Essays on Social Preferences

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2011

vorlegt
von
Diplom-Volkswirt Björn Hartig (M.A.)
aus
Osnabrück

Referent: Prof. Dr. Axel Ockenfels

Korreferent: Prof. Dr. Dirk Sliwka

Tag der Promotion: 10. Juni 2013

Contents

List of Figures

List of Tables

How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortunes of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it.

Adam Smith, The Theory of Moral Sentiments

A house may be large or small; as long as the neighboring houses are likewise small, it satisfies all social requirements for a residence. But let there arise next to a little house a palace, and the little house shrinks into a hut. [...] and however high it may shoot up in the course of civilization, if the neighboring palace rises in equal or even greater measure, the occupant of the relatively little house will always find himself more uncomfortable, more dissatisfied, more cramped within his four walls.

Karl Marx, Wage-Labor and Capital

1 Introduction

1.1 Historical Background

The pioneer economist Francis Y. Edgeworth (1881) states in his seminal article *Mathematical Psychics* that "the first principle of economics is that every agent is actuated only by self-interest". For the most part of the 20th century, virtually all economic models understood self-interest exclusively as concern for own material payoff. This had not always been the case, however, as most early economists had in fact a much broader view on the matter. Self-interest was, quite generally, the pursuit of pleasure and the avoidance of pain (Bentham 1789). Utility, therefore, was a real psychological substance and consequently, understanding the determinants of hedonic sensations and finding a way to reliably measure them was central to early economics. It was only about one hundred years ago that the concept of self-interest was narrowed down substantially.

At the end of the 19th century, economics attempted to associate itself more with the natural sciences, which had made significant advancements by focusing attention on mathematical analysis of objective facts instead of searching for the metaphysical essence of matter as scholars had done previously. Jevons (1871), Edgeworth (1881), and others had already mathematically formalized the idea of hedonic utility as the force behind behavior by re-applying concepts from physics to economic problems, but some economists like Pareto argued that economics also had to rid itself of its metaphysical and psychological baggage and pay attention only to observed behavior (Pareto 1971). In anticipation of what later became known as the revealed preference principle, Pareto argued that whatever goes on under the surface must inevitably reveal itself through behavior (Bruni and Sugden 2007). As a result, the main attention of economic research increasingly turned away from experienced sensations toward observed choices, eventually leading to an universal "as-if" approach concerned solely with the accuracy of its models' predictions, but uninterested in the correctness of the underlying assumptions (Friedman 1953).

This development alone would not have made material self-interest the only commonly accepted motivation for economic behavior, but Pareto and others took it a step further. To distance economics from other social sciences, they restricted economic theory to the analysis of logical actions, i.e., instrumental actions resulting from objectively true premises through deduction, pitted against the standard of rational choice. As such, the science of economics had to be primarily unconcerned with the application of its principles in the real world (Pareto 1971). The aspiration was not to explain human behavior universally, but only a limited range of behavior conforming to the theory. Pareto therefore restricted the analysis to repeated actions, ensuring that theoretically, learning would eventually eliminate all logical mistakes. When later generations dropped this restriction and applied the theory to all kinds of non-repeated actions, they nevertheless kept dismissing violations of the rational model as errors that would eventually vanish with repetition (Binmore and Samuelson 1999).

Still, even the focus on (repeated) rational choice does not necessarily entail that the only source of motivation has to be own material payoff. Unlike Mill (1844), who wanted to restrict people's motivation in political economy to the pursuit of wealth, Pareto explicitly left the door open for altruism or other forms of other-regarding preferences, which in principle are compatible with his approach (Pareto 1971). In the end, there were probably several reasons why pure material self-interest eventually prevailed. For one thing, the idea that human nature is inherently selfish already had a long philosophical history, with Hobbes, Locke, Mandeville, and Rousseau providing major contributions to this school of thought.

Critics from outside of economics seem to have played a crucial role, too. In the early 20th century, psychologists attacked the hedonic approach as too restrictive, therefore unrealistic and also unreliable (McDougall 1910). This led many economists to abandon references to hedonistic sensations entirely (Lewin 1996), paving the way for material self-interest to take over. And finally, a rather mundane reason is that the abstractness of the selfish model simply works well with the mathematical approach of rational choice theory, making material self-interest the most convenient proxy for other possible motives (Mullainathan and Thaler 2001). So convenient in fact that it gradually lost its proxy designation and became a key characteristic of the *homo oeconomicus*, the protagonist of rational behavior.[1]

1.2 The Rise of Social Preferences

Even though selfishness was never an integral part of rationality (Sen 1977), it was the last property of the *homo oeconomicus* to come under serious fire. Early on, Simon (1955) challenged the perfect information processing capability while Kahneman, Tversky, and associates attacked perfect rationality on several fronts (e.g. Tversky and Kahneman 1974; Kahneman and Tversky 1979). But although the psychologist Adams (1963) had long ago developed equity theory – a psychological model assuming that workers dislike being compensated unequally – other-regarding preferences mostly stayed under the radar even as behavioral economics slowly became institutionalized as a field during the early 1980s. That finally started to change – albeit slowly – when Güth et al. (1982) published their results of the ultimatum game experiment.

The ultimatum game is a two-player-game where the first player proposes an allocation of a pie and the second player either accepts or rejects the proposal. If he accepts, the proposal is paid out, but if he rejects, both players receive nothing. There are many Nash-equilibria in this game, but the only sub-game perfect equilibrium has the second player accept any positive offer. Anticipating this, the first player chooses the allocation in which the second player receives the lowest possible positive outcome. In reality, however, proposals by the first player generally are much higher – the modal offer is

[1] For a more detailed discussion of the paradigm shift in economics, see Lewin (1996) and Bruni and Sugden (2007), who give quite divergent accounts of the era.

often 50% of the pie – and the second player often rejects positive offers when receiving less than the proposer.

The ultimatum game experiment was replicated and modified many times over the next few years (reviews in Thaler 1988; Güth 1995), but the general results persisted. However, disagreement over how to interpret the findings soon arose. While some regarded them as prove for fairness concerns, others insisted that they were erroneous anomalies (Binmore et al. 1985; Gale et al. 1995). However, control experiments eventually showed that neither low stakes (Slonim and Roth 1998; Cameron 1999), missing learning opportunities (Slonim and Roth 1998), anonymity (Bolton and Zwick 1995), nor strategic uncertainty (Forsythe et al. 1994) could completely account for the divergence from the game theoretic equilibrium. In the last study, Forsythe et al. developed a modification of the ultimatum game called dictator game, which made it obvious that many people did in fact care about the outcome of the other player.

Like the ultimatum game, the dictator game is a two-player-game, although the second player is only in a passive role. Again, the first player decides on the division of a pie, but unlike in the ultimatum game, the second player has no say in the matter and the chosen allocation is immediately paid out. The only individually rational decision for the first player in the game is of course to keep the whole pie. However, although the second player's share of the pie is usually smaller on average than in ultimatum game offers, many people – often the majority – do give a positive amount of up to 50% of the pie.

The ultimatum game and dictator game studies ultimately spawned a series of outcome-based models of social preferences that tried to reconcile the results of the two games with utility maximization. The first attempt was probably made by Bolton (1991), who complemented monetary utility with some "relative utility". The idea eventually evolved into the model of inequity aversion by Bolton and Ockenfels (2000), who presumed that people suffer inequity costs when their own payoff differs from the equal share. A related model was developed by Fehr and Schmidt (1999), in which individuals compare their own payoff with the payoff of others and suffer from compassion or envy if they are better or worse off, respectively. Another model with a different approach was later developed by Charness and Rabin (2002). They assume that people are not motivated by some form of inequity version, but instead by welfare concerns for total efficiency and for the individual with the lowest payoff.

A different line of social preferences developed around the so-called gift-exchange game. The idea for the game goes back to Akerlof (1982, 1984), who hypothesized that firms are paying workers salaries above the marked-clearing wage and workers are repaying this "kind" behavior with higher effort. Fehr et al. (1993) found support for this fair-wage-theory in a highly stylized, yet very influential experiment. In a one-sided auction, buyers made price offers and upon acceptance, sellers determined the quality of the good. The majority of offers were well above the market clearing level and sellers on average responded to higher prices with higher quality. The experiment is widely viewed as the first clear evidence for the importance of reciprocal behavior in economic interactions.

Rabin (1993) was the first to translate the concept of reciprocity into a mathematical model. Drawing on psychological game theory (Geanakoplos et al. 1989; Battigalli and Dufwenberg 2009), he postulated that individuals form beliefs about the kindness of others (which are confirmed in equilibrium) and then prefer to reward those they perceive as kind and punish those they perceive as unkind. The main idea of the model was refined and extended several times, most notably by Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006). A slightly different approach was proposed by Levine (1998), who assumed that the weight put on another person's payoff depends on an individual's inherent type as well as on (the belief about) the type of the other person. However, Rabin's model and its successors were more influential and prompted a series of other experiments highlighting the importance of intentionality in social interaction (e.g. Blount 1995; Falk et al. 2003) as well as a re-evaluation of the ultimatum game results in the light of reciprocity (e.g. Kagel et al. 1996; Bolton et al. 1998b).

Another front where the prevalence of selfishness was challenged were social dilemma games. Social dilemma games are characterized by the conflict of individually optimal behavior and socially optimal behavior. For example, in the standard public goods game, each player receives an endowment which he can either keep for himself or invest in the public good. The total amount invested in the public good is multiplied with a factor greater than 1 (but less than the number of players). The resulting amount is then distributed evenly among all players. The socially efficient decision is to invest the whole endowment into the public good, whereas the individually rational decision is of course to keep the whole endowment. Any individual motivated only by own material gain would therefore always choose the latter strategy.

Provisions to public goods have been analyzed experimentally since the

1970s (Bohm 1972; Dawes et al. 1977; Marwell and Ames 1979). The results reject the assumption of purely self-regarding behavior by all players in virtually all instances. In one-shot public goods game, the majority of players usually makes significant contributions to the public good, although free-riders, i.e., players who do not contribute at all, of course also exist. In repeated interactions, contributions typically decline gradually over time (e.g. Kim and Walker 1984; Isaac et al. 1985) with a particularly steep drop in the last round – the so called end-game-effect. Yet, contributions generally do not vanish completely and even increase again after a surprise re-start (Andreoni 1988), suggesting that contributions are not a behavioral error vanishing with experience (see also e.g. Marwell and Ames 1980; Palfrey and Prisbrey 1997).

Two modifications of the standard public goods game have been particularly influential. The first one introduces a punishment mechanism which allows players to pay some costs to decrease another player's payoff after contributions to the public good have been made (Yamagishi 1986; Fehr and Gächter 2000). Since punishment is costly, no rational player should ever engage in it. However, the punishment mechanism is typically used by high contributors against low contributors, which usually has a disciplining effect, stabilizing players' contributions on a relatively high level in repeated public goods game (although punishment by low contributors exists, too; see Herrmann et al. 2008). The second modification turns the public goods game into a sequential game, allowing players to condition their own contribution on the contributions of the other players. Fischbacher et al. (2001) found that most players are conditional contributors, i.e., they increase their own contribution as the contributions of other players go up. Both punishment and conditional contribution are of course incompatible with material self-interest, yet are very much in line with both models of inequity aversion and models of reciprocity.

1.3 Recent and Future Developments

Over the last two decades, the experimental literature around the ultimatum game, dictator game, gift-exchange game and public goods game has grown tremendously, producing countless variants and modifications of the four games as well as completely new kind of decision-making situations that shall remain unmentioned here because there are simply too many of them by now. The assertion that most people do in fact care very much about how their own wealth compares to that of others, about fairness of outcomes and procedures, and about the intentions of others cannot be seriously questioned any more. Of course, this does not rule out that competition or other (market) forces may subdue or crowd out social preferences in certain situations, making it look as if people were only concerned for their own material payoff (e.g. Smith 1962). Despite the extensive theoretical and empirical research, there are still many aspects of social preferences that are not understood completely (or at all). Some facets may very well not even be on our radar yet until somebody points them out to the world, whereupon they will probably seem quite obvious all of a sudden. All in all, it appears that the field is in no danger of becoming saturated any time soon.

Naturally, nobody knows what the next hot topic regarding social preferences will be, but there have been a few recent experiments spurring new discussions among researchers that might give us a clue of where the field is heading toward in the next years. For example, Dana and colleagues have conducted experiments on other-regarding preferences involving information asymmetries (Dana et al. 2006, 2007). Their subjects curtail altruistic behavior when their actions are not fully observable by the receivers, which is difficult to explain with the current models. Meanwhile, Charness and Dufwenberg (2006) find that individuals seem to try to live up to others' expectations, apparently to avoid letting others down. However, the results of Vanberg (2008) indicate that in fact individuals seem to care little about others' disappointment, but instead are just inclined to keep promises they made. These kind of findings may seem rather marginal compared to equality and reciprocity concerns in general, however, properly interpreting such results may very well be the key to a deeper understanding of the true nature of other-regarding preferences.

Another topic that has recently arrived on the scene is social risk taking, i.e., risk preferences in situations when both own and other persons' payoffs are at stake. In an inter-cultural study, Bohnet et al. (2008) play stylized trust games[2] and do not find much influence of general social motivations like inequality aversion or altruism on social risk taking. In their comment to this article, Bolton and Ockenfels (2010) address this issue by comparing one-person- and dictator-game-choice-problems by varying the second player's payoff for safe and risky options. Disadvantageous inequality in the safe

[2]The trust game is a variant of the gift-exchange game where the first player "invests" into the positive reciprocal response of the second player, see Blount 1995.

option seems to increase risk taking, but inequality in the risky choice does not. Furthermore, Bradler (2009) measures individuals' social preferences under certainty, regular risk preferences and interpersonal risk preferences separately, concluding that depending on the relative standings, many people are willing to accept more risk or forgo own payoff to increase others' payoffs. Overall, at the current state of research, the question to what extend social preferences under uncertainty are more than the straightforward combination of principles of other-regarding behavior and risk preferences is still largely unresolved.

Some researchers have also started to carry experiments on social preference from the laboratory into the real world. For example, Gneezy and List (2006) conduct a real-life gift-exchange game by paying some student workers more than initially advertised. As predicted by the theory, those workers receiving the higher salary produce more output, although the effect wears off after some hours and the net-profit of the gift is negative for the employer in the experiment. In another study, Falk (2007) sends out letters asking for donations to charity and attaches gifts of various value to them. In line with reciprocity, the average donation increase with the value of the gift. These two examples show that social preferences are not just artifacts of escapist games, restricted to the obscurity of economic laboratories, but that they do in fact not only exist in the real world, but are responsible for sizable economic effects, too.

Another subfield of economics is receiving a lot of public attention recently: Neuroeconomics aims to – quite literally – look inside people's heads while making decisions. Using neuroimaging techniques like functional magnetic resonance imaging (fMRI) or magnetoencephalography (MEG), neuroeconomists measure activity in distinct brain regions or trace the chronological sequence of different neural events, respectively. Transcranial magnetic stimulation (TMS) can even be used to temporarily disrupt specific brain regions to study functioning and interconnections of the regions in question.

Many neuroeconomic studies on social preferences focus on the brain's reward system. Rilling et al. (2002, 2004) find that subjects have stronger activation in reward-related brain regions when cooperating with human counterparts than with a computer. The rewards system is also activated when making charitable donations (Moll et al. 2006; Harbaugh et al. 2007) and receiving fair offers in the ultimatum game (Tabibnia et al. 2008). Furthermore, de Quervain et al. (2004) show that effective punishment of defectors in a prisoner's dilemma leads to stronger neural responses than symbolic punishment. In the study by Fliessbach et al. (2007), two subjects simultaneously perform a task over many rounds, receiving varying prizes when successful. Activation in reward-related brain areas generally is lower when receiving a relatively low prize compared to the other subject and vice versa.

However, the reward system is of course not the only brain region related to social preferences. Sanfey et al. (2003) demonstrate that both regions concerned with emotions and with cognition are active when receiving unfair ultimatum game offer. Furthermore, Koenigs and Tranel (2007) find that subjects with lesions in the prefrontol cortex – a region associated with conflict resolution – were more likely to reject unfair ultimatum game offers, suggesting that the region is involved in moderating emotions with material costs and benefits. Moreover, not all activations are created equal. For example, cooperators in simple trust-and-reciprocity games have more activation in prefrontal regions when playing with human partners than with computers, but non-cooperators show no difference (McCabe et al. 2001). Additionally, Singer et al. (2006) find that empathic neural responses depend on the perceived fairness of others.

Part of the appeal of neuroeconomics certainly stems from the intrigue of the ever increasing technological possibilities and the general fascination of the brain as such, but another part comes from its virtual promise to do what economists have desired since the 19th century: to allow direct and reliable measurement of an individual's utility. However, as alluring as this promise most certainly is, restraint and caution are warranted. On the one hand, equating neural activation with utility may be tempting, but it would surely be shortsighted to ignore all other sources of happiness and motivation except for momentary peaks in archaic regulatory systems. On the other hand, neuroeconomics also faces substantial methodological criticism. For example, Rubinstein (2006) argues that conclusions in neuroeconomic studies are often drawn based on rather scant data, but use colorful graphics to create the illusion of hard evidence. He also claims that most neuroeconomic studies produce very little – if any – new economic knowledge, because records of neural activation alone offer very little explanatory value (Rubinstein 2008). Despite these reservations, neuroeconomics is not going away anytime soon.

Finally, as behavioral economics increasingly makes its way from the laboratories into the real world, its protagonists will also more frequently have to deal with ethical questions. While marketing has been applying psychological insights for decades – not necessarily to the benefit of the consumers (e.g. Slovic 2001) – behavioral economists have only recently started to use their knowledge to "nudge" people to make "better" decisions (Thaler and Sunstein 2008). However, this inevitability leads to the delicate issue which forms of intervention are ethically acceptable. First of all, the idea of paternalistic intervention rests on the view that on the one hand, individuals make suboptimal decisions – a claim not universally accepted (Berg and Gigerenzer 2010) – and that on the other hand, economists know what would be better for them. Unfortunately, economists, too, are imperfect human beings with their own shortcomings, agenda and biases. While there may be examples where the benefits seem obvious and the interference negligible, e.g. when employing default options to steer behavior (Johnson and Goldstein 2003; Cronqvist and Thaler 2004), economists must not shy away from the scrutiny. In particular, insight about how our brains, hormones, and genes influence behavior will eventually provide social planers with tools that seemed like science fiction not too long ago. Consider the following, slightly exaggerated (?) example: Trust is known to be an important determinant of a society's economic performance (e.g. Knack and Keefer 1997), so it would be socially desirable if people trusted each other more. A simple way to accomplish this would be to release the neuromodulator oxytocin, a hormone known to increase trust (Kosfeld et al. 2005; Zak et al. 2005), into the water supply. If that sounds preposterous to you, consider that you can actually buy perfume enhanced with oxytocin and other hormones that is explicitly marketed for business meetings, negotiations and blind dates.[3]

1.4 Three Essays on Social Preferences

The following chapters present three essays dealing with different aspects of social preferences. The first essay *Social Utility Functions: Where Do We Stand, How Did We Get There, and Where to Next?* surveys the literature on social preference models. After briefly looking at elementary groundwork from economics and psychology, we review the development of interdependent utility functions in their historical contexts. The survey is divided into three main parts. The first part looks at various outcome-based models, followed by an overview of models based on the concept of reciprocity. All remaining models are subsumed in the third part. Additionally, we summarize the main criticism leveled at social preference models before assessing the current state of the research field and its future prospects.

[3]See http://www.pherolife.com/html/faqs.html (Retrieved 12. Dec. 2012).

In the second essay *A Closer Look at Inequity Aversion and Incentives in Tournaments*, we conduct a theoretical analysis of a tournament model with inequity averse agents. It extends Grund and Sliwka's (2005) article on the same topic by comparing the results of the two models of social preferences by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) and by allowing more than just two agents to compete in the tournament. While the model of Fehr and Schmidt yields an incentive effect in two-agent-tournaments, the model of Bolton and Ockenfels does not. In multi-agent-tournaments with two prize levels, inequity averse agents show a relative preference for the prize given to the majority. The principal should therefore optimally design tournaments to have many winners to exploit this effect. On the other hand, in tournaments with many losers, incentives may be distorted to such a degree that agents actually prefer losing the tournament and exert no effort.

Finally, the third essay *Does Altruism Depend on Rational Expectations?* presents an experiment on the effect of rational expectations on altruistic behavior. In a simple dictator game setting, the dictator's endowment is randomly determined by a 50-50 lottery. The main treatment condition compares the transfer of dictators who received the same endowment, but faced different expected outcomes in the lottery. The theory predicts that dictators receiving more than their expected outcome transfer more money to the receiver than dictators receiving less than their expected outcome. However, the results confirm the hypothesis only weakly. Additionally, we do not find any evidence for the "warm glow" of altruistic giving. Possible improvements for a follow-up experiment are also discussed.

2 Social Utility Functions: Where Do We Stand, How Did We Get There, and Where to Next?

2.1 Introduction

Economic models traditionally assumed that each individual's utility function is strictly independent of other people's payoffs, intentions, or behavior. Even though on occasion, the odd economist ventured beyond the narrow horizon of material self-interest, mainstream economics was mostly satisfied with the traditional *homo oeconomicus* model as long as it yielded reasonably precise predictions. However, the end of the 20th century saw a paradigm shift as results from laboratory experiments persistently contradicted the predictions of the purely self-interested model. After an initial period of skepticism, the tidal wave of empirical evidence unleashed by the first experiments on ultimatum bargaining (Güth et al. 1982; Binmore et al. 1985) eventually washed away most major resistance. Today, it is widely accepted that people are motivated by other-regarding preferences and that these preferences can have important economic consequences (Fehr and Fischbacher 2002).

However, demonstrating that people have social preference is only the first step; using this knowledge for economic analysis and deriving predictions from it is the next. Despite the close affiliation with rationality, material self-interest is actually not an essential part of rational choice theory, only a simplifying assumption (Sen 1977; Mullainathan and Thaler 2001). In fact, social preferences generally do not contradict the axioms of rationality, i.e., completeness and transitivity of preference relations or the weak axiom of revealed preference.[4] Hence, in principle it is possible to construct well-behaved utility functions accounting for other-regarding concerns.

Over the years, economists have developed many different social preference models accounting for a wide spectrum of behavioral patterns. This paper takes stock of the social utility functions proposed in these models and

[4]Sen (1993) shows that there can be instances in social choice (among other domains) when preferences may appear to violate these axioms. For example, people with positional preferences may want to receive a payoff as high as possible, but not the highest one (in order not to appear greedy). Adding a new highest payoff to the choice set changes the decision, even though the new payoff is not chosen itself, thus violating the weak axiom of revealed preferences.

reviews them in their historical and topical contexts. The survey is limited to models with functional representation because a broader approach would inevitably require skipping more models, especially if including formative influences from outside of economics. Furthermore, the utility function typically captures the essence of a model more concisely and with more precision than any verbal synopsis. Finally, it appears that the research on social utility functions is at a crossroads. Firstly, the body of work has not yet reached dimensions rendering the attempt at a relatively complete overview impossible. Secondly, the seminal articles of Rabin (1993), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), and Charness and Rabin (2002) have now been published for at least a decade and are arguably still the most influential works on the topic, making this a good time to ask the question *quo vadis?*

This survey does not cover completely "ego-centric" utility functions of social behavior found in the classical public good literature (e.g. Samuelson 1954), in the model of warm-glow giving by Andreoni (1989), or in the crowding-out model by Bénabou and Tirole (2006). Neither does it cover theories that do not employ a functional form like Konow's theory of economic fairness (1996). It also cannot attempt to give an exhaustive overview of the experimental evidence for social preferences in general or for any model in particular. Experiments are discussed briefly if they are relevant for understanding the historical context or motivation of a model, but such occurrences (or lack thereof) are no indication of an experiment's quality or overall significance.

The survey is divided into three main categories: Outcome-based models, reciprocity models, and other models that to not belong of either of the first two categories. Within the categories, the presentation is mainly chronological, but departures from a strict timeline occur to group related models together. This survey thus adds another chapter to the review literature on social behavior in economics, which includes topics such as distributive justice (Schokkaert 1992; Konow 2003), philanthropy (Andreoni 2006), or gift exchange (Fehr et al. 1998), to name just a few. Fehr and Schmidt (2006) and Cooper and Kagel (2009) give more general overviews on social preferences, while handbooks on behavioral economics (Gilad and Kaish 1986; Camerer 2003; Altman 2006) or experimental economics (Roth and Erev 1995) offer an even broader perspective on the role of social preferences in economics.

The remainder of the survey is structured as follows. Section 2.2 introduces the notation used throughout the text and briefly outlines the key experimental games on which much of the literature on social preferences relies. The following four sections review social utility models. First, section 2.3 presents work by economists and psychologists predating the paradigm shift in the 1980s. Next, section 2.4 and 2.5 survey outcome-based models and reciprocity models, respectively. All remaining models are presented in section 2.6. The critics of social preferences models have their say in section 2.7 before section 2.8 summarizes the status quo and attempts to forecast what lies ahead.

2.2 Preliminary Remarks

2.2.1 Notation

Hardly any two models use the same kind of notation, so keeping each model's original notation would create a lot of confusion. We therefore adjust recurring mathematical expressions to fit into the same notational framework. Model-specific variables, however, usually remain as in the original paper unless overlapping requires renaming. Generally, $\{1, 2, \ldots, N\}$ denotes the set of individuals and $\{x_1, x_2, \ldots, x_N\}$ denotes an allocation of material payoffs out of the set X of possible allocations. Individual i has a social utility function if under any circumstances, his utility U_i is affected by another individual's payoff, i.e., formally

$$\exists\ \{x_i, x_{j\neq i}\}, \{x_i, x'_{j\neq i}\} \in X : U_i(x_i, x_{j\neq i}, s) \neq U_i(x_i, x'_{j\neq i}, s), \tag{1}$$

with $x_{j\neq i} \neq x'_{j\neq i}$ and s denoting the vector of relevant circumstances. For outcome-based models, the vector s is irrelevant, so the existential quantification (1) simplifies to $U_i(x_i, x_{j\neq i}) \neq U_i(x_i, x'_{j\neq i})$. For reciprocity models, s usually contains some measurement of kindness of j, which can for example depend on i's beliefs about j's intentions or inherent traits or on observed behavior of j. In other models, the relevant element of s might be a social norm, self perception, observability, etc.

2.2.2 Games

Over the last decades, experimental economists have devised a myriad of different games to analyze certain aspects of social preferences. However, the most fundamental games in the literature are the following: The ultimatum game, the dictator game, the public goods game, the prisoner's dilemma, the trust game, and the gift-exchange game. The following paragraph briefly introduces the rules of these games.

In the ultimatum game, a proposer i proposes an allocation $\{x_i, x_j\}$ of an endowment $E > 0$ between himself and a receiver j so that $x_j = E - x_i$. If the receiver accepts, the proposed allocation is implemented, otherwise both players receive nothing. The only subgame-perfect Nash-equilibrium is $\{E - \varepsilon, \varepsilon\}$, *accept*, i.e., the proposer offers the receiver the smallest possible positive amount and the receiver accepts. In the dictator game, the dictator i also chooses an allocation $\{x_i, x_j)$ of an endowment $E > 0$ between himself and the receiver j, but the allocation is immediately implemented without any reaction by the receiver. Purely self-interested dictators always keep the whole endowment.

In the standard public goods game, $N \geq 2$ players receive an endowment E which they can invest into the public good or keep for themselves. The total invested amount is multiplied with $m > 1$ and distributed evenly among all N players. The public goods game constitutes a social dilemma because it is individually optimal to invest nothing, but socially optimal to invest the whole endowment. Another social dilemma game is the prisoner's dilemma, which is a two player normal-form game in which each player has two strategies, *cooperate* and *defect*. *Defect* strictly dominates *cooperate*, i.e., it yields a higher payoff irrespective of the other player's strategy, yet each player's payoff from mutual cooperation is higher than from mutual defection.

In the trust game, the investor i and the trustee j initially receive an endowment E. In the first period, the investor can send any amount of money up to E to the trustee. The amount is then multiplied with a factor $m > 1$ and added to the trustee's endowment. In the second period, the trustee can send any amount of money back to the investor. Standard theory predicts that in the second period, j keeps all money for himself and in anticipation of that behavior, i does not send any money in the first period. Finally, the gift-exchange game is essentially a variant of the trust game in which the efficiency gains are induced by the second mover's action. In the standard version, buyers make price offers for a good and upon acceptance, sellers make a costly effort choice determining the good's quality and thereby the buyer's payoff. The game comes in a simple two-player version and a competitive market version, but the standard prediction always remains that the sellers choose the lowest possible effort level and buyers offer the lowest possible price.

2.3 Early Social Preference Models

2.3.1 Early Works from Economics

Back in 1881, Edgeworth devotes a section in the appendix of his *Mathematical Psychics* to "mixed modes of utilitarianism", noting that between pure selfishness and pure utilitarianism there are those "for whom ... his neighbor's happiness as compared with his own neither counts for nothing, not yet 'counts for one', but counts for a fraction."[5] He also explicitly expresses this idea earlier in a footnote commenting on the utilitarian arrangement on the contract curve, remarking that "if contractors [are] in a sensible degree *not* 'economic' agents, but actuated ... by a sympathy with each other's interests ... we might suppose that the object which X ... tends ... to maximize, is not P [own utility], but P + $\lambda\Pi$ [other's utility]; where λ is a *coefficient of effective sympathy*" (emphasis as in the original).[6] Edgeworth continues to describe the effects of such preferences on the contract curve, which remains the same but with narrower limits. As λ approaches 1, preferences become perfectly utilitarian and the contract curve collapses into a single point.

Following the paradigm shift toward the selfish model propelled by Pareto and his followers (see Bruni and Sugden 2007), Edgeworth's remarks are mostly forgotten until Frisch (1971) and Collard (1975) resume and extend Edgeworth's original idea.[7] Until the 1970s, however, interdependent utility functions largely disappear from economics, with Johnson (1952), who looks at the effects of relative income on consumption, one of the few noticeable exceptions. At this point, most economists who formally deal with social preferences typically keep utility functions nonspecific and discuss only indifference curves.[8] Some of the more influential examples of such endeavors include the works of Schwartz (1970), Scott (1972), and Becker (1974).

Schwartz notices that even when person i, ceteris paribus, prefers that person j consumes more, person i's behavior in equilibrium under budget constraint does not necessarily have to reflect that positive attitude. A similar point is stressed by Scott: Even when individual i generally prefers the

[5] Appendix IV, page 102.

[6] Page 16.

[7] It is unclear whether Frisch was aware of Edgeworth's remarks as his article (written in German) does not mention the *Mathematical Psychics*.

[8] While this can provide certain insights, it is generally a limited approach because it is difficult to derive predictions from indifference curves, especially when the parameters used to estimate them change.

other enjoying an equal level of consumption when holding constant own consumption, i is nevertheless unwilling to implement an equal allocation given a fixed budget.[9] Becker examines a model of household behavior in which individual i's utility depends both on own welfare and on the welfare of i's spouse. In equilibrium, i transfers income to the spouse until i's marginal utilities from both own and the spouse's consumption are equal. Therefore, a change in distribution of household income has no effect on the consumption and welfare of either member as long as i is still willing to make transfers. Becker also discusses welfare implications if individuals are envious, i.e., if they are worse off the higher the other person's payoff and are therefore willing to incur costs to harm others.[10]

Other approaches even break with traditional utility theory. For example, Margolis (1982) proposes that individuals have two utility functions S and G; one representing self-interest and one representing welfare concerns for the individual's social group. The allocation of resources to both domains occurs such that the ratios of marginal utilities in the group domain G' and in the self-interest domain S' are equal to a weight W. The weight W itself is a function of the ratio of the resources allocated to the group and to the self, respectively, so that the weight on group welfare increases the more resources are allocated to the self and vice versa. Margolis uses the model to solve traditional puzzles of economics such as public good provision or voting.

Public good provision is also the main concern of Sugden's model (1984). The model subjects utility maximization to a constraint which Sugden calls the *principle of reciprocity*. It demands that an individual contributes either the minimum of all other contributions to the public good or the amount the individual would most prefer everyone else to contribute, whichever is lower.

2.3.2 Early Works from Other Disciplines

Much of the initial groundwork on social utility functions is laid by psychologists and sociologists whose analysis is not manacled by the assumption of narrowly self-interested behavior. At the outset, Festinger (1954) demon-

[9]Such indifferent curves are created by Bolton and Ockenfels' ERC-model (2000), for example.

[10]Contrary to models in which envy is only experienced toward others whose payoffs exceed one's own, envy in Becker's model increases in the other's income irrespective of the relative standing.

strates that individuals engage in social comparison to form judgments, although he focuses only on the evaluation of opinions and abilities. Later, Adams (1963, 1965) develops equity theory, which assumes that workers dislike not being compensated equitably compared to their co-workers and therefore strive to maintain the following relation

$$\frac{\text{own outcome}}{\text{own input}} = \frac{\text{other's outcome}}{\text{other's input}}. \tag{2}$$

Arguably the first concrete social utility function after Edgeworth is proposed by Wyer (1969), who estimates individual utility for outcomes in several normal-form two-person-games. He propose that people maximize a utility function given by

$$U_i = w_1 \delta x_i + w_2(1-\delta)x_i + w_3 x_j, \tag{3}$$

where δ is a step function equaling 1 if $x_i > 0$ and 0 if $x_i \leq 0$. The two weights on own payoff $w_1 > 0$ and $w_2 > 0$ allow for loss aversion and the weight w_3 on other's payoff may be positive (altruistic), zero (individualistic), or negative (competitive). However, Wyer only looks at strategic games in which own behavior strongly depends on expectations about other's behavior and in which reciprocity concerns are very likely to play a crucial role. He finds a fairly linear correlation between outcomes and their desirability, but the model's explanatory power is rather weak even in its specific context. Nevertheless, Wyer's model can in principle explain both positive transfers in dictator games (if $w_3 > 0$) and rejections in ultimatum games (if $w_3 < 0$), though not individuals who do both.

A more elaborate utility function emerges from the literature on social value orientations (Messick and McClintock 1968; Griesinger and Livingston Jr. 1973). This theory assumes that individuals can have five different motivations when making decisions over own and other's payoffs; the desire to maximize the payoff of others (altruism), maximize joint payoff (cooperation), maximize own payoff (individualism), maximize own payoff relative to other's payoff (competition), or minimizing other's payoff (aggression or sadism).[11] Motivations can be elicited and mapped onto a circle in an own-other outcome space similar to Figure 1, with the resulting motivational

[11] Griesinger and Livingston Jr. (1973) initially also include the desire to minimize joint payoff (sadomasochism), minimize own payoff (masochism) and minimize own payoff relative to other's payoff (martyrdom). However, such motivations are virtually non-existing and are therefore usually dropped later.

vector determining the individual's type. Based on that framework, Lurie (1987) suggests the following utility function:

$$U_i = x_i \cos\alpha + x_j \sin\alpha, \tag{4}$$

where α is the angle of the motivational vector to the horizontal line. When choosing between two allocations, an individual with such preferences picks the allocation with the longer orthogonal projection onto his motivational vector. Additionally, Lurie discusses a social orientation involving both utilitarian and egalitarian concerns and eventually proposes a function that includes both an allocation's projection onto the motivational vector and the distance to it:

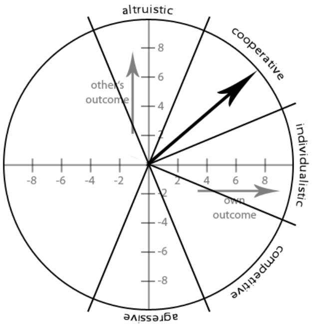

Figure 1: Social Value Orientation Classifications

$$U_i = f(P) - \delta k_1 E^{n_1} - (1-\delta) k_2 E^{n_2}, \tag{5}$$

where $n_1, n_2, k_1, k_2 > 0$. For an allocation $\{x_i, x_j\}$, the projection P is $x_i \cos\alpha + x_j \sin\alpha$ and the distance E is $|x_i \sin\alpha - x_j \cos\alpha|$. The step function δ equals 1 if $\{x_i, x_j\}$ lies above the motivational vector and 0 if it lies below, allowing for different sensitivities toward favorable and unfavorable deviation from the preferred vector. Despite the originality of Lurie's approach, it has not received much attention from economists or psychologist.

