L2 players' believed sum of other L1 players' best response functions. The maximization of U_{iL2}^t is solved by starting in the last period T. Maximizing U_{iL2}^t with respect to x_{iL2}^T yields

$$x_{iL2}^{*T} = \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N} - 1) + \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^T(\cdot, x_i^{T-1}). \tag{4.11}$$

First note that there is neither over- nor under-contribution in T and L2 players contribute according to their myopic contributions.

If $x_{iL2}^{*T} > 0$, we can plug (4.11) into the first-order condition specifying x_{iL2}^{T-1}. This yields

$$\begin{aligned} x_{iL2}^{*T-1} \quad = \quad & \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N} - 1) + \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^{T-1}(\cdot, x_i^{T-2}) \\ & + \tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}] \sum_{j\neq i} \tfrac{\partial x_j^T(\cdot)}{\partial x_i^{T-1}}. \end{aligned} \tag{4.12}$$

The first line in (4.12) represents myopic contributions and second line strategic over-contributions. Here, (4.10) is sufficient for under-contribution not to occur.

If we get a corner solution in T so that $x_{iL2}^{*T} = 0$, we can plug 0 into the first-order condition specifying x_{iL2}^{T-1}. This yields

$$\begin{aligned} x_{iL2}^{*T-1} \quad = \quad & \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N} - 1) + \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^{T-1}(\cdot, x_i^{T-2}) \\ & + \sum_{j\neq i} \tfrac{\partial x_j^T(\cdot)}{\partial x_i^{T-1}} \left[\tfrac{\delta_i}{2(1-\delta_i)} \tfrac{A}{N} - (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^2 \sum_{j\neq i} x_j^T(\cdot, x_i^{T-1}) \right], \end{aligned} \tag{4.13}$$

where the first line is myopic contribution and the second line represents over-contribution. A necessary and sufficient condition for the corner solution in T is that

$$\tfrac{\delta_i}{2(1-\delta_i)}(1 - \tfrac{A}{N}) \geq \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^T(\cdot, x_i^{T-1}). \tag{4.14}$$

Multiplying both sides of (4.14) by $\alpha_i \frac{E_i}{\sum_{j\neq i} E_j}$ shows that the term in brackets in (4.13) is weakly positive (hence no under-contribution) if $\frac{A}{N} \geq (1 - \frac{A}{N})(\alpha_i \frac{E_i}{\sum_{j\neq i} E_j})$.

For $T-2$, we first assume that $x_{iL2}^{*T-1} > 0$ after $x_{iL2}^{*T} > 0$, so that we can plug (4.12) into the first-order condition specifying x_{iL2}^{T-2}. This yields

$$\begin{aligned} x_{iL2}^{*T-2} &= \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N}-1) + \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^{T-2}(\cdot, x_i^{T-3}) \\ &\quad + \tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N}-1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}] \sum_{j\neq i} \tfrac{\partial x_j^{T-1}(\cdot)}{\partial x_i^{T-2}} (1 + \sum_{j\neq i} \tfrac{\partial x_j^{T}(\cdot)}{\partial x_i^{T-1}}). \end{aligned} \tag{4.15}$$

Again, (4.10) is sufficient for under-contribution not to occur.

If $x_{iL2}^{*T-1} > 0$ after $x_{iL2}^{*T} = 0$, we can plug (4.13) into the first-order condition specifying x_{iL2}^{T-2}. This yields

$$\begin{aligned} x_{iL2}^{*T-2} &= \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N}-1) + \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^{T-2}(\cdot, x_i^{T-3}) \\ &\quad + \sum_{j\neq i} \tfrac{\partial x_j^{T-1}(\cdot)}{\partial x_i^{T-2}} \Bigg[\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} \tfrac{\partial x_j^{T}(\cdot)}{\partial x_i^{T-1}} [\tfrac{\delta_i}{2(1-\delta_i)} \tfrac{A}{N} - (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^2 \sum_{j\neq i} x_j^{T}(\cdot, x_i^{T-1})] \\ &\quad + \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N} + \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}(\tfrac{A}{N}-1)) \Bigg], \end{aligned} \tag{4.16}$$

where the second and third lines represent over-contribution in period $T-2$, denoted x_i^{OCT-2}. Again, $\frac{A}{N} \geq (1-\frac{A}{N})(\alpha_i \frac{E_i}{\sum_{j\neq i} E_j}) \Rightarrow x_i^{OCT-2} \geq 0$, because it implies that the third line is weakly positive and via (4.14) that the second line is weakly positive.

If $x_{iL2}^{*T-1} = 0$, we can plug 0 into the first-order condition specifying x_{iL2}^{T-2}. This yields

$$\begin{aligned} x_{iL2}^{*T-2} &= \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N}-1) + \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^{T-2}(\cdot, x_i^{T-3}) \\ &\quad + \sum_{j\neq i} \tfrac{\partial x_j^{T-1}(\cdot)}{\partial x_i^{T-2}} \left[\tfrac{\delta_i}{2(1-\delta_i)} \tfrac{A}{N} - (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^2 \sum_{j\neq i} x_j^{T-1}(\cdot, x_i^{T-2}) \right]. \end{aligned} \tag{4.17}$$

Suppose $x_{iL2}^{*T-1} = 0$ after $x_{iL2}^{*T} > 0$. Then, if (4.10) holds, a necessary condition for $x_{iL2}^{*T-1} = 0$ is that

$$\tfrac{\delta_i}{2(1-\delta_i)}(1-\tfrac{A}{N}) \geq \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^{T-1}(\cdot, x_i^{T-2}). \tag{4.18}$$

Suppose $x_{iL2}^{*T-1} = 0$ after $x_{iL2}^{*T} = 0$. Then again, if (4.10) holds, (4.18) is a necessary condition for $x_{iL2}^{*T-1} = 0$. Multiplying both sides of (4.18) by $\alpha_i \frac{E_i}{\sum_{j\neq i} E_j}$ shows that the term in brackets in (4.17) is weakly positive (hence no under-contribution) if $\frac{A}{N} \geq (1-\frac{A}{N})(\alpha_i \frac{E_i}{\sum_{j\neq i} E_j})$.

Suppose $x^{*T-2}_{iL2} > 0$ after $x^{*T-1}_{iL2} > 0$ and $x^{*T}_{iL2} > 0$. Then, we can plug (4.15) into the first-order condition specifying x^{T-3}_{iL2}. This yields

$$\begin{aligned} x^{*T-3}_{iL2} &= \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N}-1) + \alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i} x^{T-3}_j(\cdot, x^{T-4}_i) \\ &\quad + \tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N}-1)\alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}] \sum_{j\neq i} \frac{\partial x^{T-2}_j(\cdot)}{\partial x^{T-3}_i} \\ &\quad (1 + \sum_{j\neq i} \frac{\partial x^{T-1}_j(\cdot)}{\partial x^{T-2}_i}(1 + \sum_{j\neq i}\frac{\partial x^{T}_j(\cdot)}{\partial x^{T-1}_i})). \end{aligned} \tag{4.19}$$

The second and third line in (4.19) determine over-contribution for which (4.10) is sufficient for being non-negative.

Now suppose $x^{*T-2}_{iL2} > 0$ after $x^{*T-1}_{iL2} > 0$ and $x^{*T}_{iL2} = 0$, Then, we can plug (4.16) into the first-order condition specifying x^{T-3}_{iL2}. This yields

$$\begin{aligned} x^{*T-3}_{iL2} &= \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N}-1) + \alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i} x^{T-3}_j(\cdot, x^{T-4}_i) \\ &\quad + \sum_{j\neq i}\frac{\partial x^{T-2}_j(\cdot)}{\partial x^{T-3}_i}\left[\alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}\frac{\partial x^{T-1}_j(\cdot)}{\partial x^{T-2}_i}\left(\alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}\frac{\partial x^{T}_j(\cdot)}{\partial x^{T-1}_i}[\tfrac{\delta_i}{2(1-\delta_i)}\tfrac{A}{N} - (\alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j})^2\right.\right. \\ &\quad \left.\left.\sum_{j\neq i} x^{T}_j(\cdot, x^{T-1}_i)] + \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N} + \alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}(\tfrac{A}{N}-1))\right) + \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N} + \alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}(\tfrac{A}{N}-1))\right]. \end{aligned} \tag{4.20}$$

Again, the second and third lines determine the amount of over-contribution, denoted x^{OCT-3}_i and $\frac{A}{N} \geq (1-\frac{A}{N})(\alpha_i \frac{E_i}{\sum_{j\neq i}E_j}) \Rightarrow x^{OCT-3}_i \geq 0$.

We can continue this until period t.[78] If all remaining periods yield positive (interior) solutions, we derive at

$$\begin{aligned} x^{*t}_{iL2} &= \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N}-1) + \alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i} x^{t}_j(\cdot, x^{t-1}_i) \\ &\quad + \tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N}-1)\alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}] \sum_{j\neq i} \frac{\partial x^{t+1}_j(\cdot)}{\partial x^{t}_i}(1 + \alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}\frac{\partial x^{t+2}_j(\cdot)}{\partial x^{t+1}_i} \\ &\quad (1 + \alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}\frac{\partial x^{t+3}_j(\cdot)}{\partial x^{t+2}_i}(\ldots(1 + \alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}\frac{\partial x^{T}_j(\cdot)}{\partial x^{T-1}_i})))). \end{aligned} \tag{4.21}$$

The first line of the RHS in (4.21) identifies the myopically motivated amount of i's contribution and the second and third lines determine the amount that is driven by strategic considerations. So, there is strategic over-contribution in the optimum if and only if the last two lines of the RHS in (4.21) are positive. From (4.21) it can be seen that given believed $\sum_{j\neq i}\frac{\partial x^{t+1}_j(\cdot)}{\partial x^{t}_i} > 0$ (which is implied by our assumption of interior solutions in all periods as shown below)

[78] Alternatively, we could also plug x^{T}_{iL2} from the maximization of U^{T}_{iL2} into U^{T-1}_{iL2} and then maximize with respect to x^{T-1}_{iL2} in order to derive at (4.12). Continuing this reasoning instead makes no difference in our case because the first-order condition defining x^{*T}_{iL2} is the same no matter whether it is derived via U^{T-1}_{iL2} or U^{T}_{iL2}.

and $0 < \delta_i < 1$, there is (weak) over-contribution if and only if (4.10) holds (weakly). If $\delta_i = 0$, there is no over-contribution. And if $\delta_i = 1$ we get a corner solution in period T. Suppose believed $\sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} = 0$. This can only happen if believed $\sum_{j\neq i} x_j^{t+1}(\cdot, x_i^t) = 0$ (because $\alpha_j > 0$) which (as shown in the proof of Proposition 4.5) implies that $x_{iL2}^{*t+1} = 0$ and also that believed $\sum_{j\neq i} x_j^{t+2}(\cdot, x_i^{t+1}), \sum_{j\neq i} x_j^{t+3}(\cdot, x_i^{t+2}), \ldots, \sum_{j\neq i} x_j^{T}(\cdot, x_i^{T-1}) = 0$ and $x_{iL2}^{*t+2}, x_{iL2}^{*t+3}, \ldots, x_{iL2}^{*T} = 0$. So, we get a corner solution in period $t+1$, and thus an interior solution in all periods implies that believed $\sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} > 0$.

While our results assume $0 < \delta_i < 1$, we can also look at the special cases. If $\delta_i = 0$ there is no over- or under- contribution. $\delta_i = 1$ leads to over-contribution (then contributing everything is optimal) if believed $\sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} > (1 - \frac{A}{N})/\frac{A}{N}$, and to no over-contribution (then contributing nothing) is optimal if believed $\sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} < (1 - \frac{A}{N})/\frac{A}{N}$.

If at least the next period yields a positive (interior) solution, but there is a corner solution at some later period, we derive at

$$\begin{aligned} x_{iL2}^{*t} \;&=\; \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N} - 1) + \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1}) \\ &\quad + \sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} \left[\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} x_i^{OCt+1} + \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N} + \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}(\tfrac{A}{N} - 1))\right] \end{aligned} \quad (4.22)$$

where x_i^{OCt+1} is the amount of over-contribution one period after t and the second line, denoted x_i^{OCt}, represents over-contribution in period t. Given that believed $\sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} > 0$, $\frac{A}{N} > (1 - \frac{A}{N})(\alpha_i \frac{E_i}{\sum_{j\neq i} E_j})$ implies that $x_i^{OCt} > 0$ and $x_i^{OCt+1} > 0$. Again, there is no over-contribution in period t if believed $\sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} = 0$. However, we saw before that this would imply that believed $\sum_{j\neq i} x_j^{t+1}(\cdot, x_i^t) = 0$ (because $\alpha_j > 0$) which in turn implies that $x_{iL2}^{*t+1} = 0$. This is a contradiction to our assumption that we have an interior solution in the next period for the considered case.

And if the next period yields a corner solution, we derive at

$$\begin{aligned} x_{iL2}^{*t} \;&=\; \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N} - 1) + \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1}) \\ &\quad + \sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} \left[\tfrac{\delta_i}{2(1-\delta_i)} \tfrac{A}{N} - (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^2 \sum_{j\neq i} x_j^{t+1}(\cdot, x_i^t)\right]. \end{aligned} \quad (4.23)$$

Again, the second line determines over-contributions and the term in brackets in (4.23) is weakly positive if (4.10) holds. Thus, we have weak over-contribution in period t if (4.10) holds. Since the next period yields a corner solution, believed $\sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} \geq 0$ leading to no over- or under- contribution in case of equality.

The last step for L2 players is to show that the second-order conditions are negative.

$$\begin{aligned}\frac{\partial^2 U_{iL2}^t}{(\partial x_i^t)^2} &= -2(1-\delta_i) + \delta_i \tfrac{A}{N}\sum_{j\neq i}\frac{\partial^2 x_j^{t+1}(\cdot)}{(\partial x_i^t)^2} - 2(1-\delta_i)(\alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}\frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t})^2 \\ &\quad +2(1-\delta_i)[x_i^{t+1} - \alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i} x_j^{t+1}(\cdot, x_i^t)](\alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}\frac{\partial^2 x_j^{t+1}(\cdot)}{(\partial x_i^t)^2}), \quad (4.24)\\ \frac{\partial^2 U_{iL2}^T}{(\partial x_i^T)^2} &= -2(1-\delta_i). \quad (4.25)\end{aligned}$$

Clearly, (4.25) is negative. A sufficient condition for (4.24) being negative is that believed $\sum_{j\neq i}\frac{\partial^2 x_j^{t+1}(\cdot)}{(\partial x_i^t)^2} = 0$. We therefore need to determine the behavior of L1 players. Each other player j maximizes U_{iL1}^t as being player i with respect to x_i^t. This determines

$$x_{iL1}^{*t} = \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N}-1) + \alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i} x_j^{t-1}, \quad (4.26)$$

which corresponds to $x_j^t(\cdot, x_i^{t-1})$ of L2 players i given an interior solution ($x_{iL1}^{*t} = 0$ in the corner solution). Thus,

$$\frac{\partial x_{iL1}^{*t}}{\partial x_j^{t-1}} = \alpha_i \tfrac{E_i}{\sum_{j\neq i}E_j} > 0 \quad (4.27)$$

in the interior solution ($\frac{\partial x_{iL1}^{*t}}{\partial x_j^{t-1}} = 0$ in the corner solution) corresponding to $\frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t}$ of L2 players i. First, note that (4.26) is an optimum since $\frac{\partial^2 U_{iL1}^t}{(x_i^t)^2} = -2(1-\delta_i) < 0$. From (4.27) (and of course also in the corner solution) it follows that $\frac{\partial^2 x_{iL1}^{*t}}{(\partial x_j^{t-1})^2} = 0$, and therefore believed $\sum_{j\neq i}\frac{\partial^2 x_j^{t+1}(\cdot)}{(\partial x_i^t)^2}$ for L2 players i.

We now turn to the case of L1 players. The first-order conditions for L1 players are identical no matter whether U_{iL1}^t or u_{iL1}^t is maximized. This holds for all t, so that each period can be considered in isolation and (4.26) describes the optimal contribution in period t. There are no strategic considerations and L1 players contribute according to their myopic contributions in all periods. The second-order condition is negative since $\frac{\partial^2 U_{iL1}^t}{(\partial x_i^t)^2} = -2(1-\delta_i)$. ■

Proof. **[Lemma 4.3]** We will show that if Assumptions 4.2, 4.3 and 4.4 are satisfied, then Assumption 4.1 holds. First note that in $t \neq T$ planned positive contributions in the last period ($x_i^{*T} > 0$ in Assumption 4.1) occur if and only if Assumption 4.2 holds, since there are only myopic contributions in the last period as over-contribution is zero. Assumption 4.3 implies that (planned) contributions (x_i^{*t}) are not lower than (planned) myopic contributions (x_i^{M*t}) in period $t \neq T$, i.e. $x_i^{*t} \geq x_i^{M*t}$ $\forall t \neq T$. In $t = T$, there is neither over- nor under-contribution by Proposition 4.1, hence contributions are equal to myopic contributions, i.e.

$x_i^{*T} = x_i^{M*T}$. Thus, $x_i^{M*t} > 0 \ \forall t \Rightarrow x_i^{*t} > 0 \ \forall t$. Assumption 4.2 implies that $x_i^{M*t} > 0 \ \forall t \neq T$ if the believed

$$\sum_{j\neq i} x_j^t \geq \sum_{j\neq i} x_j^{t+1} \geq \sum_{j\neq i} x_j^{t+2} \geq \ldots \geq \sum_{j\neq i} x_j^T. \tag{4.28}$$

For L1 players (4.28) is satisfied, since L1 players believe that all others are L0 players, who contribute in $t+1$ the same as in t. So, in t they believe that $\sum_{j\neq i} x_j^t = \sum_{j\neq i} x_j^{t+1} = \sum_{j\neq i} x_j^{t+2} = \ldots = \sum_{j\neq i} x_j^T$. For L2 players (4.28) is also satisfied, since L2 players believe that all others are L1 players, who best respond in $t+1$ to what they observed in t. These best responses can be at most perfect matches (since $0 \leq \delta_j \leq 1$ and $0 < \alpha_j \leq 1$) and by Assumption 4.4 L2 players also never (plan to) cooperate more than perfectly. So, what L2 players in t think L1 players will do is $\sum_{j\neq i} x_j^t(\cdot, x_i^{t-1}) \geq \sum_{j\neq i} x_j^{t+1}(\cdot, x_i^t) \geq \sum_{j\neq i} x_j^{t+2}(\cdot, x_i^{t+1}) \geq \ldots \geq \sum_{j\neq i} x_j^T(\cdot, x_i^{T-1})$. ■

Proof. [Proposition 4.4] Because of Proposition 4.1, Proposition 4.4 is only relevant for L2 players. By Assumption 4.1 (via Lemma 4.3), optimal contributions in $t \neq T$ are given by (4.21) and in T by (4.11). Further, Assumption 4.1 implies that in $t \neq T$ the believed

$$\sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t}, \sum_{j\neq i} \frac{\partial x_j^{t+2}(\cdot)}{\partial x_i^{t+1}}, \sum_{j\neq i} \frac{\partial x_j^{t+3}(\cdot)}{\partial x_i^{t+2}}, \ldots, \sum_{j\neq i} \frac{\partial x_j^{T}(\cdot)}{\partial x_i^{T-1}} > 0. \tag{4.29}$$

Plugging (4.29) into (4.21) we directly get that planned over- (or under-) contribution is larger for smaller t if $\frac{A}{N} + (\frac{A}{N} - 1)\alpha_i \frac{E_i}{\sum_{j\neq i} E_j} \neq 0$ and $0 < \delta_i < 1$. With $\frac{A}{N} = (1 - \frac{A}{N})\alpha_i \frac{E_i}{\sum_{j\neq i} E_j}$ or $\delta_i = 0$, there is no over- (or under-) contribution. If $\delta_i = 1$, we have $x_{iL2}^{*T} = 0$ and thus Assumption 4.1 is violated. In this case, planned over-contribution is either zero or positive but dropping to zero at some point. From (4.11) we can also directly see that (planned) over-contribution is zero in period T and since it is positive in period $T-1$ (by Proposition 4.1), it is also decreasing from the second to last to the last period (actual and planned). Note that (4.29) is not sufficient for actual over-contributions to be decreasing in other periods. However, by the proof of Lemma 4.3 it follows that in $t \neq T$ the believed

$$\sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} \geq \sum_{j\neq i} \frac{\partial x_j^{t+2}(\cdot)}{\partial x_i^{t+1}} \geq \sum_{j\neq i} \frac{\partial x_j^{t+3}(\cdot)}{\partial x_i^{t+2}} \geq \ldots \geq \sum_{j\neq i} \frac{\partial x_j^{T}(\cdot)}{\partial x_i^{T-1}}, \tag{4.30}$$

because L2 players think all others are L1 players who best respond to observed contributions from the last period. By Assumptions 4.4 and 4.3, these observed contributions cannot be increasing and the believed $(\tilde{\delta}_j, \tilde{\alpha}_j)$ is constant over time. The relations in (4.30) may hold strictly because the believed contribution path may push some best responses of others in the support of $F_{ij}(\cdot)$ in the corner solution. This means that in $t+1$, the believed $\sum_{j\neq i} \frac{\partial x_j^{t+2}(\cdot)}{\partial x_i^{t+1}}, \sum_{j\neq i} \frac{\partial x_j^{t+3}(\cdot)}{\partial x_i^{t+2}}, \ldots, \sum_{j\neq i} \frac{\partial x_j^{T}(\cdot)}{\partial x_i^{T-1}}$ may be higher than from the perspective of t, but it cannot be higher than $\sum_{j\neq i}\frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t}, \sum_{j\neq i}\frac{\partial x_j^{t+2}(\cdot)}{\partial x_i^{t+1}}, \ldots, \sum_{j\neq i}\frac{\partial x_j^{T}(\cdot)}{\partial x_i^{T-1}}$ from the perspective of t. Hence, by looking at (4.21) actual over-contributions are decreasing over time. ■

Proof. [Proposition 4.5] By the proof of Lemma 4.3, we know that planned myopic contributions are stable for L1 players and weakly decreasing for L2 players. We now look at actual myopic contributions over time. First, note that actual myopic contributions of L1 players are lower in period t than in $t-1$ if and only if the actual $\sum_{j\neq i} x_j^{t-1} < \sum_{j\neq i} x_j^{t-2}$. A sufficient condition for this to hold is that actual contributions of all players are weakly decreasing. For L2 players actual myopic contributions are lower in t than in $t-1$ if and only if the believed $\sum_{j\neq i} x_j^{t}(\cdot, x_i^{t-1}) < \sum_{j\neq i} x_j^{t-1}(\cdot, x_i^{t-2})$. Since there is no updating of beliefs about others' preferences, again a sufficient condition for this to hold is that actual contributions of all players are weakly decreasing. In the first period, our assumption about L0 players is that they contribute a certain percentage of their endowment. All players have the same belief about L0 players' contributions in the first period. This means that all players' 'home-grown' belief about contributions of L0 players is the same in the first period, namely that L0 players contribute a certain relative amount.

Case I: Suppose that this percentage is zero, i.e. L1 and L2 players believe in $t = 1$ that L0 players contribute nothing. For L1 players this means that $x_j^0 = 0\ \forall j \neq i \Rightarrow \sum_{j\neq i} x_j^0 = 0$ and her best response is $x_{iL1}^{*1} = 0$, which follows immediately since there is no over-contribution for L1 players. For L2 players this means that the believed $x_j^1(\cdot, x_i^0) = 0\ \forall j \neq i \Rightarrow \sum_{j\neq i} x_j^1(\cdot, x_i^0) = 0$, since L2 players think that all others are L1 players. Given that the believed $\sum_{j\neq i} x_j^1(\cdot, x_i^0) = 0$ it follows that $x_{iL2}^{*1} = 0$, because $x_{iL2}^{*1} > 0$ would mean that L2 player i contributes more than what is conditionally perfect (see Assumption 4.4). In all remaining periods $t > 1$, L0 players contribute the same as in the previous period. So, L1 player i's best response in the second period is $x_{iL1}^{*2} = 0$ since actual $x_j^1 = 0\ \forall j \neq i \Rightarrow \sum_{j\neq i} x_j^1 = 0$ and there is no over-contribution for L1 players. From the perspective of L2 player i, believed $x_j^2(\cdot, x_i^1) = 0\ \forall j \neq i \Rightarrow \sum_{j\neq i} x_j^2(\cdot, x_i^1) = 0$. Given that the believed $\sum_{j\neq i} x_j^2(\cdot, x_i^1) = 0$ it follows again that $x_{iL2}^{*2} = 0$. This reasoning continues until the last period where Assumption 4.4 is not needed as there is no over-contribution in period T for L2 players. Hence, for the initial belief $x_i^0 = 0$, the actual contribution path is $x_i^{*t} = 0\ \forall t$ for both L1 and L2 players and hence actual $\sum_{j\neq i} x_j^{t-1} = 0\ \forall t$ for L1 players and believed $\sum_{j\neq i} x_j^{t}(\cdot, x_i^{t-1}) = 0\ \forall t$ for L2 players.

Case II: Now, we consider the case where the 'home-grown' belief is that L0 players contribute a positive percentage of their endowment, which implies $x_i^0 > 0$ for all L0 players. From the perspective of L1 player i, $x_j^0 > 0\ \forall j \neq i \Rightarrow \sum_{j\neq i} x_j^0 > 0$ and her best response is $x_{iL1}^{*1} \geq 0$, since $\sum_{j\neq i} x_j^0$ may be too small (given her δ_i and α_i). From the perspective of L2 player i, this means that believed $x_j^1(\cdot, x_i^0) \geq 0\ \forall j \neq i \Rightarrow \sum_{j\neq i} x_j^1(\cdot, x_i^0) \geq 0$. The believed $\sum_{j\neq i} x_j^1(\cdot, x_i^0) \geq 0$ may happen, since L2 players' believe L1 players' preferences may be too 'selfish' ($\tilde{\alpha}_j$ and $\tilde{\delta}_j$). Her best response is $x_{iL2}^{*1} \geq 0$ and even if believed $\sum_{j\neq i} x_j^1(\cdot, x_i^0) > 0$,

$x^{*1}_{iL2} = 0$ may be a best response when L2 players are themselves too 'selfish' (α_i and δ_i). This means we can consider four possible sub-cases. First, $x^{*1}_{iL1} = 0 \wedge x^{*1}_{iL2} = 0$, second, $x^{*1}_{iL1} > 0 \wedge x^{*1}_{iL2} = 0$, third, $x^{*1}_{iL1} = 0 \wedge x^{*1}_{iL2} > 0$, and fourth, $x^{*1}_{iL1} > 0 \wedge x^{*1}_{iL2} > 0$.

Sub-Case II.1 ($x^{*1}_{iL1} = 0 \wedge x^{*1}_{iL2} = 0$): In this case, $\sum_{j\neq i} x^1_j = 0$ and hence $x^{*2}_{iL1} = 0$. Also, believed $\sum_{j\neq i} x^2_j(\cdot, x^1_i) = 0 \Rightarrow x^{*2}_{iL2} = 0$. This continues until the last period, thus $x^{*t}_i = 0\ \forall t$ for both L1 and L2 players and hence actual $\sum_{j\neq i} x^{t-1}_j = 0\ \forall t$ for L1 players and believed $\sum_{j\neq i} x^t_j(\cdot, x^{t-1}_i) = 0\ \forall t$ for L2 players.

Sub-Case II.2 ($x^{*1}_{iL1} > 0 \wedge x^{*1}_{iL2} = 0$): First, note that what L1 player i reacts on in $t = 2$ is less than (weakly less if $\delta_i = 0$) she reacted on in $t = 1$ (given by the 'home-grown' belief about L0), so that $0 < \sum_{j\neq i} x^1_j < \sum_{j\neq i} x^0_j$. This implies that $0 \leq x^{*2}_{iL1} < x^{*1}_{iL1}$. Suppose $x^{*1}_{iL2} = 0$ because L2 players' beliefs about L1 players' preferences were too 'selfish'. Since there is no updating about these preferences and L2 players had the same belief about the 'home-grown' contribution of L0 players than L1 players, it has to hold that the believed $\sum_{j\neq i} x^2_j(\cdot, x^1_i) = \sum_{j\neq i} x^1_j(\cdot, x^0_i) = 0$ because L1 players best responded to L0 and Assumption 4.4 holds. It follows that $x^{*2}_{iL2} = 0$. Now, suppose that $x^{*1}_{iL2} = 0$ because L2 players' preferences were too 'selfish'. Because all players never cooperate more than perfectly, $0 \leq \sum_{j\neq i} x^2_j(\cdot, x^1_i) < \sum_{j\neq i} x^1_j(\cdot, x^0_i)$ has to hold and therefore $x^{*2}_{iL2} = 0$ if over-contributions in $t = 2$ are not higher than in $t = 1$ (see Proposition 4.4). In all remaining periods either *Sub-Case II.1* or *II.2* applies.

Sub-Case II.3 ($x^{*1}_{iL1} = 0 \wedge x^{*1}_{iL2} > 0$): What L1 player i reacts on in $t = 2$ is less than she reacted on in $t = 1$, because, given beliefs about L0 players' 'home-grown' contribution, all players never cooperate more than perfectly. So, $0 \leq \sum_{j\neq i} x^1_j < \sum_{j\neq i} x^0_j$. This implies that $x^{*2}_{iL1} = x^{*1}_{iL1} = 0$. For L2 players $x^{*1}_{iL2} > 0$ can only happen if the believed $\sum_{j\neq i} x^1_j(\cdot, x^0_i) > 0$. Since all players never cooperate more than perfectly, $0 \leq \sum_{j\neq i} x^2_j(\cdot, x^1_i) \leq \sum_{j\neq i} x^1_j(\cdot, x^0_i)$ has to hold and therefore $0 \leq x^{*2}_{iL2} \leq x^{*1}_{iL2}$ if over-contributions in $t = 2$ are not higher than in $t = 1$. In all remaining periods either *Sub-Case II.1* or *II.3* applies.

Sub-Case II.4 ($x^{*1}_{iL1} > 0 \wedge x^{*1}_{iL2} > 0$): What L1 player i reacts on in $t = 2$ is less than (weakly less if $\delta_i = 0$) she reacted on in $t = 1$ (given by the 'home-grown' belief about L0), so that $0 < \sum_{j\neq i} x^1_j < \sum_{j\neq i} x^0_j$. This implies that $0 \leq x^{*2}_{iL1} < x^{*1}_{iL1}$. For L2 players $x^{*1}_{iL2} > 0$ can only happen if the believed $\sum_{j\neq i} x^1_j(\cdot, x^0_i) > 0$. Since all players never cooperate more than perfectly, $0 \leq \sum_{j\neq i} x^2_j(\cdot, x^1_i) \leq \sum_{j\neq i} x^1_j(\cdot, x^0_i)$ has to hold and therefore $0 \leq x^{*2}_{iL2} \leq x^{*1}_{iL2}$ if over-contributions in $t = 2$ are not higher than in $t = 1$. In all remaining periods either *Sub-Case II.1* or *II.2* or *II.3* or *II.4* applies. ■

Proof. [Proposition 4.7] By Lemma 4.3 we can use (4.26) and (4.21) to derive the effects of one's own minority bias (i.e. decreased α_i) on optimal contributions on L1 and L2 players, respectively. We start with L1 players, where

$$\frac{\partial x_{iL1}^{*t}}{\partial \alpha_i} = \tfrac{E_i}{\sum_{j \neq i} E_j} \sum_{j \neq i} x_j^{t-1}. \tag{4.31}$$

So, for a given contribution of others, L1 player's contribution is lower with her own minority bias being present in all periods.