In a more influential study, Loewenstein et al. (1989) estimate several social utility functions using subjects' stated satisfaction with different hypothetical outcomes for the self and a partner while manipulating relationship background and status between treatments. The function with the best fit depends on the differences in payoffs and allows for different weights on advantageous and disadvantageous inequality:

$$\begin{aligned} U_i = c + &b_1 x_i + b_2\ max[x_j - x_i, 0] + b_3\ (max[x_j - x_i, 0])^2 + \\ &b_4\ max[x_i - x_j, 0] + b_5\ (max[x_i - x_j, 0])^2 \end{aligned} \tag{6}$$

On average over all treatments, utility decreases as the difference between payoffs increases ($b_k < 0$ for $k = \{2, 4\}$), but with diminishing sensitivity ($b_k > 0$ for $k = \{3, 5\}$) and disadvantageous deviations looming larger than advantageous deviations ($b_2 < b_4$). In some treatments, however, utility actually increases on average when receiving more than the partner ($b_4 > 0$), namely when the relationship is framed as negative or as a business relationship. Generally, while disadvantageous deviations are disliked across-the-board, individual reactions to advantageous deviations are much more context-dependent. Although the study has several shortcomings[12], it foreshadows the development of inequity aversion models and adds to their validity.

2.4 Outcome-Based Models

In the late 1970s, economists become increasingly interested in bargaining research (Selten 1978; Roth et al. 1981; Rubinstein 1982) after a series of experiments conducted by social psychologists (e.g. Fouraker and Siegel 1963; Nydegger and Owen 1974; Rapoport et al. 1977) appears to at least partly contradict the game-theoretic bargaining model by Nash (1950). Economists initially try to reconcile the experimental results with the theory by focusing on the availability of information to the bargainers (Roth and Malouf 1979; Roth and Murnighan 1982), but the importance of "sociological factors" is already hinted at by Roth et al. (1981). In 1982, Güth et al. publish the results of their ultimatum game experiments, which triggers a series of follow-up studies (e.g. Binmore et al. 1985; Neelin et al. 1988; Ochs and Roth 1989), eventually leaving economists with several systematic deviations from standard theory:

- Proposers' offers are typically closer to the equal split than predicted by subgame perfect equilibrium.
- A high number of positive but unequal offers is rejected.
- In two-period games with alternating roles and individual discount rates
  - counteroffers often leave the second-round proposer with a lower payoff than the previously rejected offer.
  - the discount rate of the first-round proposer influences the outcome, even though it should be irrelevant.
  - the first-round proposer consistently receives at least as much as the second-round proposer, even if the first offer is rejected.

[12]Most notably, the study lacks any monetary incentives, which raises social desirability and demand effect concerns.

While some ascribe the results to bounded rationality (see section 2.7), others instead call the assumption of material self-interest into question. For example, Ochs and Roth (1989) speculate that subjects might have a monetary threshold and refuse all lower offers. However, that explanation does not account for disadvantageous counteroffers, prompting them to vaguely suggest to include some measurement of unfairness as deviation from the equal division into the utility function.

Ochs and Roth's suggestion is eventually implemented by Bolton (1991). In his model, individuals care about their relative payoff as well as their monetary payoff, i.e.

$$U_i = U_i(x_i, p), \tag{7}$$

where p is the proportional index, which is 1 if both individuals receive nothing and $\frac{x_i}{x_j}$ otherwise.[13] According to the model's assumption, U_i is strictly increasing in own monetary payoff x_i and strictly increasing in p as long as $p < 1$ (if $p \geq 1$, $U_{i2} = 0$). In other words, ceteris paribus, individuals prefer more money over less, prefer to be closer to the equal split when they receive less than the other, and do not care about changes in relative payoff as long as they receive at least as much as the other. Notably, the model is able to explain all five puzzles from the ultimatum game experiments. In particular, disadvantageous counteroffers can now be explained by subjects trading off lower monetary payoff for higher relative payoff.[14]

Kirchsteiger (1994) proposes an alternative explanation. He assumes that people are envious such that their utility is strictly decreasing in other's payoff, i.e., $\partial U_i / \partial x_j < 0$. However, people care more about their own payoff than the other's, i.e., $U_i(x, x) > U_i(0, 0) \; \forall \, x > 0$. In other words, any positive equal split is better than zero payoff for both players. Kirchsteiger initially suggests the utility function

$$U_i = x_i - a x_j, \tag{8}$$

[13]Bolton primarily develops his model in the context of a two-period alternating-offer bargaining game and his notation is geared toward this scenario. The simplified notation accounts for all central aspects of his model, but is more accessible when coming from a general point-of-view.

[14]Bolton also extends his model further to a game where individuals compare their payoff not with their direct bargaining partner, but with another person in the same role. However, the extension is very specific and relatively complex, so it is skipped here. A similar idea is later formalized by Ho and Su (2009).

with $0 < a < 1$, which only explains four of the five ultimatum game puzzles. He then adds the assumption that the envy parameter a is actually a function of own payoff x_i with $a(0) < 1$, $\partial a(x_i)/\partial x_i < 0$, and $a(x_i) > 0 \; \forall \; x_i \geq 0$. As own payoff increases, the negative weight on the other's payoff decreases, although it never changes signs. This addition allows Kirchsteiger to explain the first mover advantage and, additionally, that offers with a given ratio of own to other's payoff are less likely to be rejected as the stakes are increased.[15]

Shortly thereafter, the dictator game experiments by Forsythe et al. (1994) further substantiate the assumption that many individuals are motivated by other-regarding concerns, yet they also bring along a new puzzle: The same people who are willing to pay to *decrease* others' payoffs as receivers in the ultimatum game are also willing to pay to *increase* others' payoffs as dictators, with some transferring up to half of their endowment. Unable to explain the dictators' behavior, Bolton's and Kirchsteiger's models are already rendered obsolete. Yet the puzzle eventually gives rise to two models of inequity aversion which refine the concepts of the two early models to reconcile ultimatum game rejections with dictator game giving. The first of these models is the theory of fairness, competition, and cooperation by Fehr and Schmidt (1999). Fehr and Schmidt assert that individuals experience self-centered inequity aversion, i.e., they suffer when another person's payoff differs from theirs, but do not care about inequity among others' payoffs. In the N-player case, the model's utility function is

$$U_i = x_i - \alpha_i \frac{1}{N-1} \sum_{i \neq j} max[x_j - x_i, 0] - \beta_i \frac{1}{N-1} \sum_{i \neq j} max[x_i - x_j, 0], \quad (9)$$

with α the strength of disadvantageous inequity aversion (envy) and β the strength of advantageous inequity aversion (compassion). Fehr and Schmidt assume $\alpha \geq \beta$, i.e., envy looms larger than compassion, and $0 \leq \beta < 1$, i.e., people do not destroy their own payoff to decrease their costs of compassion. The selfish case is nested within the model for $\alpha = \beta = 0$.

The second model of inequity aversion is the theory of equity, reciprocity, and competition (ERC) by Bolton and Ockenfels (2000). Like Bolton's comparative model of bargaining, it is based on monetary and relative payoff, but ERC makes some crucial modifications. It assumes that individuals maximize

[15] The empirical evidence for the latter is actually relatively weak, see for example Hoffman et al. (1996), Slonim and Roth (1998), or Cameron (1999).

a "motivation function" given by

$$U_i = U_i(x_i, \sigma_i), \tag{10}$$

where σ_i is i's relative share of the total payoff $\bar{X}$ of all N individuals, i.e., $\sigma_i = x_i/\bar{X}$ or $\sigma_i = 1/N$ if $\bar{X} = 0$. The function is increasing in x_i, i.e., individuals prefer more money over less when holding their relative share constant. Furthermore, U_i is concave in σ_i and has its maximum at $\sigma_i = \frac{1}{N}$, i.e., individuals prefer to receive the equal split when holding constant own monetary payoff. The main differences to the Fehr-Schmidt-model are the following: The ERC-model is an incomplete information model, individuals do not care about the distribution of others' payoffs, and the model is silent on differences between advantageous and disadvantageous inequality.[16] To help intuitive understanding of their model, Bolton and Ockenfels also offer an exemplary function for N-player games:

$$U_i = a_i x_i - \frac{b_i}{2}\left(\frac{x_i}{\bar{X}} - \frac{1}{N}\right)^2, \tag{11}$$

with $a_i \geq 0$ and $b_i \geq 0$.[17] For $a/b = 0$, individuals are strictly relativistic and for $a/b \to \infty$ strictly self-interested.

What distinguishes both models is that they are not just consistent with dictator and ultimatum games, but with a wide range of different experimental results, including altruistic punishment in public goods games (Ostrom et al. 1992; Fehr and Gächter 2000), reciprocal responses in the trust game or the gift-exchange game (Fehr et al. 1993; Berg et al. 1995), and, in particular, seemingly self-interested behavior in competitive market environments (Smith 1962; Roth et al. 1991; Kachelmeier and Shehata 1992). There is, however, one systematic behavioral pattern that both models cannot explain. In dictator games with efficient giving, i.e., when giving up x increases the other's payoff by more than x, a significant number of people are willing to

[16] The irrelevance of the distribution of others' payoffs is motivated by the results of three-person ultimatum games conducted by Güth and Van Damme (1998) and results of the solidarity game by Selten and Ockenfels (1998). In the former, subjects show little concern for the inactive third player, and in the latter, voluntary giving to "losers" with low payoffs is largely independent of the number of "losers", suggesting that individuals are primarily concerned with the fairness of their own payoff.

[17] Individuals with these preferences will keep $(0.5 + a/b)$ in the dictator game and reject any offer lower than 0.5 - $\frac{\sqrt{ab+a^2}-a}{b}$ in the ultimatum game.

increase the other's payoffs even when it increases disadvantageous inequity (Charness and Grosskopf 2001; Kritikos and Bolle 2001).

Such behavior is particularly evident in a study by Andreoni and Miller (2002), who examine whether people's social preferences are well-behaved, i.e., if they adhere to the generalized axiom of revealed preferences (GARP[18]) and if they can be represented by a continuous, convex, and monotonic utility function. They let subjects play modified dictator games with varying endowments and transfer rates. The resulting budget lines cross to allow potential violations of GARP. Nearly half of their subjects are perfectly described by selfish, Rawlsian, or utilitarian preferences. To estimate preferences for the remaining subjects, they employ a social utility function with constant elasticity of substitution (CES)

$$U_i = (a x_i^{\rho} + (1-a) x_j^{\rho})^{\frac{1}{\rho}}, \tag{12}$$

where a is a measure of selfishness and ρ captures the convexity of preferences. The function allows for perfectly selfish ($a = 1$), perfectly Rawlsian ($a = 0.5$; $\rho = -\infty$), and perfectly utilitarian ($a = 0.5$; $\rho = 1$) preferences as well as mixed types. Andreoni and Miller conclude that people have heterogeneous but typically well-behaved preferences as less than 2% of their subjects violate GARP, demonstrating that altruism is in principle compatible with neoclassical preferences.

Around the same time, Charness and Rabin (2002) design a range of experiments to directly compare social preference models and conclude that welfare concerns explain behavior better than inequality aversion because reducing others' payoff to reduce disadvantageous inequality appears to be rare, even when relatively cheap. They propose a social utility function which includes concerns for the individual with the lowest payoff as well as concerns for efficiency,

$$\begin{aligned} U_i = &(1-\lambda)x_i + \\ &\lambda \left[\delta \cdot min[x_1, x_2, \ldots, x_n] + (1-\delta) \cdot (x_1 + x_2 + \ldots + x_n)\right], \end{aligned} \tag{13}$$

[18]To explain GARP, it is first required to define the terms *directly revealed preferred* and *indirectly revealed preferred*. If B is in the choice set when A is chosen, A is *directly revealed preferred* to B. If A is *directly revealed preferred* to C, and C is *directly revealed preferred* to B, then A is *indirectly revealed preferred* to B. GARP says that if A is *indirectly revealed preferred* to B, then B is not *strictly directly revealed preferred* to A, i.e., A is not strictly within the budget set when B is chosen. GARP allows for non-strict convexity of preferences. For a more detailed discussion, see the article by Andreoni and Miller.

where $0 \leq \lambda \leq 1$ measures the degree of social welfare concern versus self-interest and $0 \leq \delta \leq 1$ the degree of Rawlsian concern (maximin) versus efficiency concern. The model generally predicts transfers of own payoff to those worst off, but can also explain transfers to those better off when such a transfer is efficiency increasing.

Straightaway, researchers set out to test the predictive power of inequity aversion against social welfare concerns (e.g. Güth et al. 2003; Engelmann and Strobel 2004), spurring on a spirited debate between advocates for both sides (Fehr et al. 2004; Bolton and Ockenfels 2006; Engelmann and Strobel 2006). To some degree, the issue is still unresolved as the controversy about relative importance of each objective continues (e.g. Daruvala 2010; Blanco et al. 2011).

From this point on, most new outcome-based models are just modifications of one of the three major models (Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Charness and Rabin 2002). Typically, the original function receives at least one additional component, which allows the new model to better explain certain experimental data, albeit at the cost of additional free parameters. For example, Frohlich et al. (2004) add the concepts of "just desert" and "doubt" to the Fehr-Schmidt-model. When the dictator's endowment is produced by both players, the "just desert" is the proportion of the endowment generated by each player. "Doubt" is the degree to which the dictator doubts the authenticity of the game.[19] A dictator's utility function is

$$\begin{aligned} U_i = x_i - \alpha\ max[x_j - x_i, 0] - \pi\ \beta\ max[x_i - x_j, 0] \\ - \gamma\ max[p_i - x_i, 0] - \pi\ \psi\ max[p_j - x_j, 0], \end{aligned} \tag{14}$$

where p_i is player i's just desert, $\gamma > 0$ and $\psi > 0$ the weight on negative deviation from own and other's just desert, respectively, and π the "doubt" discount with $0 \leq \pi \leq 1$.

Another extension is undertaken by Tan and Bolle (2006), who again look at dictator games with varying transfer rates to disentangle altruistic and fairness concerns. They employ an ERC-type function which depends exclusively on relative payoffs and add an altruism term to the function. They arrive at the following utility function

$$U_i = \frac{x_i}{E} + a\frac{x_j}{E} - \frac{b}{2}\left(\frac{x_i}{E} - \frac{x_j}{E}\right)^2, \tag{15}$$

[19]For example, the dictator may doubt that there really is a recipient or that he really is responsible for his portion the endowment.

where $a < 1$ and $b \geq 0$ are weights on altruistic concern and fairness concern, respectively. E is the normalized endowment, which is given by $E = \frac{E_1 + tE_1}{2}$, where t is the transfer rate and E_1 is the initial endowment.[20] Tan and Bolle conclude that their model describes their experimental data better than altruism and fairness concerns alone.

Furthermore, Ho and Su (2009) look at peer-induced fairness concerns by sequentially playing two independent ultimatum games with one proposer and two responders. They find that the second responder is averse to receiving less than the first receiver. This result prompts Ho and Su to propose a utility function for the second responder k given by

$$U_k = \begin{cases} x_k - \delta \; max[0, x_i - x_k] - \rho \; \hat{p}(z) \; max[0, \hat{x}_j - x_k] & \text{if offer is accepted.} \\ 0 & \text{if offer is rejected.} \end{cases} \tag{16}$$

x_i is proposer i's payoff from the ultimatum game played with k, z is the imperfect signal of i's previous offer to the first responder j, $\hat{p}(z)$ is k's subjective probability that j accepted i's offer, and $\hat{x}_j$ is k's expectation about j payoff given the signal and acceptance of the offer. $\delta \geq 0$ and $\rho \geq 0$ represent k's strength of envy toward i and j, respectively. Removing uncertainty, function (16) simplifies to

$$U_k = x_k - \delta \; max[0, x_i - x_k] - \rho \; max[0, \hat{x}_j - x_k], \tag{17}$$

which is essentially the Fehr-Schmidt-model with different α-parameters toward proposer and second receiver.

Chen and Li (2009) analyze the results of their experiment on in-group out-group differences of social preferences by taking Charness and Rabin's two-person utility function and adding group identity to it, arriving at

$$U_i = (1 - \rho r - \sigma s)x_i + [\rho(1 + Ia)r + \sigma(1 + Ib)s]x_j, \tag{18}$$

where r and s are a step functions that are 1 if $x_i > x_j$ and if $x_i < x_j$, respectively, and 0 otherwise. ρ and σ are weights put on the other's payoff compared to own payoff when i is ahead and behind, respectively. I is also

[20] A transfer rate of t means that for each point the dictator gives up, the receiver's payoff increase by t. For $t = \infty$, the dictator's payoff is always 0 and the recipient's payoff lies between 0 and 8. For $t = 0$, it is the other way around.

a step function that is 1 if players i and j belong to the same group and 0 otherwise. Hence, through a and b, people can put different weights on the other's payoff depending on group affiliation. Chen and Li then use their data to estimate the parameters, finding that subjects generally put a significantly higher weight on the other's payoff if the two players belong to the same group, both in distribution games and in a reciprocity game.

Before we conclude this section, let us be reminiscent of two models that essentially were already obsolete upon their publication. Some theories, such as Bolton's and Kirchsteiger's, have largely sunken into oblivion, but they constitute important interim stages in the development of more complete models and provided original concepts that advanced the academic discourse. In contrast, the following two models have had very little impact on the field. They are presented here to illustrate that the research field at that time was wide open. Even though in hindsight, the ideas may appear idiosyncratic, at the time they were considered worthwhile enough to warrant publication.

Building primarily on ultimatum game experiments, Burnell et al. (1999) suggest that players view each other as rivals and, ceteris paribus, prefer to maximize the distance between own and other's payoff. The following utility function represents these preferences:

$$U_i = x_i + a(x_i - x_j), \tag{19}$$

with $a \geq 0$. It follows that responders in the ultimatum game reject an offer below the equal split if the payoff difference becomes too large. Proposers make higher offers than predicted by standard theory to avoid rejections, although 50-50 offers require very strong risk aversion. Burnell et al. are already aware of the dictator game results, which are incompatible with their model because it predicts that dictators always keep the whole pie. They therefore argue that players try to build a positive reputation for future interaction (even though games are played anonymously) and adjust their model to include expected payoffs from these future interactions.

Costa-Gomes and Zauner (2001), too, are predominantly concerned with two-player bargaining games. In their model, a player's utility consists of a linear combination of both player's payoffs and a noise term, i.e.,

$$U_{i,k} = x_{i,k} + a_i x_{j,k} + \varepsilon_{i,k}, \tag{20}$$

where k is the terminal node of the game, $\varepsilon_{i,k}$ an independent (across players and nodes) random variable with mean 0, and $a_i \in \mathbb{R}$ player i's weight on player j's payoff. Costa-Gomes and Zauner use data from Roth et al. (1991) to estimate a_i, but their estimates postulate that most people are spiteful (i.e., $a_i < 0$), which is inconsistent with dictator game results.

2.5 Reciprocity Models

Reciprocity models are based on the assumption that people want to pay like with like, i.e., be kind toward others who treat them kindly and unkind toward others who treat them unkindly. Reciprocity is not, however, based on expectations of future material benefits. Compared to outcome-based models, the development of reciprocity models sees less dramatic change over time because the main experimental evidence for reciprocal behavior is relatively consistent. Five experimental games constitute the main body of evidence. First, many people make positive contributions in one-shot public good games, but their willingness to contribute is typically not driven by altruism, but contingent on their beliefs about other people's behavior (reviews in Dawes and Thaler 1988; Ledyard 1995; Chaudhuri 2011). Second, rejections in the ultimatum game can be viewed not (only) as an expression of discontent with an unequal payoff distribution per se, but as punishment for an unkind offer (Falk et al. 2003). Third, in one-shot or finitely repeated prisoner's dilemma games, a large portion of people cooperate as long as their partner also cooperates (Krebs et al. 1982; Andreoni and Miller 1993). Fourth, in the gift-exchange game by Fehr et al. (1993), responders on average choose higher efforts the higher the buyer's offer. And finally, in the trust game, the trustee's back transfer typically increases in the investor's initial transfer (Berg et al. 1995).

Among the different approaches proposed to capture utility from reciprocity, the earliest concepts are based on intentions.[21] Standard utility theory assumes that an action is only judged depending on its consequences, not on the intentions behind it. Intention-based models do away with that

[21] There is some ambiguity in the word *intention*. Sometimes it describes what a person plans to do, sometimes what the person plans to achieve. The difference can be relevant for how others perceive the person's kindness. For example, a proposer in the ultimatum game may offer the equal split because she likes to be fair or because she fears rejection and would actually prefer to keep the whole pie. According to the first interpretation, the proposer's offer is kind in both cases, whereas according to the second interpretation, only the first proposer is kind. The intention-based models in this section tend to follow the first interpretation.

assumption and therefore constitute a more radical departure from economic tradition than outcome-based models. To incorporate intentions into the utility function, the models employ psychological game theory (Geanakoplos et al. 1989; Battigalli and Dufwenberg 2009). In psychological games, utility depends on payoffs and on own beliefs about others' strategies and their beliefs. If all criteria of a Nash equilibrium are fulfilled and in addition, all beliefs match actual behavior, it constitutes a so-called psychological Nash equilibrium.

The first intention-based reciprocity model is developed by Rabin (1993) for normal form two-player games like the prisoner's dilemma. A player receives subjective utility depending on his strategy a_i, his first order beliefs b_i about the other player's strategy choice and his second order beliefs c_i about the other player's beliefs about his strategy. Rabin defines a "kindness function"

$$f_i(a_i, b_j) = \frac{x_j(b_j, a_i) - x_j^e(b_j)}{x_j^h(b_j) - x_j^{min}(b_j)}, \tag{21}$$

where $x_j^h(b_j)$ and $x_j^{min}(b_j)$ are the highest and lowest possible payoffs for player j, respectively, given belief b_j, and x_j^e is the "equitable payoff", i.e., the mean of $x_j^h(b_j)$ and the lowest not Pareto-dominated possible payoff for player j, again given belief b_j. Since the function is normalized, it ranges from -1 (most unkind) to $\frac{1}{2}$ (kindest). Player i's beliefs about player j's kindness is given by

$$\widetilde{f}_j(b_j, c_i) = \frac{x_i(c_i, b_j) - x_i^e(c_i)}{x_i^h(c_i) - x_i^{min}(c_i)}. \tag{22}$$

Finally, the players are maximizing the following utility function

$$U_i = x_i(a_i, b_j) + \widetilde{f}_j(b_j, c_i) \cdot [1 + f_i(a_i, b_j)]. \tag{23}$$

Therefore, if player i believes player j wants to give him less than the equitable payoff ($\widetilde{f}_j(\cdot) < 0$), he wants to give player j as little as possible (holding constant own monetary payoff) and vice versa. Since both kindness functions are bounded, behavior becomes less sensitive to reciprocal concerns the bigger the material stakes get. Rabin refers to psychological Nash equilibiria in his model as fairness equilibria and notes that any Nash equilibrium that is either mutual-max (each player maximizing the other's payoff) or mutual-min (each player minimizing the other's payoff) is also a fairness equilibrium. Roughly, if material payoff is small, all fairness equilibria are mutual-max or mutual-min and if material payoff is large, all fairness equilibria are Nash equilibria.

One noteworthy application of Rabin's model is undertaken by Dickinson (2000). Though not a normal-form game, he applies the model with minor formal adjustments to the ultimatum game to test it against alternative explanations (among others Ochs and Roth 1989 and Slonim and Roth 1998). The model predicts responder behavior well, although proposer behavior is only weakly supported.

Dufwenberg and Kirchsteiger (2004) extend Rabin's model to extensive form games with N players. They take into account how beliefs and strategies can change after the conclusion of a game node, so that strategies are best responses at all stages. Each player has a set of strategies A_i. $B_{ij} = A_j$ is the set of possible beliefs of player i about player j's strategy and $C_{ijk} = B_{jk} = A_k$ is the set of possible beliefs of player i about player j's beliefs about player k's strategy. H is the set of possible histories. Then, with $a_i \in A_i$ and $h \in H$, $a_i(h)$ is player i's updated strategy after history h. Analogously, $b_{ij}(h)$ and $c_{ijk}(h)$ are player i's updated beliefs after h. All updating is Bayesian.

Given player i's beliefs about all other players' strategies $(b_{ij})_{j \neq i}$, the set of feasible payoffs for player j from i's perspective is $\{x_j(a_i^{'}, (b_{ij})_{j \neq i}) | a_i^{'} \in A_i\}$. Then, the kindness of player i toward player j at history h when choosing strategy a_i is defined as

$$\kappa_{ij}(a_i(h), (b_{ij}(h))_{j \neq i}) = x_j(a_i(h), (b_{ij}(h))_{j \neq i}) - x_j^{e_i}((b_{ij}(h))_{j \neq i}), \tag{24}$$

with $x_j^{e_i}((b_{ij}(h))_{j \neq i})$ the "equitable payoff", i.e., the mean of the highest and lowest material payoff of player j in the feasible set ignoring inefficient strategies.[22] Analogously, player i's beliefs about player j's kindness toward him at history h are

$$\lambda_{iji}(b_{ij}(h), (c_{ijk}(h))_{k \neq j}) = x_j(b_{ij}(h), (c_{ijk}(h))_{k \neq j}) - x_j^{e_j}(b_{ij}(h), (c_{ijk}(h))_{k \neq j}). \tag{25}$$

[22]A strategy is inefficient if there exists another strategy that makes every player at least as well off in every history and at least one player better off in at least one history.

This leads to the utility function

$$U_i = x_i(a_i(h), (b_{ij}(h))_{j \neq i}) + \sum_{j \in N \setminus \{i\}} \left(Y_{ij} \cdot \kappa_{ij}(a_i(h), (b_{ij}(h))_{j \neq i}) \cdot \lambda_{iji}(b_{ij}(h), (c_{ijk}(h))_{k \neq j}) \right), \tag{26}$$

where $Y_{ij} \geq 0$ is player i's reciprocity sensitivity regarding player j. Dufwenberg and Kirchsteiger define a *sequential reciprocity equilibrium* which requires that each player in each history chooses optimally given his beliefs and that following each history, beliefs are updated correctly, and that players' initial beliefs are correct.

Another modification of Rabin's model is made by Falk and Fischbacher (2006). Like Dufwenberg and Kirchsteiger, they look at extensive form games with N players, but they also introduce distributional concerns into their model. In their two-player model, $\tilde{N}$ denotes the set of nodes, $\tilde{F}$ the set of endnotes, $\tilde{N}_i$ the set of nodes at which player i moves, and $A_{\tilde{n}}$ the set of actions at node $\tilde{n}$.[23] $P(A_{\tilde{n}})$ is the set of probability distributions over the set of actions in node $\tilde{n}$, $S_i = \prod_{\tilde{n} \in \tilde{N}_i} P(A_{\tilde{n}})$ is player i's behavior strategy space, and $s_i|\tilde{n}$ his strategy conditional on node $\tilde{n}$. The player's expected payoff at node $\tilde{n}$ can be defined as $x_i(\tilde{n}, s_i, s_k) := x_i(s_i|\tilde{n}, s_j|\tilde{n})$. Finally, s_i' are player i's beliefs about player j's behavior strategy s_j and s_i'' are player i's beliefs about player j's beliefs about player i's strategy. Beliefs are consistent if $s_i = s_j' = s_i''$. Beliefs are not updated, only initial beliefs enter utility.

One problem with both Rabin's model and the Dufwenberg-Kirchsteiger model arises from their definition of kindness, which is independent of the second player's payoff so that equitable allocations may be perceived as unkind. To address this problem, Falk and Fischbacher measure kindness differently. In their model, player i's kindness φ_i is composed of an outcome measurement Δ_i with $\Delta_i = x_j - x_i$ and an intention factor ϑ_i with $0 \leq \vartheta_i \leq 1$. If $\vartheta_i = 1$, player i's action is fully intentional.[24] The kindness term φ_i is the

[23] By adding the tilde, we deviate from the original notation of Falk and Fischbacher because N is used for the number of players throughout the text.

[24] Intentionality depends on the options player i has. For example, if player i makes a choice resulting in $x_i > x_j$, but another choice would have let to a higher payoff for player j, then the intention factor is 1. If the action is not intentional, the variables takes the value of the player's pure outcome concern parameter ε_j ($0 \leq \varepsilon_j \leq 1$), which measures player j's pure concern for equitable outcome. If $\varepsilon_j > 0$, it is possible, for example, that player j rejects an unequal ultimatum offer even if player i had a singleton choice set when making the offer.

product of Δ_i and ϑ_i, so the kindness at node $\tilde{n}$ is

$$\varphi_i(\tilde{n}, s''_j, s'_j) = \vartheta_i(\tilde{n}, s''_j, s'_j) \cdot \Delta_i(\tilde{n}, s''_j, s'_j). \tag{27}$$

Additionally, the model also contains a reciprocation term σ_i which measures how much player i alters the payoff of player j with his choice at node $\tilde{n}$ leading to end node $\tilde{f}$. It is given by

$$\sigma_i(\tilde{n}, \tilde{f}, s''_i, s'_i) = x_j(v(\tilde{n}, \tilde{f}), s''_i, s'_i) - x_j(\tilde{n}, s''_i, s'_i), \tag{28}$$

with $v(\tilde{n}, \tilde{f})$ the unique node that directly follows $\tilde{n}$ on the path to $\tilde{f}$. Player i's utility is then given by

$$U_i = x_i(\tilde{f}) + \rho_i \sum_{\substack{\tilde{n} \to \tilde{f} \\ \tilde{n} \in \tilde{N}_i}} \underbrace{\varphi_j(\tilde{n}, s''_i, s'_i)}_{\text{kindness}} \cdot \underbrace{\sigma_i(\tilde{n}, \tilde{f}, s''_i, s'_i)}_{\text{reciprocation}}, \tag{29}$$

where ρ_i is the degree of player i's reciprocity concern with $\rho_i \geq 0$. Ceteris paribus, players want to decrease the other player's payoff if the other player has been unkind or if the other player receives a higher payoff. If the other player has been kind or if the other player receives a lower payoff, players want to increase the other player's payoff. The subgame perfect psychological Nash equilibrium is called *reciprocity equilibrium*.

A different approach to reciprocity is proposed by Levine (1998), who assumes that people are inherently altruistic or spiteful and that altruistic people are generally treated more kindly by others. He postulates the following utility function

$$U_i = x_i + \sum_{j \neq i} \frac{a_i + \lambda a_j}{1 + \lambda} x_j, \tag{30}$$

where a_i is individual i's coefficient of altruism with $-1 < a_i < 1$ and λ the degree of reciprocal concern with $0 \leq \lambda \leq 1$. Individuals can be altruistic ($a_i > 0$), selfish ($a_i = 0$), or spiteful ($a_i < 0$). If $\lambda > 0$, individuals have a higher regard for altruistic individuals than for selfish and spiteful others. While Levine's model offers improved tractability compared to intention-based reciprocity models, it is difficult to reconcile with many empirical results. In the ultimatum game, for example, the majority of people appears to be spiteful, which contradicts results from dictator games and public goods games.

Sethi and Somanathan (2001) employ Levine's model to develop an evolutionary theory of reciprocity. However, because altruistic individuals can never put a negative weight on the payoff of selfish others in the original model, they adjust the utility function to

$$U_i = x_i + \sum_{j \neq i} \frac{a_i + \lambda_i (a_j - a_i)}{1 + \lambda_i} x_j. \tag{31}$$

Additionally, they allow $\lambda_i > 1$, so that altruists may be spiteful to selfish others and even to others with weaker altruism than themselves. Sethi and Somanathan then show that given certain environments and parameters, reciprocal altruists ($a_i > 0, \lambda > 0$) can survive in a population that also contains opportunists ($a_i = \lambda_i = 0$).

Rotemberg (2008) further modifies Levine's idea by assuming that people are mildly altruistic and additionally expect others to be at least mildly altruistic, too. If another person fails to be sufficiently altruistic, people become angry and enjoy punishing the deviant. Ex-ante, individual i puts weight $\lambda^i > 0$ on person j's payoff. However, he expects that j's weight on i's payoff λ^j is at least $\bar{\lambda}^i > 0$. People generally assume that this is the case. However if person j's actions reveal $\lambda^j < \bar{\lambda}^i$, person i will become angry, which decreases λ^i by a certain amount $\bar{\xi}$. This leads to the utility function

$$U_u = E(x_i + [\lambda^i - \xi(\hat{\lambda}^j, \bar{\lambda}^i)]x_j)^\gamma, \tag{32}$$

where E is the expectation operator, γ represents risk attitude, and $\hat{\lambda}^j$ is the highest possible λ^j consistent with j's actions. The function $\xi(\cdot)$ is zero if $\hat{\lambda}^j \geq \bar{\lambda}^i$ and $\bar{\xi}$ otherwise. The inclusion of γ allows for risk-loving preferences or "gambling utility" (Conlisk 1993), which allows the model to explain low offers by proposers in ultimatum games. However, risk-aversion cannot account for the large portion of 50-50 offers typical for ultimatum games; this requires people to be more altruistic themselves than they demand others to be. Rotemberg also remarks that his utility function allows large changes in attitude, which is supported by neurological findings (Singer et al. 2006).

The third approach to reciprocity is an extension of the outcome-based model of Charness and Rabin (2002), which captures both social welfare concerns and negative reciprocity.[25] In the extension, each player has a

[25]Some experiments suggest that intentions do not matter much for positive reciprocity (e.g. Bolton et al. 1998a; Offerman 2002; Cox 2004), although others disagree (e.g. Falk et al. 2008).

demerit profile $d = (d_1, ..., d_N)$ with $d_j \in [0, 1]$. The greater d_j, the less weight other players put on player j's payoff. The outcome-based utility function (13) becomes

$$U_i = (1-\lambda)x_i + \lambda \Bigg[\delta \cdot min\,[x_i, min_{j \neq i}\{x_j + bd_j\}] + \\ (1-\delta) \cdot \left(x_i + \sum_{j \neq i} max[1 - kd_j, 0]x_j \right) - f \sum_{j \neq i} d_j x_j \Bigg], \tag{33}$$

where b, k, and f are nonnegative parameters. Player j's demerit d_j is derived by comparing the "appropriate" weight put on social welfare λ^* with the highest possible weight g_j that is consistent with player i's chosen strategy given the profile of strategies and demerits of all other players, i.e., $d_j = max[\lambda^* - g_j, 0]$.[26] The strategy profile for which each players' strategy is optimal in expectations and for which a consistent demerit profile d exists is called *reciprocal-fairness equilibrium*.