We now turn to L2 players and start with contribution behavior in period $t \neq T$. First, note that (4.21) can also be written as

$$\begin{aligned} x_{iL2}^{*t} &= \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N} - 1) + \alpha_i \tfrac{E_i}{\sum_{j \neq i} E_j} \sum_{j \neq i} x_j^t(\cdot, x_i^{t-1}) \\ &\quad + \tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j \neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j \neq i} E_j})^0 \sum_{j \neq i} \tfrac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} \\ &\quad + \tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j \neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j \neq i} E_j})^1 \sum_{j \neq i} \tfrac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} \sum_{j \neq i} \tfrac{\partial x_j^{t+2}(\cdot)}{\partial x_i^{t+1}} \\ &\quad + \tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j \neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j \neq i} E_j})^2 \sum_{j \neq i} \tfrac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} \sum_{j \neq i} \tfrac{\partial x_j^{t+2}(\cdot)}{\partial x_i^{t+1}} \sum_{j \neq i} \tfrac{\partial x_j^{t+3}(\cdot)}{\partial x_i^{t+2}} \\ &\quad \vdots \\ &\quad + \tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j \neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j \neq i} E_j})^{T-(t+1)} \prod_{l=1}^{T-t} \sum_{j \neq i} \tfrac{\partial x_j^{t+l}(\cdot)}{\partial x_i^{t+l-1}}. \end{aligned} \tag{4.32}$$

Now, we can use (4.32) to derive the marginal effect of increasing α_i on L2 player i's optimal contribution:

$$\begin{aligned}
\frac{\partial x^{*t}_{iL2}}{\partial \alpha_i} &= \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x^t_j(\cdot, x^{t-1}_i) \\
&+ \tfrac{\delta_i}{2(1-\delta_i)} (\tfrac{A}{N} - 1) \tfrac{E_i}{\sum_{j\neq i} E_j} (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^0 \sum_{j\neq i} \tfrac{\partial x^{t+1}_j(\cdot)}{\partial x^t_i} + 0 \\
&+ \tfrac{\delta_i}{2(1-\delta_i)} (\tfrac{A}{N} - 1) \tfrac{E_i}{\sum_{j\neq i} E_j} (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^1 \sum_{j\neq i} \tfrac{\partial x^{t+1}_j(\cdot)}{\partial x^t_i} \sum_{j\neq i} \tfrac{\partial x^{t+2}_j(\cdot)}{\partial x^{t+1}_i} \\
&+ 1 \tfrac{\delta_i}{2(1-\delta_i)} [\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}] (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^0 \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} \tfrac{\partial x^{t+1}_j(\cdot)}{\partial x^t_i} \sum_{j\neq i} \tfrac{\partial x^{t+2}_j(\cdot)}{\partial x^{t+1}_i} \\
&+ \tfrac{\delta_i}{2(1-\delta_i)} (\tfrac{A}{N} - 1) \tfrac{E_i}{\sum_{j\neq i} E_j} (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^2 \sum_{j\neq i} \tfrac{\partial x^{t+1}_j(\cdot)}{\partial x^t_i} \sum_{j\neq i} \tfrac{\partial x^{t+2}_j(\cdot)}{\partial x^{t+1}_i} \sum_{j\neq i} \tfrac{\partial x^{t+3}_j(\cdot)}{\partial x^{t+2}_i} \\
&+ 2 \tfrac{\delta_i}{2(1-\delta_i)} [\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}] (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^1 \tfrac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} \tfrac{\partial x^{t+1}_j(\cdot)}{\partial x^t_i} \sum_{j\neq i} \tfrac{\partial x^{t+2}_j(\cdot)}{\partial x^{t+1}_i} \sum_{j\neq i} \tfrac{\partial x^{t+3}_j(\cdot)}{\partial x^{t+2}_i} \\
&\vdots \\
&+ \tfrac{\delta_i}{2(1-\delta_i)} (\tfrac{A}{N} - 1) \tfrac{E_i}{\sum_{j\neq i} E_j} (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^{T-(t+1)} \prod_{l=1}^{T-t} \sum_{j\neq i} \tfrac{\partial x^{t+l}_j(\cdot)}{\partial x^{t+l-1}_i} \\
&+ (T-(t+1)) \tfrac{\delta_i}{2(1-\delta_i)} [\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}] (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^{T-(t+1)-1} \tfrac{E_i}{\sum_{j\neq i} E_j} \prod_{l=1}^{T-t} \sum_{j\neq i} \tfrac{\partial x^{t+l}_j(\cdot)}{\partial x^{t+l-1}_i}
\end{aligned} \quad (4.33)$$

Note that $\frac{E_i}{\sum_{j\neq i} E_j}$ can be canceled in (4.33) in order to determine whether the RHS of (4.33) is positive or negative. Increasing α_i has a positive effect on myopic contributions (first line of the RHS in (4.33)), but an ambiguous effect on over-contributions (remaining lines of the RHS in (4.33)). If there was only one period left, i.e. $t = T-1$, the effect on over-contribution would be clearly negative (determined by the second line of the RHS in (4.33) only).

For the last period T we can use (4.11) to derive the corresponding marginal effect.

$$\frac{\partial x^{*T}_{iL2}}{\partial \alpha_i} = \frac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x^t_j(\cdot, x^{t-1}_i), \quad (4.34)$$

where the RHS in (4.34) is positive. So, for a given believed sum of others' contributions, L2 players' myopic contribution is lower with her own minority bias being present in all periods. ■

Proof. [Proposition 4.8] Since L1 players react to L0 players, who in turn simply contribute the same as last period and do not best respond, L1 players cannot anticipate others' minority bias. By Lemma 4.3 we can use (4.21) and (4.11) to derive the effect of anticipating the minority bias of others for L2 players in period $t \neq T$ and T, respectively. If L2 player i anticipates the minority bias of another player, she will weakly decrease own contributions in period $t \neq T$ if and only if

$$\frac{\partial x^{*t}_{iL2}}{\partial x^t_j} \frac{\partial x^t_j(\cdot)}{\partial \alpha_j} + \frac{\partial x^{*t}_{iL2}}{\partial x'^{t+1}_j} \frac{\partial x'^{t+1}_j(\cdot)}{\partial \alpha_j} + \frac{\partial x^{*t}_{iL2}}{\partial x'^{t+2}_j} \frac{\partial x'^{t+2}_j(\cdot)}{\partial \alpha_j} + \ldots + \frac{\partial x^{*t}_{iL2}}{\partial x'^{T}_j} \frac{\partial x'^{T-1}_j(\cdot)}{\partial \alpha_j} \geq 0, \quad (4.35)$$

where $x_j'^{t+l} = \frac{\partial x_j^{t+l}(\cdot)}{\partial x_i^{t+l-1}}$. Now, $\frac{\partial x_{iL2}^{*t}}{\partial x_j^t} = \alpha_i \frac{E_i}{\sum_{j\neq i} E_j} > 0$. For $\frac{\partial x_j^t(\cdot)}{\partial \alpha_j}$ we can look at L1 players' behavior (as player i) and observe that either $\frac{\partial x_{iL1}^{*t}}{\partial \alpha_i} = \frac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^{t-1} > 0$ or $\frac{\partial x_{iL1}^{*t}}{\partial \alpha_i} = 0$, depending on whether $(\tilde{\delta}_j, \tilde{\alpha}_j)$ is such that the believed best response of the other L1 player is an interior solution with positive support in $F_{ij}(\cdot)$. For $\frac{\partial x_j'^{t+l}(\cdot)}{\partial \alpha_j}$, we can again look at L1 players' behavior (as player i) and observe that either $\frac{\partial x_{iL1}^{*t}/\partial x_j^{t-1}}{\partial \alpha_i} = \frac{E_i}{\sum_{j\neq i} E_j} > 0$ or $\frac{\partial x_{iL1}^{*t}/\partial x_j^{t-1}}{\partial \alpha_i} = 0$, again depending on the believed $(\tilde{\delta}_j, \tilde{\alpha}_j)$. Now, the last step is to show that $\frac{\partial x_{iL2}^{*t}}{\partial x_j'^{t+l}} \geq 0$, which by looking at (4.32) is clearly fulfilled if and only if (4.10) holds.

In the last period T, anticipating the minority bias of another player weakly decreases own contributions if and only if $\frac{\partial x_{iL2}^{*T}}{\partial x_j^T} \frac{\partial x_j^T(\cdot)}{\partial \alpha_j} \geq 0$ where $\frac{\partial x_{iL2}^{*T}}{\partial x_j^T} = \alpha_i \frac{E_i}{\sum_{j\neq i} E_j} > 0$ and $\frac{\partial x_j^T(\cdot)}{\partial \alpha_j} \geq 0$ holds as well. ■

Proof. [Proposition 4.9] Players have only beliefs about others endowments, so that formally $\sum_{j\neq i} E_j$ becomes $\int \sum_{j\neq i} E_j \; dG_i(\sum_{j\neq i} E_j)$, where others' endowments $\sum_{j\neq i} E_j$ are drawn according to the distribution $G_i(\cdot)$. In order to isolate the pure effect of deviating beliefs, we abstract from pure uncertainty effects, so that respective utilities of L1 and L2 players in period t become

$$u_{iL1}^t = \delta_i \; \Pi_i^t - (1-\delta_i)(x_i^t - \alpha_i \tfrac{\sum_{j\neq i} x_j^{t-1}}{\mathbb{E}[\sum_{j\neq i} E_j]} E_i)^2, \tag{4.36}$$

and

$$u_{iL2}^t = \delta_i \; \Pi_i^t - (1-\delta_i)(x_i^t - \alpha_i \tfrac{\mathbb{E}[\sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})]}{\mathbb{E}[\sum_{j\neq i} E_j]} E_i)^2, \tag{4.37}$$

where $\mathbb{E}[\sum_{j\neq i} E_j] = \sum_{j\neq i} E_j$ so that we can directly compare certain with uncertain endowments. In Proposition 4.9, we are interested in the marginal effect of believing others are richer or poorer in a way that determines $\mathbb{E}[\sum_{j\neq i} E_j]$, and without loss of generality we can consider the marginal effect of increasing $\sum_{j\neq i} E_j$.

Lemma 4.3 also holds under uncertain endowments as beliefs about others' endowments are fixed over time. With uncertain endowments, Assumption 4.4 becomes

$$x_{iL1}^{*k} \leq \tfrac{E_i}{\mathbb{E}[\sum_{j\neq i} E_j]} \sum_{j\neq i} x_j^{k-1} \leq E_i$$

for L1 players, and

$$x_{iL2}^{*k} \leq \tfrac{E_i}{\mathbb{E}[\sum_{j\neq i} E_j]} \int \sum_{j\neq i} x_j^k(\cdot) \; dF_i(\sum_{j\neq i} x_j^k(\cdot)) \leq E_i$$

for L2 players. Condition (4.10) becomes $\frac{A}{N} \geq (1-\frac{A}{N})\alpha_i \frac{E_i}{\mathbb{E}[\sum_{j\neq i} E_j]}$. Note that under uncertainty, Assumption 4.4 implicitly restricts beliefs about the sum of others' endowments in the following way. It can never happen that the believed sum of others' contributions exceeds the believed sum of others' endowments in any period k. This holds for L1 and L2 players, and also for L2 players' beliefs about L1 players. Hence, we can use (4.26) to derive the effect of L1 players in period t, and (4.32) and (4.11) to derive the effect of L2 players in period $t \neq T$ and T, respectively. We start with L1 players, where

$$\frac{\partial x^{*t}_{iL1}}{\partial \sum_{j\neq i} E_j} = -\alpha_i \tfrac{E_i}{(\sum_{j\neq i} E_j)^2} \sum_{j\neq i} x_j^{t-1}, \tag{4.38}$$

since L1 players best respond to L0 players who just contribute the same as last period so that second-order beliefs do not play a role. Clearly, thinking others are richer decreases contributions of L1 players.

The effect of thinking others are richer in period $t \neq T$ is determined for L2 players by

$$\begin{aligned}
&\frac{\partial x^{*t}_{iL2}}{\partial \sum_{j\neq i} E_j} = \\
&-\alpha_i \tfrac{E_i}{(\sum_{j\neq i} E_j)^2} \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1}) + \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \tfrac{\partial \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})}{\partial \sum_{j\neq i} E_j} \\
&- \tfrac{\delta_i}{2(1-\delta_i)} (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{(\sum_{j\neq i} E_j)^2} (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^0 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot) \right)^1 - 0 \\
&+1 \tfrac{\delta_i}{2(1-\delta_i)} [\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^0 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot) \right)^0 \tfrac{\partial \sum_{j\neq i} x_j'^{t+l}(\cdot)}{\partial \sum_{j\neq i} E_j} \\
&- \tfrac{\delta_i}{2(1-\delta_i)} (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{(\sum_{j\neq i} E_j)^2} (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^1 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot) \right)^2 \\
&-1 \tfrac{\delta_i}{2(1-\delta_i)} [\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^0 \alpha_i \tfrac{E_i}{(\sum_{j\neq i} E_j)^2} \left(\sum_{j\neq i} x_j'^{t+l}(\cdot) \right)^2 \\
&+2 \tfrac{\delta_i}{2(1-\delta_i)} [\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^1 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot) \right)^1 \tfrac{\partial \sum_{j\neq i} x_j'^{t+l}(\cdot)}{\partial \sum_{j\neq i} E_j} \\
&- \tfrac{\delta_i}{2(1-\delta_i)} (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{(\sum_{j\neq i} E_j)^2} (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^2 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot) \right)^3 \\
&-2 \tfrac{\delta_i}{2(1-\delta_i)} [\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^1 \alpha_i \tfrac{E_i}{(\sum_{j\neq i} E_j)^2} \left(\sum_{j\neq i} x_j'^{t+l}(\cdot) \right)^3 \\
&+3 \tfrac{\delta_i}{2(1-\delta_i)} [\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^2 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot) \right)^2 \tfrac{\partial \sum_{j\neq i} x_j'^{t+l}(\cdot)}{\partial \sum_{j\neq i} E_j} \\
&\vdots \\
&- \tfrac{\delta_i}{2(1-\delta_i)} (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{(\sum_{j\neq i} E_j)^2} (\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^{T-(t+1)} \left(\sum_{j\neq i} x_j'^{t+l}(\cdot) \right)^{T-t} \\
&-(T-(t+1)) \tfrac{\delta_i}{2(1-\delta_i)} [\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^{T-(t+1)-1} \alpha_i \tfrac{E_i}{(\sum_{j\neq i} E_j)^2} \left(\sum_{j\neq i} x_j'^{t+l}(\cdot) \right)^{T-t} \\
&+(T-t) \tfrac{\delta_i}{2(1-\delta_i)} [\tfrac{A}{N} + (\tfrac{A}{N} - 1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^{T-(t+1)} \left(\sum_{j\neq i} x_j'^{t+l}(\cdot) \right)^{T-t-1} \tfrac{\partial \sum_{j\neq i} x_j'^{t+l}(\cdot)}{\partial \sum_{j\neq i} E_j},
\end{aligned} \tag{4.39}$$

where $\int \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1}) dF_i(\sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})) \equiv \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})$ (as first stated in the proof of Proposition 4.1) and $\sum_{j\neq i} x_j'^{t+l}(\cdot) = \sum_{j\neq i} \frac{\partial x_j^{t+l}(\cdot)}{\partial x_i^{t+l-1}}$. Note that for simplicity we assumed that $\prod_{l=1}^{T-t} \sum_{j\neq i} x_j'^{t+l}(\cdot) = \left(\sum_{j\neq i} x_j'^{t+l}(\cdot)\right)^{T-t}$. This can be done since Lemma 4.3 implies an interior solution in all periods, thus $\sum_{j\neq i} x_j'^{t+l}(\cdot) > 0$, and because even if (4.27) does not hold with equality, it does not affect the qualitative comparison of $\frac{\partial x_{iL2}^{*t}}{\partial \alpha_i}$ and $\frac{\partial x_{iL2}^{*t}}{\partial \sum_{j\neq i} E_j}$ because both are equally affected. By making the same assumption, we can rewrite (4.33) as

$$\frac{\partial x_{iL2}^{*t}}{\partial \alpha_i} = \frac{E_i}{\sum_{j\neq i} E_j}[X_I'^M + X_I'^{OC}], \tag{4.40}$$

where

$$X_I'^M = \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1}), \tag{4.41}$$

$$\begin{aligned}
X_I'^{OC} = \; & \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N}-1)(\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^0 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot)\right)^1 \\
& + \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N}-1)(\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^1 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot)\right)^2 \\
& + 1\tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N}-1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^0 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot)\right)^2 \\
& + \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N}-1)(\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^2 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot)\right)^3 \\
& + 2\tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N}-1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^1 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot)\right)^3 \\
& \vdots \\
& + \tfrac{\delta_i}{2(1-\delta_i)}(\tfrac{A}{N}-1)(\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^{T-(t+1)} \left(\sum_{j\neq i} x_j'^{t+l}(\cdot)\right)^{T-t} \\
& + (T-(t+1))\tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N}-1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^{T-(t+1)-1} \left(\sum_{j\neq i} x_j'^{t+l}(\cdot)\right)^{T-t}.
\end{aligned} \tag{4.42}$$