Erlei (2008) also uses the Charness-Rabin approach to model reciprocity by combining the outcome-based version of the model with Fehr-Schmidt-type inequity aversion. However, players are not motivated by welfare concerns and inequity aversion at the same time; instead, they are of different types. In two-player games, players' preferences are represented by the following utility function

$$U_i = \begin{cases} (1 - \sigma_t - \theta_t r)x_i + (\sigma_t + \theta_t(\cdot)r)x_j, & x_i < x_j \\ (1 - \rho_t - \theta_t r)x_i + (\rho_t + \theta_t(\cdot)r)x_j, & x_i > x_j \\ (1 - \theta_t r)x_i + \theta_t(\cdot)rx_j, & x_i = x_j. \end{cases} \tag{34}$$

Each player's type t can be strictly egoistic (SE), i.e., $\sigma_{\text{SE}} = \rho_{\text{SE}} = \theta_{\text{SE}} = 0$, inequity averse (IA), i.e., $\sigma_{\text{IA}} < 0 < \rho_{\text{IA}} < 1$, or can have welfare preferences (WP), i.e., $0 < \sigma_{\text{WP}} \leq \rho_{\text{WP}} \leq 1$. Furthermore, $\theta_t \geq 0$ is the reciprocity parameter, which is determined by the type of player, the payoffs, and r, which is the reciprocity variable with $r = -1$ if the other player misbehaved and $r = 0$ otherwise. Player i regards player j as misbehaving if j's action violates i's norm and if i cannot obtain at least the same utility as when j

[26]The non-negativity of d_j rules out positive reciprocity.

had abode by i's norm. Norms are best responses in games in which both players have identical types. Erlei then applies his model to 43 games and generally receives better predictions than his three base models.

Finally, Segal and Sobel (2007) propose an axiomatic approach to reciprocity by assuming that players have preferences over strategies.[27] In their model, player i has a complete, continuous, independent, and transitive preference relation over his space of mixed strategies Σ_i. These preferences depend on the strategies of other players σ_{-i} and possibly on player i's interpretation of σ_{-i} or alternatively on the "context" of the game. Context is summarized in the mixed strategy profile $\sigma^* = (\sigma_i^*, \sigma_{-i}^*)$, which describes the "conventional" way the game is played or player i's beliefs about how the other players play.[28] Furthermore, players have preferences over outcomes that satisfy the assumption of expected utility theory and are independent of the strategic context. All players are self-interested in the sense that when holding others' material payoffs constant, they prefer higher monetary payoff. If $v_i(\sigma_i, \sigma_{-i}^*)$ is player i's utility from monetary payoff, then his preferences over strategies can be represented by the following utility function

$$U_i = v_i(\sigma_i, \sigma_{-i}^*) + \sum_{j \neq i} a_{i,\sigma^*}^j v_j(\sigma_i, \sigma_{-i}^*), \tag{35}$$

where a_{i,σ^*}^j is the weight player i gives to player j's monetary utility depending on the entire strategy profile σ^*. The model nests most outcome-based models and can account for reciprocity based on actions. Without restrictions on (35), behavior may appear irrational or off-equilibrium to outside observers who only know players' preferences over payoffs. The set of potential Nash equilibria in the model is rather large, ruling out only dominated strategies.[29]

2.6 Other Models

Not all models are (solely) based on outcomes or reciprocity. The theories in this section add factors like social norms, social identity, self-image, or emo-

[27]For an axiomatic analysis of the Fehr-Schmidt model, see Neilson (2006).

[28]Unlike in psychological game theory, there are no higher-order beliefs in the context σ^*.

[29]For example, in 2x2 games with four different outcomes, all four pure strategy pairs can be Nash-equilibria even though in the standard theory, there could be at most two Nash equilibria in pure strategies.

tions to the utility function, although outcomes naturally remain pivotal. By including economically intangible variables like social identity, some of these models borrow strongly from psychological and sociological concepts while at the same time, they remain firmly rooted in the rationality framework. Many of these hybrid models are relatively isolated approaches outside the main academic discourse and therefore have not undergone much progression (yet). Whether they constitute groundbreaking ideas pulling down disciplinary walls or ill-fated interdisciplinary potpourris, neither fish nor fowl, still remains to be seen.

Largely independent of the contemporary ultimatum game and dictator game literature, Akerlof and Kranton (2000) develop a model around social identity, i.e., a person's sense of self. They assume that following the behavioral descriptions of the own social category affirms self-image, whereas violations evoke anxiety and discomfort, so that self-identity changes payoffs from different actions. In the model, there are sets of social categories C and prescriptions P which indicate appropriate behavior, physical characteristics, and other attributes of people belonging to the categories. Each person i assigns a social category out of C to all people, including herself. This mapping is denoted c_i.[30] They propose the utility function

$$U_i = U_i(a_i, a_{-i}, I_i), \tag{36}$$

where I_i is person i's identity, a_i is the vector of i's actions and a_{-i} the actions of all other individuals. Person i's identity is represented by $I_i = I_i(a_i, a_{-i}; c_i, \epsilon_i, P)$, where ϵ_i are i's characteristics. Person i will then choose the action a_i that maximizes her utility taking a_{-i}, c_i, ϵ_i, and P as given. Akerlof and Kranton then use this framework to discuss various identity-related behaviors like gender specific job choice, donations to one's alma mater, or advertising. In two separate papers, they extend the discussion to education (Akerlof and Kranton 2002) and contract theory (Akerlof and Kranton 2005).

In an approach that is more closely related to purely outcome-based models, Cox et al. (2007) introduce an emotional state θ into a CES utility function. The emotional state influences the marginal rate of substitution

[30] The category person i assigns to person j is not necessarily the same that person j assigns to himself.

between own and other's payoff.

$$U_i = \begin{cases} (x_i^\alpha + \theta x_j^\alpha)/\alpha, & \alpha \in (-\infty, 0) \cup (0, 1]; \\ x_i x_j^\theta, & \alpha = 0, \end{cases} \tag{37}$$

with α indicating convexity. The marginal rate of substitution is $\frac{\partial U_i / \partial x_i}{\partial U_i / \partial x_j} = \theta^{-1} \left(\frac{x_j}{x_i} \right)^{1-\alpha}$, so the willingness to pay for increasing the other's payoff is $\frac{1}{MRS}$ if $x_i = x_j$. The emotional state θ depends on the reciprocity variable r and the status variable s. The reciprocity variable r is the difference between the maximum payoff i can secure for himself after j's earlier choice and the payoff i would get in a "neutral" situation. The status variable s represents relative status (other than relative payoff), i.e., asymmetries in individuals' entitlements or obligations, e.g. due to age, job title, or exerted effort. The model can account for a relatively wide spectrum of phenomena, e.g. both efficiency concerns (if $\theta > 0$) and ultimatum game rejections (if f $r < 0$ leads to negative $\theta < 0$) can be explained by it.

A very specialized model originates from an experiment using a variant of the trust game. Charness and Dufwenberg (2006) compare subjects' beliefs and behavior with and without prior communication between investor and trustee. They observe that communication increases both the investor's beliefs in the trustee's trustworthiness and the likelihood that the trustee behaves trustworthily, prompting Charness and Dufwenberg to conclude that trustees are guilt averse, i.e., they try to not let others down (Baumeister et al. 1994).[31] Battigalli and Dufwenberg (2007) formalize guilt aversion in extensive-form games with the help of psychological game theory. In their setting, T denotes the set of nodes with $Z \subset T$ the set of endnotes. H_i is player i's information structure, which is a partition of T. The pure strategy s_i contains all choices of player i for each $h \in H_i$. Player i has first-order beliefs α_i about the other players' strategies, second-order beliefs β_i about other players' first-order beliefs, third-order beliefs γ_i about other players' second-order beliefs and so on. All beliefs are Bayesian. Given s_j and α_j at the initial information set h^0, player j has an expected material payoff $E_{s_j,\alpha j}$. The difference between this expected material payoff and actual material payoff at z is denoted $D_j(z, s_j, \alpha_j)$ and measures how much player j

[31] There is some evidence that people actually just have a preference for keeping their promises, but do not care much about disappointing others' beliefs (Vanberg 2008).

feels “let down”. Finally, $G_{ij}(z, s_{-i}, \alpha_j)$ measures how much of D_j is player i’s responsibility given s_{-i}, α_j, and z.

Within this framework, Battigalli and Dufwenberg propose two concepts of guilt. First, they assume player i experiences *simple guilt* when letting down player j, so i maximizes the expected value of the following utility function

$$U_i^{\mathrm{SG}} = x_i(z) - \sum_{j \neq i} \theta_{ij} G_{ij}(z, s_{-i}, \alpha_j), \tag{38}$$

where $\theta_{ij} \geq 0$ is player i’s guilt sensitivity. Second, they assume player i dislikes being blamed by others for letting them down. Given s_i and initial beliefs α_i^0 and β_i^0, i derives $G_{ij}^0(s_i, \alpha_i^0, \beta_i^0)$ which measures how much he expects to let j down ex-ante. At z, $E_{\alpha_j,\beta_j,\gamma j}[G_{ij}^0|H_j(z)]$ is j’s inference about how much i intentionally let j down. Disliking *guilt from blame*, player i’s utility is given by

$$U_i^{\mathrm{GB}} = x_i(z) - \sum_{j \neq i} \theta_{ij} E_{\alpha_j,\beta_j,\gamma j}[G_{ij}^0|H_j(z)]. \tag{39}$$

In two-player games with simultaneous moves and without chance, both concepts yield the same equilibria, but in other games, they may not.

Andreoni and Bernheim (2009) look at the dictator game from a novel perspective, assuming that people care about fairness to a certain degree, but also like to be perceived as fair. Concretely, dictators care about their own payoff x_i, their social image m as perceived by some audience A, and the difference between the receiver’s payoff x_j and the fair payoff x^F which they assume is 50% when both players are equally meritorious. Dictator i’s utility function is given by

$$U_i = F(x_i, m) + tG(x_j - x^F), \tag{40}$$

with t the weight on fairness. $F(x_i, m)$ is unbounded, strictly increasing, and concave in x_i. $G(x_j - x^F)$ is strictly concave and reaches its maximum at 0. Both functions are twice continuously differentiable. The distribution of t in the population is atomless with full support between 0 and $\bar{t}$ and is private information. The cumulative distribution function (CDF) of t is H. Social image m is normalized such that if A perceives i to put weight $\widehat{t}$ on fairness, m is $\widehat{t}$. Φ is the CDF representing A’s beliefs about $\widehat{t}$ and $B(\Phi)$ is the corresponding social image. For example, $B(\Phi)$ may be the mean

t given Φ. Hence, the dictator signals t to A through the transfer x_j. In equilibrium, i chooses the optimal transfer x_j such that given the mappings of types t to transfers x_j and transfers x_j to beliefs Φ, the transfer maximizes the dictator's utility.

Andreoni and Bernheim (2009) show that if the population contains sufficiently fair-minded players, there exists a pool at the equal split.[32] This is consistent with experimental evidence from dictator games where transfers typically exhibit a spike at 50% of the pie, but are non-existing slightly below 50%.

Both aversion of guilt from blame (Battigalli and Dufwenberg 2007) and Andreoni and Bernheim's model offer possible explanations for why people tend to behave more selfishly when their actions are not clearly attributable as several experiments have shown (e.g. Kagel et al. 1996; Güth et al. 1996; Dana et al. 2006, 2007). A player may want to "hide" behind outcomes that may also be due to chance, i.e., she implements an outcome that does not allow observers to attribute the result to her actions with certainty. This way, the player can avoid negative inferences about her behavior while securing a higher material payoff than she would have implemented under full supervision in order to avoid blame or to appear fair.

The last model in this section is based on social norms. López-Pérez (2008) argues that people have internalized fairness or distributive justice norms and that their emotional reactions to violations of these norms compel them to certain behavior. He defines a norm ψ as a correspondence $\psi : h \to A(h)$, i.e., for any information set h of the game, action $a \in A(h)$ is consistent with the norm ψ if $a \in \psi(h)$. Principled players endure a cost when violating a norm they have internalized. These costs depend inversely on the number of transgressors, so violating the norm is less costly when many players do so. A principled player's utility at node z is given by

$$U_i = \begin{cases} x_i(z) & \text{if } i \in R(z), \\ x_i(z) - \gamma \cdot r(z) & \text{if } i \notin R(z), \end{cases} \tag{41}$$

where $\gamma > 0$ measures how intensely the player has internalized the norm,

[32]This result is primarily due to the violation of the single-crossing property of the dictator's preferences over fairness and social image. If $t_1 > t_2$, player 1 incurs lower costs than player 2 when raising his transfer if the transfer is below 50%. However, if the transfer is above 50%, he incurs higher costs. Hence, player 1's willingness to increase the transfer to improve his social image drops sharply at 50%, which therefore becomes a "natural boundary" for transfers.

$R(z)$ is the set of players who have respected the norm and $r(z)$ the number of players in $R(z)$. López-Pérez applies the model using a *efficiency and equity norm* and a *honesty and fairness norm* (López-Pérez 2012) and emphasizes that his model can explain various socials norms including dressing norms, codes of etiquette, etc. However, although the model can ex-post explain observed behavior by postulating a corresponding norm, it says very little about how a norm is chosen and therefore has difficulties predicting behavior in unprecedented situations.

2.7 The Case Against

Most people would probably not find it presumptuous to say that social preference models have established themselves in mainstream economics. Various journals publish theoretical or empirical papers on the topic and the number of researchers involved in such studies has grown tremendously over the last decades. To young researchers, it may seem inconceivable that not too long ago, many economists vehemently rejected the idea of interrelated utility and truly believed that "when self-interest and ethical values [...] are in conflict, much of the time, most of the time in fact, self-interest theory [...] will win" (Stigler 1980).

Accordingly, the prevalent early response to the ultimatum game puzzles was to explain them with bounded rationality. Because people do not engage in backward induction (Binmore 1987; Binmore et al. 1988) – so the argument goes – they first need to learn how to behave. Since the rejection of low offers is more costly for proposers than for responders, proposers learn more quickly to make higher offers than responders learn to accept low offers (Roth and Erev 1995; Gale et al. 1995). Players therefore arrive at a Nash-equilibrium that is not sub-game perfect. Eventually, however, the argument does not hold up too well. In the ultimatum game, recipients are faced with such a simple decision (that does not require backward induction) that it is inconceivable that they make mistakes contradicting their true preferences. Furthermore, it seems that in very simple two-stage games, players are generally able to predict others' behavior quite well (Suleiman 1996; Fehr et al. 1998). With the emergence of the even simpler dictator game, the argument eventually faded away.

A related argument claims that when subjects in the laboratory are confronted with an unfamiliar situation, they resort to general social norms to make their decisions.[33] However, these norms supposedly evolved in an environment of repeated interaction and are therefore ill-adapted for one-shot interactions like the ultimatum game (Binmore and Samuelson 1994; Binmore 1988). For example, a fairness norm may have evolved "as an efficient solution to the equilibrium selection problem" (Binmore and Samuelson 1999) which also requires a punishment mechanism to sustain it. That is why players "who are offered unfairly small amounts [...] feel resentful and [...] want to punish the proposer by refusing" (ibid.). However, following this argument, it becomes difficult to maintain that the rejection of an unfair offer constitutes something else than a revealed social preference. If players choose *rejection* when *acceptance* is available, then the mutual rejection payoff of zero is directly revealed preferred to the unfair offer. Labeling such behavior an error calls fundamental economic principles into question. The fact that such behavior is not an evolutionary stable strategy in an environment of anonymous ultimatum game interactions is irrelevant. Furthermore, subjects typically seem to be aware of the differences between one-shot and repeated interactions (e.g. Andreoni and Miller 1993; Fehr and Fischbacher 2003) and decrease their demand for fairness as it becomes more expensive (Andreoni and Vesterlund 2001), suggesting that they treat fairness as a normal good.

Most other objections against the assumption of social preferences have also eventually been refuted, including, for example, low stake size (Hoffman et al. 1996; Slonim and Roth 1998; Cameron 1999) or the representativity of student subjects (Fehr and Fischbacher 2002; Bellemare and Kröger 2003). Even Binmore and Shaked (2010) "acknowledge that the accumulation of experimental evidence can be regarded as an informal proof that such preferences exist". However, critics continue to raise methodological concerns. Primarily, Binmore and Shaked (2010) criticize that experimental and behavioral economists usually explain deviations from standard theory ex-post by modifying the self-interested model, but fail both to properly define their theory's domain of application and to make testable predictions. In particular, they focus their criticism on the inequality model by Fehr and Schmidt (1999) as an example, reproaching them for inadequately estimating the distribution of their α and β parameters, failing to make out-of-sample prediction, and selectively reporting experimental results that fit their model

[33]Elster (1989) offers a neutral perspective on the relationship of social norms and economic theory.

(also Shaked 2005). Fehr and Schmidt (2005, 2010) firmly reject the criticism, stating that "a model of social behavior is always an idealization that focuses on some forces affecting individual behavior and abstracts from many others. Whether a particular model is a good one depends on the situation under consideration and on the question that is being addressed." In this they have also received support from Eckel and Gintis (2010).

A different line of argument – again primarily aimed at inequality aversion – concerns the explanatory power of social preference models. For example, Bergh (2008) argues that outcome-based models are too simplistic compared to distributive justice theories from political philosophy, that they ignore aspects like procedural concerns or property rights, and – most importantly – that they fail to explain why people exhibit certain preferences in the first place. Similar criticism comes from Berg and Gigerenzer (2010), who generally object to *as-if* behavioral economics as neoclassical economics in disguise. Specifically, they criticize that social preference models take an unrealistic assumption – utility maximization – and make it even more complicated by adding more parameters to the utility function, ignoring all psychological insight on decision making. However, such arguments typically go unheeded among most economists. Following Friedman (1953), even behavioral economists primarily care about the quality of their models' predictions, but not necessarily about how well it represents psychological insights.

2.8 Concluding Remarks: Where Do We Stand and Where to Next?

Would the right social preference model please stand up! As this title of an article by Daruvala (2010) suggests, there is still no consensus among behavioral economists about the precise nature of people's social preferences. On the one hand, the relative importance of intentionality and outcomes is still under discussion (Sutter 2007; Falk et al. 2008). On the other hand, even within the outcome-based approach, the controversy between advocates of inequity aversion and welfare concerns continues, too (Daruvala 2010; Blanco et al. 2011; Graf et al. 2012). Given that social preferences appear to be decidedly heterogeneous among individuals (Andreoni and Miller 2002; Fisman et al. 2007; Blanco et al. 2011) and contexts (Kahneman et al. 1986; Loewenstein et al. 1989; Eckel and Grossman 1996a), it is unlikely that a conclusive consensus in favor of a single model will ever be reached. In the meantime, more and more models have entered the stage. After a period in which the main approaches underwent continuous development – arguably culminating in the unifying model of Falk and Fischbacher (2006) – the recent past has seen several models going beyond outcomes and reciprocity (e.g. Battigalli and Dufwenberg 2007; Andreoni and Bernheim 2009). So instead of moving toward a consensus or at least toward any form of universal framework, the field appears to drift more and more apart, leaving behind a patchwork of various, sometimes completely isolated, approaches.

However, among the plethora of available models, the theories of Fehr and Schmidt (1999), Bolton and Ockenfels (2000), and Charness and Rabin (2002) continue to dominate the economic analysis of social preferences even more than a decade after their publication (see e.g. chapter 3 and 4). It appears that these three models possess characteristics that make them particularly "useful" for economists. Compared to other models, they distinguish themselves by "simplicity" and "fruitfulness"; all three are among the most traceable models available, allowing straightforward qualitative predictions and requiring only the knowledge of two free parameters to make quantitative predictions.[34] In addition, their universality gives them a wide domain of application which extends way past the basic experimental games. Other models either have more free parameters, are less traceable, or apply to a narrower domain. However, it is not true that all other models lead an unheeded niche existence. For example, Rabin's intention-based reciprocity model and its successors are difficult to trace because they require information about beliefs (among other things), but they allow straightforward qualitative predictions in many different situations. However, some models have such a specific domain of application that their primary value clearly does not lie in their applicability. Instead, they mainly constitute formalizations of concepts rather than tools for data analysis.

This leads to the question which development the research on social preference models will (or should) take in the future. A possible direction is to continue the trend toward more and more specific models with relatively narrow domains of application. Given that the list of factors known to influence behavior in social games already includes items such as age (Fehr and Fischbacher 2002; Bellemare and Kröger 2003), gender (Eckel and Grossman

[34]The ERC-model has no specific functional form, but the exemplary function requires has two free parameters.

1996b; Croson and Buchan 1999; Dufwenberg and Muren 2006), social relationship (Burnham et al. 2000; Leider et al. 2009), social distance (Hoffman et al. 1996; Branas-Garza et al. 2010), personality traits (Brandstätter and Königstein 2001; Schmitt et al. 2008), procedural fairness (Bolton and Ockenfels 2006), communication (Brosig et al. 2003; Fiedler and Haruvy 2009; Balliet 2009), status (Ball et al. 2001), entitlement (Hoffman et al. 1996; Cherry et al. 2002; Bosman et al. 2005), verbal punishment opportunities (Xiao and Houser 2005), effort (Fong 2001), institutional environment (Frey and Bohnet 1995; Elliott et al. 1998; Ostrom 2000), peer effects (Cason and Mui 1998), expressed expectations (Charness and Rabin 2005), persuasion (Andersson et al. 2010), promises (Vanberg 2008), culture (Roth et al. 1991; Buchan et al. 2004; Henrich et al. 2005, 2010), and so on, there appears to be no shortage of starting points for further work.[35] A second possible direction is toward a universal model of social preferences that unifies the three elements inequity aversion, welfare concerns, and reciprocity in a single framework and – ideally – is also able to accommodate aspects of other models. While attempts at such a unifying theory have been undertaken, none has been widely recognized as successful so far (Engelmann 2012).

To those who focus on applications and predictions, the latter way is preferable in principal.[36] However, to quote Von Neumann and Morgenstern (1953), "such fortunate occurrences are rare and happen only after each field has been thoroughly explored". And therein lies the potential of the first way. Formalizing behavioral patterns clarifies concepts and allows predictions to be made for experimental testing to resolve unanswered question. This approach may not create impact theories like the seminal papers did, but it may serve as an important intermediate step toward a deeper understanding of human behavior. Therefore, as long as modification is not an end in itself, it is misguided to accuse those who add new components to the utility function of working in a "neo-classical repair shop" (Güth 1995). Even if, as Binmore and Samuelson (1999) argue, social preference models only do well as descriptions of behavioral patterns, but do not explain why individuals develop certain preferences in a certain situation, it does not mean they are a priori useless. The more systematic behavioral patterns are brought to light

[35] A particular area that still appears to be under-represented in models are preferences toward (inactive) third parties where existing models do not predict behavior well (e.g. Kagel and Wolfe 2001; Bereby-Meyer and Niederle 2005).

[36] Assuming, of course, that a unification is indeed attainable, which we will not know until somebody actually accomplishes the task.

and described formally, the likelier an eventual unifying theory will indeed be universal. Therefore, as Eckel and Gintis (2010) put it, "economists need more, not fewer, speculative and creative models of human behavior".

3 A Closer Look at Inequity Aversion and Incentives in Tournaments

3.1 Introduction

In their seminal article, Lazear and Rosen (1981) showed that tournaments are effective incentive mechanisms in various settings. Their major advantage is that only ordinal information on agents' performances is needed, reducing the principal's need for costly monitoring. At the same time, they allow credible commitment to a fixed prize structure even when performance signals are unverifiable by third parties. Furthermore, tournaments reduce the impact of common shocks in productivity, so agents do not need to be insured against them.

However, tournaments inevitably create inequality, something people usually dislike (Loewenstein et al. 1989). If agents have social preferences as modeled by Bolton and Ockenfels (2000) or Fehr and Schmidt (1999), they suffer from disadvantageous inequality (having less than others) as well as from advantageous inequality (having more than others). This consequently results in tournaments being less attractive for inequity averse agents than for agents without social preferences. This has been systematically demonstrated by Grund and Sliwka (2005). Using the model of Fehr and Schmidt (FS), who assume that disadvantageous inequality is worse for agents than advantageous inequality, they show that implemented effort levels are lower for inequity averse agents compared to purely self-regarding ones when the principal chooses winner prize and loser prize endogenously.

If, however, prizes for winning and losing the tournament are exogenously given, advantageous and disadvantageous inequality have diametrical effects on incentives. On the one hand, disutility from advantageous inequality decreases incentives to win the tournament, but on the other hand, disutility from disadvantageous inequality increases incentives to avoid losing the tournament. Grund and Sliwka find that the overall incentive effect is positive, leading to higher efforts by inequity averse agents compared to the standard model in tournaments with two agents.

In this paper, we employ the ERC-model of Bolton and Ockenfels to review the results of Grund and Sliwka and to examine to what extend those depend on the assumptions of the FS-model. While the main finding that the principal's payoff is reduced when agents are inequity averse persists, the incentive effect of inequity aversion vanishes in the ERC-model. Additionally, we extend the model to tournaments with any number of participants and winners. While we do not find any incentive effects of the tournament structure for standard agents, both types of inequity averse agents display a general tendency to prefer being in the majority group where they suffer less from the tournament's inequality. The principal can use this effect to her advantage by designing tournaments with many winners, although she can never reach her first best payoff. Conversely, if tournaments have very few winners and many losers, incentives may be distorted to such a degree that agents are better off losing the tournament and do not exert any effort at all.

This paper contributes to the analysis of tournament incentives established by Lazear and Rosen (1981). In particular, it adds to a strand of literature assuming that tournament participants are not exclusively motivated by their monetary payoff. Building on an idea introduced by Stark (1987, 1990), Kräkel (2000) was the first to formally introduce a form of social utility to an agent's utility function in a tournament setting. In his model, agents are trying to avoid relative deprivation. The concept is related to disadvantageous inequality insofar as agents dislike receiving a lower income than their reference group. Kräkel shows that agents who dislike relative deprivation exert more effort than agents in the standard model. Independent of Grund and Sliwka, Demougin and Fluet (2003) also examine tournament incentives under FS-inequity aversion, although in their model, they assume limited liability and the principal can costly improve performance assessment. They find that the principal prefers more envious agents when monitoring costs are higher. Another related study has been done by Kräkel (2008), where agents in a tournament compare not their monetary outcome, but their performance, experiencing positive or negative emotions when outperforming or falling behind their rival.

The paper is also related to the study of multi-agent tournaments. Like tournaments with two agents, multi-agent tournaments are formally equivalent to piece-rate contracts as Malcomson (1986) shows. Ferrall (1996) uses a multi-agent tournament model to estimate the significance of promotions in law firms, but does little else than confirm a general incentive effect. Kräkel (2000) finds that in multi-agent tournaments with either a single winner prize or a single loser prize, the incentive effect of relative deprivation becomes stronger with a single loser prize and weaker with a single winner prize. Kalra and Shi (2001) theoretically analyze sales contests and conclude that the optimal design has a single winner if agents are risk-neutral, but more than one winner if agents are risk-averse. Lim et al. (2009) confirm these predictions with laboratory data and field data. Orrison et al. (2004) look at multi-agent tournaments both theoretically and experimentally. They find that although the theory predicts no differences, subjects tend to decrease their effort when there are more winner prizes than loser prizes and subjects exert the highest effort when the numbers of winners and losers are equal. Harbring and Irlenbusch (2008) get similar experimental results in a tournament setting with sabotage. However, neither Orrison et al. nor Harbring and Irlenbusch find systematic differences of the total number of agents. Chen et al. (2011) consider multi-agent tournaments with asymmetric contestants who receive different initial endowments. In their experiment, both advantaged and disadvantaged agents exert more effort than the standard theory predicts. Similar to Kräkel (2008), Chen et al. extend the standard model so that favorites experience additional negative utility from losing whereas underdogs experience additional positive utility from winning, which provides an explanation for their experimental results.

The analysis of fairness concerns in the work environment is a highly relevant topic. Survey studies by Blinder and Choi (1990), Campbell III and Kamlani (1997), and Agell and Lundborg (2003) show that co-workers in firms usually have strong equity concerns, which constitutes one cause for downwards wage rigidity. Itoh (2004), Dur and Glazer (2008), and Englmaier and Wambach (2010) examine optimal contracts under moral hazard when agents are inequity averse toward the principal, whereas Neilson and Stowe (2010), Goukasian and Wan (2010), and Bartling and von Siemens (2010) derive optimal contracts when agents compare themselves to each other. Rey-Biel (2008) considers models of team incentives and Demougin et al. (2006) examine a two-task environment. Desiraju and Sappington (2007) introduce inequity aversion into the adverse selection model and Goel and Thakor (2006) and Bartling (2011) extend the analysis to other-regarding agents with risk-aversion. Finally, Kölle et al. (2011) investigate how inequality aversion changes contributions to public goods. Experimental studies are headlined by Fehr and Fischbacher (2002), who examine bonus and trust contracts under moral hazard both experimentally and theoretically. Cabrales et al. (2007) study contract and effort decisions between principals and teams of agents in a market experiment. Furthermore, Mohnen et al. (2008) explain peer pressure in teams with inequity aversion and confirm their findings in a real-effort experiment.

Virtually all of the theoretical studies of inequity aversion employ the

FS-model or closely related concepts for their analysis. The ERC-model, on the other hand, is hardly ever used in principal-agent-models. This lack of comparative studies is very unfortunate because both models do not just take a different mathematical form, they represent two different concepts of social comparison. In the FS-model, an individual compares his own outcome to the outcome of each other individual separately, experiencing envy or compassion to every single one of them. In the ERC-model, an individual is just concerned with how his own outcome compares to a social reference point and does not care about the individual outcomes of others. When only two agents interact, both approaches overlap to a certain degree. Yet in scenarios with many agents, both models may differ significantly in their predictions. Even in two-person-settings, however, a comparison of both models can be worthwhile. Not only does the ERC-model not explicitly distinguish between advantageous and disadvantageous inequality like the FS-model does, the ERC-model's very general form contrasts the linearity of the FS-model, which makes the latter mathematically easier to manage, but also more prone to corner solutions. If nothing else, employing both models serves as a general robustness check which may lead to new (experimental) research questions when both models make different predictions.

3.2 Two-Agent Tournaments and Inequity Aversion

3.2.1 The Model of Grund and Sliwka

Grund and Sliwka use a single-round tournament in which two agents compete for one prize (e.g. a promotion or a bonus) given to the agent producing the higher output. Each agent i's output is given by his production function $q_i = h(e_i) + \varepsilon_i$, where e_i is the agent's effort, $h(e_i)$ a concave function, and ε_i a random component which is independent and identically distributed for both agents. The costs of effort are $C(e_i)$ with $C'(e_i) > 0$ and $C''(e_i) > 0$. The prizes for the winner and the loser of the tournament are w_1 and w_2, respectively, with $w_1 > w_2$ and $\Delta w = w_1 - w_2$. Negative wages are allowed and a limited liability assumption is not imposed.

The agents have social preferences as proposed by Fehr and Schmidt (1999). Therefore, agent i's utility from his tournament prize w_i is given by

$$u_i = w_i - \alpha\ max\{w_j - w_i; 0\} - \beta\ max\{w_i - w_j; 0\} - C(e_i), \tag{42}$$

where w_i and w_j are the agents' wages.[37] According to the assumptions made by Fehr and Schmidt, $\alpha \geq 0$, $\beta \geq 0$, and $\alpha > \beta$, so agents suffer from inequality and disadvantageous inequality is worse than advantageous inequality. Furthermore, the model assumes $\beta < 1$, i.e., an agent always prefers receiving more money if the other agent's payoff is held constant. The utility of the tournament winner is therefore given by

$$u_i^W = w_1 - \beta\,\Delta w - C(e_i) = w_2 + (1-\beta)\,\Delta w - C(e_i) \tag{43}$$

and the utility of the loser by

$$u_i^L = w_2 - \alpha\,\Delta w - C(e_i). \tag{44}$$

Agents are assumed to evaluate lotteries by comparing expected utility. With $G(\cdot)$ the distribution function of the random variable $\varepsilon_i - \varepsilon_j$ and $g(\cdot)$ its density function, agent i's winning probability is

$$1 - G(h(e_j) - h(e_i)) = G(h(e_i) - h(e_j)) \tag{45}$$

and his expected utility is given by

$$\begin{aligned} EU_i =& G(h(e_i) - h(e_j))[w_1 - \beta\Delta w - C(e_i)] \\ &+ (1 - G(h(e_i) - h(e_j))[w_2 - \alpha\Delta w - C(e_i)] \\ =& w_2 - \alpha\Delta w + G(h(e_i) - h(e_j))[\Delta w(1-\beta+\alpha)] - C(e_i). \end{aligned} \tag{46}$$

Symmetric agents choose identical efforts in equilibrium.[38] This equilibrium effort is characterized by

$$\Delta w(1-\beta+\alpha)g(0) = \frac{C'(e^*)}{h'(e^*)} \tag{47}$$

and it increases in the prize spread Δw as well as in the strength of envy α, but decreases in the strength of compassion β. From the assumption $\alpha > \beta$ follows that $(1-\beta+\alpha)$ is always greater than 1. Therefore, for any given prize structure, inequity averse agents exert higher effort than purely self-regarding agents with $\alpha = \beta = 0$.

[37]It is assumed that the agents' utility is not affected by the principal's payoff because she does not belong to the agents' reference group.

[38]For a discussion about when such an equilibrium exists, see section 3.3.1.

However, Grund and Sliwka go on to show that despite this incentive effect, the principal is actually worse off with inequity averse agents when participation is voluntary and the principal can freely set the tournament's prizes to extract all surplus from the agents. The reason is that inequity averse agents need to be compensated for their inequity costs with a higher loser prize in order to be willing to participate in the tournament. The principal therefore reduces the price spread and implements lower effort levels compared to the first best solution.

3.2.2 The ERC-Model and Tournaments with Two Agents

The Model: The ERC-Model (Equity, Reciprocity, and Competition) by Bolton and Ockenfels (2000) is an alternative model of social preferences. According to the ERC-model, players are motivated only by their monetary payoff and their relative standing. Each agent i maximizes the expected value of his motivation function $v_i = v_i(y_i, \sigma_i)$, where y_i is the agent's monetary payoff and σ_i is his share y_i/Y of the combined total payoff Y of all agents.[39] The authors assume $v_{iy_i}(y_i, \sigma_i) \geq 0$, i.e., for any given relative payoff, agents prefer more money over less. Furthermore, $v_{i\sigma_i}(y_i, 1/N) = 0$ and $v_{i\sigma_i\sigma_i}(y_i, \sigma_i) < 0$, i.e., for any given monetary payoff, the motivation function is strictly concave in the relative payoff with its maximum at the average share. Finally, v_i is assumed to be continuous and twice differentiable on the whole domain for mathematical convenience.

Bolton and Ockenfels do not stipulate a fixed form for their motivation function, so we employ a general utility function consistent with the ERC-model. Agent i's utility is the sum of the monetary utility from his tournament prize minus inequity costs.[40] The agent's utility function is given by

$$U_i = w_i - W\kappa(\varsigma_i), \tag{48}$$

where W is the sum of the tournament's prizes (here w_1 and w_2) and $\kappa(\varsigma_i)$ is the inequity cost function. The inequity costs depend on ς, which is the

[39] If $Y = 0$, $\sigma_i = 1/N$, where N is the number of players. However, $Y = 0$ cannot occur in tournaments.

[40] Unlike this utility function, the original ERC-model also allows for preferences putting zero weight on the monetary payoff. Such agents, however, would never voluntarily participate in a tournament since they do not care about their tournament prize, but suffer inequity costs for sure.

difference of the received share σ_i and the equal share 1/2, i.e., $\varsigma_i = \sigma_i - 1/2$. $\kappa(\varsigma_i)$ is a twice differentiable function with its minimum at $\kappa(0) = 0$ and $\kappa''(\varsigma_i) > 0$. From $\varsigma_i = \frac{w_i}{2w_2+\Delta w} - \frac{1}{2}$ follows $\kappa_{\Delta w}(\varsigma_i) > 0$ (if $w_2 > 0$) and $\kappa_{w_2}(\varsigma_i) < 0$, i.e., inequity costs increase in the prize spread and decrease in the loser prize. Furthermore, we assume $\kappa(x) = \kappa(-x)$, i.e., inequity costs are equal for winner and loser of the tournament. Bolton and Ockenfels do not explicitly make this assumption, although in their article, they offer an exemplification for an ERC-utility function exhibiting this property (see (49)).