We can also rewrite (4.39) as

$$\frac{\partial x_{iL2}^{*t}}{\partial \sum_{j\neq i} E_j} = -\alpha_i \frac{E_i}{(\sum_{j\neq i} E_j)^2}[X_I'^M + X_I'^{OC}] + X_{II}'^M + X_{II}'^{OC}, \tag{4.43}$$

where

$$
\begin{aligned}
X'^{M}_{II} &= \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \tfrac{\partial \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})}{\partial \sum_{j\neq i} E_j}, \qquad (4.44)\\
X'^{OC}_{II} &= 1\tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N}-1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^0 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot)\right)^0 \tfrac{\partial \sum_{j\neq i} x_j'^{t+l}(\cdot)}{\partial \sum_{j\neq i} E_j}\\
&\quad +2\tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N}-1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^1 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot)\right)^1 \tfrac{\partial \sum_{j\neq i} x_j'^{t+l}(\cdot)}{\partial \sum_{j\neq i} E_j}\\
&\quad +3\tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N}-1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^2 \left(\sum_{j\neq i} x_j'^{t+l}(\cdot)\right)^2 \tfrac{\partial \sum_{j\neq i} x_j'^{t+l}(\cdot)}{\partial \sum_{j\neq i} E_j}\\
&\quad \vdots\\
&\quad +(T-t)\tfrac{\delta_i}{2(1-\delta_i)}[\tfrac{A}{N} + (\tfrac{A}{N}-1)\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j}](\alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j})^{T-(t+1)} \left(\sum_{j\neq i} x_j'^{t+l}(\cdot)\right)^{T-t-1} \tfrac{\partial \sum_{j\neq i} x_j'^{t+l}(\cdot)}{\partial \sum_{j\neq i} E_j}, \qquad (4.45)
\end{aligned}
$$

By looking at (4.38) we know that $\frac{\partial \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})}{\partial \sum_{j\neq i} E_j} < 0$ since L2 players have the same first- and second-order belief and hence $X'^{M}_{II} < 0$. Taking the derivative of (4.27) with respect to $\sum_{j\neq i} E_j$ shows that $\frac{\partial \sum_{j\neq i} x_j'^{t+l}(\cdot)}{\partial \sum_{j\neq i} E_j} < 0$ which implies $X'^{OC}_{II} < 0$. Therefore, if the RHS in (4.40)> 0, then it holds that the RHS in (4.43)< 0. The marginal effect on myopic contributions in (4.43) is negative and on (over-) contributions it is negative if the marginal effect on (over-) contributions is positive in (4.40).

For the last period T, we can use (4.11) to derive the corresponding marginal effects.

$$
\frac{\partial x_{iL2}^{*T}}{\partial \sum_{j\neq i} E_j} = -\alpha_i \tfrac{E_i}{(\sum_{j\neq i} E_j)^2} \sum_{j\neq i} x_j^T(\cdot, x_i^{T-1}) + \alpha_i \tfrac{E_i}{\sum_{j\neq i} E_j} \tfrac{\partial \sum_{j\neq i} x_j^T(\cdot, x_i^{T-1})}{\partial \sum_{j\neq i} E_j}, \qquad (4.46)
$$

where the RHS in (4.46) is negative since $\frac{\partial \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})}{\partial \sum_{j\neq i} E_j} < 0$ because L2 players have the same first- and second-order belief. ∎

Proof. [Proposition 4.10] We start with L1 players. Taking the derivative with respect to $\frac{A}{N}$ in (4.26) yields $\frac{\partial x_{iL1}^{*t}}{\partial (A/N)} = \frac{\delta_i}{2(1-\delta_i)} > 0$. How exactly the derivative with respect to $\frac{A}{N}$ looks like for L2 players in period $t \neq T$ depends on whether there are planned corner solutions in later periods. (4.21) determines optimal contributions if there are no planned corner solutions in later periods, (4.22) when the next period is no corner solution planned but at some later period, and (4.23) when the next period already yields a corner solution. It is straightforward to check that marginally increasing $\frac{A}{N}$ yields a positive effect on myopic contributions in (4.21), (4.22), and (4.23). Therefore the effect on x_{iL2}^{*t} is clearly positive in all cases. Moreover, the effect on over-contribution is positive in (4.21), positive in (4.22) because $\frac{\partial X_i^{OCt+1}}{\partial (A/N)} \geq 0$, and weakly positive in (4.23) because the believed $\sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} \geq 0$ in this case. In the last period T, taking the derivative with respect to $\frac{A}{N}$ in (4.11) yields $\frac{\partial x_{iL2}^{*T}}{\partial (A/N)} = \frac{\delta_i}{2(1-\delta_i)} > 0$. ∎

Proof. [Proposition 4.11] In order to determine the effect of increasing N for L1 players, we can look at (4.26). If the number of players N increases and A increases in such a way as to hold the marginal per capita return from the public good $\frac{A}{N}$ constant, then there is no effect for L1 players in period t if $\frac{\sum_{j\neq i} x_j^{t-1}}{\sum_{j\neq i} E_j}$ remains constant. By looking at (4.11), the same holds true for L2 players' contributions in period T if the believed $\frac{\sum_{j\neq i} x_j^T(\cdot, x_i^{T-1})}{\sum_{j\neq i} E_j}$ stays constant. In period $t \neq T$ and given Assumption 4.1, we can use (4.21) to determine the effect of increasing N for L2 players. Again, their myopic contributions in period $t \neq T$ are not affected if the believed $\frac{\sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})}{\sum_{j\neq i} E_j}$ stays constant. This changes, however with respect to their over-contribution. Under heterogeneous endowments, the believed $\sum_{j\neq i} \frac{\partial x_j^{t+l}(\cdot)}{\partial x_i^{t+l-1}}$ may change with additional players depending on their endowments. Under homogeneous endowments, the believed $\sum_{j\neq i} \frac{\partial x_j^{t+l}(\cdot)}{\partial x_i^{t+l-1}}$ is constant so that increasing N has no effect if the believed preferences of others remain constant. Still, $\alpha_i \frac{E_i}{\sum_{j\neq i} E_j} = \alpha_i \frac{1}{N-1}$ in (4.21) decreases with N, so that the term $[\frac{A}{N} + (\frac{A}{N} - 1)\alpha_i \frac{E_i}{\sum_{j\neq i} E_j}]$ increases with N. So, even if there was under-contribution, under-contributions would decrease with N and hence contributions of L2 players would increase. Also, over-contributions increase with N, so L2 players contribute more in period $t \neq T$ if the number of players N increases. ■

Proof. [Proposition 4.12] By Lemma 4.3 we can observe the effect of increasing T in period $t \neq T$ by looking at (4.26) for L1 players and (4.21) for L2 players. Increasing T does not affect L1 players' contributions as they are completely myopic. L2 players' over-contribute and these over-contributions increase since the finite series of the believed sum of slopes of others believed best response functions is longer. In period T, L2 players (but not L1 players) also increase contributions as already with one more period left they will start to over-contribute. ■

Proof. [Proposition 4.13] With $T = 1$ we can use (4.26) for L1 players and (4.11) for L2 players in order to see what players contribute in a one-shot game. First, note that there is no over-contribution for L2 players, so both L1 and L2 players contribute according to their myopic contributions. For L1 players if $\sum_{j\neq i} x_j^0 = 0$, they would not contribute. Similarly, L2 players with believed $\sum_{j\neq i} x_j^1(\cdot, x_i^0) = 0$ would not contribute. L1 players with $\sum_{j\neq i} x_j^0 > 0$ would have $x_{iL1}^{*1} \geq 0$ and $x_{iL1}^{*1} > 0$ if δ_i is sufficiently small for a given value of α_i, E_i, $\sum_{j\neq i} E_j$, and $\frac{A}{N}$. L2 players with believed $\sum_{j\neq i} x_j^1(\cdot, x_i^0) > 0$ would also have $x_{iL2}^{*1} \geq 0$ and $x_{iL2}^{*1} > 0$ if δ_i is sufficiently small for a given value of α_i, E_i, $\sum_{j\neq i} E_j$, and $\frac{A}{N}$. Since L2 players react on believed $\sum_{j\neq i} x_j^1(\cdot, x_i^0)$, what they believe others will contribute depends additionally on the believed $\tilde{\delta}_j$ and $\tilde{\alpha}_j$ of L1 players. Since this belief is not necessarily correct, it can well happen that L2 players contribute more than L1 players (for a common believed relative 'home-grown' contribution of L0 players) despite they best respond and imperfectly match other L1 players. ■

Proof. [Proposition 4.14] By Lemma 4.3, we can use (4.21) to derive the effect of increased selfishness on L2 players' contribution behavior in period $t \neq T$, (4.11) to derive their effect in period T, and (4.26) to derive the effect for L1 players in period t. We start with L2 players in period $t \neq T$.

$$\begin{aligned}\frac{\partial x_{iL2}^{*t}}{\partial \delta_i} &= \frac{\left[\frac{A}{N} - 1 + (\frac{A}{N} + (\frac{A}{N} - 1)\alpha_i \frac{E_i}{\sum_{j\neq i} E_j}) \sum_{j\neq i} x'_j\right] 2(1-\delta_i)}{4(1-\delta_i)^2} \\ &\quad - \frac{\left[\delta_i(\frac{A}{N} - 1) + \delta_i(\frac{A}{N} + (\frac{A}{N} - 1)\alpha_i \frac{E_i}{\sum_{j\neq i} E_j}) \sum_{j\neq i} x'_j\right] (-2)}{4(1-\delta_i)^2} \\ &= \frac{(\frac{A}{N} - 1) + (\frac{A}{N} + (\frac{A}{N} - 1)\alpha_i \frac{E_i}{\sum_{j\neq i} E_j}) \sum_{j\neq i} x'_j}{2(1-\delta_i)^2}, \end{aligned} \tag{4.47}$$

where $\sum_{j\neq i} x'_j = \sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t}(1 + \alpha_i \frac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} \frac{\partial x_j^{t+2}(\cdot)}{\partial x_i^{t+1}}(1 + \alpha_i \frac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} \frac{\partial x_j^{t+3}(\cdot)}{\partial x_i^{t+2}}(\ldots(1 + \alpha_i \frac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} \frac{\partial x_j^{T}(\cdot)}{\partial x_i^{T-1}}))))$. The denominator in (4.47) is positive, the first term of the numerator is negative and the second term of the numerator is positive if and only if (4.10) holds strictly. So, if there was no over-contribution or even under-contribution, clearly more selfishness would decrease contributions. Now, suppose that (4.10) holds strictly. Then, more selfishness does not increase contributions if and only if

$$\sum_{j\neq i} x'_j \leq \frac{1 - \frac{A}{N}}{\frac{A}{N} - (1 - \frac{A}{N})\alpha_i \frac{E_i}{\sum_{j\neq i} E_j}}. \tag{4.48}$$

Since we already showed that second-order conditions are negative (in the proof of Proposition 4.1), the assumption that contributions in period t are not higher than under perfect relative conditional cooperation translates into the condition that the first-order condition specifying x_{iL2}^{*t} is weakly negative at $x_i^t = \frac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})$ (where $\sum_{j\neq i} x_j^t(\cdot, x_i^{t-1}) \equiv \int \sum_{j\neq i} x_j^t(\cdot)\, dF_i(\sum_{j\neq i} x_j^t(\cdot)))$, so that only a decrease of x_i^t can equalize the first-order condition.

$$\begin{aligned} &\left.\frac{\partial U_{iL2}^t}{\partial x_i^t}\right|_{x_i^t = \frac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1})} \\ = \ &\delta_i(\tfrac{A}{N} - 1) - 2(1-\delta_i)(1-\alpha_i)\frac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} x_j^t(\cdot, x_i^{t-1}) \\ &+ \delta_i \tfrac{A}{N} \sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t} - 2(1-\delta_i)(x_i^{t+1} - \alpha_i \frac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t}) \\ &(-\alpha_i \frac{E_i}{\sum_{j\neq i} E_j} \sum_{j\neq i} \frac{\partial x_j^{t+1}(\cdot)}{\partial x_i^t}) \\ \leq \ &0. \end{aligned} \tag{4.49}$$

Plugging in x^{*t+1}_{iL2} yields

$$\frac{\delta_i}{2(1-\delta_i)}\sum_{j\neq i}\frac{\partial x^{t+1}_j(\cdot)}{\partial x^t_i} \leq \frac{\frac{\delta_i}{2(1-\delta_i)}(1-\frac{A}{N})+\frac{E_i}{\sum_{j\neq i}E_j}(1-\alpha_i)\sum_{j\neq i}x^t_j(\cdot,x^{t-1}_i)}{(\frac{A}{N}+(\frac{A}{N}-1)\alpha_i\frac{E_i}{\sum_{j\neq i}E_j})(1+\alpha_i\frac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}\frac{\partial x^{t+2}_j(\cdot)}{\partial x^{t+1}_i}(\ldots(1+\alpha_i\frac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}\frac{\partial x^T_j(\cdot)}{\partial x^{T-1}_i})))} \tag{4.50}$$

and hence

$$\sum_{j\neq i}x'_j \leq \frac{1-\frac{A}{N}}{\frac{A}{N}-(1-\frac{A}{N})\alpha_i\frac{E_i}{\sum_{j\neq i}E_j}}+\frac{2(1-\delta_i)(1-\alpha_i)\frac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}x^t_j(\cdot,x^{t-1}_i)}{\delta_i(\frac{A}{N}-(1-\frac{A}{N})\alpha_i\frac{E_i}{\sum_{j\neq i}E_j})}. \tag{4.51}$$

The LHS in (4.51) is equal to the LHS in (4.48). The first term of the RHS in (4.51) is equal to the RHS in (4.48). The second term of the RHS in (4.51) is positive since the believed $\sum_{j\neq i}x^t_j(\cdot,x^{t-1}_i)>0$ (by Lemma 4.3). So, (4.48) implies (4.51) but not vice versa.

In the last period T,

$$\frac{\partial x^{*T}_{iL2}}{\partial\delta_i}=\frac{\frac{A}{N}-1}{2(1-\delta_i)}<0. \tag{4.52}$$

So, clearly more selfishness decreases L2 players' contributions in the last period.