Additionally, by multiplying $\kappa(\varsigma_i)$ with W, we obtain another useful characteristic of the inequity costs. If $\frac{w_2+\Delta w}{w_2} = \frac{v_2+\Delta v}{v_2}$ and $w_2 > v_2$, then $\kappa(\varsigma_w) = \kappa(\varsigma_v)$, but $W\kappa(\varsigma_w) > V\kappa(\varsigma_v)$ with $V = 2v_2 + \Delta v$. This means that for a given ratio of winner to loser prize, inequity costs increase in the sum of both prizes, but the ratio of inequity costs to the prizes remains constant. The reason for this assumption may not be immediately clear. However, consider an ultimatum game and assume a player would reject an offer of 2|8. This means that his inequity costs from a 20% share are higher than the monetary utility of 2. Lets say the inequity costs are 3. If inequity costs only depended on the share, but did not increase in total payoff, the same player would accept an offer of 4|16 because his inequity costs would still be 3 since he still receives 20% of the pie. Under the new assumption, however, the player's inequity costs would increase to 6 ($3/10 = 6/20$). Therefore, the assumption assures that if a player rejects an allocation with a given share in the ultimatum game, he will reject every other allocation with the same share, too. Bolton and Ockenfels do not explicitly make this assumption because they only consider games with fixed total payoffs and can always normalize the total payoff to 1, which is not possible here. We therefore strongly believe that the additional assumption is very much in the spirit of the ERC-model.

As mentioned above, Bolton and Ockenfels give an exemplary ERC-utility function in their article:

$$U_i(\sigma_i, Y, \sigma_i) = a\sigma_i Y - \frac{b}{2}\left(\sigma_i - \frac{1}{2}\right)^2 \tag{49}$$

A player's type is characterized by a/b ($a \geq 0$, $b \geq 0$). Players with $a/b = 0$ are strictly relativistic while those with $a/b = \infty$ display narrow self-interest. The former type would give half of the endowment in the dictator game and reject any offer lower than half in the ultimatum game while the latter would behave according to standard economic theory in both games, keeping the whole endowment in the dictator game and accepting any positive offer in the ultimatum game. With the total payoff Y normalized to 1, players keep $(0.5+a/b)$ in the dictator game and reject any offer lower than $0.5-\frac{\sqrt{ab+a^2}-a}{b}$ (see A.1.1).

Adjusting (49) to our setting and fixing $a=1$ yields

$$\begin{aligned} U_i(w_i, W) =& w_i - W\frac{b}{2}\left(\frac{w_i}{w_2+w_1}-\frac{1}{2}\right)^2 \\ =& w_i - (2w_2+\Delta w)\frac{b}{2}\left(\frac{w_2}{2w_2+\Delta w}-\frac{1}{2}\right)^2 \\ =& w_i - \frac{b(\Delta w)^2}{8(2w_2+\Delta w)}, \end{aligned} \tag{50}$$

which exhibits all characteristics of the general ERC-utility function.

Tournament Incentives: An agent i with ERC-preferences receives the following utility from his prize w_i in the tournament:

$$U_i = w_i - W\kappa(\varsigma_i) = w_i - k(w_2, \Delta w). \tag{51}$$

The inequity costs $W\kappa(\varsigma_i)$ depend only on the loser prize w_2 and the prize spread Δw, but not on w_i, hence we can express $W\kappa(\varsigma_i)$ as $k(\Delta w, w_2)$ instead. Accordingly, the utility of winning the tournament and the utility of losing it are given by

$$U_i^W = w_1 - k(w_2, \Delta w) - C(e_i) \tag{52}$$

and

$$U_i^L = w_2 - k(w_2, \Delta w) - C(e_i), \tag{53}$$

respectively. Agent i's probability of winning the tournament is still given by equation (45), therefore, agent i's expected utility is

$$\begin{aligned} EU_i =& G(h(e_i)-h(e_j))[w_1 - k(w_2, \Delta w)] \\ &+ (1-G(h(e_i)-h(e_j))[w_2 - k(w_2, \Delta w)] - C(e_i), \end{aligned} \tag{54}$$

which can be simplified to

$$EU_i = w_2 - k(w_2, \Delta w) + G(h(e_i)-h(e_j))\Delta w - C(e_i). \tag{55}$$

Maximizing expected utility with respect to effort yields the first order condition

$$g(h(e_i) - h(e_j))\ h'(e_i)\ \Delta w - C'(e_i) = 0, \tag{56}$$

which is equal to the standard result for purely self-regarding agents. Since inequity costs do not depend on the success in the tournament, they do not influence the effort decision of the agents.

Proposition 1: *For any given prize spread in two-agent-tournaments, agents with ERC-preferences choose the same equilibrium effort level as agents with standard preferences. The optimal effort level e^* is characterized by*

$$\Delta w\ g(0) = \frac{C'(e^*)}{h'(e^*)}. \tag{57}$$

The equilibrium effort is increasing in the prize spread Δw, but is independent of the degree of the agent's inequity aversion.

This result stands in contrast to Grund and Sliwka's proposition that for any given prize spread, inequity averse agents exert more effort than purely self-regarding agents. The difference is driven by the assumption of Fehr and Schmidt's model that disadvantageous inequality looms larger than advantageous inequality.

Corollary 1: *If the prize structure is fixed, the principal's profit is independent of the degree of inequity aversion of agents with ERC-preferences. A tournament between two such agents leads to the same profits as one between two purely self-regarding agents.*

The Optimal Prize Structure: When the principal can design the tournament's prize structure, she maximizes her payoff which consists of the output produced by the agents minus the prizes she has to pay, i.e.,

$$U_P = 2h(e) - w_1 - w_2 = 2h(e) - 2w_2 - \Delta w. \tag{58}$$

With unlimited liability, the principal maximizes the total surplus and extracts all rents from the agents. However, the agents will only participate if their expected utility from the tournament (55) is at least as high as their reservation utility U_0. Following Grund and Sliwka, we assume that both agents' reservation utility U_0 is independent of their degree of inequity aversion. The agents' participation constraint is therefore given by

$$w_2 + \frac{1}{2}\Delta w - C(e^*) - k(w_2, \Delta w) \geq U_0. \tag{59}$$

Increasing the prize spread Δw increases the monetary utility of the tournament, but also increases inequity costs. Since inequity costs are convex, for small prize spreads, the appeal of the tournament increases when the prize spread increases. At some point, however, inequity costs grow faster than utility from money, i.e., $k_{\Delta w}(\Delta w, w_2) > \frac{1}{2}$, so increasing the prize spread beyond this point decreases the appeal of the tournament. Grund and Sliwka encounter a similar issue. In their model, the participation constraint is given by

$$w_2 + \frac{1}{2}(1 - \beta - \alpha)\Delta w - C(e^*) \geq U_0. \tag{60}$$

If $\alpha + \beta > 1$, the utility from participating in the tournament strictly decrease in the prize spread. The authors therefore assume $\alpha + \beta < 1$ for all agents. However, this assumption is a bit problematic as it suggests $\beta < 0.5$ even though $\beta = 0.5$ is the threshold for positive transfers in the dictator game. Using an ERC-utility function, at least some range of $\Delta w > 0$ exists where the utility of participating in the tournament increases in the prize spread Δw.

Assuming the parameters are such that the principal is indeed interested in arranging the tournament, she maximizes her expected payoff (58) while taking into account the agents' incentive constraint (56) and participation constraint (59). As the incentive constraint (56) shows, changing the loser prize w_2 does not change the agents' effort levels, so under unlimited liability, the principal can always lower w_2 to make the participation constraint binding and extract all surplus. Therefore, the principal chooses the loser prize

$$w_2 = U_0 - \frac{1}{2}\Delta w + C(e^*) + k(w_2, \Delta w). \tag{61}$$

Compared to the loser prize in the standard model with purely self-regarding agents, which is given by

$$w_2 = U_0 - \frac{1}{2}\Delta w + C(e^*), \tag{62}$$

the loser prize of agents with ERC-preferences is always higher for any given prize spread because the agents need to be compensated for their inequity costs.

Proposition 2: *For any given prize spread and reservation utility, agents with ERC-preferences have to be paid a higher loser prize in order to participate in the tournament. The difference in loser prizes is equal to the inequity costs endured by the agents.*

Inserting the loser prize (61) into the principal's payoff (58) and expressing e^* as a function of Δw yields

$$2h(e(\Delta w)) - 2C(e(\Delta w)) - 2k(w_2, \Delta w) - 2U_0. \tag{63}$$

Maximizing this expression yields the first order condition

$$h'(e(\Delta w)) - \underbrace{\frac{k_{\Delta w}(w_2, \Delta w)}{e'(\Delta w)}}_{>0} = C'(e(\Delta w)). \tag{64}$$

Compared to the first order condition of the standard model given by

$$h'(e(\Delta w)) = C'(e(\Delta w)), \tag{65}$$

the optimal prize spread Δw and thereby the equilibrium effort level is lower for agents with ERC-preferences.

Proposition 3: *In two-agent-tournaments, the principal chooses a lower prize spread for inequity averse agents with ERC-preferences, implementing lower effort levels compared to the first best solution with standard agents.*

Solving the principal's maximization problem is decidedly less straightforward than in the standard model or in the model of Grund and Sliwka because adjusting the prize spread changes the optimal loser prize and in turn, adjusting the loser prize changes inequity costs, which requires adjusting the prize spread and so on. For an approach to a solution and a numerical example, see A.2.1 in the appendix.

3.3 Multi-Agent-Tournaments

3.3.1 The Standard Model

Tournament Incentives: We look at a simple variant of the two-agent-tournament discussed in section 3.2. First, we consider agents with standard preferences (without inequity aversion) to establish benchmark results. A number of agents $N \geq 2$ competes for $m \geq 1$ winner prizes w_1, which are given to the m agents producing the m highest outputs, i.e., agent i is among the m winners if his output exceeds $y_{N-m:N-1}$, which is the $N-m$-th order statistic of outputs produced by the other agents. For better readability, we henceforth denote $N-m$ as l and $N-1$ as n. Agent i's expected utility is given by

$$P_w \Delta w + w_2 - C(e_i), \tag{66}$$

where P_w is his probability to be among the m winners. To maximize his expected utility, agent i will choose an effort level that satisfies

$$\frac{\partial P_w}{\partial e_i} = C'(e_i) \tag{67}$$

and

$$\frac{\partial^2 P_w}{\partial e_i^2} < C''(e_i). \tag{68}$$

For $N = 2$, agent i's probability of winning depends on the difference of both agents' effort-depending output $h(e_j) - h(e_i)$ and the difference of their realization of the random component $\varepsilon_i - \varepsilon_j$, i.e., $P_w = P(\varepsilon_i - \varepsilon_j > h(e_j) - h(e_i))$. If $F(x)$ is the cumulative distribution function of ε and $f(x)$ its density function, the density of the random variable $\varepsilon_i - \varepsilon_j$ is the convolution

$$(f^- * f)(\varepsilon_i - \varepsilon_j) = \int_{-\infty}^{+\infty} f(y) f([\varepsilon_i - \varepsilon_j] + y) dy \ , \tag{69}$$

where f^- stands for $f(-x)$. Since ε_i and ε_j are identically distributed, $(f^- * f)(\varepsilon_i - \varepsilon_j) = (f^- * f)(\varepsilon_j - \varepsilon_i)$. Therefore, in order to maximize his expected utility, agent i will choose an effort satisfying

$$(f^- * f)(h(e_j) - h(e_i)) h'(e_i) \Delta w - C'(e_i) = 0. \tag{70}$$

In a symmetric equilibrium, each agent's effort e^* in this case is characterized by the following equation:

$$\Delta w \int_{-\infty}^{+\infty} f(y)f(y)dy = \frac{C'(e^*)}{h'(e^*)}, \tag{71}$$

which is the original result from Lazear and Rosen (with the addition of specifying the distribution of $\varepsilon_i - \varepsilon_j$). However, as Lazear and Rosen noted, a symmetric equilibrium does not necessarily exist for all possible cumulative distribution functions $F(x)$. Such an equilibrium implies $e_1 = e_2 = e^*$ and (71) holding. Given that the left hand side is a positive constant and the right hand side is 0 for $e = 0$ and increasing in effort, there exists a unique value e^* satisfying (71). However, this value is not necessarily the best response given the other agent chooses e^* because the reaction function may be discontinuous in the relevant range so that (66) is not concave. Therefore, the symmetric equilibrium only exists if the variance of the random variable is sufficiently large.[41]

For $N > 2$, agent i only cares about surpassing $q_{l:n}$, the l-th highest output of the other agents. All other outputs are irrelevant for his effort decision. If ε has limited support with lower bound $\varepsilon_L \neq -\infty$ and upper bound $\varepsilon_U \neq \infty$ and if no other agent has chosen an effort level within $2(\varepsilon_U - \varepsilon_L)$ of the l-th highest effort $e_{l:n}$, agent i's problem is reduced to the two-agent-case. Otherwise, however, the problem gets significantly more complicated because the l-th highest output does not necessarily come from the agent choosing the l-th highest effort. Agent i's winning probability $P_w = P(q_i > q_{l:n})$ depends on the effort choices of all agents that (before the realization of the random component) have a positive probability of producing $q_{l:n}$ and the subsequent realization of their respective random components. We denote the cumulative distribution function of $q_{l:n}$ as $\hat{G}(\cdot)$, which is determined by the vector of effort choices made by the other agents $\bar{e}_{j\neq i}$ and ε's density function $f(x)$. The expression for agent i's expected utility is now

$$EU_i = P(h(e_i) > q_{l:n} - \varepsilon_i)\Delta w + w_2 - C(e_i) = \hat{G}(h(e_i))\Delta w + w_2 - C(e_i), \tag{72}$$

[41] For better intuition, consider a uniformly distributed random variable with a support marginally greater than 0. There exists a unique e^* satisfying (71). However, if one agent chooses e^*, the other agent may be better off exerting enough additional effort to ensure 100% winning probabily instead of also choosing e^*.

and the first order condition in turn becomes

$$
\begin{aligned}
&(f^{-} * \hat{g})(h(e_i))\ h'(e_i)\Delta w - C'(e_i) = 0 \\
\Leftrightarrow &\int_{-\infty}^{+\infty} f(y)\hat{g}(h(e_i) + y)dy\Delta w = \frac{C'(e_i)}{h'(e_i)}.
\end{aligned} \tag{73}
$$

For the equilibrium analysis, we only consider the symmetric equilibrium in pure strategies. A symmetric equilibrium implies $e_i = e^* \ \forall\ i$ and $\int_{-\infty}^{+\infty} f(y)\hat{g}(h(e_i)+y)dy\Delta w = \frac{C'(e^*)}{h'(e^*)}$. Assuming the second order condition can be met[42], a symmetric equilibrium exists because each agent's effort choice enters the objective function of the other agents through $\hat{G}(\cdot)$ via $\bar{e}_{j\neq i}$. So in equilibrium, all agents face the same distribution $G(\cdot)$ and therefore have identical objective functions. Mixed equilibria and asymmetric equilibria in pure strategies may exists for certain distributions of the random variable, but a general statement about existence or necessary conditions is virtually impossible to make due to the complexity and possible range of $\hat{G}(\cdot)$.

If all other agents choose the equilibrium effort level e^*, each produces an output of $h(e^*) + \varepsilon$. Hence, agent i is among the tournament winners if his output exceeds the sum $h(e^*) + \varepsilon_{l:n}$. The distribution function of the order statistic $\varepsilon_{l:n}$ is given by

$$
F_{\varepsilon_{l:n}}(x) = \sum_{k=l}^{n} \binom{n}{k} F(x)^k \left[1 - F(x)\right]^{n-k} \tag{74}
$$

or

$$
F_{\varepsilon_{l:n}}(x) = F_{\varepsilon_{l-1:n}}(x) - \binom{n}{l-1} F(x)^{l-1} \left[1 - F(x)\right]^{n-l+1}. \tag{75}
$$

Using the latter, we derive the probability density function of $\varepsilon_{l:n}$:

$$
f_{\varepsilon_{l:n}}(x) = \frac{n!}{(l-1)!(n-l)!} F(x)^{l-1} \left[1 - F(x)\right]^{n-l} f(x). \tag{76}
$$

Agent i's expected utility is now given by

$$
EU_i = P(h(e_i) - h(e^*) > \varepsilon_{l:n} - \varepsilon_i)\Delta w + w_2 - C(e_i), \tag{77}
$$

[42] As discussed above, for some distributions of the random variable, the second order condition cannot be met.

yielding the first order condition

$$
\begin{aligned}
&(f^{-} * f_{\varepsilon_{l:n}})(h(e_i) - h(e^*))\ h'(e_i)\Delta w - C'(e_i) = 0 \\
\Leftrightarrow &\int_{-\infty}^{+\infty} f(h(e_i) - h(e^*) + y) \cdot \\
&\frac{n!}{(l-1)!(n-l)!} F(y)^{l-1} \left[1 - F(y)\right]^{n-l} f(y) dy \Delta w = \frac{C'(e_i)}{h'(e_i)}.
\end{aligned}
\tag{78}
$$

In equilibrium, agent i chooses $e_i = e^*$ and (78) becomes

$$
\Delta w \int_{-\infty}^{+\infty} \frac{n!}{(l-1)!(n-l)!} F(y)^{l-1} \left[1 - F(y)\right]^{n-l} f(y)^2 dy = \frac{C'(e^*)}{h'(e^*)}. \tag{79}
$$

The integral is the marginal winning probability in equilibrium $f_{\varepsilon_{l:n}}(0)$.

Furthermore, if ε is uniformly distributed on the interval $(0, \widetilde{\varepsilon})$, (78) becomes

$$
\frac{\Delta w}{\widetilde{\varepsilon}} \frac{n!}{(l-1)!(n-l)!} \int_0^{\widetilde{\varepsilon}} \left(\frac{y}{\widetilde{\varepsilon}}\right)^{l-1} \left[1 - \frac{y}{\widetilde{\varepsilon}}\right]^{n-l} dy = \frac{C'(e^*)}{h'(e^*)}. \tag{80}
$$

By substituting $z = \frac{y}{\widetilde{\varepsilon}}$, the integral becomes the beta integral[43]

$$
\int_0^1 z^{l-1} \left[1 - z\right]^{n-l} dz = \frac{\Gamma(l)\Gamma(n-l+1)}{\Gamma(l+n-l+1)} = \frac{(l-1)!(n-l)!}{n!} \tag{81}
$$

and the first order condition simply becomes

$$
\frac{\Delta w}{\widetilde{\varepsilon}} = \frac{C'(e^*)}{h'(e^*)}. \tag{82}
$$

Hence, the equilibrium effort depends neither on the number of agents nor on the number of winners in the tournament. Since every uniform distribution can be made to correspond to any given interval by linear transformation, this result generally holds for all uniform distributions $U(\varepsilon_L, \varepsilon_U)$. For other distributions, however, this result typically does not hold.

[43] The beta integral (also called the Eulerian integral of the first kind) $\int_0^1 x^{p-1}(1-x)^{q-1}$ is equal to the beta function $B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$. The Γ-function is the factorial function with the argument decreased by 1, i.e., $\Gamma(p) = (p-1)!$

Proposition 4: *In the symmetric equilibrium, purely self-regarding agents in the multi-agent-tournament choose identical effort levels characterized by*

$$f_{\varepsilon_{l:n}}(0)\ \Delta w = \frac{C'(e^*)}{h'(e^*)}, \tag{83}$$

If the random component is uniformly distributed, the equilibrium effort does not depend on the number of participants or on the number of winners, and $f_{\varepsilon_{l:n}}(0) = \frac{1}{\varepsilon_U - \varepsilon_L}$.

Intuitively, it may seem sensible to assume that with many (few) winners, the base probability to win is high (low) so that the optimal effort level is low (high).[44] However, this intuition is not correct. The optimal effort choice depends on the marginal winning probability $f_{\varepsilon_{l:n}}(0)$, not on the total winning probability, because increasing the own effort level only changes the outcome of an agent under one condition: He is not among the winners before the increase, but the increase moves him past the "marginal winner", i.e., the winner with the lowest output. Only the probability of this event is relevant for effort choice.

In the symmetric equilibrium when all agents choose the effort level e^*, the random variable of the marginal winner is $\varepsilon_{l:n}$ (ignoring agent i). The distribution of $\varepsilon_{l:n}$ depends both on the number of agents and on the number of winners. For example, for a given number of agents, the skewness of $F_{\varepsilon_{l:n}}$ becomes larger with each additional winner. For random variables that are not uniformly distributed, this usually entails that the marginal winning probability also depends on the number of agents and the number of winners. However, this is not true for uniformly distributed random variables because their density is constant. For intuition, consider a normally distributed random variable. For simplicity, we only look at the expected value of the relevant order statistic, $E(\varepsilon_{l:n})$, instead of its whole distribution. If the number of winners decreases (keeping N constant), the expected value of the relevant order statistic increases ($E(\varepsilon_{l+1:n}) > E(\varepsilon_{l:n})$). The density of the random variable at both positions is usually different, i.e., $f_\varepsilon(\varepsilon_{l+1:n}) \neq f_\varepsilon(\varepsilon_{l:n})$. Therefore, the likelihood that a marginal increase in effort changes the agent's winner status is also different. However, if ε is uniformly distributed, the expected value of the order statistic still increases,

[44]The experimental results of Orrison et al. (2004) suggest that their subjects actually followed this reasoning because subjects' effort levels decreased in the number of winners.

yet the density of ε remains the same. Hence the marginal winning probability does not change. Therefore, the number of agents and the number of winners is irrelevant for the marginal winning probability in equilibrium if the random variable is uniformly distributed.

Similar to the two-player case discussed above, the existence of the symmetric equilibrium is not guaranteed. Although a unique e^* exists, it is not necessarily the best response to all other agents choosing e^*. As before, the larger the variance of the random variable, the more likely an interior symmetric equilibrium exists. It is not possible to give general conditions which are sufficient for the existence because the relevant distribution functions can take various forms.[45].

The Optimal Prize Structure: The principal maximizes her payoff $U_P = Nh(e) - Nw_2 - m\ \Delta w$ under the binding participation constraint

$$\begin{aligned} & w_2 + \frac{m}{N}\Delta w - C(e^*) = U_0 \\ \Leftrightarrow & w_2 = U_0 - \frac{m}{N}\Delta w + C(e^*), \end{aligned} \tag{84}$$

which gives us the maximization problem of the principal

$$\max_{e}\ Nh(e^*) - NU_0 - NC(e^*). \tag{85}$$

This yields the standard result of first best efforts

$$h'(e^{FB}) = C'(e^{FB}). \tag{86}$$

Proposition 5: *In multi-agent-tournaments, the principal chooses a prize spread that induces first best efforts defined by*

$$h'(e^{FB}) = C'(e^{FB}),$$

then chooses the loser prize to extract all surplus from the agents. Neither the implemented effort level nor the principal's payoff depends on the number of winners in the tournament.

[45] Orrison et al. (2004) impose a functional form on both $C(e)$ and $h(e)$ and choose parameters to ensure the existence of an interior symmetric equilibrium in their experiment. Yet even in their setting, it is possible to create a counterexample without symmetric equilibrium by choosing a distribution with narrow support.

Unlike Proposition 4, this result holds for all distributions of ε. Since the principal can always extract all rents from the agents via the loser prize, she always implements the first best effort level to maximize total surplus.[46]

3.3.2 The FS-Model in Multi-Agent-Tournaments

Tournament Incentives: In their article, Grund and Sliwka briefly mention that their findings hold for tournaments with N agents as long as there are at least $N/2$ winner prizes and no intermediate outcomes. In tournaments with less than $N/2$ winners, they note, envy becomes weaker and effort levels may drop below those of purely self-regarding agents. In any case, their key result, that the principal is worse off with inequity averse agents, always holds. The following section shows that those assertions are in fact correct.

Consistent with Fehr and Schmidt (1999), in a tournament with N agents, m winners and prizes of w_1 for the m winners and w_2 for the $N-m$ losers, agent i's utility is given by

$$u_i = w_i - \frac{m}{N-1}\,\alpha\,(w_1 - w_i) - \frac{N-m}{N-1}\,\beta\,(w_i - w_2) - C(e_i). \tag{87}$$

The extend of envy experienced by the loser(s) increases in the number of winners but decreases in the number of total participants. Inversely, the extend of compassion experienced by the winner(s) decreases in the number of other winners and increases in the number of total participants (if there are at least two winners). The original model is embedded for $N = 2$ and $m = 1$. Agent i's expected utility is given by

$$\begin{aligned} EU_i = {} & \hat{G}(h(e_i))\left(w_1 - \frac{N-m}{N-1}\,\beta\,(w_1 - w_2)\right) + \\ & \left(1 - \hat{G}(h(e_i))\right)\left(w_2 - \frac{m}{N-1}\,\alpha\,(w_1 - w_2)\right) - C(e_i) \\ = {} & w_2 - \frac{m}{N-1}\,\alpha\,\Delta w \\ & + \hat{G}(h(e_i))\Delta w(1 - \frac{N-m}{N-1}\,\beta + \frac{m}{N-1}\,\alpha) - C(e_i). \end{aligned} \tag{88}$$

[46] Kräkel (2000) finds a similar result. He shows that with endogeneous prizes, the principal implements first best effort levels for both standard workers and workers suffering from relative deprivation, irrespective of whether the winner prize is given to one agent or to $N-1$ agents. However, note that proposition 5 depends on the agents' risk-neutrality. With risk-averse agents, the principal will typically not implement first-best effort levels (Kräkel 2008).

The agent chooses his effort level to maximize his expected utility, which yields the first order condition

$$\hat{g}(h(e_i))h'(e_i)\Delta w(1 - \frac{N-m}{N-1}\,\beta + \frac{m}{N-1}\,\alpha) - C'(e_i) = 0 \tag{89}$$

When all other agents choose the equilibrium effort e^*, these two expressions become

$$\begin{aligned} EU_i &= F_{\varepsilon_{l:n}}(h(e_i) - h(e^*))\left(w_1 - \frac{N-m}{N-1}\,\beta\,(w_1 - w_2)\right) + \\ &\qquad (1 - F_{\varepsilon_{l:n}}(h(e_i) - h(e^*)))\left(w_2 - \frac{m}{N-1}\,\alpha\,(w_1 - w_2)\right) - C(e_i) \\ &= w_2 - \frac{m}{N-1}\,\alpha\,\Delta w \\ &\qquad + F_{\varepsilon_{l:n}}(h(e_i) - h(e^*))\Delta w(1 - \frac{N-m}{N-1}\,\beta + \frac{m}{N-1}\,\alpha) - C(e_i). \end{aligned} \tag{90}$$

and

$$f_{\varepsilon_{l:n}}(h(e_i) - h(e^*))h'(e_i)\Delta w(1 - \frac{N-m}{N-1}\,\beta + \frac{m}{N-1}\,\alpha) - C'(e_i) = 0. \tag{91}$$

In equilibrium, $e_i = e^*$ and $f_{\varepsilon_{l:n}}(h(e_i) - h(e^*))h'(e_i) = f_{\varepsilon_{l:n}}(0)$.

Proposition 6: *In equilibrium, all agents with FS-preferences choose identical effort levels characterized by*

$$\Delta w(1 - \frac{N-m}{N-1}\,\beta + \frac{m}{N-1}\,\alpha)f_{\varepsilon_{l:n}}(0) = \frac{C'(e^*)}{h'(e^*)}. \tag{92}$$

The equilibrium effort is increasing in the prize spread Δw and in the strength of envy α. It decreases in the strength of compassion β. If ε is uniformly distributed, the equilibrium effort increases in the number of winners and decreases in the number of losers $N-m$ and $f_{\varepsilon_{l:n}}(0) = \frac{1}{\varepsilon_U - \varepsilon_L}$. For a given prize spread, agents do not necessarily exert more effort than purely self-regarding agents.

Corollary 2: *If the prize structure is fixed, the principal's profit is the higher the stronger the agents' envy and the more agents win the tournament. The principal's profit is the lower the stronger the agents' compassion and the more participants are in the tournament. Compared to purely self-regarding agents, a tournament with inequity averse agents does not necessarily lead to a higher profit if there are more losers than winners.*

Envy and compassion have the same effect on incentives as in two-agent-tournaments. However, contrary to the original results, agents do not necessarily exert higher efforts than purely self-regarding agents. Agents only do so if the term in brackets is larger than 1, i.e., if the costs of envy toward winners exceed the costs of compassion toward losers. Since $\alpha > \beta$, inequity averse agents with FS-preferences will always exert more effort than purely self-regarding agents if the winner prize is given to at least half of the participants in the tournament (as indicated by Grund and Sliwka). Otherwise, if the agents' degree of compassion β is large enough, i.e., $\beta > \frac{m}{N-m}\alpha$, inequity averse agents actually exert less effort than standard agents. However, even though compassion may outweigh envy in multi-agent-tournaments, the incentive effect of the prize spread can never become negative. Since $\beta < 1$, the term in square brackets is always greater than zero and therefore a positive effort level e^* exists that solves the incentive condition (92).

The Optimal Prize Structure: As usual, the principal chooses the overall prize structure that maximizes her payoff under the participation constraint, which is given by

$$w_2 + \frac{m}{N}\left(1 - \frac{N-m}{N-1}(\alpha+\beta)\right)\Delta w - C(e^*) \geq U_0. \tag{93}$$

If we want to uphold Grund and Sliwka's condition that given the loser prize, increasing the prize spread increases the overall appeal of the tournament, their original assumption $\alpha+\beta < 1$ can be relaxed to $\alpha+\beta < \frac{N-1}{N-m}$. Individual inequity costs are lower in multi-agent-tournaments because when there are more than two agents, there is a positive probability that after the conclusion of the tournament, agent i will be in the same position (winner or loser) as at least one other agent. Since that agent receives the same monetary payoff, agent i neither feels compassion nor envy toward him.

The principal can still make the participation constraint binding, so she chooses the loser prize

$$w_2 = U_0 - \frac{m}{N}\left(1 - \frac{N-m}{N-1}(\alpha+\beta)\right)\Delta w + C(e^*). \tag{94}$$

Inserting this loser prize into the principal's payoff function yields the following maximization problem

$$\max_e \; Nh(e) - NC(e) - NU_0 - \frac{m(N-m)}{N-1}(\alpha+\beta)\Delta w \tag{95}$$

Solving the incentive constraint (92) for the prize spread Δw yields

$$\Delta w = \frac{C'(e^*)}{h'(e^*)\left(1 + \frac{m}{N-1}\,\alpha - \frac{N-m}{N-1}\,\beta\right) f_{\varepsilon_{l:n}}(0)}, \tag{96}$$

which we can insert into the principal's payoff function to obtain her maximization problem

$$\begin{aligned} &\max_e \; Nh(e) - NC(e) - NU_0 - \\ &\qquad \frac{m(N-m)}{N-1}(\alpha+\beta)\frac{C'(e)}{h'(e)\left(1 + \frac{m}{N-1}\,\alpha - \frac{N-m}{N-1}\,\beta\right) f_{\varepsilon_{l:n}}(0)}. \end{aligned} \tag{97}$$

This leads to the first order condition

$$h'(e) - \underbrace{\frac{m(N-m)(\alpha+\beta)(C''(e)h'(e) - C'(e)h''(e))}{N\left(N-1)(1 + \frac{m}{N-1}\alpha - \frac{N-m}{N-1}\beta\right) f_{\varepsilon_{l:n}}(0)(h'(e))^2}}_{>0} = C'(e). \tag{98}$$

Proposition 7: *Compared to the first best solution $h'(e) = C'(e)$, the principal chooses to implement a lower effort level when agents are inequity averse with FS-preferences.*

Making use of the envelope theorem, we obtain the partial derivatives of (97) with respect to β

$$-\frac{n(N-m)(N-1+N\alpha)}{(N-1+m\alpha-(N-m)\beta)^2}\frac{m(N-m)C'(e)}{h'(e)f_{\varepsilon_{l:n}}(0)},$$

which is always negative. This is of course not surprising because a higher degree of compassion not only decreases the agents' effort for a given prize spread, it also increases inequity costs for which the agents have to be compensated with a higher loser prize. However, the effect is less straightforward for envy. The partial derivative of (97) with respect to α is

$$\frac{1+N(\beta-1)}{(N-1+m\alpha-(N-m)\beta)^2}\frac{m(N-m)C'(e)}{h'(e)f_{\varepsilon_{l:n}}(0)}, \tag{99}$$

which is negative if $1+N(\beta-1)<0$, so the principal's payoff decreases in the degree of envy as long as $\beta<\frac{N-1}{N}$ and increases otherwise.[47] If compassion is relatively weak, the cost effect outweighs the incentive effect, but vice versa when compassion is relatively strong.

Corollary 3: *When the principal can design the tournament's prize structure, her profits are smaller with inequity averse FS-agents than with purely-self-regarding agents. Her payoff decreases in the strength of compassion and decrease in the strength of envy if* $\beta<\frac{N-1}{N}$ *and increases otherwise.*

3.3.3 The ERC-Model in Multi-Agent-Tournaments

Tournament Incentives: In tournaments with $N>2$ participants, the agents' most preferred share where inequality costs are zero moves from $1/2$ to $1/N$. Agent i's utility from receiving his prize w_i in a tournament with N participants is

$$u_i = w_i - W\kappa(\varsigma_i), \tag{100}$$

where w_i is the agent's monetary payoff and $W\kappa(\varsigma_i)$ his inequity costs. W is the sum of all prizes $w_1+...+w_N$ or $Nw_2+m\ \Delta w$ and ς_i the difference of received share σ_i and the equal share $1/N$, i.e., $\varsigma_i=\sigma_i-1/N$. $\kappa(\varsigma_i)$ is a twice differentiable function with $\kappa(0)=0$ and $\kappa_{\varsigma_i\varsigma_i}(\varsigma_i)>0$. We still assume that $\kappa(x)=\kappa(-x)$, i.e., inequity costs are the same for advantageous and disadvantageous deviations from the equal share. Therefore, if $m=\frac{N}{2}$ ($m>\frac{N}{2}$, $m<\frac{N}{2}$), then $k^W=k^L$ ($k^W<k^L$, $k^W>k^L$), i.e., if the winner prize is given to half (more than half, less than half) of the participants, inequity costs are equal for losers and winners (lower for winners, lower for losers). Intuitively speaking, ERC-agents "feel better" when they are part of the majority group (ignoring monetary payoffs, of course). From $\varsigma_i=$

[47] Note that $\beta>\frac{N-1}{N}$ also implies $\beta>\frac{N-1}{N-m}$, i.e., increasing the prize spread makes the tournament less attractive to the agent.

$\frac{w_i}{Nw_2+m\Delta w} - \frac{1}{N}$ follows $\kappa_{\Delta w}(\varsigma_i) > 0$ and $\kappa_{w_2}(\varsigma_i) < 0$, i.e., inequity costs increase in the prize spread, but decrease in the loser prize.