For L1 players in period t, we also get

$$\frac{\partial x^{*t}_{iL1}}{\partial\delta_i}=\frac{\frac{A}{N}-1}{2(1-\delta_i)}<0, \tag{4.53}$$

so that more selfishness decreases L1 players' contributions in period t. ■

Proof. [Corollary 4.15] Contributions of L2 players in period $t\neq T$ are not higher than under full relative conditional cooperation if and only if the first-order condition specifying x^{*t}_{iL2} is weakly negative at $x^t_i=\alpha_i\frac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}x^t_j(\cdot,x^{t-1}_i)$ (where $\sum_{j\neq i}x^t_j(\cdot,x^{t-1}_i)\equiv\int\sum_{j\neq i}x^t_j(\cdot)\,dF(\sum_{j\neq i}x^t_j(\cdot))$), so that only a decrease of x^t_i can equalize the first-order condition (since the second-order conditions are negative as shown in the proof of Proposition 4.1).

$$\begin{aligned}\left.\frac{\partial U^t_{iL2}}{\partial x^t_i}\right|_{x^t_i=\alpha_i\frac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}x^t_j(\cdot,x^{t-1}_i)} &= \delta_i(\tfrac{A}{N}-1)+\delta_i\tfrac{A}{N}\sum_{j\neq i}\frac{\partial x^{t+1}_j(\cdot)}{\partial x^t_i}-2(1-\delta_i)\\ &\quad (x^{t+1}_i-\alpha_i\frac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}\frac{\partial x^{t+1}_j(\cdot)}{\partial x^t_i})(-\alpha_i\frac{E_i}{\sum_{j\neq i}E_j}\sum_{j\neq i}\frac{\partial x^{t+1}_j(\cdot)}{\partial x^t_i})\\ &\leq 0.\end{aligned} \tag{4.54}$$

Plugging in x_{iL2}^{*t+1} yields

$$\sum_{j \neq i} x'_j \leq \frac{1 - \frac{A}{N}}{\frac{A}{N} - (1 - \frac{A}{N})\alpha_i \frac{E_i}{\sum_{j \neq i} E_j}}. \tag{4.55}$$

(4.55) is equal to (4.48), so more selfishness decreases contributions if and only if contributions are lower than under full relative conditional cooperation. ■

Proof. [Proposition 4.16] By Assumption 4.1, we can look at (4.26) for L1 players, (4.21) for L2 players in period $t \neq T$, and (4.11) for L2 players in period T. Increasing the endowments of all players by the same factor $z > 1$ (weakly) increases L1 players' contributions in period t if and only if $\sum_{j \neq i} x_j^{t-1}$ (weakly) increases. This is because $\frac{zE_i}{z \sum_{j \neq i} E_j} = \frac{E_i}{\sum_{j \neq i} E_j}$. Similarly, L2 players' contributions in period t (weakly) increase if and only if the believed $\sum_{j \neq i} x_j^t(\cdot, x_i^{t-1})$ (weakly) increases because over-contributions stay constant as the believed $\frac{\partial x_j^{t+l}(\cdot)}{\partial x_i^{t+l-1}}$ is not affected.

In the first period, all players have the same 'home-grown' belief about L0 players' contributions, which is simply some percentage of their endowment. Increasing all endowments by some factor z means that absolute 'home-grown' contributions also increase by z. Best responding to that means that L1 players contribute more in the first period, but also L2 players who best respond to L1 players. These higher contributions in the first period shift players to a higher contribution path, so that $\sum_{j \neq i} x_j^{t-1}$ as well as the believed $\sum_{j \neq i} x_j^t(\cdot, x_i^{t-1})$ increases in all periods. ■

4.9 Appendix C: Instructions (translated from German)

General Instructions[79]

Thank you very much for your appearance. In the next 90 minutes you are going to take part in an experiment at the laboratory of MELESSA.
If you read the following instructions carefully, you can earn money depending on your decisions. The money earned during the experiment will be added to the show-up fee and paid out in cash at the end of the experiment.

During the experiment you are not allowed to communicate with the other participants. If you have any questions, please raise your arm. One of the experimenters will then come to your seat to answer your questions. You are not allowed to use your mobile phones or to open any programs on the computer. In the case of violation of these rules we have to exclude you from the experiment including all payments.

During the experiment we will refer to Experimental Monetary Units (EMU) and not to Euros. Your income will be calculated in EMU. At the end of the experiment, the total sum of earnings in EMU will be exchanged into Euros.

The exchange rate is: **1EMU = 10 Eurocents**.

In this experiment there are **two different types**: **L** and **H**. Type L and type H differ in their initial endowment. The total number of both types in the experiment is equal, hence there are as many L as there are H types.
At the beginning of the experiment you are randomly assigned a type. All participants have an equal chance to be of either type (L or H).
Then all participants are, again randomly, distributed into groups of three. None of the participants knows the identity of his or her group partners. Thus, all your decisions remain completely anonymous.

The Decision Situation:

We will explain the details of the experiment later. First, we want you to get familiar with the decision problem.

You are in a group together with two other participants (hence there are three group members in total). Every group member receives an initial endowment which depends on the type (L or H). Remember: the type is assigned randomly. Type L and type H have the following **initial endowments**.

[79] All participants; both types; both treatments.

Type L has an **initial endowment** of **4 EMU**.
Type H has an **initial endowment** of **8 EMU**.

All group members have to decide how many EMU of their initial endowment they want to transfer to a project. All EMU that are not transferred to the project are kept in private (by the respective participant). In the following we will describe how your total income is calculated.

Income from Private Property:
Every unit (EMU) of the initial endowment that has not been transferred to the project is kept in private property. For instance, if you are an L type (you have an initial endowment of 4 EMU) and decide to transfer 3 EMU to the project, 1 EMU will be kept as your private property. If you decide, for example, to transfer nothing (0 EMU) to the project, your private property will contain 4 EMU.

Income from the Project:
All group members benefit equally from the contributions made toward the project. This means that your income from the project is equal to that of the two other group members. Only the sum of contributions determines this income. Your income (and that of the two other group members) is calculated as follows.

Income from the project = Sum of all 3 contributions $*$ 0.6.

Example:
If you contribute 3 EMU to the project and the other group members contribute 1 EMU and 6 EMU, the sum of contributions is 10 EMU ($3 + 1 + 6 = 10$). Then, all group members receive 6 EMU ($10 * 0.6$) as their income from the project.

Total income:
Your total income is calculated as the sum of the income from private property and the income from the project:

Your private income (= Your endowment - Your contribution to the project)
+ Your income from the project (= 0.6 $*$ Sum of all contributions)
= Your total income.

The following **control questions** will help you to better understand the calculation of your total income.

Control Questions:

Please answer the following questions. They will help you to better understand the calculation of your income. Please fill in the solutions of all questions and raise your arm as soon as you have finished this task. Then, one of the experimenters will come to your seat in order to check your answers.

1. Assume that all three group members have an initial endowment of 4 EMU. Assume further that neither you nor any of the other group members has contributed anything to the project.
a) What will your total income be?____________
b) What will the total income of each of the other two group members be? ____________

2. Assume that all three group members have an initial endowment of 8 EMU. Assume further that you have contributed 8 EMU and the two other group members have also contributed 8 EMU.
a) What will your total income be? ____________
b) What will the total income of each of the other two group members be? ____________

3. Assume that you have an initial endowment of 4 EMU. The two other group members have an initial endowment of 8 EMU. The two other group members contribute 5 EMU each.
a) What will your total income be if you contribute nothing (0 EMU)? ____________
b) What will your total income be if you contribute 2 EMU?____________
c) What will your total income be if you contribute 4 EMU?____________

4. Assume that all group members have an initial endowment of 4 EMU. Assume further that you contribute 2 EMU.
a) What will your total income be if the two other group members contribute the sum of 6 EMU?____________
b) What will your total income be if the two other group members contribute the sum of 8 EMU?____________
c) What will your total income be if the two other group members contribute the sum of 2 EMU?____________

After you have answered all four control questions or in case you have any other questions, please raise your arm. Then, one of the experimenters will check your answers of the control questions and will answer further questions.

Instructions Part 1[80]

The Experiment

The experiment includes the decision situations that you have just become acquainted with. At the end of the experiment you will receive a payment according to your decisions.

[80] P-Experiment; H type; both treatments.

The experiment consists of **two parts**. These two parts are completely independent from each other. Your decisions in the first part of the experiment have no influence on the process or the income of the second part of the experiment, and vice versa. During the experiment you have to confirm your decisions by pressing the "OK" button. Please note that decisions that have been confirmed are final and cannot be reversed.

You have been randomly assigned the type H. This means you are **type H for the entire duration of the experiment** and you can decide over the **initial endowment** of **8 EMU**.

The **procedure** of the experiment is the following:

1. Please read the instructions for the first part of the experiment carefully and put them on your desk in front of you (up-side down) when you have finished reading.
2. Next, you will exercise Part 1 of the experiment on the computer.
3. As soon as all participants have completed Part 1, you will receive the instructions for Part 2 of the experiment. Again, you should read them carefully and put them up-side down on your desk when you have understood the set-up and in case you have no further questions.
4. Then, you will complete Part 2 of the experiment on the computer.
5. As soon as all participants have finished Part 2, you will be asked to complete a questionnaire on the computer. Your answers to this questionnaire have no influence on your income.
6. At the end of the experiment you will get information about your income of both parts of the experiment.

In the following we describe the first part of the experiment in detail.

Instructions Part 1

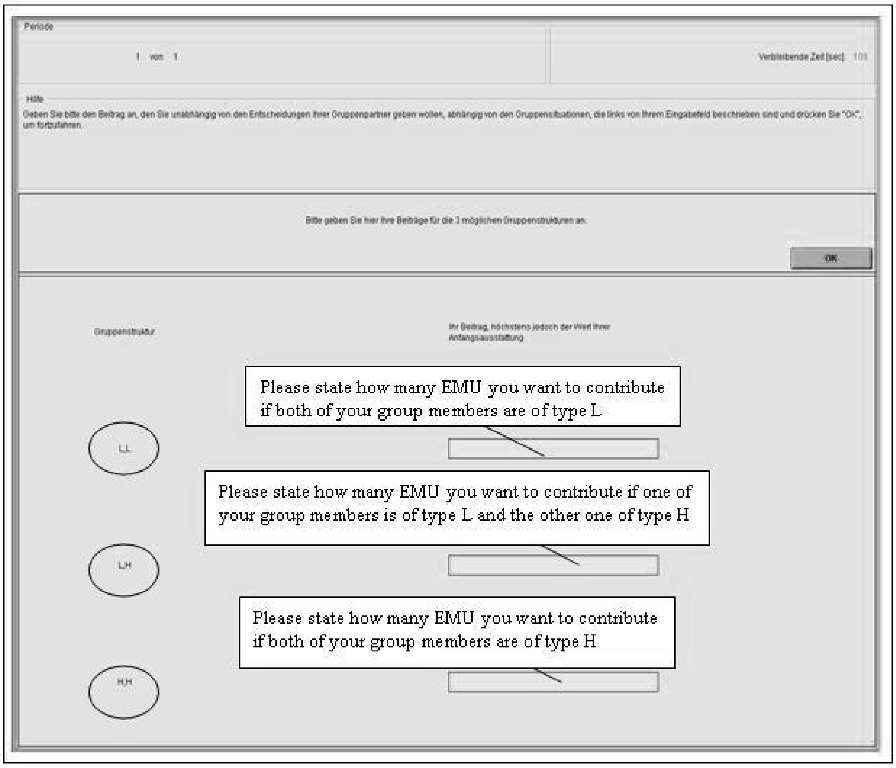
Since you do not know the types (L or H) of the other two members of your group, you may find yourself in one of three possible situations. The two other group members may be both of type L, they may both be of type H or one group member may be a L type and the other may be an H type.

The first part of the experiment consists of **two stages**:

Stage 1: unconditional decisions
Stage 2: conditional decisions

1. Unconditional Decisions:

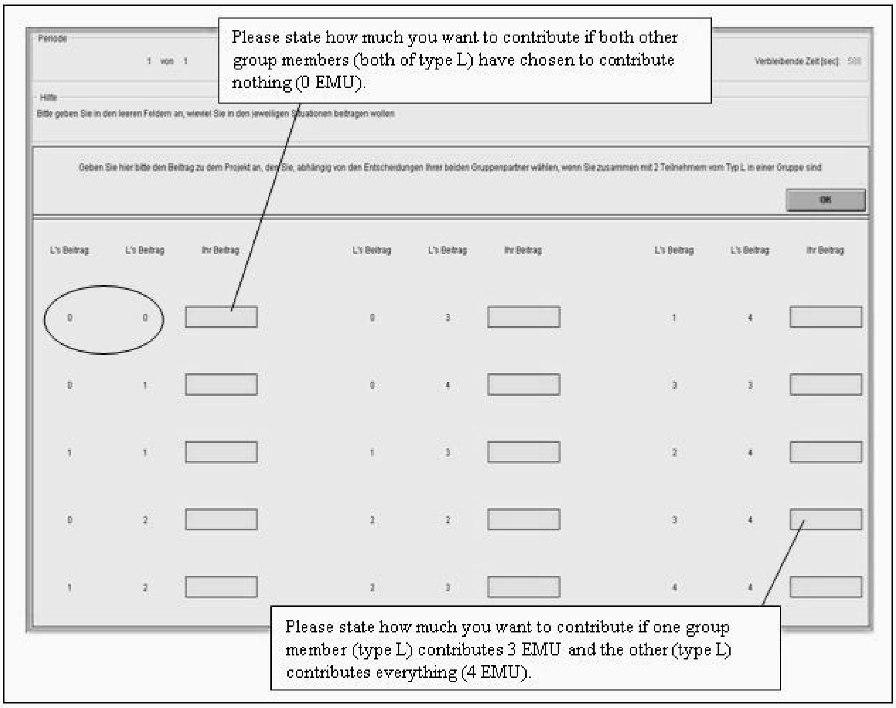
Here you decide about your contribution to the project **independently** from the decisions of the other group members. In this table you have to state your contribution to the project for all three possible group structures in which you may find yourself (see above). Since all participants of the experiment decide simultaneously, neither you nor the two other participants in your group know the level of contributions of the other group members.



2. Conditional Decisions:

Here you decide about your contribution, **conditional** on the decisions of your group members. To do this you have to fill in the following three tables. For each of the three possible group structures you get one table. The rows in the tables tell you what the contributions of your group members are. You have to state in each row and column how much you want to contribute given the contributions of the other group members. This means you know how much the other group members have contributed before you decide about your own contributions in the different situations.

In the following example you see the table for the case where both other group members are of type L:



Two **Examples**:

1. In the first situation (upper left side of the screenshot) the two other group members (in this table both are of type L) have decided to contribute 0 EMU. Hence you have to decide how many EMU you want to contribute to the project (between 0 and 8, since you are of type H) and enter the amount into the respective cell.

2. In the second situation (lower right side) one of the group members has contributed 3 EMU, and the other has contributed 4 EMU. Again, you have to decide about your own contribution and enter the number into the respective cell.

These decisions have to be made for all presented situations.

Note that every single situation may be relevant for your income.

After all the participants have made all the decisions, conditional and unconditional, they will be allocated randomly into groups of three. As mentioned above, you may be matched with two group members of type L, both group members of type H, or one group member of type L and the other one of type H.
Then, one member of the group is chosen randomly. For this participant, the three tables of the conditional contribution decisions are relevant (which of these three tables is relevant, depends on the randomly determined group structure). Which particular situation (entry) of the respective table is in fact relevant, depends now on the unconditional contributions of the other two group members for the randomly determined group structure.

So, the income of all three group members depends on the conditional contribution of the randomly chosen participant and on the unconditional contributions of the other two group members.
At the time when you make your decisions, you will not know, whether you will be one of the participants for whom the unconditional contribution is relevant or one for whom the conditional contributions are relevant. Therefore you should think carefully about every single decision (in both stages), since every one of them may determine your income (of this part of the experiment).

For an illustration consider the following **example**:

Assume you are randomly allocated to a group in which the other group members are of type L. Assume further that you are randomly chosen to be the participant for whom the conditional contribution is relevant. Consequently, the unconditional contribution is relevant for the other two group members. Assume that the unconditional contributions of the other two group members (both type L) are 2 EMU and 3 EMU for the relevant situation (for both, the situation is relevant in which one group member is of type L and the other is of type H). For yourself, the table for the group structure with two type L participants is relevant. Your income is now determined by the decision on your contribution in case one group member has contributed 2 EMU and the other has contributed 3 EMU. Assume you have chosen to contribute 5 EMU in this situation.
In this case, the sum of contributions is $2+3+5 = 10$ EMU. Hence, all group members receive $0.6 * 10 = 6$ EMU from the project. In total you receive 3 EMU from your private property ($8 - 5 = 3$ EMU) and 6 EMU from the project leading to the sum of 9 EMU.
The group member (type L) who has contributed 2 EMU receives 2 EMU from his/her private property ($4 - 2 = 2$ EMU) and 6 EMU from the project and hence has a total income of 8 EMU.
The other group member (type L), who has contributed 3 EMU, receives 1 EMU from his/her private property ($4 - 3 = 1$ EMU) and 6 EMU from the project, making it a total income of 7 EMU.

If you have any further questions, please raise your arm. An experimenter will then come to your seat in order to answer the questions. If you have no further questions, please put the instructions up-side down on your desk.

Instructions Part 2[81]

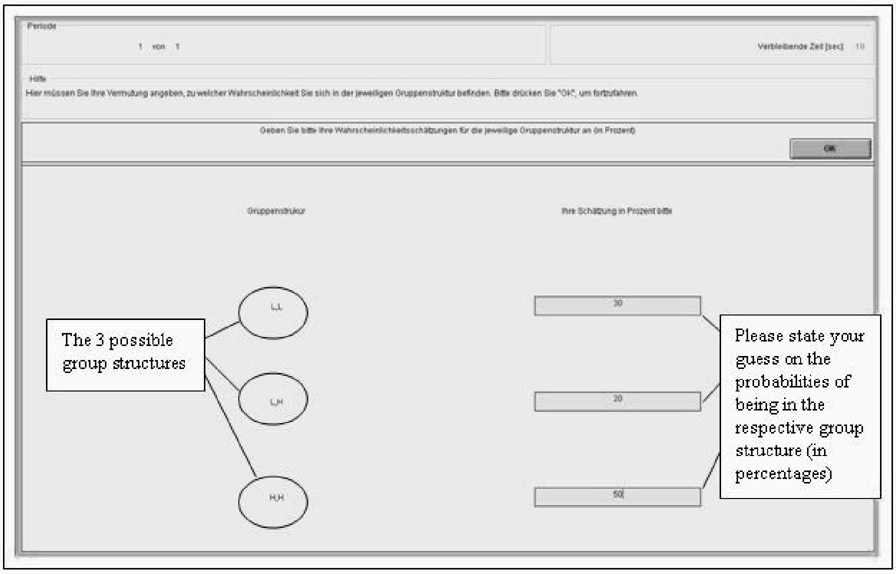
In the second part of the experiment, the now familiar decision situation is repeated 10 times (structured in 10 periods).

[81] C-Experiment; H type; uncertainty treatment

You are still **type H** and have an initial endowment of **8 EMU** (you will receive these 8 EMU in each period).
You are still together with the two other group members from the first part of the experiment. You still do not know the types of the other group members. You will stay together with these two participants over all 10 periods.
The decision situation remains unchanged. You decide in each period about the allocation of your initial endowment of 8 EMU. The calculation of your income remains also unchanged and is repeated in the following:

$$\begin{aligned} &\text{Private income } (= 8 \text{ - contribution to the project}) \\ + &\text{ Income from the project } (= 0.6 * \text{Sum of all contributions}) \\ = &\text{ Total income.} \end{aligned}$$

At the beginning of each period, we will ask you to indicate your expectations about the initial endowments of the other two group members. So, you are required to make a guess about the type of the other participants in your group. There are three possible compositions of your group: you are together with two participants of type L, with two participants of type H, or with one participant of type H and one of type L. Please allocate probabilities (in percentage points) to each of these group structures.

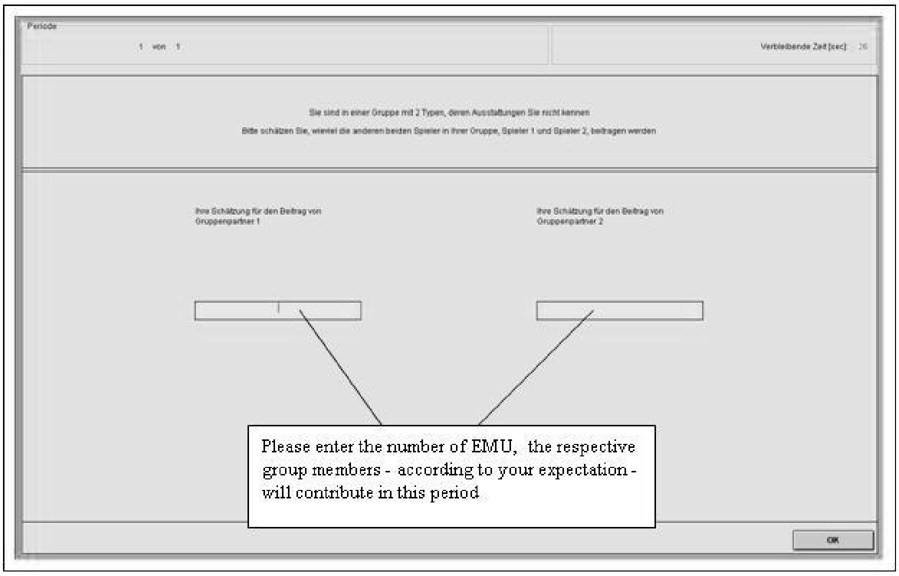


For example, if you are 100% sure that you are in a group in which both other group members are of type L, you will fill in 100 in the first row and 0 in each of the other two rows.
If you rather think the chances are 50% to be in a group in which both other group members are of type L, 25% to be in a group in which one group member is of type L and the other is of type H, and 25% to be in a group with two participants of type H, you will need to fill in 50 in the first, 25 in the second and 25 in the third row.

Note that the probabilities of your expectations (the numbers you fill in) have to add up to 100.

If you believe that the probability to be in each of the three group structures is equal, please fill in 34, 33, and 33 in the three rows.

Moreover, you are asked how many EMU you think the other two group members will contribute in the respective period. So, you have to guess how many EMU each of your group members will contribute.



You will be paid for the accuracy of this guess in the following way:

- If your expectation is completely correct, you will receive 2 EMU (for each group member).
- If your expectation deviates by not more than 1 EMU from the real value, you will receive 1 EMU (for each group member).
- If your expectation deviates by 2 EMU or more, you will receive nothing.

In a last step you have to decide how many EMU of your initial endowment of 8 EMU (which you receive at the beginning of each of the 10 periods) you actually want to contribute to the project in the respective period.

As soon as all group members have stated their decisions, you will be shown a table that states how many EMU all other group members have contributed and how many EMU you have earned in the respective period (excluding the income from correct expectations about the group structure). Then, a new period starts.

All group members get the same information about the contributions of the other group members.
None of the group members receives information about the type (L or H) of any of the other group members.

The income of all periods is summed up. After the 10^{th} period you receive the payment in private and separately from other participants.

The **final income** then consists of the following:

1. Show-up fee of 4 Euros.
2. The income from the first part of the experiment.
3. The sum of incomes from the 10 periods of the second part of the experiment, including the payments for correct expectations.

If you have any questions, please raise your arm. An experimenter will come to your seat in order to answer your questions. If you do not have any further questions, please put the instructions up-side down on your desk, so that we can start with the second part of the experiment on the computer.

Bibliography

[1] Abeler, Johannes, Armin Falk, Lorenz Götte, and David Huffman (2011). "Reference Points and Effort Provision", *American Economic Review* 101(2), pp470-492.

[2] Altmann, Steffen, Thomas Dohmen, and Matthias Wibral (2008). "Do the Reciprocal Trust Less?", *Economics Letters* 99(3), pp454-457.

[3] Ambrus, Attila, and Parag A. Pathak (2011). "Cooperation Over Finite Horizons: A Theory and Experiments", *Journal of Public Economics* 95(7/8), pp500-512.

[4] Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and E. Elisabet Rutström (2006). "Elicitation Using Multiple Price List Formats", *Experimental Economics* 9(4), pp383-405.

[5] Anderson, Lisa R., Jennifer M. Mellor, and Jeffrey Milyo (2008). "Inequality and Public Good Provision: An Experimental Analysis", *Journal of Socio-Economics* 37(3), pp1010-1028.

[6] Andreoni, James (1988). "Why Free Ride? Strategies and Learning in Public Goods Experiments", *Journal of Public Economics* 37(3), pp291-304.

[7] Andreoni, James (1995a). "Cooperation in Public-Goods Experiments: Kindness or Confusion", *American Economic Review* 85(4), pp891-904.

[8] Andreoni, James (1995b). "Warm Glow Versus Cold Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments", *Quarterly Journal of Economics* 105(1), pp1-21.

[9] Andreoni, James, and Rachel Croson (2008). "Partners Versus Strangers: Random Matching in Public Goods Experiments", in C.R. Plott and V.L. Smith, eds., 'Handbook of Experimental Economics Results', Vol.1, North-Holland, Amsterdam, pp776-783.

[10] Andreoni, James, and Charles Sprenger (2009). "Certain and Uncertain Utility: The Allais Paradox and Five Decision Theory Phenomena", Working Paper, University of California, San Diego.

[11] Andreoni, James, and Charles Sprenger (2010). "Uncertainty Equivalents: Linear Tests of the Independence Axiom", Working Paper, University of California, San Diego.

[12] Aquino, Karl, Victoria Steisel, and Avi Kay (1992). "The Effects of Resource Distribution, Voice, and Decision Framing on the Provision of Public Goods", *Journal of Conflict Resolution* 36(4), pp665-687.

[13] Aumann, Robert J., and Roberto Serrano (2008). "An Economic Index of Riskiness", *Journal of Political Economy* 116(5), pp810-836.

[14] Aurélie, Dariel, and Arno Riedl (2010). "Different Social Preferences, Similar Selfish Actions: The Case of Public Goods Game and Gift Exchange Game", mimeo.

[15] Ballinger, T. Parker, Michael G. Palumbo, and Nathaniel T. Wilcox (2003). "Precautionary Saving and Social Learning across Generations: An Experiment", *Economic Journal* 113(7), pp920-947.

[16] Ballinger, T. Parker, and Nathaniel T. Wilcox (1997). "Decisions, Error and Heterogeneity", *Economic Journal* 107(4), pp1090-1105.

[17] Bar-Hillel, Maya (1973). "On the Subjective Probability of Compound Events", *Organizational Behavior and Human Performance* 9(3), pp396-406.

[18] Bell, David E. (1982). "Regret in Decision Making under Uncertainty", *Operations Research* 30(5), pp961-981.

[19] Bell, David E. (1983). "Risk Premiums for Decision Regret", *Management Science* 29(10), pp1156-1166.

[20] Bell, David E. (1985). "Disappointment in Decision Making under Uncertainty", *Operations Research* 33(1), pp1-27.

[21] Bernasconi, Michele (1994). "Nonlinear Preferences and Two-Stage Lotteries: Theories and Evidence", *Economic Journal* 104(1), pp54-70.

[22] Bernheim, B. Douglas (1984). “Rationalizable Strategic Behavior”, *Econometrica* 52(4), pp1007-1028.

[23] Binswanger, Hans P. (1980). “Attitudes Toward Risk: Experimental Measurement in Rural India”, *American Journal of Agricultural Economics* 62(3), pp395-407.

[24] Binswanger, Hans P. (1981). “Attitudes Toward Risk: Theoretical Implications of an Experiment in Rural India”, *Economic Journal* 91(4), pp867-890.

[25] Blanco, Marianna, Dirk Engelmann, and Hans Theo Normann (2011). “A Within-Subject Analysis of Other-Regarding Preferences”, *Games and Economic Behavior* 72(2), pp321-338.

[26] Bleichrodt, Han, Alessandra Cillo, and Enrico Diecidue (2010). “A Quantitative Measurement of Regret Theory”, *Management Science* 56(1), pp161-175.

[27] Bolton, Gary E., and Axel Ockenfels (2000). “ERC: A Theory of Equity, Reciprocity, and Competition”, *American Economic Review* 90(1), pp166-193.

[28] Bombardini, Matilde, and Francesco Trebbi (2010). “Risk Aversion and Expected Utility Theory: An Experiment with Large and Small Stakes”, Working Paper, University of British Columbia.

[29] Bosch-Domènech, Antoni, José G. Montalvo, Rosemarie Nagel, and Albert Satorra (2002). “One, Two, (Three), Infinity, . . . : Newspaper and Lab Beauty-Contest Experiments”, *American Economic Review* 92(5), pp1687-1701.

[30] Bosch-Domènech, Antoni, and Joaquim Silvestre (2006). “Reflections on Gains and Losses: A $2 \times 2 \times 7$ Experiment”, *Journal of Risk and Uncertainty* 33(3), pp217-235.

[31] Botelho, Anabela, Glenn W. Harrison, Ligia M. Costa Pinto, and Elisabet E. Rutström (2009). “Testing Static Game Theory with Dynamic Experiments: A Case Study of Public Goods”, *Games and Economic Behavior* 67(1), pp253-265.

[32] Bramoullé, Yann, and Nicolas Treich (2009). “Can Uncertainty Alleviate the Commons Problem”, *Journal of the European Economic Association* 7(5), pp1042-1067.

[33] Brandts, Jordi, and Arthur Schram (2001). "Cooperation and Noise in Public Goods Experiments: Applying the Contribution Function Approach", *Journal of Public Economics* 79(2), pp399-427.

[34] Brockett, Patrick L., and Linda L. Golden (1987). "A Class of Utility Functions Containing all the Common Utility Functions", *Management Science* 33(8), pp955-964.

[35] Brown, Alexander L., Colin F. Camerer, and Dan Lovallo (2011). "To Review or Not to Review? Limited Strategic Thinking at the Movie Box Office", Working Paper, Texas A&M University.

[36] Browning, Martin, and Annamaria Lusardi (1996). "Household Saving: Micro Theories and Micro Facts", *Journal of Economic Literature* 34(4), pp1797-1855.

[37] Bruner, David (2007). "Multiple Switching Behavior in Multiple Price Lists", Working Paper, Appalachian State University.

[38] Bruner, David (2009). "Changing the Probability versus Changing the Reward", *Experimental Economics* 12(4), pp367-385.

[39] Bruner, David, Michael McKee, and Rudy Santore (2008). "Hand in the Cookie Jar: An Experimental Investigation of Equity-Based Compensation and Managerial Fraud", *Southern Economic Journal* 75(1), pp261-278.

[40] Brunnermeier, Markus K., and Stefan Nagel (2008). "Do Wealth Fluctuations Generate Time-Varying Risk Aversion? Micro-Evidence on Individuals' Asset Allocation", *American Economic Review* 98(3), pp713-736.

[41] Buckley, Edward, and Rachel Croson (2006). "Income and Wealth Heterogeneity in the Voluntary Provision of Linear Public Goods", *Journal of Public Economics* 90(4/5), pp935-955.

[42] Burlando, Roberto, and John D. Hey (1997). "Do Anglo-Saxons Free-Ride More?", *Journal of Public Economics* 64(1), pp41-60.

[43] Caballé, Jordi, and Alexey Pomansky (1996). "Mixed Risk Aversion", *Journal of Economic Theory* 71(2), pp485-513.

[44] Calvet, Laurent E., and Paolo Sodini (2010). "Twin Picks: Disentangling the Determinants of Risk-Taking in Household Portfolios", NBER Working Paper, No. w15859.

[45] Camerer, Colin F. (1989). "An Experimental Test of Several Generalized Expected Utility Theories", *Journal of Risk and Uncertainty* 2(1), pp61-104.

[46] Camerer, Colin F. (2003). "Behavioral Game Theory: Experiments in Strategic Interaction", Princeton University Press, Princeton, NJ.

[47] Camerer, Colin F., Linda Babcock, George Loewenstein, Richard H. Thaler (1997). "Labor Supply of New York City Cabdrivers: One Day at a Time", *Quarterly Journal of Economics* 112(2), pp407-441.

[48] Camerer, Colin, Teck-Hua Ho, and Juin Kuan Chong (2004). "A Cognitive Hierarchy Model of Games", *Quarterly Journal of Economics* 119(3), pp861-898.

[49] Camille, Nathalie, Giorgio Coricelli, Jerome Sallet, Pascale Pradat-Diehl, Jean-René Duhamel, and Angela Sirigu (2004). "The Involvement of the Orbitofrontal Cortex in the Experience of Regret", *Science* 304, pp1167-1170.

[50] Card, David, and Gordon B. Dahl (2011). "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior", *Quarterly Journal of Economics* 126(1), pp103-143.

[51] Cardenas, Juan-Camilo (2003). "Real Wealth and Experimental Cooperation: Experiments in the Field Lab", *Journal of Development Economics* 70(2), pp263-289.

[52] Carroll, Christopher D., and Miles S. Kimball (2008). "Precautionary Saving and Precautionary Wealth", in Steven N. Durlauf, and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*, 2nd ed., vol. 6, New York: Palgrave MacMillan, pp576-585.