For better readability, we denote $W\kappa(\varsigma_i)$ as k^W if agent i is a winner (i.e., $w_i = w_1$) and as k^L if agent i is a loser (i.e., $w_i = w_2$). Then, agent i's expected utility is given by

$$\begin{aligned} EU_i &= G(h(e_i))(w_1 - k^W) \\ &\quad + (1 - G(h(e_i)))\,(w_2 - k^L) - C(e) \\ &= w_2 - k^L + G(h(e_i))(\Delta w - k^W + k^L) - C(e), \end{aligned} \tag{101}$$

which leads to the first order condition

$$\left(\Delta w - k^W + k^L\right)\gamma(h(e_i), H_{-i})h'(e_i) - C(e_i) = 0. \tag{102}$$

When all other agents choose the equilibrium effort e^*, these two expressions become

$$\begin{aligned} EU_i &= F_{\varepsilon_{l:n}}(h(e_i) - h(e^*))(w_1 - k^W) \\ &\quad + (1 - F_{\varepsilon_{l:n}}(h(e_i) - h(e^*))))\,(w_2 - k^L) - C(e) \\ &= w_2 - k^L + F_{\varepsilon_{l:n}}(h(e_i) - h(e^*))(\Delta w - k^W + k^L) - C(e), \end{aligned} \tag{103}$$

and

$$\left(\Delta w - k^W + k^L\right) f_{\varepsilon_{l:n}}(h(e_i) - h(e^*))h'(e_i) - C(e_i) = 0. \tag{104}$$

In equilibrium, $e_i = e^*$ and $f_{\varepsilon_{l:n}}(h(e_i) - h(e^*))h'(e_i) = f_{\varepsilon_{l:n}}(0)$.

Proposition 8: *In equilibrium, all agents with ERC-preferences choose identical effort levels characterized by*

$$(\Delta w - k^W + k^L)f_{\varepsilon_{l:n}}(0) = \frac{C'(e^*)}{h'(e^*)}. \tag{105}$$

Compared to purely self-regarding agents, for any given prize spread agents with ERC-preferences will exert ...

- *... the same effort if the winner prize is given to half of the participants.*
- *... more effort if the winner prize is given to more than half of the participants.*

- *... less effort if the winner prize is given to less than half of the participants.*

If $m = N/2$, $k^W = k^L$ and both expressions cancel each other out. If $m > \frac{N}{2}$, it follows $k^W < k^L$, so there is an additional incentive effect to avoid the loser's higher inequity costs which increases effort. If $m < \frac{N}{2}$, it follows $k^W > k^L$, so winning becomes relatively less attractive because it entails higher inequity costs than losing.

Additionally, if the difference of the inequity costs for winning and losing exceeds the prize spread, i.e., if $k^W \geq k^L + \Delta w$, incentives are distorted to such an extend that the agents actually prefer losing the tournament over winning it. In this case, since effort is costly, agents would choose the minimal possible effort level. Since inequity costs are convex, this becomes more likely the higher the prize spread Δw (see also A.2.2).

Proposition 9: *If the winner prize is given to less than half of the participants in the tournament, incentives may be distorted to such a degree that agents prefer to loose the tournament and do not exert any effort. This become more likely the higher the prize spread and the stronger the degree of inequality aversion and less likely the higher loser prize.*

The Optimal Prize Structure: Agent i's participation constraint is

$$w_2 + \frac{m}{N}\Delta w - C(e^*) - \frac{m}{N}k^W - \frac{N-m}{N}k^L \geq U_0, \tag{106}$$

which leads to the loser prize

$$w_2 = U_0 - \frac{m}{N}\Delta w + C(e^*) + \frac{m}{N}k^W + \frac{N-m}{N}k^L. \tag{107}$$

The principal's maximization problem is then given by

$$\max_e Nh(e) - NC(e) - mk^W - (N-m)k^L - NU_0, \tag{108}$$

which yields the first order condition

$$\begin{aligned} &Nh'(e^*) - NC'(e^*) - mk^W_{e^*} - (N-m)k^L_{e^*} = 0 \\ \Leftrightarrow &h'(e^*) - \underbrace{\frac{mk^W_{e^*} + (N-m)k^L_{e^*}}{N}}_{>0} = C'(e^*) \end{aligned} \tag{109}$$

Proposition 10: *Compared to the first best solution $h'(e) = C'(e)$, the principal chooses to implement a lower effort level when agents are inequity averse with ERC-preferences.*

Since both the inequity costs of winners k^W and the inequity costs of losers k^L increase in the prize spread Δw and the prize spread in turn strictly increases in e, both inequity costs strictly increase in the inplemented effort level, i.e., $k_e^W > 0$ and $k_e^W > 0$. Therefore, irrespective of how many winners the tournament has, the implemented effort levels are lower than in the first best solution.

Corollary 4: *When the principal can design the tournament's prize structure, her profits are smaller with inequity averse ERC-agents than with purely-self-regarding agents.*

3.3.4 Diminishing Sensitivity and FS-Preferences

We have seen that when agents have ERC-preferences and losers are in the majority, if the prize spread becomes relatively large, incentives may be distorted to such a degree that agents are not interested in winning the tournament anymore. With FS-preferences on the other hand, we did not find a similar result. However, if we allow the marginal utility of the monetary payoff to be decreasing, scenarios may arise in which agents with FS-preferences are not motivated to exert effort either.

To introduce diminishing marginal utility to the FS model, we adjust the agents' utility function (87) in the following way

$$u_i = v_w(w_i) - \frac{m}{N-1} v_\alpha(w_1 - w_i) - \frac{N-m}{N-1} v_\beta(w_i - w_2) - C(e_i), \qquad (110)$$

where v_w, v_α, and v_β are monotonously increasing concave functions with $v_w(0) = v_\alpha(0) = v_\beta(0) = 0$. To comply with the original assumptions that envy is stronger than compassion and that compassion is not stronger than monetary utility for any given amount, we additionally assume $v_\alpha(x) > v_\beta(x)$ and $v_w(x) > v_\beta(x)$ for all $x > 0$. The latter assumption assures that a player would never reject an ultimatum game offer where he receives more than the other player. Note that although we introduce concave inequity costs here, this is not required for the distortion of incentives, which can also occur with linear inequity costs.

When all other agents choose e^*, agent i's expected utility is given by

$$\begin{aligned}
EU_i = {} & F_{\varepsilon_{l:n}}(h(e_i) - h(e^*)) \left(v_w(w_1) - \frac{N-m}{N-1} v_\beta(w_1 - w_2) \right) \\
& + (1 - F_{\varepsilon_{l:n}}(h(e_i) - h(e^*))) \left(v_w(w_2) - \frac{m}{N-1} v_\alpha(w_1 - w_2) \right) \\
& - C(e_i) \\
= {} & v_w(w_2) - \frac{m}{N-1} v_\alpha(\Delta w) + F_{\varepsilon_{l:n}}(h(e_i) - h(e^*)) \cdot \\
& \left(v_w(w_1) - v_w(w_2) + \frac{m}{N-1} v_\alpha(\Delta w) - \frac{N-m}{N-1} v_\beta(\Delta w) \right) - C(e_i),
\end{aligned} \tag{111}$$

which yields the first order condition

$$\begin{aligned}
& f_{\varepsilon_{l:n}}(h(e_i) - h(e^*)) \cdot \\
& \left(v_w(w_1) - v_w(w_2) + \frac{m}{N-1} v_\alpha(\Delta w) - \frac{N-m}{N-1} v_\beta(\Delta w) \right) = \\
& \frac{C'(e_i)}{h'(e_i)},
\end{aligned} \tag{112}$$

which means the equilibrium effort is characterized by

$$\begin{aligned}
& f_{\varepsilon_{l:n}}(0) \left(v_w(w_1) - v_w(w_2) + \frac{m}{N-1} v_\alpha(\Delta w) - \frac{N-m}{N-1} v_\beta(\Delta w) \right) = \\
& \frac{C'(e^*_{FS})}{h'(e^*_{FS})}.
\end{aligned} \tag{113}$$

By way of comparison, purely self-regarding agents with diminishing sensitivity of monetary utility but without inequity aversion would choose equilibrium effort characterized by

$$f_{\varepsilon_{l:n}}(0) \left(v_w(w_1) - v_w(w_2) \right) = \frac{C'(e^*_S)}{h'(e_S)^*}. \tag{114}$$

Compared to standard agents with linear monetary utility, for given prizes w_1 and w_2, agents with diminishing sensitivity would choose lower effort levels because $v_w(w_1) - v_w(w_2) < \Delta w$ (although the principal would implement the same first best effort levels if she can choose the prizes). Comparing agents with diminishing sensitivity with and without inequity aversion, the inequity averse agents would exert less effort if their compassion is relatively strong and the winner prize is given to less than half of the participants, i.e., $v_\beta(\Delta w) > \frac{m}{N-m} v_\alpha(\Delta w)$. Otherwise, inequity averse agents exert more effort. This result mirrors the one without diminishing sensitivity.

If the winner prize is given to less than half of the participants, incentives may be distorted to such a degree that agents prefer losing the tournament over winning it. This is possible because if the loser prize is strictly positive, the difference in monetary utility of the two prizes is smaller than the monetary utility of the prize spread as we have seen above. This, in turn, makes it possible that the costs of compassion exceed the monetary utility gain of receiving the winner prize, i.e., $v_\beta(\Delta w) > v_w(w_1) - v_w(w_2)$ despite the assumption $v_\beta(x) < v_w(x)$ for all $x > 0$. If additionally envy is relatively weak, this can lead to the overall incentive effect of the prize structure becoming negative, i.e.,

$$\left(v_w(w_1) - v_w(w_2) + \frac{m}{N-1} v_\alpha(\Delta w) - \frac{N-m}{N-1} v_\beta(\Delta w) \right) < 0,$$

in which case the agents prefer to loose the tournament and do not exert any effort at all. This happens if

$$v_\beta(\Delta w) > \frac{N-1}{N-m}(v_w(w_1) - v_w(w_2)) + \frac{m}{N-m} v_\alpha(\Delta w), \tag{115}$$

i.e., if compassion is relatively strong, the loser prize relatively high and the number of winners low (necessarily below $N/2$).

Consider the following illustrative example: A firm hires ten agents ($N = 10$), who receive a fix wage of \$975 (w_2). After contracts are signed, the principal posts a reward of additional \$49 (Δw), which will be given to the agent producing the highest output. Assuming $v_w = v_\alpha = \sqrt{x}$ and $v_\beta(x) = 0.5\sqrt{x}$, the agents utility from winning the tournament is

$$u_{win} = \sqrt{w_1} - \frac{N-m}{(N-1)} 0.5\sqrt{\Delta w} = \sqrt{1024} - \frac{9}{18}\sqrt{49} = 32 - 3.5 = 28.5$$

while the utility from losing is

$$u_{loss} = \sqrt{w_2} - \frac{m}{(N-1)}\sqrt{\Delta w} = \sqrt{975} - \frac{1}{9}\sqrt{49} = 31.22 - 0.78 = 30.44.$$

Since effort is costly, the agents choose zero effort to maximize their probability to not win the tournament.

3.3.5 The Optimal Tournament Structure

Both the FS-model and the ERC-model generally agree that for a given prize spread, effort of an inequity averse agent increases the more winners the tournament has. Should the principal therefore always construct tournaments with the maximum number of winners $m = N - 1$?

Recall that in the model with standard agents, the number of winners is irrelevant for the principal's payoff. The principal's payoff is

$$\max_{e} Nh(e^{FB}) - NU_0 - NC(e^{FB}), \tag{116}$$

with e^{FB} characterized by $h'(e^{FB}) = C'(e^{FB})$, which is independent of the number of winners.

Next, let us consider the FS-model. In this model, the principal's payoff is

$$\begin{aligned} U_P &= Nh(e) - NC(e) - NU_0 \\ &\quad - \frac{m(N-m)}{N-1}(\alpha+\beta)\frac{C'(e)}{h'(e)\left(1+\frac{m}{N-1}\,\alpha-\frac{N-m}{N-1}\,\beta\right) f_{\varepsilon_{l:n}}(0)}. \end{aligned} \tag{117}$$

Making use of the envelope theorem, the partial derivative of the principal's payoff with respect to m is the partial derivative of the fraction

$$-\frac{m(N-m)(\alpha+\beta)}{N-1}\frac{C'(e)}{h'(e)\left(1+\frac{m}{N-1}\,\alpha-\frac{N-m}{N-1}\,\beta\right) f_{\varepsilon_{l:n}}(0)}. \tag{118}$$

The direct derivative of this expression is very inaccessible (see A.3.1), so instead, we look at two parts of the expression separately. First, consider the first fraction $\frac{m(N-m)(\alpha+\beta)}{N-1}$. This part of the expression represents the inequity costs effect. When m agents receive the winner prize, there are m agents suffering compassion costs of $\frac{(N-m)\beta\Delta w}{N-1}$ and $N-m$ agents suffering envy costs of $\frac{m\alpha\Delta w}{N-1}$, so total inequity costs in the tournament are given by

$$\frac{m(N-m)(\alpha+\beta)\Delta w}{N-1}. \tag{119}$$

The partial derivative with respect to m is $\frac{(N-2m)(\alpha+\beta)}{N-1}$, so (119) has its maximum at $m = N/2$ and decreases equally to both sides as m moves away from $N/2$. Second, consider the fraction

$$\frac{C'(e)}{h'(e)\left(1+\frac{m}{N-1}\,\alpha-\frac{N-m}{N-1}\,\beta\right) f_{\varepsilon_{l:n}}(0)}. \tag{120}$$

The partial derivative with respect to m is

$$-\frac{(N-1)(\alpha-\beta)C'(e)}{h'(e)\left(N-1+m\alpha+\beta(N-m)\right)^2 f_{\varepsilon_{l:n}}(0)} < 0. \tag{121}$$

This part represents the incentive effect, which increases in the number of winners because $\alpha > \beta$. So overall, both fractions are decreasing in m for $m \geq N/2$ and therefore the principal's payoff (117) increases as the number of winners increases beyond half of the participants. Therefore, when designing tournaments with agents with FS-preferences, the principal should give the winner prize to $N-1$ agents. Then, the principal has to pay the agents the lowest compensation possible for their inequity costs and agents exert the most effort (see proposition 6) for a given prize spread. Since under unlimited liability, the principal can get back all additional money paid to the winners by lowering the loser prize accordingly, $N-1$ winners it is clearly the principal's preferred tournament structure.

Proposition 11: *When the principal can design the tournament structure, she will choose the maximum number of winners if agents have FS-preferences.*

For agents with ERC-preferences, the results are less straightforward. The principal's payoff is

$$\begin{aligned} U_P &= Nh(e) - NC(e) - mk^W - (N-m)k^L - NU_0 \\ &= Nh(e) - NC(e) - mW\kappa(\varsigma^W) - (N-m)W\kappa(\varsigma^L) - NU_0 \end{aligned} \tag{122}$$

The partial derivative with respect to m is

$$-W\kappa(\varsigma^W) + W\kappa(\varsigma^L) - mW\kappa_{\varsigma^W}(\varsigma^W)\varsigma^W_m - (N-m)W\kappa_{\varsigma^L}(\varsigma^L)\varsigma^L_m, \tag{123}$$

but we can ignore W to simplify to

$$-\kappa(\varsigma^W) + \kappa(\varsigma^L) - m\kappa_{\varsigma^W}(\varsigma^W)\varsigma^W_m - (N-m)\kappa_{\varsigma^L}(\varsigma^L)\varsigma^L_m. \tag{124}$$

Again, the expression consists of two parts. The first one is $-\kappa(\varsigma^W) + \kappa(\varsigma^L)$, which represents the incentive effect. It is positive when there are more winners than losers in the tournament (because $-\varsigma^L > \varsigma^W$ for $m > N/2$). The second part represents the inequity costs effect. The partial derivative of ς_i with respect to m is $-\frac{\Delta w w_i}{(Nw_2+m\Delta w)^2} < 0$. Increasing ς_i increases the inequality costs of winners, as their share moves further away from $1/N$. On the other hand, it decreases the inequality costs of the losers as they move further toward $1/N$. Therefore, we know

$$-m\underbrace{\kappa_{\varsigma^W}(\varsigma^W)}_{>0}\underbrace{\varsigma_m^W}_{<0}-(N-m)\underbrace{\kappa_{\varsigma^L}(\varsigma^L)}_{<0}\underbrace{\varsigma_m^L}_{<0}. \tag{125}$$

The first part is always positive, the second one always negative. Although the expression is positive for $m = N/2$ because at this point $-\varsigma^L = \varsigma^W$ and $\varsigma_m^W < \varsigma_m^L$ generally, it does not necessarily stay positive. If m increases further, ς^L moves further away from the equal share $1/N$ while ς^W moves closer toward it, and if $\kappa(\varsigma)$ is very steep, $\kappa_{\varsigma^L}(\varsigma^L)$ may overpower all other effects and make expression (124) negative. In that case, the principal's payoff would decrease if m increases. So depending on the steepness of the inequity costs, the principal will either give the winner prize to $N/2$ of the participants or to $N-1$ participants.

Let us also consider the exemplary ERC-utility function (50), which adjusts to the following expression in multi-agent-tournaments

$$\begin{aligned}U_i &= w_i - W\frac{b}{2}\left(\frac{w_i}{Nw_2+mw_1}-\frac{1}{N}\right)^2\\ &= w_i - (Nw_2+mw_1)\frac{b}{2}\left(\frac{w_i}{Nw_2+mw_1}-\frac{1}{N}\right)^2.\end{aligned} \tag{126}$$

In a tournament with N participants and m winners, the total inequity costs are

$$\begin{aligned}IC &= m(Nw_2+m\Delta w)\frac{b}{2}\left(\frac{w_2+\Delta w}{Nw_2+m\Delta w}-\frac{1}{N}\right)^2\\ &\quad + (N-m)(Nw_2+m\Delta w)\frac{b}{2}\left(\frac{w_2}{Nw_2+m\Delta w}-\frac{1}{N}\right)^2.\end{aligned} \tag{127}$$

The partial derivative with respect to m (see A.3.2) can be simplified to

$$\frac{bN^2w_2(\Delta w)^2 - 2bmNw_2(\Delta w)^2 - bm^2(\Delta w)^3}{2N^3(w_2)^2 + 4mN^2w_2\Delta w + 2m^2N(\Delta w)^2}. \tag{128}$$

For $m = N/2$, the numerator becomes

$$-b\frac{N^2}{4}(\Delta w)^3 < 0 \tag{129}$$

Therefore, inequity costs strictly decrease in the number of winners m independent of the degree of inequity aversion b. As a result, if agents have ERC-preferences expressed by (126), the principal will always give the winner prize to $N-1$ agents. Note, however, that this results does not necessarily hold for a more general utility function

$$U_i = w_i - W\frac{b}{2}\left(\frac{w_i}{Nw_2 + mw_1} - \frac{1}{N}\right)^x \tag{130}$$

if $x > 2$.

Proposition 12: *When the principal can design the tournament structure, she will choose the maximum number of winners if agents have ERC-preferences with inequity costs that are not overly steep. Otherwise, she will give the winner prize to half of the agents.*

Remember that those results pertain only when the random variable is uniformly distributed and $f_{\varepsilon_{l:n}}(0)$ is independent of m. It is easy to see from (117) that the principal's payoff in the FS-model increases as $f_{\varepsilon_{l:n}}(0)$ increases. The same applies for the ERC-model, but it is very difficult to show this without a concrete function for the inequity costs. However, the same intuition applies for both models: From (92) and (122), it follows that the necessary prize spread to induce a certain level of effort decreases in $f_{\varepsilon_{l:n}}(0)$. A lower prize spread implies lower inequity costs, so the principal is better off for each level of effort she wants to implement.

Recall that $f_{\varepsilon_{l:n}}(0)$ is the integral in (79), which consists of three parts. Replacing l with $N-m$ and n with $N-1$, these three parts are $\frac{N-1!}{(N-m-1)!(m-1)!}$, $F(y)^{N-m-1}\left[1-F(y)\right]^{m-1}$ and $f(y)^2$. The latter expression is independent of m and the other two are obviously both maximized if $m = \frac{N}{2}$, so $f_{\varepsilon_{l:n}}(0)$ is maximized if $m = \frac{N}{2}$, too. Therefore, all other things equal, the principal prefers to give the winner prize to $\frac{N}{2}$ agents if the random component is not uniformly distributed.

Taken together, whether the aggregate effect is such that the principal prefers to give the winner prize to the maximum number of agents or to half the agents is impossible to say generally because of the complex expression of $f_{\varepsilon_{l:n}}(0)$. However, it is clear that giving the winner prize to only one agent is suboptimal.

3.4 Discussion

Both the FS-model and the ERC-model agree that the principal is always worse off with inequity averse agents compared to purely self-regarding ones. However, this is no surprise because both models introduce costs to the standard model which somebody has to pay. Since the agents never receive any rent under unlimited liability anyway, it inevitably has to be the principal who pays these costs.

The first disagreement between both models arises in tournaments with two agents. Under the FS-model, the prevalence of envy over compassion creates an incentive effect as agents try to avoid the more costly emotion of envy by winning the tournament. No such incentive effect occurs in the ERC-model because it does not distinguish between advantageous and disadvantageous inequality.[48]

While it stands to reason that being ahead is preferable to being behind, it is far from clear that the reason for this extends beyond the tangible benefits. For example, while everyone should prefer an allocation of 12 for oneself and 8 for the other (12|8) over 8 for oneself and 12 for the other (8|12), is it necessarily true that people prefer 12|8 over 12|16 as the FS-model would predict? Some studies have found evidence that disagrees with the FS-model on this.[49] If Fehr and Schmidt's assumption $\alpha > \beta$ turned out to be incorrect, all models owing their main results to this assumption would immediately become obsolete.

However, the point is not to question the validity of the FS-model's assumptions, but to call for more robustness checks when the results of an analysis depend on such a pivotal assumption. For example, in the tournament model, the incentive effect of inequity aversion vanishes if we dismiss the assumption that envy is stronger than compassion. Yet, at the same time, the incentive effect returns if we instead consider status seeking preferences, i.e., $\beta < 0$.[50] In fact, with status seeking agents, the principal would be even better off than with purely self-regarding agents if the utility from

[48] To be precise, the original ERC-model does neither postulate nor preclude differences between advantageous and disadvantageous inequality, as both instances are theoretically compatible with the general ERC-value function.

[49] For example, Engelmann and Strobel (2004) do not find much evidence for such preferences when examining dictator choices over three-person-allocations to compare the social preference models of Fehr and Schmidt, Bolton and Ockenfels, and Charness and Rabin (2002).

[50] The incentive effect can easily be deduced from (47).

high status exceeds any disutility from low status (e.g. envy), i.e., if $-\beta > \alpha$ (see A.1.2).

Extending the analysis to more forms and functions of social preferences does not only verify the universality of results, it also allows to examine which type of agent is more likely to self-select into a particular incentive mechanism, which type is preferred by the principal, and whether this creates a natural match or mismatch. For example, both compassionate and envious types tend to dislike tournaments, while status seekers tend to like them. Principals generally prefer status-seekers with low envy, so agents' and principals' preferences are in accord with each other. Looking at real life firms utilizing tournament incentive schemes, we should, therefore, expect to find people who enjoy being better off than others but who can cope well when they are worse off.[51] Strongly envious and particularly compassionate types, however, should be rare in real-life tournaments.

Regarding multi-agent-tournaments, we find that although the FS-model and the ERC-model yield slightly different results in detail, they both generally agree that agents tend to prefer being in the majority group. Losing the tournament is less bad when most of the other agents also lose and worse when most of them win. From the principal's perspective, the best tournaments have many winners, the worst have many losers. The two models paint different worst case scenarios though. On the one hand, in the FS-model, a small prize spread in association with a high loser prize (and diminishing sensitivity) can distort incentives to such a degree that agents do not want to win the tournament. On the other hand, the ERC-model predicts that agents may lose interest in the tournament when the prize spread exceeds a certain size. Although seemingly contradictory, both results may be conceivable depending on the interpretation of compassion. For the first case, consider a department where the manager tenders a very small monetary incentive to the best performer. The employees may care more about being perceived as a "bootlicker" by their colleagues when winning than they care about the small prize money. If the manager increases the award significantly, the employees' attitude will eventually change. For the second case, consider a senior manager deciding whether to enter a promotion tournament for a leadership position. If the position's salary is exorbitantly high, she may consider it unethical and decline to participate in the tournament, choosing

[51] Alternatively, a status seeker could also be fairly envious, but have a low subjective probability of losing, e.g. because of overconfidence.

instead to compete for a position with a more moderate salary increase.

If we take the result that many-winner-tournaments are better at face value, we cannot help but wonder why tournaments with a majority of winners are seemingly rare in reality? On the one hand, more often than not, it may just be unfeasible to have that many winners. In promotion tournaments, for example, there are simply not enough positions to fill. Alternatively, principals may face restrictions in the design of the prize structure, e.g. unlimited liability may not be possible. On the other hand, if tournaments are interpreted as competition on a societal level - much like Lazear and Rosen (1981) do in their article - we could argue that people actually are engaged in a tournament with many winners where the winner prize is not suffering social decline (e.g. being fired and becoming unemployed or something worse). After all, it may not be completely outlandish to argue that the presence of a large middle class in a society creates an incentive effect for the general population to belong at least to that class.

Regarding the external validity of the theoretical results, it should be noted that Orrison et al. (2004) experimentally analyze agents' effort decisions in six-agent-tournaments over many rounds. Their results are essentially contrary to the predictions of the inequity aversion models. They find that effort choices are lower with four winners than with two or three winners.[52] The authors conjecture that agents shirk more when prizes are given to many participants because they anchor on the high winning probability and do not thoroughly understand the difference between marginal and total winning probabilities. Their result suggests that there are other psychological factors working contrary to the effects of inequity aversion which may render many-winner-tournaments unsuitable in practice. However, it would be interesting to see if the results of Orrison et al. extend to real effort tasks and different distributions of the random variable.

3.5 Conclusion

In this essay, we apply the inequity aversion models of Fehr and Schmidt as well as Bolton and Ockenfels to a simple tournament model. We find that irrespective of the model, inequity aversion generally reduces the principal's payoff when she can design the tournament's structure. With respect to incentive effects for a given prize structure, both models yield different

[52]Similar results are obtained by Harbring and Irlenbusch (2008).

prediction. The increase in effort under the FS-model is not found under the ERC-model for two-agent-tournaments. In tournaments with more than two agents, both models largely agree that incentives are distorted toward the majority group. Under certain circumstances, this effect can lead to the elimination of all incentives to win the tournament.

4 Does Altruism Depend on Rational Expectations?

4.1 Introduction

Reference points are values or states of nature that act as benchmarks or indicators for judgments, comparisons, and appraisals. As such, they constitute a pivotal element of decision making. Their influence extends to various spheres of economic behavior, including – but not limited to – risk attitudes (Kahneman and Tversky 1979; Koszegi and Rabin 2007), stock trading (Odean 1998; Meng 2010), labor supply (Camerer et al. 1997; Fehr and Camerer 2007; Crawford and Meng 2010), effort provision (Mas 2006; Abeler et al. 2011), consumption behavior (Bell and Bucklin 1999; Koszegi and Rabin 2009), brand choice (Hardie et al. 1993; Mazumdar et al. 2005), social judgments (Holyoak and Gordon 1983), negotiations (Kristensen 1997), and job satisfaction (Ockenfels et al. 2010). While some aspects of reference points are relatively well understood – for example, that negative deviations usually loom larger than positive deviations (Tversky and Kahneman 1991) or how reference points anchor judgments and decisions (Mussweiler and Strack 1999; Epley and Gilovich 2006) – other issues are still unresolved. In particular, the question which reference point is chosen when more than one is available is still a much debated topic. Some models adopt the status quo as the reference point (Samuelson and Zeckhauser 1988; Kahneman et al. 1991; Masatlioglu and Ok 2005) whereas others endorse rational expectations (Loomes and Sugden 1986; Koszegi 2006). When the status quo is also the best predictor of the future, both reference points naturally coincide, but when they do not, which one do decision makers pick? Since it is very difficult to reliably observe expectations in the field, laboratory experiments provide a good opportunity to gain insights into the importance of expectations for the formation of reference points.

A couple of recent studies directly address this issue. Abeler et al. (2011) manipulate expectations in a real-effort experiment. Their subjects work on a tedious task and can quit whenever they desire. After they have finished working, the subjects either receive a piece rate for each solved tasked or a fixed payment, each with 50% probability. By manipulating the size of the fixed payment, the authors vary the subjects' expectations about their payment and find that more effort is provided when the fixed payment is higher.

Moreover, Ericson and Fuster (2010) endow subjects with an item which they are allowed to trade with some probability. The lower this probability, the more likely the subjects are to keep the item. Ericson and Fuster also elicit willingness-to-pay prizes for items that the subjects may also receive for free in a lottery. The higher the winning probability in the lottery, the higher a subject's valuation of the item.

The idea for this particular study is based on psychological research on outcome favorability and pro-social behavior. Some psychological studies find that subjects who have experienced success in a task are more willing to work on behalf of someone else (Berkowitz and Connor 1966), to donate to charity, and to help others (Isen 1970). These results have also been replicated with children who donate more (Isen et al. 1973; Barnett and Bryan 1974) and share more with others (Bryant 1983) after being successful in bowling. Pleasant surprises also have a similar effect. Subjects who have found a dime in the coin return of a telephone booth or who have received cookies while studying are more willing to mail a lost letter (Levin and Isen 1975) or help picking up dropped papers (Isen and Levin 1972). The increased pro-social behavior is usually attributed to positive mood induced by the favorable event (Isen 1999). However, from an economics perspective, mood is a rather vague and intangible concept (although Hermalin and Isen 2007 have developed an intertemporal economic model of mood). Instead, the pleasant events described above can also be understood as positive deviations from a reference point and the evidence suggests that such deviations increase pro-social behavior. Recently, Gneezy and List (2006) have used pleasant surprise in a field experiment with a gift-exchange setting by paying higher salaries than previously announced. Subjects receiving a higher payment than expected show increased effort for a few hours, although the net benefit is negative compared to regularly paid subjects. Also related is the study of Falk (2007), who attaches gifts to solicitation letters asking for charity donations, finding that the likelihood of donations increases with the value of the gift.

We test the hypothesis that expectation-based reference-dependent preferences influence pro-social behavior by letting subjects play a dictator game in which the endowment is determined by a lottery. In our study, we compare two groups of subjects who receive the same endowment, but have faced different expected outcomes in the lottery, i.e., in one group subjects could also have received a higher endowment, whereas in the other group, they could have received a lower endowment. This way, we can test whether expectations affect the subsequent dictator transfer while controlling for income effects. We get some weak confirmation of our hypothesis, but also find evidence for other reference points interfering with the effect we try to identify.

4.2 Theoretical Predictions

The theoretical predictions regarding reference-dependent utility are based on the model by Koszegi (2006), which to our knowledge is not only the most widely used model of expectation-based reference points, but also easy to comprehend mathematically. To model altruistic preferences, we fall back to the CES-utility function of Andreoni and Miller (2002). This decision may appear unusual in the face of the supply for more commonly used models of social preferences. However, we require a framework that does not produce corner solutions which rules out linear models like Fehr and Schmidt (1999) or Charness and Rabin (2002). We could have utilized the ERC-model by Bolton and Ockenfels (2000), but it seems improper to combine expectancy-dependent utility with the relative payoff expression of the model because this would entail that depending on the reference point, an individual's most preferred share lies above or below the equal share. That leaves us with the CES-utility function of Andreoni and Miller (2002), which works well in our context.

Koszegi (2006) assume that an individual's utility $u(c|r)$ from an outcome c is the sum of his classical outcome-based utility $m(c)$ and gain-loss-utility $n(c|r)$ depending on the person's reference point r, i.e., $u(c|r) = m(c) + n(c|r)$. The reference point r is the person's recent probabilistic belief and the gain-loss-utility is a function of the difference between the outcome and the reference point, i.e., $n(c|r) = \mu(m(c) - m(r))$. The function $\mu(\cdot)$ satisfies the properties of Kahneman and Tversky's (1979) value function. Therefore, the individual receives additional positive utility when the realized outcome is higher than the reference point and suffers from additional disutility when it is lower.

Consider the two lotteries L_L and L_H, each with two equally likely outcomes E_L, E_M and E_M, E_H, respectively. The three outcomes are ranked $E_H > E_M > E_L$. The reference points in both lotteries are the expected values, i.e., $r_L = 0.5E_L + 0.5E_M$ and $r_H = 0.5E_M + 0.5E_H$, respectively. The middle outcome E_M is higher than the reference point r_L in lottery L_L, but lower than the reference point r_H in lottery L_H. Therefore, with reference dependent utility, receiving E_M in lottery L_L yields a higher utility than receiving E_M in lottery L_H, i.e.

$$\begin{aligned} & r_L = 0.5E_L + 0.5E_M < 0.5E_M + 0.5E_H = r_H \\ \Leftrightarrow \quad & n(E_M|r_L) > n(E_M|r_H) \\ \Leftrightarrow \quad & m(E_M) + n(E_M|r_L) > m(E_M) + n(E_M|r_H) \\ \Leftrightarrow \quad & u(E_M|r_L) > u(E_M|r_H). \end{aligned}$$

In the dictator game, we assume that individuals receiving an endowment E make their transfer decision x according to their utility function

$$u_i = (a[E - x]^\rho + (1 - a)x^\rho)^{\frac{1}{\rho}}, \tag{131}$$

which is the CES-utility function of Andreoni and Miller (2002), where a is the degree of selfishness with $0 \leq a \leq 1$ and ρ the convexity of preferences, yielding the elasticity of substitution $\sigma = 1/(\rho - 1)$. An individual maximizes his utility by choosing the transfer x that satisfies

$$x = E\frac{1}{1 + A}$$

with $A = \left(\frac{a}{1-a}\right)^{\frac{1}{1-\rho}}$ and $A \geq 1$ if we assume $a \geq 0.5$, i.e., if the own payoff is valued at least as high as the other person's payoff.

If individuals get their endowment from a lottery, we can introduce reference-dependent-utility from the realization of the endowment to the utility function, i.e., (131) changes to

$$u_i = (a[E - x + n(E|r)]^\rho + (1 - a)x^\rho)^{\frac{1}{\rho}} \tag{132}$$

and the individual will make a transfer given by

$$x = [E + n(E|r)]\frac{1}{1 + A}.$$

Therefore, if the endowment is determined by the lotteries L_L and L_H, we should expect individuals receiving E_M from lottery L_L to make higher transfers than those receiving the same endowment from lottery L_H.

4.3 Experimental Design

The experiment was conducted using the software zTree (Fischbacher 2007) at the Cologne Laboratory for Economic Research at the University of Cologne, Germany. 264 subjects, who were recruited via the Online Recruitment System ORSEE (Greiner 2004), participated in twelve separate sessions. The average age was 24 years and 52.2% of the participants were female. The subjects stayed at the lab for about 25 minutes and earned €6.73 on average, including a €2.50 show-up fee and a €0.50 bonus that was given to ensure that each participant received more than the extra invitees who could not participate and were given the show-up fee, so that receivers would not be too frustrated in case they were not given any points by their senders. Once the experiment had started, the subjects were divided into senders and receivers to play a standard dictator game where the endowment was determined by a lottery. In the *LOW*-treatment, the possible endowments were 100 points (*L100*) and 150 points (*L150*) and in the *HIGH*-Treatment, they were 150 points (*H150*) and 200 points (*H200*). One point equaled five cents.

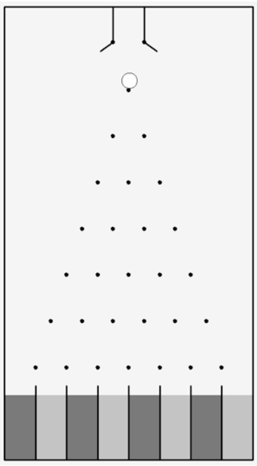

Figure 2: The Random Mechanism

The main treatment comparison occurs betweens the subjects who receive 150 points in the *LOW*-treatment (*L150*) and those who receive 150 points in the *HIGH*-treatment (*H150*). The amount of 150 points was chosen so that the equal split does not align with a natural focal point like 100 or 50. Overall, we had 132 senders, 66 in each treatment. In the *LOW*-treatment, 35 subjects received 100 points and 31 received 150 points. In the *HIGH*-treatment, 33 senders received 150 and 200 points, respectively. The disparity in the *LOW*-treatment is due to some variation in session size and chance.