[53] Chan, Kenneth S., Stuart Mestelman, Robert Moir, and R. Andrew Muller (1999). "Heterogeneity and the Voluntary Provision of Public Goods", *Experimental Economics* 2(1), pp5-30.

[54] Charness, Gary, and Matthew Rabin (2002). "Understanding Social Preferences with Simple Tests", *Quarterly Journal of Economics* 117(3), pp817-869.

[55] Chen, Yan, and Sherry Xin Li (2009). "Group Identity and Social Preferences", *American Economic Review* 99(1), pp431-457.

[56] Cherry, Todd L., Stephan Kroll, and Jason F. Shogren (2005). "The Impact of Endowment Heterogeneity and Origin on Public Good Contributions: Evidence

from the Lab", *Journal of Economic Behavior and Organization* 57(3), pp357-365.

[57] Chetty, Raj (2006). "A New Method of Estimating Risk Aversion", *American Economic Review* 96(5), pp1821-1834.

[58] Chew, Soo Hong, Edi Karni, and Zvi Safra (1987). "Risk Aversion in the Theory of Expected Utility with Rank Dependent Probabilities", *Journal of Economic Theory* 42(2), pp370-381.

[59] Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv (2007). "Consistency and Heterogeneity of Individual Behavior under Uncertainty", *American Economic Review* 97(5), pp1921-1938.

[60] Coricelli, Giorgio, Hugo D. Critchley, Mateus Joffily, John P. O'Doherty, Angela Sirigu, and Raymond J. Dolan (2005). "Regret and its Avoidance: A Neuroimaging Study of Choice Behavior", *Nature Neuroscience* 8, pp1255-1262.

[61] Costa-Gomes, Miguel A., and Vincent P. Crawford (2006). "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study", *American Economic Review* 96(5), pp1737-1768.

[62] Costa-Gomes, Miguel A., Vincent P. Crawford, and Bruno Broseta (2001). "Cognition and Behavior in Normal-Form Games: An Experimental Study", *Econometrica* 69(5), pp1193-1235.

[63] Costa-Gomes, Miguel A., Vincent P. Crawford, and Nagore Iriberri (2009). "Comparing Models of Strategic Thinking in Van Huyck, Battalio, and Beil's Coordination Games", *Journal of the European Economic Association* 7(2-3), pp365-376.

[64] Costa-Gomes, Miguel A., and Georg Weizsäcker (2008). "Stated Beliefs and Play in Normal-Form Games", *Review of Economic Studies* 75(3), pp729-762.

[65] Courbage, Christophe, and Béatrice Rey (2006). "Prudence and Optimal Prevention for Health Risks", *Health Economics* 15(12), pp1323-1327.

[66] Cox, James C., and Vjollca Sadiraj (2007). "On Modeling Voluntary Contributions to Public Goods", *Public Finance Review* 35(2), pp311-332.

[67] Crawford, Vincent P., Miguel A. Costa-Gomes, and Nagore Iriberri (2010). "Strategic Thinking", Working Paper, University of California, San Diego.

[68] Crawford, Vincent P., and Juanjuan Meng (2011). "New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income", *American Economic Review* 101(5), pp1912-1932.

[69] Crawford, Vincent P., and Nagore Iriberri (2007a). "Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?", *Econometrica* 75(6), pp1721-1770.

[70] Crawford, Vincent P., and Nagore Iriberri (2007b). "Fatal Attraction: Salience, Naivete, and Sophistication in Experimental 'Hide-and-Seek' Games", *American Economic Review* 97(5), pp1731-1750.

[71] Crawford, Vincent P., Tamar Kugler, Zvika Neeman, and Ady Pauzner (2009). "Behaviorally Optimal Auction Design: An Example and Some Observations", *Journal of the European Economic Association* 7(2/3), pp377-387.

[72] Croson, Rachel (1996). "Partners and Strangers Revisited", *Economics Letters* 53(1), pp25-32.

[73] Croson, Rachel (2007). "Theories of Commitment, Altruism, and Reciprocity: Evidence from Linear Public Goods Games", *Economic Inquiry* 45(2), pp199-216.

[74] Croson, Rachel, Enrique Fatas, and Tibor Neugebauer (2005). "Reciprocity, Matching and Conditional Cooperation in Two Public Goods Games", *Economics Letters* 87(1), pp95-101.

[75] Dawes, Robyn M., and Richard H. Thaler (1988). "Anomalies: Cooperation", *Journal of Economic Perspectives* 2(3), pp187-197.

[76] Deck, Cary, and Harris Schlesinger (2010). "Exploring Higher Order Risk Effects", *Review of Economic Studies* 77(4), pp1403-1420.

[77] Delquié, Philippe, and Alessandra Cillo (2006). "Disappointment without Prior Expectation: A Unifying Perspective on Decision under Risk", *Journal of Risk and Uncertainty* 33(3), pp197-215.

[78] Diamond, Peter A., and Joseph E. Stiglitz (1974). "Increases in Risk and in Risk Aversion", *Journal of Economic Theory* 8(3), pp337-360.

[79] Dionne, Georges, and Louis Eeckhoudt (1984). "Insurance and Saving: Some Further Results", *Insurance: Mathematics and Economics* 3(2), pp101-110.

[80] Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2010). "Are Risk Aversion and Impatience Related to Cognitive Ability?", *American Economic Review* 100(2), pp1238-1260.

[81] Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner (2011). "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences", *Journal of the European Economic Association* 9(3), pp522-550.

[82] Doran, Kirk Bennett (2009). "Reference Points, Expectations, and Heterogeneous Daily Labor Supply", Working Paper, University of Notre Dame.

[83] Dynan, Karen E. (1993). "How Prudent are Consumers?", *Journal of Political Economy* 101(6), pp1104-1113.

[84] Ebert, Sebastian, and Daniel Wiesen (2011). "Testing for Prudence and Skewness Seeking", *Management Science* 57(7), pp1334-1349.

[85] Eeckhoudt, Louis, and Christian Gollier (2005). "The Impact of Prudence on Optimal Prevention", *Economic Theory* 26(4), pp989-994.

[86] Eeckhoudt, Louis, Christian Gollier, and Harris Schlesinger (1995). "The Risk-Averse (and Prudent) Newsboy", *Management Science* 41(5), pp786-794.

[87] Eeckhoudt, Louis, Christian Gollier, and Harris Schlesinger (1996). "Changes in Background Risk and Risk Taking Behavior", *Econometrica* 64(3), pp683-689.

[88] Eeckhoudt, Louis, and Harris Schlesinger (2006). "Putting Risk in its Proper Place", *American Economic Review* 96(1), pp280-289.

[89] Eeckhoudt, Louis, and Harris Schlesinger (2008). "Changes in Risk and the Demand for Saving", *Journal of Monetary Economics* 55(7), pp1329-1336.

[90] Eckel, Catherine C., and Philip J. Grossman (2008). "Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices", *Journal of Economic Behavior and Organization* 68(1), pp1-17.

[91] Ekern, Steinar (1980). "Increasing N^{th} Degree Risk", *Economics Letters* 6(4), pp329-333.

[92] Ericson, Keith M. Marzilli, and Andreas Fuster (2010). "Expectations as Endowments: Reference-Dependent Preferences and Exchange Behavior", Working Paper, Harvard University.

[93] Eső, Péter, and Lucy White (2004). "Precautionary Bidding in Auctions", *Econometrica* 72(1), pp77-92.

[94] Fagart, Marie-Cécile, and Bernard Sinclair-Desgagné (2007). "Ranking Contingent Monitoring Systems", *Management Science* 53(9), pp1501-1509.

[95] Farber, Henry S. (2005). "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers", *Journal of Political Economy* 113(1), pp46-82.

[96] Farber, Henry S. (2008). "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers", *American Economic Review*, 98(3), pp1069-1082.

[97] Fehr, Ernst, and Klaus M. Schmidt (1999). "A Theory of Fairness, Competition, and Cooperation", *Quarterly Journal of Economics* 114(3), pp817-868.

[98] Fehr-Duda, Helga, Adrian Bruhin, Thomas Epper, and Renate Schubert (2010). "Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size", *Journal of Risk and Uncertainty* 40(2), pp147-180.

[99] Fei, Wenan, and Harris Schlesinger (2008). "Precautionary Insurance Demand with State-Dependent Background Risk", *Journal of Risk and Insurance* 75(1), pp1-16.

[100] Fischbacher, Urs (2007). "Z-Tree: Zurich Toolbox for Ready-Made Economics Experiments", *Experimental Economics* 10(2), pp171-178.

[101] Fischbacher, Urs, and Simon Gächter (2010). "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments", *American Economic Review* 100(1), pp541-556.

[102] Fischbacher, Urs, Simon Gächter, and Ernst Fehr (2001). "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment", *Economics Letters* 71(3), pp397-404.

[103] Friedman, Zachary G. (2005). "Testing the Reduction of Compound Lotteries Axiom: Violations in Decision Theory", Working Paper, Washington University, St. Louis.

[104] Friend, Irwin, and Marshall E. Blume (1975). "The Demand for Risky Assets", *American Economic Review* 65(5), pp900-922.

[105] Gächter, Simon (2007). "Conditional Cooperation: Behavioral Regularities from the Lab and the Field and their Policy Implications", in B.S. Frey and A. Stutzer, eds., 'Economics and Psychology. A Promising New Cross-Disciplinary Field', MIT Press, Cambridge, MA, pp19-50.

[106] Gächter, Simon, Daniele Nosenzo, Elke Renner, and Martin Sefton (forthcoming). "Who Makes a Good Leader? Cooperativeness, Optimism, and Leading-By-Example", *Economic Inquiry.*

[107] Gill, David, and Victoria Prowse (forthcoming). "A Structural Analysis of Disappointment Aversion in a Real Effort Competition", *American Economic Review.*

[108] Gollier, Christian (2003). "To Insure or Not to Insure?: An Insurance Puzzle", *Geneva Papers on Risk and Insurance Theory* 28(1), pp5-24.

[109] Gollier, Christian (2010). "Ecological Discounting", *Journal of Economic Theory* 145(2), pp812-829.

[110] Gollier, Christian, Bruno Jullien, and Nicolas Treich (2000). "Scientific Progress and Irreversibility: An Economic Interpretation of the 'Precautionary Principle' ", *Journal of Public Economics* 75(2), pp229-253.

[111] Gollier, Christian, and John W. Pratt (1996). "Risk Vulnerability and the Tempering Effect of Background Risk", *Econometrica* 64(5), pp1109-1123.

[112] Gollier, Christian, and Nicolas Treich (2003). "Scientific progress and irreversibility: an economic interpretation of the 'Precautionary Principle' ", *Journal of Risk and Uncertainty* 27(1), pp77-103.

[113] Greiner, Ben (2004). "An Online Recruitment System for Economic Experiments", in Kurt Kremer, and Volker Macho, eds., *Forschung und wissenschaftliches Rechnen 2003*, GWDG Bericht 63, pp79-93. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

[114] Guala, Francesco (2005). "The Methodology of Experimental Economics", Cambridge University Press, New York.

[115] Hack, Andreas, and Frauke Lammers (2009). "The Role of Expectations in the Formation of Reference Points over Time", mimeo.

[116] Harrison, Glenn W., Steven J. Humphrey, and Arjan Verschoor (2010). "Choice under Uncertainty: Evidence from Ethiopia, India and Uganda", *Economic Journal* 120(2), pp80-104.

[117] Harrison, Glenn W., Eric Johnson, Melayne M. McInnes, and E. Elisabet Rutsröm (2005). "Risk Aversion and Incentive Effects: Comment", *American Economic Review* 95(3), pp900-904.

[118] Harrison, Glenn W., John A. List, and Charles Towe (2007). "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion", *Econometrica* 75(2), pp433-458.

[119] Harrison, Glenn W., and E. Elisabet Rutsröm (2008). "Risk Aversion in the Laboratory", in James C. Cox and Glenn W. Harrison, eds., *Risk Aversion in Experiments*, Research in Experimental Economics vol. 12, pp41-196. Bingley: Emerald.

[120] Heidhues, Paul, and Botond Kőszegi (2008). "Competition and Price Variation when Consumers are Loss Averse", *American Economic Review* 98(4), pp1245-1268.

[121] Heidhues, Paul, and Botond Kőszegi (2010). "Regular Prices and Sales", Working Paper, ESMT Working Paper 10-008.

[122] Herrmann, Benedikt, and Christian Thöni (2009). "Measuring Conditional Cooperation: A Replication Study in Russia", *Experimental Economics* 12(1), pp87-92.

[123] Herweg, Fabian, and Mierendorff, Konrad (forthcoming). "Uncertain Demand, Consumer Loss Aversion, and Flat-Rate Tariffs", *Journal of the European Economic Association*.

[124] Herweg, Fabian, Daniel Müller, and Philipp Weinschenk (2010). "Binary Payment Schemes: Moral Hazard and Loss Aversion", *American Economic Review* 100(5), pp.2451-2477.

[125] Hey, John D., and Chris Orme (1994). "Investigating Generalizations of Expected Utility Theory Using Experimental Data", *Econometrica* 62(6), pp1291-1326.

[126] Hildreth, Clifford, and Glenn J. Knowles (1982). "Some Estimates of Farmers' Utility Functions", *Technical Bulletin* 335, Agricultural Experimental Station, University of Minnesota, Minneapolis.

[127] Hilton, Ronald W. (1989). "Risk Attitude Under Random Utility", *Journal of Mathematical Psychology* 33(2), pp206-222.

[128] Ho, Teck-Hua, Colin Camerer, and Keith Weigelt (1998). "Iterated Dominance and Iterated Best Response in Experimental p-Beauty Contests", *American Economic Review* 88(4), pp947-969.

[129] Holt, Charles A., and Susan K. Laury (2002). "Risk Aversion and Incentive Effects", *American Economic Review* 92(5), pp1644-1655.

[130] Huck, Steffen, and Georg Weizsäcker (1999). "Risk, Complexity, and Deviations from Expected Value Maximization: Results of a Lottery Choice Experiment", *Journal of Economic Psychology* 20(6), pp699-715.

[131] Inman, Jeffrey J., James S. Dyer, and Jianmin Jia (1997). "A Generalized Utility Model of Disappointment and Regret Effects on Post-Choice Valuation", *Marketing Science* 16(2), pp97-111.

[132] Isaac, R. Mark, and James M. Walker (1988). "Communication and Free-Riding Behavior: The Voluntary Contribution Mechanism", *Economic Inquiry* 26(4), pp585-608.

[133] Isaac, R. Mark, James M. Walker, and Susan H. Thomas (1984). "Divergent Evidence on Free-Riding: An Experimental Examination of Possible Explanations", *Public Choice* 43(2), pp113-149.

[134] Isaac, R. Mark, James M. Walker, and Arlington W. Williams (1994). "Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups", *Journal of Public Economics* 54(1), pp1-36.

[135] Jacobson, Sarah, and Ragan Petrie (2009). "Learning from Mistakes: What Do Inconsistent Choices over Risk Tell Us?", *Journal of Risk and Uncertainty* 38(2), pp143-158.

[136] Jappelli, Tullio, Mario Padula, and Luigi Pistaferri (2008). "A Direct Test of the Buffer-Stock Model of Saving", *Journal of the European Economic Association* 6(6), pp1186-1210.

[137] Kachelmeier, Steven J., and Mohamed Shehata (1992). "Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence from the People's Republic of China", *American Economic Review* 82(5), pp1120-1141.

[138] Kahneman, Daniel, and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk", *Econometrica* 47(2), pp263-291.

[139] Keller, L. Robin (1985). "The Effects of Problem Representation on the Sure-Thing and Substitution Principles", *Management Science* 31(6), pp738-751.

[140] Kelley, Harold, and Anthony Stahelski (1970). "Social Interaction Basis of Co-operators' and Competitors' Beliefs About Others", *Journal of Personality and Social Psychology* 16(1), pp190-219.

[141] Keser, Claudia, and Frans van Winden (2000). "Conditional Cooperation and Voluntary Contributions to Public Goods", *Scandinavian Journal of Economics* 102(1), pp23-39.