After the subjects were assigned randomly to an isolated computer terminal, instructions differing for the two roles (see B.1) were displayed on the computer screen. Subjects were given as much time as they needed to read the instructions and ask questions, although less than a handful of subjects had any inquiries. The senders were told that they would receive an endowment of points which they could allocate between themselves and the receiver as they desired. However, the exact amount of points available would be determined by a random mechanism. They were informed about the two possible amounts of points and that both realizations were equally likely. After the senders indicated that they had understood the instructions, a random mechanism (see Figure 2) determined the realized endowment. After every sender had pressed a button, a small ball fell down several levels to end up either in a light-gray or dark-gray box, determining whether the endowment was high or low. The visual representation was used for two reasons: First, to make the random draw more transparent, and second, to create more suspense for the senders, which we hoped might increase the psychological effect. The winning color alternated between subjects and so did the order in which the two possible endowments were presented. Thereby, we controlled for any effects the color might have as well as for possible anchoring effects of the first number shown. After the lottery ended, senders declared how many points they wanted to transfer to their receiver. On average, senders confirmed their transfer 34 seconds after the conclusion of the lottery (median 30 seconds).

Meanwhile, the receivers were informed about their role and that the senders would receive a randomly determined amount of points to allocate between themselves and the receiver. They were not informed about either the possible or the realized amounts and the senders knew that. To generate some data from the receivers, they were also asked to make decisions in a dictator game with random endowment, although they of course knew the decisions were only hypothetical. However, contrary to the senders, receivers answered for both possible endowments and did not encounter the visual random mechanism. When the senders were in the *LOW*-treatment, receivers were confronted with the two possible amounts of the *HIGH*-treatment and vice versa, so we could truthfully tell the receivers that the situation they faced was not the one of their assigned sender. After both receivers and senders made their decisions, they answered a questionnaire before they received their payment.

4.4 Results

4.4.1 Main Results

The main variable of interest in the analysis is the "transfer ratio", i.e., the ratio of a subject's transfer to his endowment, instead of the transfer itself.[53] For a quick overview, Figure 3 shows the histograms of the senders' transfer ratios for all four sub-treatments and Table 1 gives the basic statistics.[54]

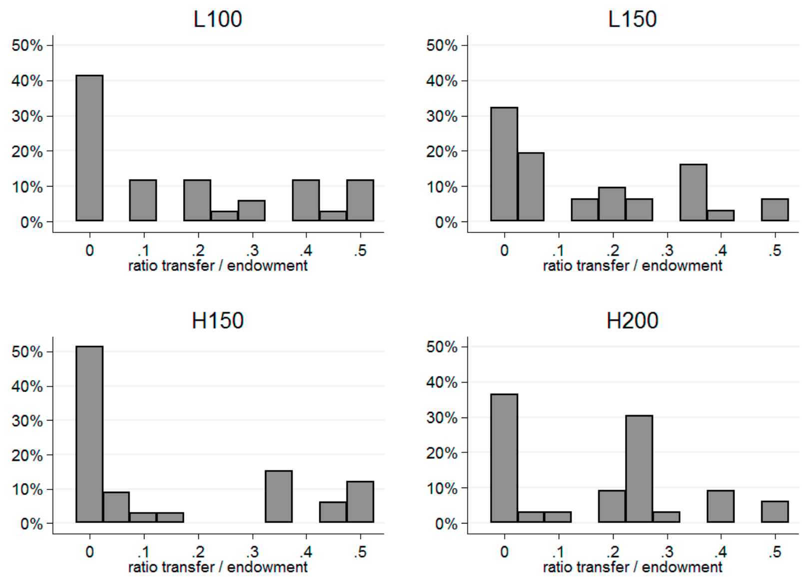

Figure 3: Ratio of Transfer/Endowment for Senders

Comparing the transfer decisions in the two main sub-treatments *L150* and *H150*, the data do not immediately show a clear support of the hypothesis that transfers are higher in *L150* than in *H150*. A one-sided Mann-Whitney-U-Test (MWU) only gives a p-value of 0.23 (z = 0.737), which

[53]For example, if a subject in *L100* transfers 25 points, the transfer ratio is 0.25 or 25% (25/100) and if a subject transfers 30 points in *L150*, the transfer ratio is 0.20 or 20% (30/150). This should make it easier to compare results across sub-treatments with different endowments.

[54]Note that in the sub-treatment *L100*, one subject gave 80% of the endowment, which is not represented in the histogram in order to retain the same dimensions across all four graphs. Table 1 contains one entry with and one without this subject. It appears that this subject simply made a mistake because in the questionnaire, the subject stated that "*instead of the fair 50 points I gave only 20 points*".

slightly points into the right direction, but is not statistically significant. Other non-parametric tests like the Kolmogorov-Smirnov-Test or a Two-Sample Randomization-Test do not indicate any treatment differences either. However, consistent with the hypothesis, subjects are more likely to keep the whole endowment in *H150* (16 of 33, 48.5%) than in *L150* (9 of 31, 29%). This result is weakly significant on a one-sided Fisher's Exact Test (FE, p=0.09). A one-sided logit-regression confirms that subjects are more likely to choose a positive transfer in *L150* than in *H150* (p=0.055) and this treatment difference remains weakly significant in all estimated alternative models (see Table 14).

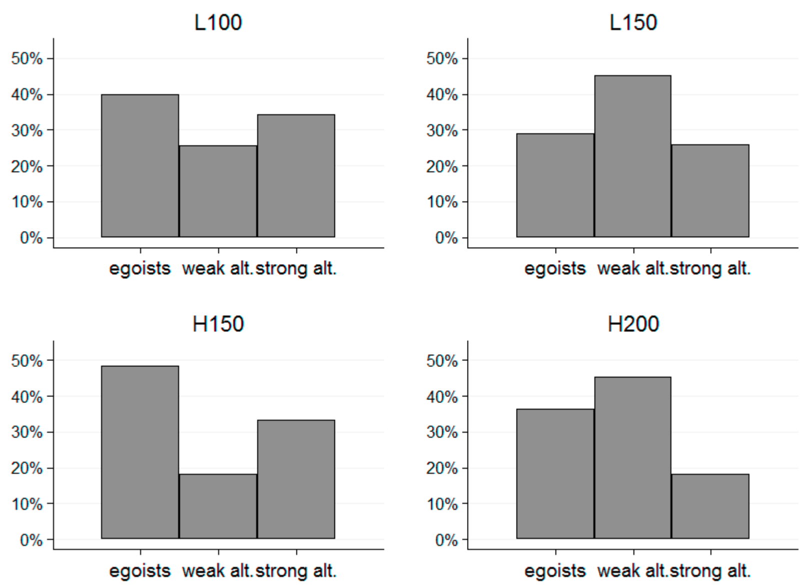

Figure 4: Categorization of Senders

From the histograms, it appears that the presumed treatment effect mostly affects low contributors. Dividing subjects into the three categories *egoist* (transfer = 0), *weak altruist* (0 < transfer ≤ 25%), and *strong altruists* (transfer > 25%), we get the distributions presented in Figure 4 and Table 2. A χ^2-test weakly confirms that the types differ in frequencies between *L150* and *H150* (p=0.062, $\chi^2 = 5.58$). A one-sided MWU-Test using only egoists and weak altruists shows highly significant differences between *L150* and *H150* (p=0.006, z=-2.50).[55] Overall, while there is some support for the presumed treatment effect in the data, the hypothesis that there are no difference between the treatments cannot be rejected with as much confidence as we would typically like.

	Mean	StD	Median	N
All	.169	.184	.100	132
L100	.197	.215	.100	35
	(.179)	(.190)	(.100)	(34)
L150	.154	160	.067	31
H150	.151	.197	.013	33
H200	.174	.159	.200	33

Table 1: Transfer Ratios of Senders

	Egoists	Altruists Weak	Altruists Strong
All	51	44	37
L100	14	9	12
L150	9	14	8
H150	16	6	11
H200	12	15	6

Table 2: Categorization of Senders

4.4.2 The Warm Glow of Giving? – Mood

In the questionnaire, senders are asked to recall their mood before the draw of the random mechanism, after the draw of the random mechanism, and after their transfer decision, and rate it on a scale from 0 (terrible) to 10 (excellent). The average mood among senders before the lottery is very similar across all four sub-treatments, with no significant differences. The success in the lottery has the expected effect on mood (see Figure 5 and also Figure 7 in the appendix). To compare mood before and after the lottery, we use a Wilcoxon Matched-Pair Signed-Rank Tests (MPSR) which yields p-values below 0.001 in all four sub-treatments, i.e., mood increases significantly if the subject wins the lottery and decreases significantly if the subject loses. However, the subjects' average mood also changes significantly again in the opposite direction after they have made their transfers. The mood decreases for subjects who have won the lottery (MPSR p = 0.003, z=3.01 for *L150*, p = 0.043, z = 2.02 for *H200*) and increases for subjects who have lost the lottery (MPSR p = 0.019, z=-2.34 for *L100*, p = 0.003, z = 3.01 for *H200*; see Figure 8 in the appendix). This effect is confirmed by a highly significant

[55] Although this (rather conveniently) ignores that *H150* has a small advantage in strong altruists, 11 to 8.

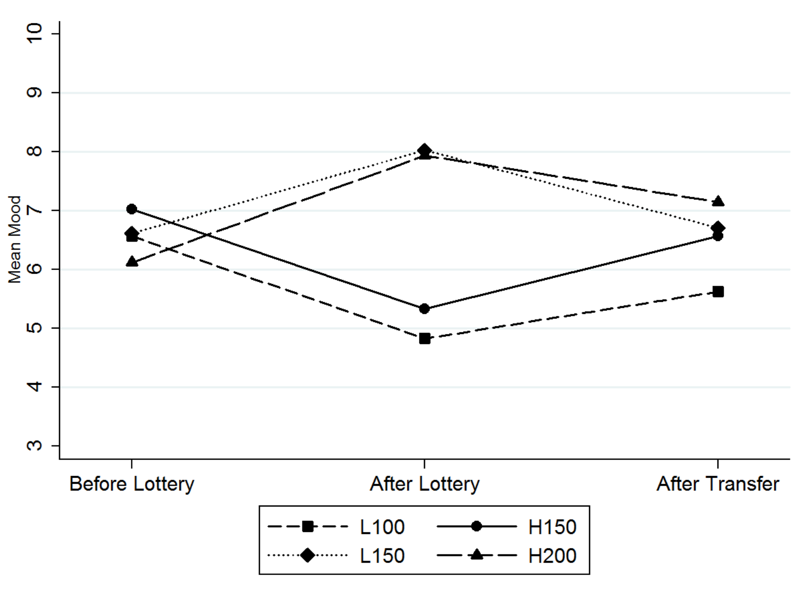

Figure 5: Mood of Senders

negative correlation between the mood change after the lottery and the mood change after the transfer (Spearman's ρ = -0.53, p $<$ 0.0001), i.e., when the mood increases after the lottery, it tends to decrease after the transfer and vice versa (Figure 9). Taken together, the overall effect does not completely balance out (Figure 10). Subjects in *L100* are still in significantly worse mood at the end of the experiment than at the start (MPSR p $<$ 0.001, z=3.43), whereas those in *H200* are in significantly better mood (MPSR p = 0.013, z=-2.49). For the two main sub-treatments, the subjects in *L150* experience a slight increase in mood on average while those in *H150* experience a slight decrease. The difference between both sub-treatments is weakly significant (MWU p=0.081, z=-1.75).

Contrary to our expectations, there is virtually no direct relationship between the subjects' mood and their transfer behavior. All correlations between absolute mood or mood change and the transfer-ratio are generally very small and insignificant (see Table 4 in the appendix). Other tests like OLS-regressions do not show any significant effect of mood or mood change on transfers either. There are not even any differences between egoists, weak altruists, and strong altruists as mood improvement and mood deterioration are essentially identically distributed among the three types (Table 5). For example, the average transfer ratio of subjects whose mood improves after the transfer is 18.9% and the average transfer ratio of subjects whose mood worsens is 18.3%. To sum up, we find no evidence that positive mood induces higher transfers (or the other way around).

4.4.3 Fairness

The surprising result of the mood analysis is that mood "changes direction" after the transfer, yet there appears to be no relationship between mood and transfer behavior at all. One explanation could be that the initial mood change caused by the lottery is simply very short lived and vanishes on its own within the approx. 30 seconds it takes the senders to make their transfer decision. However, in other studies, similar effects last from 20 minutes (Isen et al. 1976) up to several hours (Gneezy and List 2006). Another possible explanation is that the subjects in the sub-treatments perceive their "moral obligation" differently and that their change in mood after the transfer reflects that. We can test the latter explanation because subjects are asked what they consider a "fair" transfer for their own situation. Over all sub-treatments, the average fair ratio of transfer/endowment is 36.6% with a mode and median answer of 50% (65 subjects). Two other ratios are chosen more than ten times, 33.3% (14) and 0% (13), respectively. [56]

The average stated fair ratio is 39.2% (median 50%) when subjects win (*L150* and *H200*) and 32.8% (median 40%) when subjects lose (*L100* and *H150*). The difference is weakly significant (two-sided MWU p=0.062, z=-1.86). When we only consider the two main sub-treatments *L150* and *H150*, the fair ratio in *L150* has the higher mean and median (39.2% to 32.8%; 50% to 40%), but the difference is not significant (two-sided MWU p = 0.13, z=-1.52). There is some empirical evidence that positive mood increases fairness concerns (Carnevale and Isen 1986) and indeed, there is a positive and significant correlation between a subject's mood change after the lottery and the stated fair ratio ($\rho = 0.19$, p=0.03), i.e., improved mood leads to a higher fairness standard.

Can the relationship of fair ratio to actual transfer ratio help explain the mood change after the transfer? First, note that the stated fair ratio is a relative good predictor of a subject's actual transfer (Table 6). It is highly

[56] Two subjects gave answers above 50%, 52.6% (79 of 150) and 100% (150 of 150). Both answers are ignored for the remaining analysis because they are probably mistakes of some sort.

significantly correlated with the actual transfer ratio (Spearman's $\rho = 0.36$, $p < 0.001$), although if broken down for each sub-treatment, the relationship seems to be stronger for the two loss conditions (Table 7). Indeed, only 25% of subjects in *L150* and *H200* transfer the amount they consider fair, whereas 36.8% of the subjects in *L100* and *H150* do (FE p=0.13). Winners deviate more from their own fairness standard (-22.7%[57]) than losers (-16.3%), a significant difference (MWU p=0.045, z=-2.01).

In sum, subjects who win the lottery have a higher fairness standard, but also deviate more from this standard, apparently "paying" with some of their improved mood from the success in the lottery to be less fair. On the other hand, subjects who lose the lottery remain closer to their fairness standard, which makes them feel good about themselves, compensating for their mood loss after the lottery. This explanation is consistent with the fact that the treatment effect is mostly found for lower transfer ratios. Both effects may more or less balance out for strong altruists, but cannot balance out for weak altruists because their fairness standard drops too close to zero.[58] However, there is a huge variance in the data on the individual level, so we must be very careful with drawing final conclusions. At least, we can establish that the relationship between mood and pro-social behavior is not as straightforward as it appears in other studies.

4.4.4 Aspirations

Besides the status quo and expectations, Kahneman and Tversky (1979) also suggest the possibility that aspirations can act as reference points. Therefore, we also ask senders if they came to the lab with a certain monetary goal in mind. 59 senders answered in the affirmative, with the average monetary goal being €7.92 (median 7.5, mode 5 and 10). 53 of the 59 subjects have monetary goals matching or exceeding their maximum possible income, i.e., the sum of their endowment plus show-up fee and bonus. Therefore, they need to keep the whole endowment to come as close as possible to their target. However, monetary goals appear to have little to no influence on the actual transfers. Although the average and median transfer ratio of subjects with a

[57]The difference between fair ratio and actual ratio is in absolute terms, i.e., if the fair ratio is 50% and the actual transfer ratio is 40%, the difference is 10%.

[58]Note that of nine of the twelve "fair egoists", i.e., subjects who consider no transfer to be fair, lose in the lottery.

monetary goal are slightly lower than those of other subjects (mean: 0.159 to 0.178; median: 0.1 to 0.2) and their share of egoists is higher (26 of 73, 25 of 59), too, no test (MWU: p = 0.424, z=0.8; MT: p=0.698, Pearson $\chi^2 = 0.15$; FE: p=0.475) shows any significant differences. Transfer ratios of senders with high ($\geq$ €7.50) and low ($<$ €7.50) goals do not differ significantly either (mean: 0.153 to 0.165; median: 0.06 to 0.1; MWU p=0.756, z=0.32; MT p = 0.87, Pearson $\chi^2 = 0.03$; FE p=0.605; see Table 9).

Nevertheless, some subjects state in the questionnaire that their goal did in fact influence their decision as these selected quotes show:

- "*Since I saw that I would not reach the 10 Euro, I took as much as I could.*"
- "*[...] once I knew it was possible to reach [the 6 Euro], my goal was to allocate accordingly.*"
- "*Because of my goal, I kept all points for myself.*"
- "*It influenced my decision so that I gave the receiver only 25% of the endowment to reach my income goal.*"

The great majority of the 59 subjects, however, stated that their aspiration level did not influence their decision (with some adding that they would have kept the whole endowment anyway). Taken together, while it seems that some people are influenced by it, in general, the aspiration level has no noticeable effect on the aggregated results.

4.4.5 Beliefs

Another possible reference point is the belief about what other subjects in the same situation do. In the questionnaire, senders state their belief about the mean contribution of other subjects with the same endowment. Although we do not incentivize accuracy of beliefs, they are remarkably close to the actual contributions. On average, senders believe that other senders give 17.9% of the endowment, which is less than 1% higher than the actual average transfer ratio of 17%. Differences between the four sub-treatments are small and insignificant (Table 8). We do not find any correlation with mood or mood change (all p-values $\geq$ 0.285 or higher) and there is also no effect of the lottery's result on beliefs in general (MWU p = 0.56, z = -0.58) or in the two main sub-treatments *L150* and *H150* (MWU p = 0.393, z=-0.85).

A sender's beliefs about the contribution of others is – by far – the best predictor for his own transfer. The Spearman correlation coefficient is $\rho = 0.636$ (p < 0.001) and a simple regression explaining the transfer ratio with the ratio of belief/endowment (belief ratio) gives an R^2 of 0.4. However, only 48 of the 132 senders (36.4%) acknowledge that they were actually thinking about what the other participants might do when they made their transfers. Although these 48 senders have a stronger correlation between belief and transfer ($\rho = 0.75, p < 0.001$), the correlation for the other 84 senders remains quite high ($\rho = 0.57, p < 0.001$) and the aggregated beliefs do not differ significantly between both groups (means: 17%, 18.4%; MWU p = 0.81, z = -0.24).

It appears that the belief is a very strong reference point, both consciously and unconsciously. The unconscious effect may be channeled through the fairness perception as beliefs and fairness standards are highly correlated (ρ = 0.42, p < 0.001). However, since we did not incentivize accuracy of beliefs, senders might use their beliefs as a pretext to justify their own (low) transfers.

4.4.6 Receivers

Receivers are asked to make the same transfer decision as senders, only hypothetically and for both endowments at the same time, i.e., either 100 and 150 or 150 and 200.[59] An overview of the answers is given in Figure 6 and Table 3. Not surprisingly, the average hypothetical transfer ratio of 27.9% is higher than the average real transfer ratio of 17%.

However, more interesting, the receivers show a clear distinction between the good and the bad outcome. Of the 61 receivers making their decision in the *LOW*-Treatment, 48 choose a higher ratio (not amount!) in *L150* while only 13 choose the same ratio in both cases. Of the 64 receivers deciding in the *HIGH*-Treatment, 46 choose a higher ratio in *H200* and 18 choose the same ratio. Nobody has a higher ratio when losing the lottery, meaning, among other things, that nobody choose the same transfer for both outcomes. Both the within treatment differences (*L100* to *L150* and *H150* to *H200*) and the between treatment differences (*L150* to *H150*) are highly significant (MPSR

[59] In the following analysis, we ignore seven subjects who made hypothetical transfers higher than 50% in any entry. We assume these receivers either misread the instructions and entered the amount they would have kept or purposefully gave a nonsensical answer out of frustration over being in the passive role.

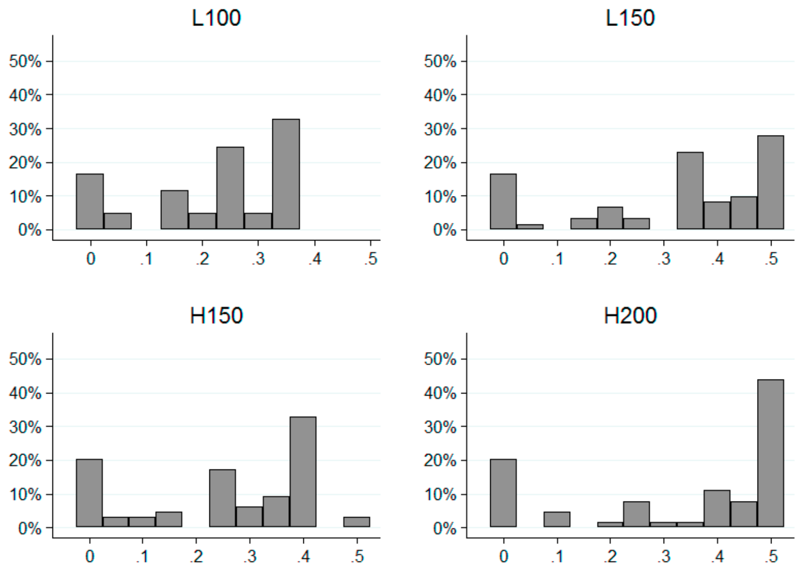

Figure 6: Ratio of Hypothetical Transfer/Endowment for Receivers

	Mean	StD	Median	N
All	.267	.168	.267	250
L100	.217	.123	.267	61
L150	.321	.178	.333	61
H150	.245	.154	.288	64
H200	.334	.200	.448	64

Table 3: Ratio of Hypothetical Transfer/Endowment for Receivers

$p < 0.001$; one-sided MWU p=0.003, z=2.73). These results clearly support the main hypothesis – with the reservation that these are non-incentivized decisions.

In the questionnaire, the receivers also state their idea of a fair transfer for both endowments. The fair ratio is nearly identical for all four situations (Table 10) as 93 of the 129 subjects who gave sensible answers chose 50% in both cases. There was no significant effect of the lottery's result on fairness standards. This is remarkable because althought the fairness standard hardly changes between sub-treatments, (hypothetical) transfers change a lot. As a sidenote, comparing the receivers' mean fair ratio of 46.7% to the senders' 36.6%, there appears to be some self-serving bias in the fairness assessment (the difference is highly significant for all sub-treatments with the exception of *H200*).[60]

Receivers are also asked about their beliefs about how much senders would transfer for each of the two possible endowments.[61] On average, the receivers believe that senders would transfer 25.3% of the endowment, which is a higher than the actual transfers (Table 11). The differences between sub-treatments are weaker compared to the hypothetical transfer, although the result remains weakly significant for *L150* and *H150* (one-sided MWU p = 0.051, z=1.64), i.e., receivers expect higher transfers in *L150* than in *H150*. Similar to the senders, the belief has a higher correlation ($\rho = 0.64$) to the hypothetical transfer than the fairness standard ($\rho = 0.40$), though both correlations are highly significant ($p < 0.001$).

4.4.7 Economics Students

Of all 132 senders, 60 are economics students. Their average transfer ratio over all four sub-treatments is 11.5%, compared to 21.5% for other participants (Tables 12 and 13). This difference is highly significant (two-sided MWU p=0.001, z = 3.25). The economics students are much more homogeneous, i.e., the variances of their transfers is significantly lower as a randomization test for differences in variance (Kaiser and Lacy 2009) confirms (0.035 to 0.027; p=0.05). In the two main sub-treatments *L150* and *H150*, there are 16 and 15 economics students, respectively, transferring on average 14.3% and 4.5% of the endowment. This difference is weakly statistically significant (one-sided MWU p=0.057, z=-1.577), supporting the main hypothesis.

[60] It is also noteworthy that nobody stated a fair standard of 0%.

[61] Again, we ignore ten receivers that gave answers above 50%.

Economics students also differ slightly in their overall assessment of a fair ratio from other participants (mean 32.6% to mean 38.8%, MWU p=0.09, z=1.69), but show similar correlation of fair transfer to actual transfer ($\rho = 0.37$, p = 0.004 to $\rho = 0.38$, p<0.001). However, they have significantly lower beliefs about the other senders' transfers, 13.8% to 21.3% (MWU p=0.003, z=2.91), which may explain the different transfer behavior. Additionally, it is noteworthy that the beliefs of a group reflect that group's behavior reasonably well, i.e., the beliefs of economics and other students are relatively closer to their actual respective transfers.[62] However, there are no significant differences between these groups with regard to mood or mood changes.

4.5 Discussion

4.5.1 Interpretation of the Results

The data do not oblige to a straightforward confirmation of the theoretical prediction, but tell a more complex story. To recapitulate, we expected that subjects with an endowment of 150 points transfer more when their expected endowment was lower (*L150*) compared to when their expected endowment was higher (*H150*). However, the differences in transfers between the two main sub-treatments are only indicative for the most part. For example, the likelihood of a sender transferring a positive amount is weakly significantly stronger in *L150*. However, the hypothesis receives stronger support from data collected from the receivers. Hypothetical transfer behavior is very much in line with the theoretical predictions as the majority of receivers chooses higher transfer ratios after winning the lottery. Assuming the receiver data is in fact representative for actual behavior, it appears that the use of the strategy method helps to remove the noise which seems to plague the senders' data.

The analysis of the questionnaire data identifies the importance of other reference points beside expectations. Fairness perception and beliefs about other people's behavior show a strong correlation with own transfer, whereas aspiration levels generally appear to have very little to no influence. Both fairness and beliefs are "soft" reference points, i.e., they are not based on an observable objective value, and may therefore themselves be expectancy-dependent. However, while there is weak evidence that fairness standards

[62]Dakkak et al. (2007) find a similar effect in an intercultural trust game where beliefs about the outgroup reflected behavior of the ingroup.

are higher after winning the lottery, beliefs do not significantly depend on the outcome of the lottery. It appears therefore that fairness standards are influenced to some degree by expectations, whereas beliefs are not, even though there is a very strong correlation between fairness standards and beliefs. In general, the strong correlation of beliefs with actual transfers and the very low sensibility of beliefs to expectations seem to some degree to overshadow the effect of expectations on transfers.

The analysis of the senders' mood does not confirm the assumption that good mood leads to more pro-social behavior or vice versa. The only way mood might play a role in the transfer decision is by presuming some kind of mental accounting where subjects "spend" some of their good mood to make transfers lower than what they consider fair. However, the individual data are too noisy to credibly validate such claims.

4.5.2 Possible Improvements

In hindsight, the experiment probably could have benefited from a few different design choices. One particular decision that appears to have backfired is to not employ the strategy method (Selten 1967) for the senders' decision. Originally, we were afraid that making the decision for two situations would prompt the subjects to choose the same transfer ratio both times. Additionally, we worried that the treatment effect would become less prominent when the subjects were "cold", i.e., when they did not actually experience the result of the lottery. Judging by the decision patterns of the receivers, those concerns were clearly unwarranted. On the other hand, the application of the strategy method may have opened the door for the supposition that the results were at least partially affected by some kind of demand effect.

Another aspect we would have liked to change was the composition of the subject pool. As it were, we had to invite a much more heterogeneous group than initially planned. In particular, we would have liked to invite subjects with similar, preferably little or no experimental experience.[63] However, we had to realize quickly that we would not get the necessary number of participants this way. The fact that the treatment effect is much more prominent among economics students than among other participants seems to indicate that a more homogeneous subject pool would have created less noise (possibly because beliefs are more homogeneous), which would have allowed the

[63] A quote from a sender's questionnaire: "I decided to give nothing because from my experience, decisions are nearly always selfish in this kind of experiment."

treatment effect to become more evident.

Furthermore, we will consider incentivizing the subjects' beliefs elicitation in a follow-up experiment, even though Gächter et al. (2006) find that while incentives increase the accuracy of beliefs slightly, in public goods games, they do not change the overall distribution of contributions and beliefs. Nevertheless, since the results using beliefs are so strong, it would be good to substantiate their validity with incentive compatibility. However, we originally decided against incentives for two reasons. First, we did not want to prolong the experiment any further. Second, with incentivized belief elicitation, the subjects would have had another source of potential income, which might have distorted their behavior in the dictator game. A possible solution to this dilemma would be to have two groups, one with and one without incentives for accurate beliefs.

Apart from the reliability of their elicitation, beliefs (or rather their variance) also pose a problem for the identification of the treatment effect. The fact that beliefs are apparently unaffected by expectations, yet are highly correlated with actual transfer behavior and are also fairly unevenly distributed between 0% and 50% appears to severely conceal the effect we are most interested in. By harmonizing subjects' beliefs over others' behavior, we might be able to expose the effect of expectations more clearly. For example, we could simply tell subjects what others have done in similar experiments or mask the clue more subtly, e.g. in a seemingly unrelated quiz. However, we would need to be careful not to create strong focal points or demand effects, which could counteract the treatment effect again. However, it would be preferable to demonstrate any treatment effect without resorting to manipulating the subjects' beliefs. A more homogeneous subject pool, for example, may already give us sufficiently uniform beliefs to accomplish that.

There is one more thing to consider for a possible follow-up experiment. By announcing beforehand that the senders will make transfers to the receivers, it is possible that senders not only judge their realized endowment with respect to their expectations, but that they also form expectations about their transfers. We could circumvent this by announcing the dictator game after the realization of the lottery, which should be ethically unproblematic because the senders have not made any decision at this point. However, the drawback is that the instructions for the dictator game would increase the delay between the realization of the lottery and the transfer decision, which may weaken the desired treatment effect. Such considerations would of course become obsolete with the implementation of the strategy method, which seems to be the most promising plan at the moment.

4.6 Conclusion

Our experiment provides some support for the theory of expectancy-dependent-preferences. It appears that for a given monetary payoff, individuals are more likely to act pro-socially after a positive deviation from their expectation than after a negative deviation. However, the evidence is not hard enough to draw definite conclusions with authority. At least, we are able to identify several points in the original design of the experiment which can be improved in a follow-up study in order to obtain more convincing data.

References

Abeler, Johannes, Armin Falk, Lorenz Götte, and David Huffman, "Reference Points and Effort Provision," *American Economic Review*, 2011, *101* (2), 470–492.

Adams, J. Stacy, "Toward an Understanding of Inequity," *Journal of Abnormal Psychology*, 1963, *67* (5), 422–436.

_ , "Inequity in Social Exchange," in Leonard Berkowitz, ed., *Advances In Experimental Social Psychology, Vol. II*, New York: Academic Press, 1965, pp. 267–299.

Agell, Jonas and Per Lundborg, "Survey Evidence on Wage Rigidity and Unemployment: Sweden in the 1990s," *The Scandinavian Journal of Economics*, 2003, *105* (1), 15–30.

Akerlof, George A., "Labor Contracts as Partial Gift Exchange," *The Quarterly Journal of Economics*, 1982, *97* (4), 543–569.

_ , "Gift Exchange and Efficiency-Wage Theory: Four Views," *American Economic Review*, 1984, *74* (2), 79–83.

_ **and Rachel E. Kranton**, "Economics and Identity," *The Quarterly Journal of Economics*, 2000, *115* (3), 715–753.

_ **and** _ , "Identity and Schooling: Some Lessons for the Economics of Education," *Journal of Economic Literature*, 2002, *40* (4), 1167–1201.

_ **and** _ , "Identity and the Economics of Organizations," *Journal of Economic Perspectives*, 2005, *19* (1), 9–32.

Altman, Morris, *Handbook of Contemporary Behavioral Economics*, New York: ME Sharpe, Inc., 2006.

Andersson, Ola, Matteo M. Galizzi, Tim Hoppe, Sebastian Kranz, Karen Van der Wiel, and Erik Wengström, "Persuasion in Experimental Ultimatum Games," *Economics Letters*, 2010, *108* (1), 16–18.

Andreoni, James, "Why Free Ride?," *Journal of Public Economics*, 1988, *37* (3), 291–304.

_ , "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *The Journal of Political Economy*, 1989, *97* (6), 1447–1458.

_ , "Philanthropy," in Serge-Christophe Kolm and Jean Mercier Ythier, eds., *Handbook of the Economics of Giving, Altruism and Reciprocity*, Amsterdam: Elsevier, 2006, pp. 1201–1269.

_ **and B. Douglas Bernheim**, "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects," *Econometrica*, 2009, *77* (5), 1607–1636.

_ **and John H. Miller**, "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence," *The Economic Journal*, 1993, *103* (418), 570–585.

_ **and** _ , "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica*, 2002, *70* (2), 737–753.

_ **and Lise Vesterlund**, "Which is the Fair Sex? Gender Differences in Altruism," *The Quarterly Journal of Economics*, 2001, *116* (1), 293–312.

Ball, Sheryl, Catherine C. Eckel, Philip J. Grossmanb, and William Zame, "Status in Markets," *The Quarterly Journal of Economics*, 2001, *116* (1), 161–188.

Balliet, Daniel, "Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review," *Journal of Conflict Resolution*, 2009, *54* (1), 39–57.

Barnett, Mark A. and James H. Bryan, "Effects of Competition with Outcome Feedback on Children's Helping Behavior," *Developmental Psychology*, 1974, *10* (6), 838–842.

Bartling, Björn, "Relative Performance or Team Evaluation? Optimal Contracts for Other-Regarding Agents," *The Journal of Economic Behavior & Organization*, 2011, *79*, 183– 193.

_ **and Ferdinand A. von Siemens**, "The Intensity of Incentives in Firms and Markets: Moral Hazard with Envious Agents," *Labour Economics*, 2010, *17* (3), 598–607.

Battigalli, Pierpaolo and Martin Dufwenberg, "Guilt in Games," *American Economic Review*, 2007, *97* (2), 170–176.

_ **and** _ , "Dynamic Psychological Games," *Journal of Economic Theory*, 2009, *144* (1), 1–35.

Baumeister, Roy F., Arlene M. Stillwell, and Todd F. Heatherton, "Guilt: An Interpersonal Approach.," *Psychological Bulletin*, 1994, *115* (2), 243–267.

Becker, Gary S., "A Theory of Social Interaction," *The Journal of Political Economy*, 1974, *82* (6), 1063–1093.

Bell, David R. and Randolph E. Bucklin, "The Role of Internal Reference Points in the Category Purchase Decision," *The Journal of Consumer Research*, 1999, *26* (2), 128–143.

Bellemare, Charles and Sabine Kröger, "On Representative Trust," *Working Paper*, 2003, *2003-47.*

Bénabou, Roland and Jean Tirole, "Incentives and Prosocial Behavior," *American Economic Review*, 2006, *96* (5), 1652–1678.

Bentham, Jeremy, "An Introduction to the Principles of Morals and Legislation," in J.H. Burns and H.L.A. Hart, eds., *Collected Works of Jeremy Bentham (1996)*, Vol. 7, New York: Oxford University Press, USA, 1789, pp. 1–341.

Bereby-Meyer, Yoella and Muriel Niederle, "Fairness in Bargaining," *Journal of Economic Behavior & Organization*, 2005, *56* (2), 173–186.

Berg, Joyce, John Dickhaut, and Kevin A. McCabe, "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 1995, *10* (1), 122–142.

Berg, Nathan and Gerd Gigerenzer, "As-if Behavioral Economics: Neoclassical Economics in Disguise?," *History of Economic Ideas*, 2010, *18* (1), 133–166.

Bergh, Andreas, "A Critical Note on the Theory of Inequity Aversion," *Journal of Socio-Economics*, 2008, *37* (5), 1789–1796.

Berkowitz, Leonard and William H. Connor, "Success, Failure, and Social Responsibility.," *The Journal of Personality and Social Psychology*, 1966, *4* (6), 664–669.

Binmore, Ken, "Modeling Rational Players: Part I," *Economics and Philosophy*, 1987, *3* (2), 179–214.

_ , "Modeling Rational Players: Part II," *Economics and Philosophy*, 1988, *4* (1), 9–55.

_ **and Avner Shaked**, "Experimental economics: Where next?," *Journal of Economic Behavior & Organization*, January 2010, *73* (1), 87–100.

_ **and Larry Samuelson**, "An Economist's Perspective On The Evolution Of Norms," *Journal of Institutional and Theoretical Economics*, 1994, *150* (1), 45–63.

_ **and** _ , "Evolutionary Drift and Equilibrium Selection," *The Review of Economic Studies*, 1999, *66* (2), 363–393.

_ **, Avner Shaked, and John Sutton**, "Testing Noncooperative Bargaining Theory: A Preliminary Study," *American Economic Review*, 1985, *75* (5), 1178–1180.

_ **,** _ **, and** _ , "A Further Test of Noncooperative Bargaining Theory: Reply," *American Economic Review*, 1988, *78* (4), 837–839.

Blanco, Mariana, Dirk Engelmann, and Hans Theo Normann, "A Within-Subject Analysis of Other-Regarding Preferences," *Games and Economic Behavior*, 2011, *72* (2), 321–338.

Blinder, Alan S. and Don H. Choi, "A Shred of Evidence on Theories of Wage Stickiness," *The Quarterly Journal of Economics*, 1990, *105* (4), 1003–1015.

Blount, Sally, "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences," *Organizational Behavior and Human Decision Processes*, 1995, *63* (2), 131–144.

Bohm, Peter, "Estimating Demand for Public Goods: An Experiment," *The European Economic Review*, 1972, *3*, 111–130.

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

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

_ and Axel Ockenfels, "ERC - A Theory of Equity, Reciprocity and Competition," *American Economic Review*, 2000, *100* (1), 166–93.

_ and _ , "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment," *American Economic Review*, 2006, *96* (5), 1906–1911.

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

_ and Rami Zwick, "Anonymity Versus Punishment in Ultimatum Bargaining," *Games and Economic Behavior*, 1995, *10* (1), 95–121.

_ , Elena Katok, and Rami Zwick, "Dictator Game Giving: Rules of Fairness Versus Acts of Kindness," *International Journal of Game Theory*, 1998, *27* (2), 269–299.

_ , Jordi Brandts, and Axel Ockenfels, "Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Game," *Experimental Economics*, 1998, *1* (3), 207–219.

Bosman, Ronald, Matthias Sutter, and Frans van Winden, "The Impact of Real Effort and Emotions in the Power-To-Take Game," *Journal of Economic Psychology*, 2005, *26* (3), 407–429.

Bradler, Christiane, "Social Preferences Under Risk - An Experimental Analysis," *Working Paper*, 2009.

Branas-Garza, Pablo, Ramón Cobo-Reyes, María Paz Espinosa, Natalia Jiménez, Jaromír Kovárík, and Giovanni Ponti, "Altruism and Social Integration," *Games and Economic Behavior*, 2010, *69* (2), 249–257.

Brandstätter, Hermann and Manfred Königstein, "Personality Influences on Ultimatum Bargaining Decisions," *European Journal of Personality*, 2001, *15* (S1), 53–70.

Brosig, Jeannette, Joachim Weimann, and Axel Ockenfels, "The Effect Of Communication Media On Cooperation," *German Economic Review*, 2003, *4* (2), 217–241.

Bruni, Luigino and Robert Sugden, "The Road Not Taken: How Psychology Was Removed From Economics, and How It Might be Brought Back," *The Economic Journal*, 2007, *117* (516), 146–173.

Bryant, Brenda K., "Context of Success, Affective Arousal, and Generosity: The Neglected Role of Negative Affect in Success Experience," *The American Educational Research Journal*, 1983, *20* (4), 553–562.

Buchan, Nancy R., Rachel T. A. Croson, and Eric J. Johnson, "When Do Fair Beliefs Influence Bargaining Behavior? Experimental Bargaining in Japan and the United States," *Journal of Consumer Research*, 2004, *31* (1), 181–190.

Burnell, Stephen J., Lewis Evans, and Shuntian Yao, "The Ultimatum Game: Optimal Strategies Without Fairness," *Games and Economic Behavior*, 1999, *26* (2), 221–252.

Burnham, Terence, Kevin McCabe, and Vernon L Smith, "Friend-or-Foe Intentionality Priming in an Extensive Form Trust Game," *Journal of Economic Behavior & Organization*, 2000, *43* (1), 57–73.

Cabrales, Antonio, Raffaele Miniaci, and Marco Piovesan, "An Experiment on Markets and Contracts: Do Social Preferences Determine Corporate Culture?," *Working Paper*, 2007.

Camerer, Colin F., "Introduction," in "Behavioral Game Theory," Princeton: Princeton University Press, 2003, pp. 1–25.

_ , Linda Babcock, George F. Loewenstein, and Richard H. Thaler, "Labor Supply of New York City Cabdrivers: One Day at a Time," *The Quarterly Journal of Economics*, 1997, *112* (2), 407–441.

Cameron, Lisa A., "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia," *Economic Inquiry*, 1999, *37* (1), 47–59.

Campbell III, Carl M. and Kunal S. Kamlani, "The Reasons for Wage Rigidity: Evidence from a Survey of Firms," *The Quarterly Journal of Economics*, 1997, *112* (3), 759–789.

Carnevale, Peter J.D. and Alice M. Isen, "The Influence of Positive Affect and Visual Access on the Discovery of Integrative Solutions in Bilateral Negotiation," *Organizational Behavior and Human Decision Processes*, 1986, *37* (1), 1–13.

Cason, Timothy N. and Vai-Lam Mui, "Social Influence in the Sequential Dictator Game," *Journal Of Mathematical Psychology*, 1998, *42* (2/3), 248–265.

Charness, Gary and Brit Grosskopf, "Relative Payoffs and Happiness: An Experimental Study," *Journal of Economic Behavior & Organization*, 2001, *45* (3), 301–328.

_ **and Martin Dufwenberg**, "Promises and Partnership," *Econometrica*, 2006, *74* (6), 1579–1601.

_ **and Matthew Rabin**, "Understanding Social Preferences with Simple Tests," *The Quarterly Journal of Economics*, 2002, *117* (3), 817–869.

_ **and** _ , "Expressed Preferences and Behavior in Experimental Games," *Games and Economic Behavior*, 2005, *53* (2), 151–169.

Chaudhuri, Ananish, "Conditional Cooperation And Social Norms In Public Goods Experiments: A Survey Of The Literature," *Experimental Economics*, 2011, *14* (1), 47–83.

Chen, Chia-Ching, I-Ming Chiu, John Smith, and Yamada Tetsuji, "Too Smart to Be Selfish? Measure of Intelligence, Social Preferences, and Consistency," *Working Paper*, 2011, (MPRA 34438).

Chen, Yan and Sherry Xin Li, "Group Identity And Social Preferences," *American Economic Review*, 2009, *99* (1), 431–457.

Cherry, Todd L., Peter Frykblom, and Jason F. Shogren, "Hardnose The Dictator," *American Economic Review*, 2002, *92* (4), 1218–1221.

Collard, David, "Edgeworth's Propositions on Altruism," *The Economic Journal*, 1975, *85* (338), 355–360.

Conlisk, John, "The Utility Of Gambling," *Journal of Risk and Uncertainty*, June 1993, *6* (3), 255–275.

Cooper, David J. and John H. Kagel, "Other Regarding Preferences: A Selective Survey Of Experimental Results," in "Handbook of Experimental Economics (Upcoming)" 2009.

Costa-Gomes, Miguel and Klaus G. Zauner, "Ultimatum Bargaining Behavior in Israel, Japan, Slovenia, and the United States: A Social Utility Analysis," *Games and Economic Behavior*, 2001, *34* (2), 238–269.

Cox, James C., "How to Identify Trust and Reciprocity," *Games and Economic Behavior*, 2004, *46* (2), 260–281.

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

Crawford, Vincent P. and Juanjuan Meng, "New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *Working Paper*, 2010.

Cronqvist, Henrik and Richard H. Thaler, "Design Choices in Privatized Social-Security Systems: Learning from the Swedish Experience," *American Economic Review*, 2004, *94* (2), 424–428.

Croson, Rachel T. A. and Nancy R. Buchan, "Gender and Culture: International Experimental Evidence from Trust Games," *American Economic Review*, 1999, *89* (2), 386–391.

Dakkak, Ihab, Heike Hennig-Schmidt, Reinhard Selten, Gari Walkowitz, and Eyal Winter, "Actions and Beliefs in a Trilateral Trust Game Involving Germans, Israelis, and Palestinians," *Working Paper*, 2007.

Dana, Jason, Daylian M. Cain, and Robyn M. Dawes, "What You Don't Know Won't Hurt Me: Costly (but Quiet) Exit in Dictator Games," *Organizational Behavior and Human Decision Processes*, 2006, *100* (2), 193–201.

_ **, Roberto Weber, and Jason Xi Kuang**, "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness," *Economic Theory*, 2007, *33* (1), 67–80.

Daruvala, Dinky, "Would the Right Social Preference Model Please Stand Up!," *Journal of Economic Behavior & Organization*, 2010, *73* (2), 199–208.

Dawes, Robyn M. and Richard H. Thaler, "Anomalies: Cooperation," *The Journal of Economic Perspectives*, 1988, *2* (3), 187–197.

_ **, Jeanne McTavish, and Harriet Shaklee**, "Behavior, Communication, and Assumptions About Other People's Behavior in a Commons Dilemma Situation," *The Journal of Personality and Social Psychology*, 1977, *35* (1), 1–11.

de Quervain, Dominique J.-F., Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Ulrich Schnyder, Alfred Buck, and Ernst Fehr, "The Neural Basis of Altruistic Punishment," *Science*, 2004, *305* (5688), 1254–1258.

Demougin, Dominique and Claude Fluet, "Inequity Aversion in Tournaments," *Working Paper*, 2003.

_ **,** _ **, and Carsten Helm**, "Output and Wages with Inequality Averse Agents," *The Canadian Journal Of Economics*, 2006, *39* (2), 399–413.

Desiraju, Ramarao and David E. Sappington, "Equity and Adverse Selection," *The Journal of Economics & Management Strategy*, 2007, *16* (2), 285–318.

Dickinson, David L., "Ultimatum Decision-Making: A Test of Reciprocal Kindness," *Theory And Decision*, 2000, *48*, 151–177.

Dufwenberg, Martin and Astri Muren, "Generosity, Anonymity, Gender," *Journal of Economic Behavior & Organization*, 2006, *61* (1), 42–49.

_ **and Georg Kirchsteiger**, "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, 2004, *47* (2), 268–298.

Dur, Robert and Amihai Glazer, "Optimal Incentive Contracts When Workers Envy Their Boss," *Journal of Law, Economics, and Organization*, 2008, *24* (1), 120–137.

Eckel, Catherine C. and Herbert Gintis, "Blaming The Messenger: Notes On The Current State Of Experimental Economics," *Journal of Economic Behavior & Organization*, 2010, *73* (1), 109–119.

_ **and Philip J. Grossman**, "Altruism in Anonymous Dictator Games," *Games and Economic Behavior*, 1996, *16*, 181–191.

_ **and** _ , "The Relative Price Of Fairness: Gender Differences In A Punishment Game," *Journal of Economic Behavior & Organization*, 1996, *30*, 143–158.

Edgeworth, Francis Y., *Mathematical Psychics - An Essay on the Application of Mathematics to the Moral Sciences*, London: C. Kegan Paul & Co., 1881.

Elliott, Catherine S., Donald M. Hayward, and Sebastian Canon, "Institutional Framing: Some Experimental Evidence," *Journal of Economic Behavior & Organization*, 1998, *35* (4), 455–464.

Elster, Jon, "Social Norms And Economic Theory," *The Journal of Economic Perspectives*, 1989, *3* (4), 99–117.

Engelmann, Dirk, "How Not to Extend Models of Inequality Aversion," *Journal of Economic Behavior & Organization*, 2012, *81* (2), 599–605.

_ **and Martin Strobel**, "Inequality Aversion, Efficiency, And Maximin Preferences In Simple Distribution Experiments," *American Economic Review*, 2004, *94* (4), 857–869.

_ **and** _ , "Inequality Aversion, Efficiency, and Maximin Preferences In Simple Distribution Experiments: Reply," *American Economic Review*, 2006, *96* (5), 1918–1923.

Englmaier, Florian and Achim Wambach, "Optimal Incentive Contracts Under Inequity Aversion," *Games and Economic Behavior*, 2010, *69* (2), 312–328.

Epley, Nicholas and Thomas Gilovich, "The Anchoring-and-Adjustment Heuristic: Why the Adjustments are Insufficient.," *Psychological Science*, 2006, *17* (4), 311–318.

Ericson, Keith M. Marzilli and Andreas Fuster, "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments," *Working Paper*, 2010.

Erlei, Mathias, "Heterogeneous Social Preferences," *Journal of Economic Behavior & Organization*, 2008, *65*, 436–457.

Falk, Armin, "Gift Exchange in the Field," *Econometrica*, 2007, *75* (5), 1501–1511.

_ **and Urs Fischbacher**, "A Theory of Reciprocity," *Games and Economic Behavior*, 2006, *54*, 293–315.

_ **, Ernst Fehr, and Urs Fischbacher**, "On the Nature of Fair Behavior," *Economic Inquiry*, 2003, *41* (1), 20–26.

_ **,** _ **, and** _ , "Testing Theories Of Fairness - Intentions matter," *Games and Economic Behavior*, 2008, *62* (1), 287–303.

Fehr, Ernst and Colin F. Camerer, "Social Neuroeconomics: The Neural Circuitry of Social Preferences," *Trends in Cognitive Sciences*, 2007, *11* (10), 419–427.

_ **and Klaus M. Schmidt**, "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 1999, *114* (3), 817–868.

_ **and** _ , "The Rhetoric Of Inequity Aversion - A Reply," *Working Paper*, 2005.

_ **and** _ , "The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories," in Serge-Christophe Kolm and Jean Mercier Ythier, eds., *Handbook of the Economics of Giving, Altruism and Reciprocity*, Vol. 1, Amsterdam: North-Holland, 2006, pp. 615–691.

_ **and** _ , "On Inequity Aversion: A Reply To Binmore And Shaked," *Journal of Economic Behavior & Organization*, 2010, *73* (1), 101–108.

_ **and Simon Gächter**, "Cooperation and Punishment in Public Good Experiments," *American Economic Review*, 2000, *90* (4), 980–994.

_ **and Urs Fischbacher**, "Why Social Preferences Matter - The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives," *The Economic Journal*, 2002, *112* (478), 1–33.

_ **and** _ , "The Nature Of Human Altruism," *Nature*, 2003, *425* (6960), 785–791.

_ **, Georg Kirchsteiger, and Arno Riedl**, "Does Fairness Prevent Market Clearing? An Experimental Investigation," *The Quarterly Journal of Economics*, 1993, *108* (2), 437–459.

_ **,** _ **, and** _ , "Gift Exchange And Reciprocity In Competitive Experimental Markets," *European Economic Review*, 1998, *42* (1), 1–34.

_ **, Michael Naef, and Klaus M. Schmidt**, "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment," *American Economic Review*, 2004, *96* (5), 1912–1917.

Ferrall, Christopher, "Promotions and Incentives in Partnerships: Evidence From Major U.S. Law Firms," *The Canadian Journal Of Economics*, 1996, *29* (4), 811–827.

Festinger, Leon, "A Theory of Social Comparison Processes," *Human Relations*, 1954, *7*, 117–140.

Fiedler, Marina and Ernan Haruvy, "The Lab Versus The Virtual Lab And Virtual Field - An Experimental Investigation Of Trust Games With Communication," *Journal of Economic Behavior & Organization*, 2009, *72* (2), 716–724.

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

_ **, Ernst Fehr, and Simon Gächter**, "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment," *Economics Letters*, 2001, *71*, 397–404.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits, "Individual Preferences For Giving," *American Economic Review*, 2007, *97* (5), 1858–1876.

Fliessbach, Klaus, Bernd Weber, Peter Trautner, Thomas Dohmen, Uwe Sunde, Christian Elger, and Armin Falk, "Social Comparison Affects Reward-Related Brain Activity in the Human Ventral Striatum," *Science*, 2007, *318* (5854), 1305–1308.

Fong, Christina, "Social Preferences, Self-Interest, and the Demand for Redistribution," *Journal of Public Economics*, 2001, *82* (2), 225–246.

Forsythe, Robert, Joel L. Horowitz, N.E. Savin, and Martin Sefton, "Fairness in Simple Bargaining Games," *Games and Economic*, 1994, *6*, 347–369.

Fouraker, Lawrence E. and Sidney Siegel, *Bargaining Behavior*, New York: McGraw-Hill, 1963.

Frey, Bruno S. and Iris Bohnet, "Institutions Affect Fairness: Experimental Investigations," *Journal of Institutional and Theoretical Economics*, 1995, *151* (2), 286–303.

Friedman, Milton, "The Methodology of Positive Economics," in "Essays In Positive Economics," Vol. II, Chicago: University Of Chicago Press, 1953, pp. 3–43.

Frisch, Helmut, "Die Kontraktkurve Bei Interdependenzen Im Konsum," *Kyklos*, 1971, *24* (4), 644–659.

Frohlich, Norman, Joe Oppenheimer, and Anja Kurki, "Modeling Other-Regarding Preferences And An Experimental Test," *Public Choice*, 2004, *119*, 91–117.

Gächter, Simon, Elke Renner, and Karina Whitehead, "The Effects of (Incentivized) Belief Elicitation in Public Good Experiments," *Experimental Economics*, 2006.

Gale, John, Kenneth G. Binmore, and Larry Samuelson, "Learning to Be Imperfect: The Ultimatum Game," *Games and Economic Behavior*, 1995, *8* (1), 56–90.

Geanakoplos, John, D. Pearce, and E. Stacchetti, "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, 1989, *1*, 60–79.

Gilad, Benjamin and Stanley Kaish, *Handbook of Behavioral Economics*, Greenwich: Jai Press Inc., 1986.

Gneezy, Uri and John A. List, "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 2006, *74* (5), 1365–1384.

Goel, Anand M. and Anjan V. Thakor, "Optimal Contracts When Agents Envy Each Other," *Working Paper*, 2006.

Goukasian, Levon and Xuhu Wan, "Optimal Incentive Contracts under Relative Income Concerns," *Mathematics and Financial Economics*, 2010, *4* (1), 57–86.

Graf, Christoph, Rudolf Vetschera, and Yingchao Zhang, "Parameters of Social Preference Functions: Measurement and External Validity," *Theory and Decision*, May 2012, (Online), 1–26.

Greiner, Ben, "An Online Recruitment System for Economic Experiments," *Forschung und Wissenschaftliches Rechnen*, 2004, *63*, 79–93.

Griesinger, Donald W. and James W. Livingston Jr., "Toward a Model of Interpersonal Motivation in Experimental Games," *Behavioral Science*, 1973, *18* (3), 173–188.

Grund, Christian and Dirk Sliwka, "Envy and Compassion in Tournaments," *The Journal of Economics & Management Strategy*, 2005, *14* (1), 187–207.

Güth, W, S Huck, and P Ockenfels, "Two-Level Ultimatum Bargaining With Incomplete Information: An Experimental Study," *The Economic Journal*, 1996, *106* (436), 593–604.

Güth, Werner, "On Ultimatum Bargaining Experiments - A Personal Review," *The Journal of Economic Behavior & Organization*, 1995, *27* (3), 329–344.

— **and Eric Van Damme**, "Information, Strategic Behavior, and Fairness in Ultimatum Bargaining: An Experimental Study," *Journal of Mathematical Psychology*, June 1998, *42*, 227–247.

— **, Hartmut Kliemt, and Axel Ockenfels**, "Fairness Versus Efficiency: An Experimental Study Of (Mutual) Gift Giving," *Journal of Economic Behavior & Organization*, 2003, *50*, 465–475.

— **, Rolf Schmittberger, and Bernd Schwarze**, "An Experimental Analysis of Ultimatum Bargaining," *The Journal of Economic Behavior & Organization*, 1982, *3*, 367–388.

Harbaugh, William T., Ulrich Mayr, and Daniel R. Burghart, "Neural Responses to Taxation and Voluntary Giving Reveal Motives for Charitable Donations.," *Science*, 2007, *316* (5831), 1622–1625.

Harbring, Christine and Bernd Irlenbusch, "How Many Winners Are Good to Have? On Tournaments with Sabotage," *Journal of Economic Behavior & Organization*, 2008, *65* (3-4), 682–702.

Hardie, Bruce G.S., Eric J. Johnson, and Peter S. Fader, "Modeling Loss Aversion and Reference Dependence Effects on Brand Choice," *Marketing Science*, 1993, *12* (4), 378–394.

Henrich, Joseph, Jean Ensminger, Richard McElreath, Abigail Barr, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank Marlowe, David Tracer, and John Ziker, "Markets, Religion, Community Size, and the Evolution of Fairness and Punishment," *Science*, 2010, *327* (5972), 1480–1484.

— **, Robert Boyd, Samuel Bowles, Colin F. Camerer, Ernst Fehr, Herbert Gintis, Richard McElreath, Michael Alvard, Abigail Barr, Jean Ensminger, Natalie Smith Henrich, Kim Hill, Francisco Gil-White, Michael Gurven, Frank W. Marlowe, John Q. Patton, and David Tracer**, "'Economic Man' In Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies," *Behavioral And Brain Sciences*, 2005, *28*, 795–855.

Hermalin, Benjamin E. and Alice M. Isen, "A Model of the Effect of Affect on Economic Decision Making," *Quantitative Marketing and Economics*, 2007, *6* (1), 17–40.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter, "Antisocial Punishment Across Societies," *Science*, 2008, *319* (5868), 1362–1367.

Ho, Teck-Hua and Xuanming Su, "Peer-Induced Fairness in Games," *American Economic Review*, 2009, *99* (5), 2022–2049.

Hoffman, Elizabeth, Kevin A. McCabe, and Vernon L. Smith, "On Expectations and the Monetary Stakes in Ultimatum Ggames," *International Journal of Game Theory*, 1996, *25*, 289–301.

Holyoak, Keith J. and Peter C. Gordon, "Social Reference Points," *The Journal of Personality and Social Psychology*, 1983, *44* (5), 881–887.

Isaac, R. Mark, K. McCue, and C. Plott, "Public Goods Provision in an Experimental Environment," *The Journal of Public Economics*, 1985, *26*, 51–74.

Isen, Alice M., "Success, Failure, Attention, and Reaction to Others: The Warm Glow of Success.," *The Journal of Personality and Social Psychology*, 1970, *15* (4), 294–301.

_ , "Positive Affect," in Tim Dalgleish and Mick J. Power, eds., *Handbook of Cognition and Emotion*, Chichester, UK: John Wiley & Sons, 1999, pp. 521–539.

_ **and Paula F. Levin**, "Effect of Feeling Good on Helping: Cookies and Kindness," *The Journal of Personality and Social Psychology*, 1972, *21* (3), 384–388.

_ **, Margaret S. Clark, and Mark F. Schwartz**, "Duration of the Effect of Good Mood on Helping: Footprints on the Sands of Time," *The Journal of Personality and Social Psychology*, 1976, *34* (3), 385–393.

_ **, Nancy Horn, and David L. Rosenhan**, "Effects of Success and Failure on Children's Generosity," *The Journal of Personality and Social Psychology*, 1973, *27* (2), 239–247.

Itoh, Hideshi, "Moral Hazard and Other-Regarding Preferences," *The Japanese Economic Review*, 2004, *55* (1), 18–45.

Jevons, W. Stanley, *The Theory of Political Economy* 1871.

Johnson, Eric J. and Daniel Goldstein, "Do Defaults Save Lives?," *Science*, 2003, *302*, 1338–1339.

Johnson, Harry G., "The Effects of Income-Redistribution on Aggregate Consumption with Interdependence of Consumers' Preferences," *Economica*, 1952, *19* (74), 131–147.

Kachelmeier, Steven J. and Mohamed Shehata, "Culture and Competition: A Laboratory Market Comparison Between China and the West," *Journal of Economic Behavior & Organization*, 1992, *19* (2), 145–168.

Kagel, John H. and Katherine Willey Wolfe, "Tests of Fairness Models Based on Equity Considerations in a Three-Person Ultimatum Game," *Experimental Economics*, 2001, *4* (2001), 203–219.

_ , Chung Kim, and Donald Moser, "Fairness in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs," *Games and Economic Behavior*, 1996, *13*, 100–111.

Kahneman, Daniel and A. Tversky, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 1979, *47* (2), 263–291.

_ , Jack L. Knetsch, and Richard H. Thaler, "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *American Economic Review*, 1986, *76* (4), 728–741.

_ , _ , and _ , "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias," *The Journal of Economic Perspectives*, 1991, *5* (1), 193–206.

Kaiser, Johannes and Michael G. Lacy, "A General-Purpose Method for Two-Group Randomization Tests," *The Stata Journal*, 2009, *9* (1), 70–85.

Kalra, Ajay and Mengze Shi, "Designing Optimal Sales Contests: A Theoretical Perspective," *Marketing Science*, 2001, *20* (2), 170–193.

Kim, Oliver and Mark Walker, "The Free Rider Problem: Experimental Evidence," *Public Choice*, 1984, *43* (1), 3–24.

Kirchsteiger, Georg, "The Role of Envy in Ultimatum Games," *Journal of Economic Behavior & Organization*, 1994, *25* (3), 373–389.

Knack, Stephen and Philipp Keefer, "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation," *The Quarterly Journal of Economics*, 1997, *112* (4), 1251–1288.

Koenigs, Michael and Daniel Tranel, "Irrational Economic Decision-Making After Ventromedial Prefrontal Damage: Evidence from the Ultimatum Game," *The Journal of Neuroscience*, 2007, *27* (4), 951–956.

Kölle, Felix, Dirk Sliwka, and Nannan Zhou, "Inequality, Inequity Aversion, and the Provision of Public Goods," *Working Paper*, 2011.

Konow, James, "A Positive Theory of Economic Fairness," *Journal of Economic Behavior & Organization*, 1996, *31*, 13–35.

_ , "Which is the Fairest One of All? A Positive Analysis of Justice Theories," *Journal Of Economic Literature*, 2003, *41*, 1188–1239.

Kosfeld, Michael, Markus Heinrichs, Paul J. Zak, Urs Fischbacher, and Ernst Fehr, "Oxytocin Increases Trust in Humans," *Nature*, 2005, *435* (7042), 673–676.

Koszegi, Botond, "Emotional Agency," *The Quarterly Journal of Economics*, February 2006, *121* (1), 121–155.

_ **and Matthew Rabin**, "Reference-Dependent Risk Attitudes," *American Economic Review*, 2007, *97* (4), 1047–1073.

_ **and** _ , "Reference-Dependent Consumption Plans," *American Economic Review*, 2009, *99* (3), 909–936.

Kräkel, Matthias, "Relative Deprivation in Rank-Order Tournaments," *Labour Economics*, 2000, *7*, 385–407.

_ , "Emotions in Tournaments," *The Journal of Economic Behavior and Organization*, 2008, *67* (1), 204–214.

Krebs, David M., Paul Milgrom, John Roberts, and Robert Wilson, "Rational Cooperation in the Finitely Dilemma Repeated Prisoners," *Journal of Economic Theory*, 1982, *27*, 245–252.

Kristensen, Henrik, "The Effects of Anchor Points and Reference Points on Negotiation Process and Outcome," *Organizational Behavior and Human Decision Processes*, 1997, *71* (1), 85–94.

Kritikos, Alexander and Friedel Bolle, "Distributional Concerns: Equity- or Efficiency-Oriented?," *Economics Letters*, 2001, *73* (3), 333–338.

Lazear, Edward P. and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *The Journal of Political Economy*, 1981, *89*, 841–864.

Ledyard, John O., "Public Goods: A Survey of Experimental Research," in Alvin E. Roth and John H. Kagel, eds., *The Handbook of Experimental Economics*, 1995.

Leider, Stephen, Markus M. Möbius, Tanya Rosenblat, and Quoc-Anh Do, "Directed Altruism and Enforced Reciprocity in Social Networks," *The Quarterly Journal of Economics*, 2009, *124* (4), 1815–1851.

Levin, Paula F. and Alice M. Isen, "Further Studies on the Effect of Feeling Good on Helping," *Sociometry*, 1975, *38* (1), 141–147.

Levine, David K., "Modeling Altruism and Spitefulness in Experiments," *The Review of Economic Dynamics*, 1998, *1* (3), 593–622.

Lewin, Shira B., "Economics and Psychology: Lessons for Our Own Day from the Early Twentieth Century," *The Journal of Economic Literature*, 1996, *34* (3), 1293–1323.

Lim, Noah, Michael J. Ahearne, and Sung H. Ham, "Designing Sales Contests: Does the Prize Structure Matter?," *Journal of Marketing Research*, 2009, *46* (3), 356–371.

Loewenstein, George F., Leigh Thompson, and Max H. Bazerman, "Social Utility and Decision Making in Interpersonal Contexts," *The Journal of Personality and Social Psychology*, 1989, *57* (3), 426–441.

Loomes, Graham and Robert Sugden, "Disappointment and Dynamic Consistency in Choice Under Uncertainty," *The Review of Economic Studies*, 1986, *53* (2), 271–282.

López-Pérez, Raúl, "Aversion to Norm-Breaking: A Model," *Games and Economic Behavior*, 2008, *64* (1), 237–267.

_ , "The Power of Words: A Model of Honesty and Fairness," *Journal of Economic Psychology*, 2012, *33* (3), 642–658.

Lurie, Stephen, "A Parametric Model of Utility for Two-Person Distributions," *Psychological Review*, 1987, *94* (1), 42–60.

Malcomson, James M., "Rank-Order Contracts for a Principal with Many Agents," *The Review of Economic Studies*, 1986, *53* (5), 807–817.

Margolis, Howard, *Selfishness, Altruims, and Rationality*, Chicago: Chicago University Press, 1982.

Marwell, Gerald and Ruth E. Ames, "Experiments on the Provision of Public Goods I: Resources, Interest, Group Size, and the Free-Rider Problem.," *The American Journal of Sociology*, 1979, *84* (6), 1335–1360.

_ **and** _ , "Experiments on the Provision of Public Goods II: Provision Points, Stakes, Experience, and the Free-Rider," *The American Journal of Sociology*, 1980, *85* (4), 926–937.

Mas, Alexandre, "Pay, Reference Points, and Police Performance," *Working Paper*, 2006.

Masatlioglu, Yusufcan and Efe A. Ok, "Rational Choice with Status Quo Bias," *The Journal of Economic Theory*, 2005, *121* (1), 1–29.

Mazumdar, Tridib, S.P. Raj, and Indrajit Sinha, "Reference Price Research: Review and Proposition," *The Journal of Marketing*, 2005, *69* (4), 84–102.

McCabe, Kevin, Daniel Houser, Lee Ryan, Vernon L. Smith, and Theodore Trouard, "A Functional Imaging Study of Cooperation in Two-Person Reciprocal Exchange," *Proceedings of the National Academy of Sciences of the United States of America*, 2001, *98* (20), 11832–11835.

McDougall, Williams, *An Introduction to Social Psychology*, London: Methuen & Co, 1910.

Meng, Juanjuan, "The Disposition Effect and Expectations as Reference Point," *Working Paper*, 2010, pp. 1–50.

Messick, David M. and Charles G. McClintock, "Motivational Bases of Choice in Experimental Games," *Journal of Experimental Social Psychology*, 1968, *4*, 1–25.

Mill, John Stuart, "On the Definition of Political Economy; and on the Method of Investigation Proper to it," in J.S. Mills, ed., *Essays on Some Unsettled Questions of Political Economy*, London: Longmans, 1844.

Mohnen, Alwine, K. Pokorny, and Dirk Sliwka, "Transparency, Inequity Aversion, and the Dynamics of Peer Pressure in Teams: Theory and Evidence," *Working Paper*, 2008.

Moll, Jorge, Frank Krueger, Roland Zahn, Matteo Pardini, Ricardo de Oliveira-Souza, and Jordan Grafman, "Human Fronto-Mesolimbic Networks Guide Decisions about Charitable Donation," *Proceedings of the National Academy of Sciences of the United States of America*, 2006, *103* (42), 15623–15628.

Mullainathan, Sendhil and Richard H. Thaler, "Behavioral Economics," in Neil J. Smelser and Paul B. Baltes, eds., *International Encyclopedia of the Social and Behavioral Sciences, Vol. 20*, Oxford and New York: Oxford University Press, 2001, pp. 1094–1100.

Mussweiler, Thomas and Fritz Strack, "Hypothesis-Consistent Testing and Semantic Priming in the Anchoring Paradigm: A Selective Accessibility Model," *The Journal of Experimental Social Psychology*, 1999, *35*, 136–164.

Nash, John F. Sr., "The Bargaining Problem," *Econometrica*, 1950, *18* (2), 155–162.

Neelin, Janet, Hugo Sonnenschein, and Matthew Spiegel, "A Further Test of Noncooperative Bargaining Theory: Comment," *American Economic Review*, 1988, *78* (4), 824–836.

Neilson, William S., "Axiomatic Reference-Dependence in Behavior Toward Others and Toward Risk," *Economic Theory*, 2006, *28*, 681–692.

_ **and Jill Stowe**, "Piece-Rate Contracts for Other-Regarding Workers," *Economic Inquiry*, 2010, *48* (3), 575–586.

Nydegger, Rudy V. and Guillermo Owen, "Two-Person Bargaining: An Experimental Test of the Nash Axioms," *International Journal of Game Theory*, 1974, *3* (4), 239–249.

Ochs, Jack and Alvin E. Roth, "An Experimental Study of Sequential Bargaining," *American Economic Review*, 1989, *79* (3), 355–384.

Ockenfels, Axel, Peter Werner, and Dirk Sliwka, "Bonus Payments and Reference Point Violations," *Working Paper*, 2010.

Odean, Terrance, "Are Investors Reluctant to Realize Their Losses?," *The Journal of Finance*, 1998, *53* (5), 1775–1798.

Offerman, Theo, "Hurting Hurts More Than Helping Helps," *European Economic Review*, 2002, *46*, 1423–1437.

Orrison, Alannah, Andrew Schotter, and Keith Weigelt, "Multiperson Tournaments: An Experimental Examination," *Management Science*, 2004, *50* (2), 268–279.

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

_ **, James Walker, and Roy Gardner**, "Covenants With and Without a Sword: Self-Governance is Possible," *The American Political Science Review*, 1992, *86* (2), 404–417.

Palfrey, Thomas R. and Jeffrey E. Prisbrey, "Anomalous Behavior in Public Goods Experiments: How Much and Why?," *American Economic Review*, 1997, *87* (5), 829–846.

Pareto, Vilfredo, "Manual of Political Economy," in "English Translation of the 1909 French Edition of the 1906 Italian Manuale D'Economia Politica Con Una Introduzione Alla Scienza Sociale," New York: Kelly, 1971.