[142] Kihlstrom, Richard E., David Romer, and Steve Williams (1981). "Risk Aversion with Random Initial Wealth", *Econometrica* 49(4), pp911-920.

[143] Kim, Oliver, and Mark Walker (1984). "The Free Rider Problem: Experimental Evidence", *Public Choice* 43(1), pp3-24.

[144] Kimball, Miles S. (1990). "Precautionary Saving in the Small and in the Large", *Econometrica* 58(1), pp53-73.

[145] Kimball, Miles S. (1992). "Precautionary Motives for Holding Assets", in Peter Newman, Murray Milgate, and John Falwell, eds., *The New Palgrave Dictionary of Money and Finance*, vol. 3, London: MacMillan, pp158-161.

[146] Knetsch, Jack L., and Wei-Kang Wong (2009). "The Endowment Effect and the Reference State: Evidence and Manipulations", *Journal of Economic Behavior and Organization* 71(2), pp407-413.

[147] Kocher, Martin G. (2008). "Robustness of Conditional Cooperation in Public Goods Experiments", mimeo.

[148] Kocher, Martin G., Todd Cherry, Stephan Kroll, Robert. J. Netzer, and Matthias Sutter (2008). "Conditional Cooperation on Three Continents", *Economics Letters* 101(3), pp175-178.

[149] Kocher, Martin G., Julius Pahlke, and Stefan T. Trautmann (2010). "An Experimental Test of Precautionary Bidding", Working Paper, University of Munich.

[150] Köbberling, Veronika, and Peter P. Wakker (2005). "An Index of Loss Aversion", *Journal of Economic Theory* 122(1), pp119-131.

[151] Kőszegi, Botond, and Matthew Rabin (2006). "A Model of Reference-Dependent Preferences", *Quarterly Journal of Economics* 121(4), pp1133-1165.

[152] Kőszegi, Botond, and Matthew Rabin (2007). "Reference-Dependent Risk Attitudes", *American Economic Review* 97(4), pp1047-1073.

[153] Kőszegi, Botond, and Matthew Rabin (2009). "Reference-Dependent Consumption Plans", *American Economic Review* 99(3), pp909-936.

[154] Krähmer, Daniel, and Rebecca Stone (2008). "Regret in Dynamic Decision Problems", Working Paper, University of Bonn.

[155] Kydland, Finn E., and Edward C. Prescott (1982). "Time to Build and Aggregate Fluctuations", *Econometrica* 50(6), pp1345-1370.

[156] Lange, Andreas, and Anmol Ratan (2010). "Multi-Dimensional Reference-Dependent Preferences in Sealed-Bid Auctions – How (Most) Laboratory Experiments Differ from the Field", *Games and Economic Behavior* 68(2), pp634-645.

[157] Ledyard, John (1995). "Public Goods: A Survey of Experimental Research", in J. Kagel and A. Roth, eds., 'The Handbook of Experimental Economics', Princeton University Press, Princeton, NJ, pp111-194.

[158] Lee, Jeong-Joon, and Yasuyuki Sawada (2007). "The Degree of Precautionary Saving: A Reexamination", *Economics Letters* 96(2), pp196-201.

[159] Leland, Hayne E. (1968). "Saving and Uncertainty: The Precautionary Demand for Saving", *Quarterly Journal of Economics* 82(3), pp465-473.

[160] Levati, M. Vittoria, Matthias Sutter, and Eline van der Heijden (2007). "Leading by Example in a Public Goods Experiment with Heterogeneity and Income Information", *Journal of Conflict Resolution* 51(5), pp793-818.

[161] Levy, Haim (1994). "Absolute and Relative Risk Aversion: An Experimental Study", *Journal of Risk and Uncertainty* 8(3), pp289-307.

[162] Loomes, Graham, and Robert Sugden (1982). "Regret Theory: An Altenative Theory of Rational Choice under Uncertainty", *Economic Journal* 92(4), pp805-824.

[163] Loomes, Graham, and Robert Sugden (1986). "Disappointment and Dynamic Consistency in Choice under Uncertainty", *Review of Economic Studies* 53(2), pp271-282.

[164] Loomes, Graham, and Robert Sugden (1987). "Testing for Regret and Disappointment in Choice Under Uncertainty", *Economic Journal* 97(Supplement), pp118-129.

[165] Loomes, Graham, and Robert Sugden (1998). "Testing Different Stochastic Specifications of Risky Choice", *Economica* 65(4), pp581-598.

[166] Macera, Rosario (2010). "Intertemporal Incentives under Loss Aversion", Working Paper, University of California, Berkeley.

[167] Machina, Mark J. (1982). ""Expected Utility" Analysis Without the Independence Axiom", *Econometrica* 50(2), pp277-323.

[168] Machina, Mark J. (1987). "Choice under Uncertainty: Problems Solved and Unsolved", *Journal of Economic Perspectives* 1(1), pp121-154.

[169] Machina, Mark J. (2008). "Non-Expected Utility Theory", in Steven N. Durlauf, and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*, 2nd ed., vol. 6, New York: Palgrave MacMillan, pp74-84.

[170] Maier, Johannes, and Maximilian Rüger (2010). "Experimental Evidence on Higher Order Risk Preferences with Real Monetary Losses", mimeo.

[171] Markowitz, Harry (1952). "The Utility of Wealth", *Journal of Political Economy* 60(2), pp151-158.

[172] Marwell, Gerald, and Ruth E. Ames (1979). "Experiments on the Provision of Public Goods. I. Resources, Interest, Group Size, and the Free-Rider Problem", *American Journal of Sociology* 84(6), pp1335-1360.

[173] Marwell, Gerald, and Ruth E. Ames (1981). "Economists Free Ride, Does Anyone Else?", *Journal of Public Economics* 15(3), pp295-310.

[174] Mas, Alexandre (2006). "Pay, Reference Points, and Police Performance", *Quarterly Journal of Economics* 121(3), pp783-821.

[175] McCord, Mark, and Richard de Neufville (1986). ""Lottery Equivalents": Reduction of the Certainty Effect Problem in Utility Assessment", *Management Science* 32(1), pp56-60.

[176] Mehra, Rajnish (2003). "The Equity Premium: Why Is It a Puzzle?", *Financial Analysts Journal* 59(1), pp54-69.

[177] Menezes, Carmen F., Charles G. Geiss, and John F. Tressler (1980). "Increasing Downside Risk", *American Economic Review* 70(5), pp921-932.

[178] Menezes, Carmen F., and X. Henry Wang (2005). "Increasing Outer Risk", *Journal of Mathematical Economics* 41(7), pp875-886.

[179] Meng, Juanjuan (2009). "The Disposition Effect and Expectations as Reference Point", Working Paper, University of California San Diego.

[180] Miller, Louis, David E. Meyer, and John T. Lanzetta (1969). "Choice Among Equal Expected Value Alternatives: Sequential Effects of Winning Probability Level on Risk Preferences", *Journal of Experimental Psychology* 79(3/1), pp419-423.

[181] Moffet, Denis (1975). "Risk-Bearing and Consumption Theory", *Astin Bulletin* 8(3), pp342-358.

[182] Nagel, Rosemarie (1995). "Unraveling in Guessing Games: An Experimental Study", *American Economic Review* 85(5), pp1313-1326.

[183] Neugebauer, Tibor, Javier Perote, Ulrich Schmidt, and Malte Loos (2009). "Selfish-Biased Conditional Cooperation: On the Decline of Contributions in Repeated Public Goods Experiments", *Journal of Economic Psychology* 30(1), pp52-60.

[184] Noussair, Charles N., Stefan T. Trautmann, and Gijs van de Kuilen (2011). "Higher Order Risk Attitudes, Demographics, and Financial Decisions", CentER Working Paper No. 2011-055, Tilburg University.

[185] Ockenfels, Axel (1999). "Fairness, Reziprozität und Eigennutz - Ökonomische Theorie und Experimentelle Evidenz. Die Einheit der Gesellschaftswissenschaften", Bd. 108, Mohr Siebeck, Tübingen.

[186] Palfrey, Thomas R., and Jeffrey E. Prisbrey (1996). "Altruism, Reputation, and Noise in Linear Public Goods Experiments", *Journal of Public Economics* 61(3), pp409-427.

[187] Pearce, David G. (1984). "Rationalizable Strategic Behavior and the Problem of Perfection", *Econometrica* 52(4), pp1029-1050.

[188] Pope, Devin G., and Maurice E. Schweitzer (2011). "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes", *American Economic Review* 101(1), pp129-157.

[189] Post, Thierry , Martijn J. van den Assem, Guido Baltussen, and Richard H. Thaler (2008). "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show", *American Economic Review* 98(1), pp38-71.

[190] Prasad, Kislaya, and Timothy C. Salmon (2010). "Self Selection and Market Power in Risk Sharing Contracts", Working Paper, Florida State University.

[191] Quiggin, John (1982). "A Theory of Anticipated Utility", *Journal of Economic Behavior and Organization* 3(4), pp323-343.

[192] Rabin, Matthew (2000). "Risk Aversion and Expected-Utility Theory: A Calibration Theorem", *Econometrica* 68(5), pp1281-1292.

[193] Rapoport, Amnon (1988). "Provision of Step-Level Goods: Effects of Inequality in Resources", *Journal of Personality and Social Psychology* 54(3), pp432-440.

[194] Rapoport, Amnon, and Ramzi Suleiman (1993). "Incremental Contribution in Step-Level Public Goods Games with Asymmetric Players", *Organizational Behavior and Human Decision Processes* 55(2), pp171-194.

[195] Reuben, Ernesto, and Arno Riedl (2011). "Enforcement of Contribution Norms in Public Good Games with Heterogeneous Populations", Working Paper, University of Maastricht.

[196] Röell, Ailsa (1987). "Risk Aversion in Quiggin and Yaari's Rank-Order Model of Choice Under Uncertainty", *Economic Journal* 97(388a), pp143-159.

[197] Ross, Stephen A. (1981). "Some Stronger Measures of Risk Aversion in the Small and the Large with Applications", *Econometrica* 49(3), pp621-638.

[198] Rothschild, Michael, and Joseph E. Stiglitz (1970). "Increasing Risk: I. A Definition", *Journal of Economic Theory* 2(3), pp225-243.

[199] Saijo, Tatsuyoshi, and Hideki Nakamura (1995). "The "Spite" Dilemma in Voluntary Contribution Mechanism Experiments", *Journal of Conflict Resolution* 39(3), pp535-560.

[200] Sandmo, Agnar (1970). "The Effect of Uncertainty on Saving Decisions", *Review of Economic Studies* 37(3), pp353-360.

[201] Schmidt, Ulrich, and Horst Zank (2008). "Risk Aversion in Cumulative Prospect Theory", *Management Science* 54(1), pp208-216.

[202] Selten, Reinhard (1967). "Die Strategiemethode zur Erforschung des Eingeschrängt Rationalen Verhaltens im Rahmen eines Oligopolexperimentes", in H. Sauermann, ed., 'Beiträge zur Experimentellen Wirtschaftsforschung', Mohr Siebeck, Tübingen, pp136-168.

[203] Sonnemans, Joep, Arthur Schram, and Theo Offerman (1999). "Strategic Behavior in Public Goods Games: When Partners Drift Apart", *Economics Letters* 62(1), pp35-41.

[204] Sonsino, Doron, Uri Benzion, and Galit Mador (2002). "The Complexity Effects on Choice with Uncertainty – Experimental Evidence", *Economic Journal* 112(482), pp936-965.

[205] Stahl, Dale O., and Paul R. Wilson (1994). "Experimental Evidence on Players' Models of Other Players", *Journal of Economic Behavior and Organization* 25(3), pp309-327.

[206] Stahl, Dale O., and Paul R. Wilson (1995). "On Players' Models of Other Players: Theory and Experimental Evidence", *Games and Economic Behavior* 10(1), pp218-254.

[207] Starmer, Chris, and Robert Sugden (1989). "Probability and Juxtaposition Effects: An Experimental Investigation of the Common Ratio Effect", *Journal of Risk and Uncertainty* 2(2), pp159-178.

[208] Sugden, Robert (2003). "Reference-Dependent Subjective Expected Utility", *Journal of Economic Theory* 111(2), pp172-191.

[209] Szpiro, George (1986). "Measuring Risk Aversion: An Alternative Approach", *Review of Economics and Statistics* 68(1), pp156-159.

[210] Tarazona-Gomez, Marcela (2004). "Are Individuals Prudent? An Experimental Approach Using Lottery Choices", Working Paper, University of Toulouse.

[211] Thaler, Richard H, and Eric J. Johnson (1990). "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice", *Management Science* 36(6), pp643-660.

[212] Tobin, James, and Walter Dolde (1971). "Wealth, Liquidity and Consumption", in *Consumer Spending and Monetary Policy: The Linkage*, Boston: Federal Reserve Bank of Boston, pp99-146.

[213] Treich, Nicolas (2010). "Risk-Aversion and Prudence in Rent-Seeking Games", *Public Choice* 145(3), pp339-349.

[214] Tversky, Amos, and Daniel Kahneman (1991). "Loss Aversion in Riskless Choice: A Reference-Dependent Model", *Quarterly Journal of Economics* 106(4), pp1039-1061.

[215] Tversky, Amos, and Daniel Kahneman (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty", *Journal of Risk and Uncertainty* 5(4), pp297-323.

[216] Van Dijk, Eric, and Malgorzata Grodzka (1992). "The Influence of Endowments Asymmetry and Information Level on the Contribution to a Public Step Good", *Journal of Economic Psychology* 13(2), pp329-342.

[217] Van Dijk, Eric, and Henk Wilke (1994). "Asymmetry of Wealth and Public Good Provision", *Social Psychology Quarterly* 57(4), pp352-359.

[218] Von Neumann, John, and Oskar Morgenstern (1944). "The Theory of Games and Economic Behavior", Princeton University Press, Princeton, NJ.

[219] Wakker, Peter P. (2005). "Formalizing Reference Dependence and Initial Wealth in Rabin's Calibration Theorem", Working Paper, Erasmus University Rotterdam.

[220] Wakker, Peter P., and Daniel Deneffe (1996). "Eliciting von Neumann-Morgenstern Utilities when Probabilities are Distorted or Unknown", *Management Science* 42(8), pp1131-1150.

[221] Wang, Stephanie W., Michelle Filiba, and Colin F. Camerer (2010). "Dynamically Optimized Sequential Experimentation (DOSE) for Estimating Economic Preference Parameters", Working Paper, California Institute of Technology.

[222] Weimann, Joachim (1994). “Individual Behavior in a Free Riding Experiment”, *Journal of Public Economics* 54(2), pp185-200.

[223] White, Lucy (2008). “Prudence in Bargaining: The Effect of Uncertainty on Bargaining Outcomes”, *Games and Economic Behavior* 62(1), pp211-231.

[224] Wilcox, Nathaniel T. (2007). “Stochastic Models for Binary Discrete Choice Under Risk: A Critical Stochastic Modeling Primer and Econometric Comparison”, in James C. Cox, and Glenn W. Harrison, eds., *Risk Aversion in Experiments*, Research in Experimental Economics vol. 12, Bingley (UK): Emerald, pp197-292.

[225] Wilcox, Nathaniel T. (2011). “‘Stochastically More Risk Averse:’ A Contextual Theory of Stochastic Discrete Choice Under Risk”, *Journal of Econometrics* 162(1), pp89-104.

[226] Wit, Arjaan P., Henk A. M. Wilke, and Harmen Oppewal (1992). “Fairness in Asymmetric Social Dilemmas”, in W.B.G. Liebrand, D. M. Messick, and H.A.M. Wilke, eds., ‘Social Dilemmas: Theoretical Issues and Research Findings’, Pergamon, New York, pp183-197.

[227] Yaari, Menahem E. (1969). “Some Remarks on Measures of Risk Aversion and on Their Uses”, *Journal of Economic Theory* 1(3), pp315-329.

[228] Yaari, Menahem E. (1987). “The Dual Theory of Choice under Risk”, *Econometrica* 55(1), pp95-115.

[229] Zelmer, Jennifer (2003). “Linear Public Goods Experiments: A Meta-Analysis”, *Experimental Economics* 6(3), pp299-310.