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

Rapoport, Anatol, Oded Frenkel, and Josef Perner, "Experiments with Cooperative 2x2 Games," *Theory And Decision*, 1977, *8* (1), 67–92.

Rey-Biel, Pedro, "Inequity Aversion and Team Incentives," *The Scandinavian Journal of Economics*, 2008, *110* (2), 297–320.

Rilling, James, David Gutman, Thorsten Zeh, Giuseppe Pagnoni, Gregory Berns, and Clinton Kilts, "A Neural Basis for Social Cooperation.," *Neuron*, 2002, *35* (2), 395–405.

Rilling, James K., Alan G. Sanfey, Jessica A. Aronson, Leigh E. Nystrom, and Jonathan D. Cohen, "Opposing BOLD Responses to Reciprocated and Unreciprocated Altruism in Putative Reward Pathways," *Neuroreport*, 2004, *15* (16), 2539–2543.

Rotemberg, Julio J., "Minimally Acceptable Altruism and the Ultimatum Game," *Journal of Economic Behavior & Organization*, 2008, *66* (3-4), 457–476.

Roth, Alvin E. and Ido Erev, "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, 1995, *8*, 164–212.

_ **and Keith J. Murnighan**, "The Role of Information in Bargaining: An Experimental Study," *Econometrica*, 1982, *50* (5), 1123–1142.

_ **and Michael W.K. Malouf**, "Game-Theoretic Models and the Role of Information in Bargaining.," *Psychological Review*, 1979, *86* (6), 574–594.

_ **,** _ **, and Keith J. Murnighan**, "Sociological Versus Strategic Factors In Bargaining," *Journal of Economic Behavior & Organization*, 1981, *2*, 153–177.

_ **, Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir**, "Bargaining And Market Behavior In Jerusalem, Ljubljana, Pittsburgh, And Tokyo: An Experimental Study," *American Economic Review*, 1991, *81* (5), 1068–1095.

Rubinstein, Ariel, "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 1982, *50* (1), 97–109.

— , "Discussion of 'Behavioral Economics'," in R. Blundell, W. K. Newey, and T. Persson, eds., *Advances in Economic Theory*, Cambridge: Cambridge University Press, 2006, pp. 246–254.

— , "Comments on Neuroeconomics," *Economics and Philosophy*, 2008, *24* (03), 485–494.

Samuelson, Paul A., "The Pure Theory Of Public Expenditure," *The Review Of Economics And Statistics*, 1954, *36* (4), 387–389.

Samuelson, William and Richard Zeckhauser, "Status Quo Bias in Decision Making," *The Journal of Risk and Uncertainty*, 1988, *1* (1), 7–59.

Sanfey, Alan G., James K. Rilling, Jessica A. Aronson, Leigh E. Nystrom, and Jonathan D. Cohen, "The Neural Basis of Economic Decision-Making in the Ultimatum Game," *Science*, 2003, *300* (5626), 1755–1758.

Schmitt, Pamela, Robert Shupp, Kurtis Swope, and Justin Mayer, "Pre-Commitment and Personality: Behavioral Explanations in Ultimatum Games," *Journal of Economic Behavior & Organization*, 2008, *66* (3-4), 597–605.

Schokkaert, Erik, "The Economics of Distributive Justice, Welfare and Freedoms," in Klaus R. Scherer, ed., *Justice: Interdisciplinary Perspectives*, Cambridge: Cambridge University Press, 1992, pp. 65–113.

Schwartz, Robert A., "Personal Philanthropic Contributions," *The Journal of Political Economy*, 1970, *78* (6), 1264–1291.

Scott, Robert Haney, "Avarice, Altruism, and Second Party Preferences," *The Quarterly Journal of Economics*, 1972, *86* (1), 1–18.

Segal, Uzi and Joel Sobel, "Tit for Tat: Foundations of Preferences for Reciprocity in Strategic Settings," *Journal of Economic Theory*, 2007, *136* (1), 197–216.

Selten, Reinhard, "Die Strategiemethode zur Erforschung des Eingeschränkt Rationalen Verhaltens im Rahmen eines Oligopolexperiments," in Heinz Sauermann, ed., *Beiträge zur experimentellen Wirtschaftsforsung*, Tuebingen: Mohr, 1967, pp. 136–168.

_ , "The Equity Principle in Economic Behavior," in H. Gottinger and W. Leinfellner, eds., *Decision Theory And Social Ethics, Issues In Social Choice*, Dordrecht: Reifel Publishing, 1978, pp. 289–301.

_ **and Axel Ockenfels**, "An Experimental Solidarity Game," *Journal of Economic Behavior & Organization*, 1998, *34* (4), 517–539.

Sen, Amartya K., "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory," *Philosophy & Public Affairs*, 1977, *6* (4), 317–344.

_ , "Internal Consistency of Choice," *Econometrica*, 1993, *61* (3), 495–521.

Sethi, Rajiv and E. Somanathan, "Preference Evolution and Reciprocity," *Journal of Economic Theory*, 2001, *97* (2), 273–297.

Shaked, Avner, "The Rhetoric of Inequity Aversion," *Working Paper*, 2005.

Simon, Herbert A., "A Behavioral Model Of Rational Choice," *The Quarterly Journal of Economics*, 1955, *69* (1), 99–118.

Singer, Tania, Ben Seymour, John P. O. Doherty, Klaas E. Stephan, Raymond J. Dolan, and Chris D. Frith, "Empathic Neural Responses are Modulated by the Perceived Fairness of Others," *Nature*, 2006, *439* (7075), 466–469.

Slonim, Robert L. and Alvin E. Roth, "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic," *Econometrica*, 1998, *66* (3), 569–596.

Slovic, Paul, "Cigarette Smokers - Rational Actors or Rational Fools?," in Paul Slovic, ed., *Smoking: Risk, Perception, and Policy*, Vol. 216, Thousand Oaks: Sage, 2001, pp. 97–120.

Smith, Vernon L., "An Experimental Study of Competitive Market Behavior," *Journal of Political Economy*, 1962, *70* (2), 111–137.

Stark, Oded, "The Design of Reward Structures in Career Games: A Relative Deprivation Approach," *Economics Letters*, 1987, *25*, 389–392.

_ , "A Relative Deprivation Approach to Performance Incentives in Career Games and Other Contests," *Kyklos*, 1990, *43* (2), 211–227.

Stigler, George J., "Economics or Ethics?," *The Tanner Lectures On Human Values*, 1980.

Sugden, Robert, "Reciprocity: The Supply of Public Goods Through Voluntary Contributions," *The Economic Journal*, 1984, *94* (376), 772–787.

Suleiman, Ramzi, "Expectations and Fairness in a Modified Ultimatum Game," *Journal of Economic Psychology*, 1996, *17* (5), 531–554.

Sutter, Matthias, "Outcomes versus intentions: On the nature of fair behavior and its development with age," *Journal of Economic Psychology*, January 2007, *28* (1), 69–78.

Tabibnia, Golnaz, Ajay B. Satpute, and Matthew D. Lieberman, "The Sunny Side of Fairness: Preference for Fairness Activates Reward Circuitry (and Disregarding Unfairness Activates Self-Control Circuitry)," *Psychological Science*, 2008, *19* (4), 339–347.

Tan, Jonathan H. W. and Friedel Bolle, "On the Relative Strengths of Altruism and Fairness," *Theory and Decision*, 2006, *60* (1), 35–67.

Thaler, Richard H., "Anomalies: The Ultimatum Game," *The Journal of Economic Perspectives*, 1988, *2* (4), 195–206.

_ **and Cass R. Sunstein**, *Nudge*, New Haven: Yale University Press, 2008.

Tversky, Amos and Daniel Kahneman, "Judgment Under Uncertainty: Heuristics and Biases," *Science*, 1974, *185* (4157), 1124–1131.

_ **and** _ , "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *The Quarterly Journal of Economics*, 1991, *106* (4), 1039–1061.

Vanberg, Christoph, "Why Do People Keep Their Promises? An Experimental Test of Two Explanations," *Econometrica*, 2008, *76* (6), 1467–1480.

Von Neumann, John and Oskar Morgenstern, *Theory of Games and Economic Behavior*, 3rd ed., Princeton: Princeton University Press, 1953.

Wyer, Robert S., "Prediction of Behavior in Two-Person-Games," *Journal of Personality and Social Psychology*, 1969, *13* (3), 222–238.

Xiao, Erte and Daniel Houser, "Emotion Expression in Human Punishment Behavior," *Proceedings of the National Academy of Sciences*, 2005, *102* (20), 7398–7401.

Yamagishi, Toshio, "The Provision of a Sanctioning System as a Public Good," *Journal of Personality and Social Psychology*, 1986, *51* (1), 110–116.

Zak, Paul J., Robert Kurzban, and William T. Matzner, "Oxytocin is Associated with Human Trustworthiness," *Hormones and Behavior*, 2005, *48* (5), 522–527.

A Appendix: A Closer Look at Inequity Aversion and Incentives in Tournaments

A.1 Proofs

A.1.1 Proof I: ERC in DG and UG

Given an endowment of $Y = 1$, maximizing the utility function

$$U_i(\sigma_i Y, \sigma_i) = a\sigma_i Y - \frac{b}{2}\left(\sigma_i - \frac{1}{2}\right)^2 \tag{133}$$

yields the first order condition

$$a - b(\sigma_i - \frac{1}{2}) = 0. \tag{134}$$

Therefore, in the dictator game, the player keeps the ratio of the endowment given by

$$\sigma_i = \frac{1}{2} + a/b. \tag{135}$$

Hence, the player only makes a positive transfer if $2a < b$.

In the ultimatum game, the player rejects an offer if the utility from the offer is less than when both players get a payoff of zero, i.e.

$$\begin{aligned}
&a\sigma_i - \frac{b}{2}\left(\sigma_i - \frac{1}{2}\right)^2 < 0\\
\Leftrightarrow &a\sigma_i - \frac{b}{2}\left(\sigma_i^2 - \sigma_i + \frac{1}{4}\right) < 0\\
\Leftrightarrow &\frac{b}{2}\sigma_i^2 - \left(a + \frac{b}{2}\right)\sigma_i > -\frac{b}{8}\\
\Leftrightarrow &\sigma_i^2 - \left(1 + \frac{2a}{b}\right)\sigma_i > -\frac{1}{4}\\
\Leftrightarrow &\sigma_i^2 - \left(1 + \frac{2a}{b}\right)\sigma_i + \left(\frac{1}{2} + \frac{a}{b}\right)^2 > -\frac{1}{4} + \left(0.5 + \frac{a}{b}\right)^2\\
\Leftrightarrow &\left(\sigma_i - \frac{1}{2} - \frac{a}{b}\right)^2 > \frac{ab + a^2}{b^2}\\
\Leftrightarrow &\sigma_i - \frac{1}{2} - \frac{a}{b} < -\frac{\sqrt{ab + a^2}}{b}\\
\Leftrightarrow &\sigma_i < \frac{1}{2} - \frac{\sqrt{ab + a^2} - a}{b}
\end{aligned} \tag{136}$$

A.1.2 Proof II: Status Seeking

We define a status seeker as an agent i who receives a positive utility of $\beta(w_i - w_j)$ when his prize exceeds the prize of the other agent. The mathematical analysis of envious and status seeking agents ($\alpha > 0$, $\beta < 0$) is obviously very closely related to the analysis of the FS-model, so we immediately skip to the first result. Envious and status seeking agents will choose effort level characterized by

$$\Delta w(1 - \beta + \alpha)g(0) = \frac{C'(e^*)}{h'(e^*)} \tag{137}$$

and will therefore always exert more effort than purely self-regarding agents and than envious and compassionate agents. When the principal makes the participation constraint binding, the envious and status seeking agent receives a loser prize of

$$w_2 = U_0 - \frac{1}{2}(1 - \beta - \alpha)\Delta w + C(e_i). \tag{138}$$

If the positive status utility of winning exceeds the utility loss of envy when losing, i.e., $-\beta > \alpha$, agents accept lower loser prizes than purely self-regarding agents. Inserting this loser price into the principal's payoff function leads to the following maximization problem

$$\underset{e}{max}\; 2h(e^*) - 2C(e^*) - (\alpha + \beta)\Delta w - 2U_0 \tag{139}$$

Expression the prize spread Δw as a function of e^* yields the first order condition

$$h'(e^*) - \frac{(\alpha + \beta)}{2(1 + \alpha - \beta)g(0)} \frac{C''(e^*)h'(^*) - C'(e^*)h''(e^*)}{(h'(e^*))^2} = C'(e^*). \tag{140}$$

If $-\beta > \alpha$, implemented effort levels are above the first best level and since agents also accept lower loser prizes, the principal is better off than with purely self-regarding agents. If $-\beta < \alpha$, implemented effort levels are below the first best levels and agents have to be paid higher loser prizes, so the principal remains worse off than with standard agents. However, for any given degree of envy, she is still better off with status seeking agents than with compassionate ones. The partial derivatives with respect to α and β are given by

$$-\frac{1 - 2\beta}{(1 + \alpha - \beta)^2} C'(e)h'(e)g(0) < 0$$

and

$$-\frac{1+2\alpha}{(1+\alpha-\beta)^2}C'(e)h'(e)g(0) < 0,$$

respectively. So the principal's payoff strictly decreases in the degree on envy and strictly increases in the degree of the status seeking (decreasing β).

A.2 Examples

A.2.1 Approach to a Solution for the Optimal Prize Structure with ERC-Agents and Numerical Example

A possible approach to a solution is, for example, the Lagrange multiplier. Maximizing the principal's payoff (63) under the constraint (59) yields the Lagrange-function

$$\begin{aligned}\Lambda(\Delta w, w_2, \lambda) = 2h(e(\Delta w)) - 2C(e(\Delta w)) - 2k(w_2, \Delta w) - 2U_0 \\ + \lambda\left[w_2 - U_0 + \frac{1}{2}\Delta w - C(e(\Delta w)) - k(w_2, \Delta w)\right]\end{aligned} \tag{141}$$

and the three necessary conditions

$$\begin{aligned}\frac{\partial\ \Lambda}{\partial\ \Delta w} = 2h'(e(\Delta w))e'(\Delta w) - 2C'(e(\Delta w))e'(\Delta w) - 2k_{\Delta w}(w_2, \Delta w) + \\ \lambda\left[\frac{1}{2} - C'(e(\Delta w))e'(\Delta w) - k_{\Delta w}(w_2, \Delta w)\right] = 0,\end{aligned} \tag{142}$$

$$\frac{\partial\ \Lambda}{\partial\ w_2} = -k_{w_2}(w_2, \Delta w) + \lambda\left[1 - k_{w_2}(w_2, \Delta w)\right] = 0 \tag{143}$$

and

$$\frac{\partial\ \Lambda}{\partial\ \lambda} = w_2 - U_0 + \frac{1}{2}\Delta w - C(e(\Delta w)) - k(w_2, \Delta w) = 0. \tag{144}$$

Solving (143) for λ and inserting the expression into (142) yields

$$\begin{aligned}2h'(e(\Delta w))e'(\Delta w) - 2C'(e(\Delta w))e'(\Delta w) - 2k_{\Delta w}(w_2, \Delta w) \\ + \frac{k_{w_2}(w_2, \Delta w)}{1 - k_{w_2}(w_2, \Delta w)}\left[\frac{1}{2} - C'(e(\Delta w))e'(\Delta w) - k_{\Delta w}(w_2, \Delta w)\right] = 0,\end{aligned} \tag{145}$$

This expression would have to be solved for either Δw or w_2 and inserted into (144), but this is not possible in the current general form because (145) contains neither Δw nor w_2 directly. However, for more intuitive insight into the mechanism, please consider the following numerical example.

Assume $h(e) = e$, $C(e) = \frac{e^2}{8}$, $U_0 = 0$ and $g(0) = \frac{1}{2}$. Agents have preferences expressed by the ERC-utility function (50) with $b = 0$ for purely self-regarding agents and $b = 2$ for inequity averse agents. The incentive constraint (57) becomes

$$\frac{1}{2}\Delta w = \frac{e^*}{4},$$

so both types of agents choose $e^* = 2\ \Delta w$ in equilibrium. With purely self-regarding agents, the loser prize is given by (62) and the principal's maximization problem (63) becomes

$$\begin{aligned}&max_{\Delta w}\ 2h(e^*) - 2U_0 + \Delta w - 2C(e^*) - \Delta w = \\ &max_{\Delta w}\ 4\ \Delta w - (\Delta w)^2.\end{aligned}$$

This yields the first order condition

$$\Delta w = 2.$$

To meet the agents' participation constraint, the principal has to pay a loser prize of

$$w_2 = U_0 - \frac{1}{2}\Delta w + C(e^*) = 0 - 1 + 2 = 1.$$

To sum up, in the standard model, the principal sets the prize spread Δw to 2 and the loser prize w_2 to 1. Agents choose first best effort levels of $e^{FB} = 4$ and earn $U_0 = 0$ in expectation. The total surplus U_P is 4.

With inequity averse agents, the participation constraint with given by

$$\begin{aligned}&w_2 + \frac{1}{2}\Delta w - C(e^*) - k(\Delta w, w_2) = U_0 \\ \Leftrightarrow &w_2 + \frac{1}{2}\Delta w - \frac{\Delta w}{4} - \frac{(\Delta w)^2}{2(2w_2 + \Delta w)} = 0,\end{aligned}$$

which leads to the principal's maximization

$$\begin{aligned}&max_{\Delta w}\ 2h(e^*) - 2U_0 + \Delta w - 2C(e^*) - 2k(\Delta w, w_2) - \Delta w = \\ &max_{\Delta w}\ 4\ \Delta w - (\Delta w)^2 - \frac{(\Delta w)^2}{2(2w_2 + \Delta w)}.\end{aligned}$$

The solution to this problem is non-trivial because the inequity costs depend both on the prize spread Δw and the loser prize w_2. Therefore, we construct the Lagrange-function

$$\begin{aligned}\Lambda(\Delta w, w_2, \lambda) =& 4\ \Delta w - (\Delta w)^2 - \frac{(\Delta w)^2}{2(2w_2 + \Delta w)} \\ & - \lambda \left[-w_2 - \frac{\Delta w}{2} + \frac{(\Delta w)^2}{2} + \frac{(\Delta w)^2}{4(2w_2 + \Delta w)} \right]\end{aligned} \tag{146}$$

and get the three necessary conditions

$$\begin{aligned}\frac{\partial\ \Lambda}{\partial\ \Delta w} =& 4 - 2\ \Delta w - \frac{(\Delta w)(4w_2 + \Delta w)}{2(2w_2 + \Delta w)^2} \\ & - \lambda \left[-\frac{1}{2} + \Delta w + \frac{(\Delta w)(4w_2 + \Delta w)}{4(2w_2 + \Delta w)^2} \right] = 0,\end{aligned}$$

$$\frac{\partial\ \Lambda}{\partial\ w_2} = \frac{(\Delta w)^2}{(2w_2 + \Delta w)^2} + \lambda \left[1 + \frac{(\Delta w)^2}{2(2w_2 + \Delta w)^2} \right] = 0$$

and

$$\frac{\partial\ \Lambda}{\partial\ \lambda} = w_2 + \frac{\Delta w}{2} - \frac{(\Delta w)^2}{2} - \frac{(\Delta w)^2}{4(2w_2 + \Delta w)} = 0,$$

Solving the second equation for λ yields

$$\lambda = -\frac{2(\Delta w)^2}{(\Delta w)^2 + 2(2w_2 + \Delta w)^2}.$$

Inserting this expression into the first equation gives us

$$\begin{aligned}& 4 - 2\ \Delta w - \frac{(\Delta w)(4w_2 + \Delta w)}{2(2w_2 + \Delta w)^2} \\ & + \frac{2(\Delta w)^2}{(\Delta w)^2 + 2(2w_2 + \Delta w)^2} \left[-\frac{1}{2} + 2\Delta w + \frac{(\Delta w)(4w_2 + \Delta w)}{4(2w_2 + \Delta w)^2} \right] = 0.\end{aligned}$$

Solving for w_2 yields

$$w_2 = \frac{\Delta w \sqrt{16\ \Delta w - 31} - 4(\Delta w)^2 - 7\ \Delta w}{8\ \Delta w - 16}$$

and inserting this expression into the third necessary condition yields

$$\begin{aligned}&\frac{\Delta w\sqrt{16\ \Delta w-31}-4(\Delta w)^2-7\ \Delta w}{8\ \Delta w-16}+\frac{\Delta w}{2}-\frac{(\Delta w)^2}{2}\\&-\frac{(\Delta w)^2(\Delta w-2)}{\Delta w\sqrt{16\ \Delta w-31}-15\ \Delta w}=0,\end{aligned} \tag{147}$$

which can only be solved numerically. Doing so yields an optimal prize spread of $\Delta w \approx 1.98$ and in turn a loser prize of $w_2 \approx 1.19$ (and $\lambda \approx 0.19$). Therefore, the agents will choose an effort level of $e^* \approx 3.96$. This prize structure results in a total surplus of $U_P \approx 3.55$ for the principal.

A.2.2 Numerical Example that ERC-Agents Do Not Exert Effort

Assuming agents have the following utility function

$$u_i = w_i - \frac{b}{2}W(\frac{w_i}{W} - \frac{1}{N})^2. \tag{148}$$

Then agents prefer losing the tournament if

$$\begin{aligned}&u_i(w_2) > u_i(w_1)\\ \Leftrightarrow & w_2 - \frac{b}{2}(Nw_2 + mDeltaw)(\frac{w_2}{Nw_2+m\Delta w} - \frac{1}{N})^2 >\\ & w_1 - \frac{b}{2}(Nw_2 + mDeltaw)(\frac{w_1}{Nw_2+m\Delta w} - \frac{1}{N})^2\\ \Leftrightarrow & -\frac{bm^2(\Delta w)^2}{2N^3w_2+2mN^2\Delta w} > \Delta w - \frac{bN^2(\Delta w)^2 - 2bNm(\Delta w)^2 + bm^2(\Delta w)^2}{2N^3w_2+2mN^2\Delta w}\\ \Leftrightarrow & \frac{bN^2(\Delta w)^2 - 2bNm(\Delta w)^2}{2N^3w_2+2mN^2\Delta w} > \Delta w\\ \Leftrightarrow & b > \frac{2N(Nw_2+m\Delta w)}{\Delta w(N-2m)}.\end{aligned}$$

The partial derivative of the right hand side with respect to w_2, Δw, N, and m are given by

$$\frac{2N^2}{\Delta w(N-2m)} > 0,$$

$$-\frac{2N^2w_2}{(\Delta w)^2(N-2m)} < 0,$$

$$\frac{2w_2N^2 - 8mw_2N - 4m^2\Delta w}{N^2\Delta w - 4mN\Delta w + 4m^2\Delta w},$$

and

$$\frac{N^2(4w_2 + \Delta w)}{N^2\Delta w - 4mN\Delta w + 4\Delta w m^2},$$

so the necessary degree of inequity aversion to prefer losing the tournament increases in the loser prize and decreases in the prize spread. For the number of participants and winners, no general results can be derived.

A.3 Derivatives

A.3.1 Derivative of (118)

$$\frac{(\alpha+\beta)(m^2(\alpha+\beta) + 2m(N(1-\beta)-1) - N(N(1-\beta)-1))}{(N-1+m\alpha-(N-m)\beta)^2}\frac{C'(e)}{h'(e)f_{\varepsilon_{l:n}}(0)}. \tag{149}$$

A.3.2 Derivative of (127)

$$\begin{aligned}
&\frac{b(Nw_2+m\Delta w)\left(\frac{w_2+\Delta w}{Nw_2+m\Delta w}-\frac{1}{N}\right)^2}{2} + \frac{bm\Delta w\left(\frac{w_2+\Delta w}{Nw_2+m\Delta w}-\frac{1}{N}\right)^2}{2}\\
&-\frac{b(Nw_2+m\Delta w)\left(\frac{w_2}{Nw_2+m\Delta w}-\frac{1}{N}\right)^2}{2} + \frac{b(N-m)\Delta w\left(\frac{w_2}{Nw_2+m\Delta w}-\frac{1}{N}\right)^2}{2}\\
&-\frac{bm\Delta w(w_2+\Delta w)\left(\frac{w_2+\Delta w}{Nw_2+m\Delta w}-\frac{1}{N}\right)^2}{Nw_2+m\Delta w}\\
&-\frac{b(N-m)w_2\Delta w\left(\frac{w_2}{Nw_2+m\Delta w}-\frac{1}{N}\right)^2}{Nw_2+m\Delta w},
\end{aligned} \tag{150}$$

B Appendix: Does Altruism Depend on Rational Expectations?

B.1 Instructions

B.1.1 General Information

You are now participating in a scientific economic experiment. In this experiment, you can earn money. Your income depends on your decision and on the decisions of other participants. Independent of the results of the experiment, each participant will receive the show-up fee of 2.50 Euro and an experiment-specific bonus of 0.50 Euro.[64] Therefore, each participant will receive at least 3.00 Euros today. All payoffs from the experiment are added to this amount. At the end of the experiment, the total amount you have earned will be paid out in cash.

These instructions are private information for you. Please cease all communication with other participants from this point on. If you have any questions, please raise your hand. We will come to your cabin and help you. Please note that because of the electronic devices, eating and drinking in the cabin is not allowed. A violation of these rules can lead to the disqualification from the experiment and all payments.

We also ask you to turn off all cellphones, MP3-players, etc. and to go without other distractions like books, newspapers or other documents. Thank you kindly!

During the experiment, all your income will be calculated in points, not in Euros. At the end of the experiment, all your points will be converted into Euros. 1 point equals 5 Cents.

Please also note the following points: All decisions are made anonymously, i.e., no other participant is informed about another participant's decision. The payment at the end of the experiment is also anonymous, i.e., no other participant is informed about the payment of another participant.

The following page explains the exact procedure of today's experiment.

[64]We did a pilot where we had an exceptionally high share of senders who transfered no points (over 50%). To keep frustration levels low we added the 50 cents to ensure that each participant would receives more money when participating than when he or she could not particpate (because we needed even numbers).

B.1.2 Senders Only

The participants of today's experiment have been randomly assigned to different groups. All members of your group will receive an identical amount of points. You can allocate these points as you desire between yourself and another participant (receiver). Each member of your group has been assigned exactly one (different) receiver. The exact amount of points that you and the other members of your group can allocate has not been determined yet. The amount will either be *LOW-Treatment:* 100 points or 150 points / *HIGH-Treatment:* 150 points or 200 points.

A random mechanism will determine which one of the two amounts you will have available. Both amounts are equally likely, i.e., both have a probability of 50%. After chance has determined the exact amount of points, you decide how many points you want to allocate to the receiver. You keep all remaining points for yourself. The receiver is informed that somebody decides about the allocation of points between him/herself and the receiver. The receiver has no information about how many points are potentially or actually available. The receiver will not receive other points except those that you allocate to him/her.

The experiment ends after you made your allocation decision. There are no further rounds. At the conclusion of the experiment, every participant is paid out his or her points at the same exchange rate (1 point = 5 cents) in addition to the 3 Euro that every participant receives for participating.

B.1.3 Receivers Only

The participants of today's experiment have been randomly assigned to different groups. You belong to the group of receivers. Each receiver has been assigned exactly one member of another group. The participant that has been assigned to you will receive a randomly determined amount of points. The participant can allocate these points freely between him- or herself and you. As a receiver, you do not have the possibility to influence your payoff. However, we would like to ask you a few questions about how you would have acted if you had been in a different position.

The experiment ends after all allocation decisions have been made. There are no further rounds. At the conclusion of the experiment, every participant is paid out his or her points at the same exchange rate (1 point = 5 cents) in addition to the 3 Euro that every participant receives for participating.

B.2 Additional Graphics

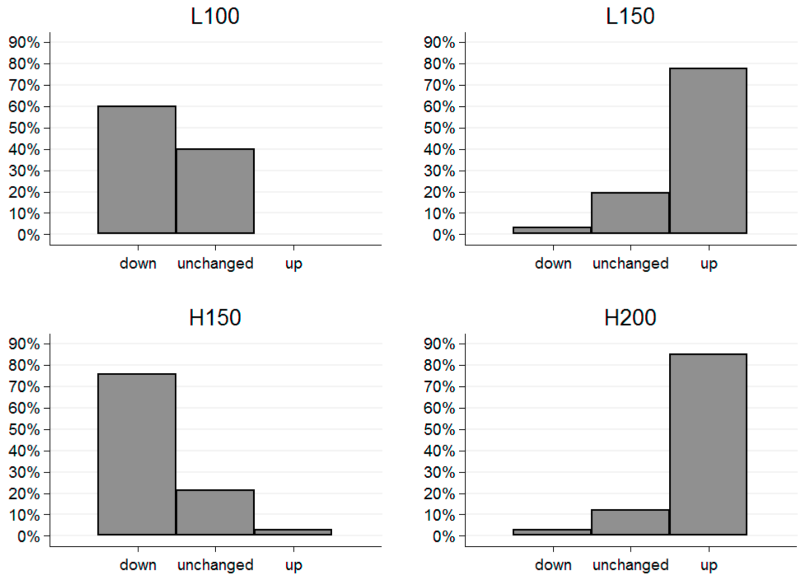

Figure 7: Mood Change After Lottery

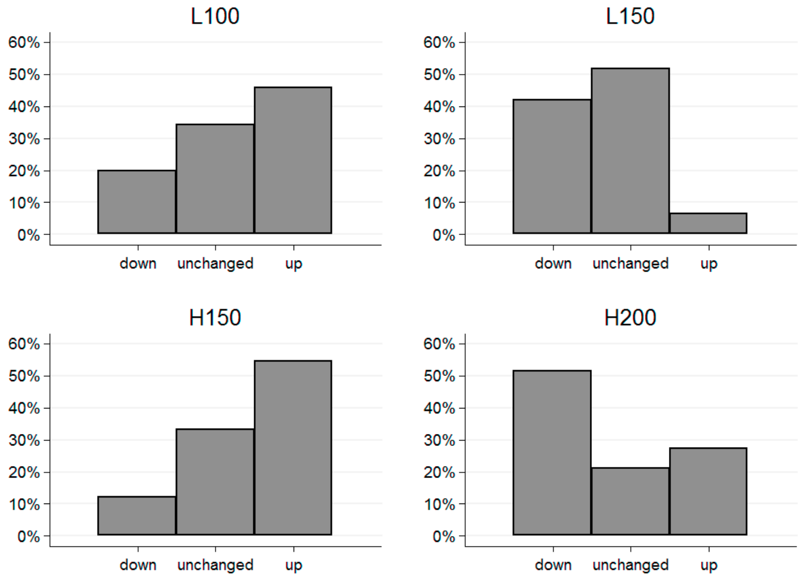

Figure 8: Mood Change After Transfer

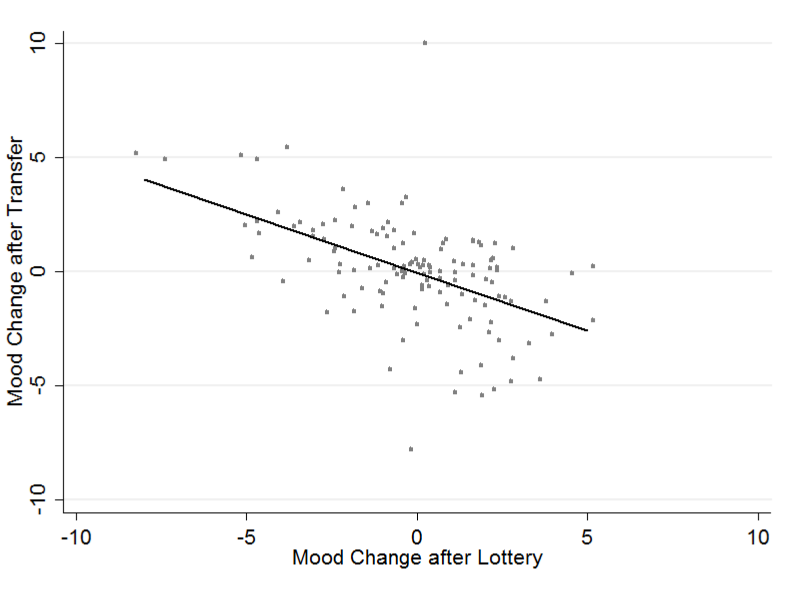

Figure 9: Mood Change Correlation
(with Noise Added for Superimposed Values)

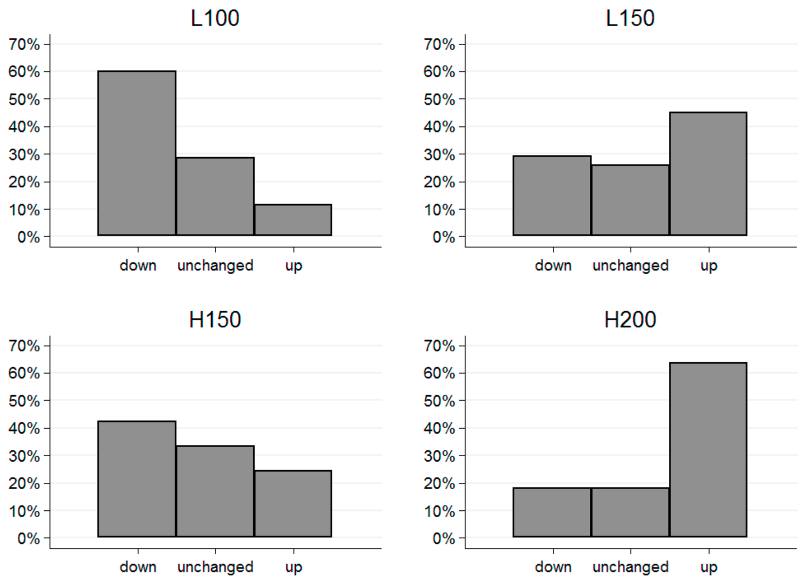

Figure 10: Total Mood Change

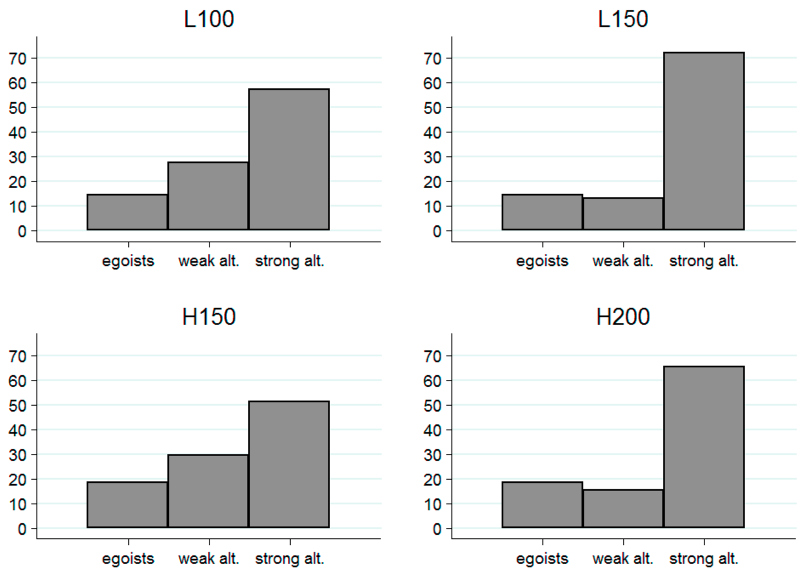

Figure 11: Categorization of Receivers's Hypothetical Transfer

B.3 Additional Tables

Mood	L100	L150	H150	H200	All
Before lottery	0.11 (.511)	0.23 (.214)	-0.18 (.314)	0.11 (.561)	0.06 (.523)
After lottery	0.25 (.141)	0.23 (.208)	0.13 (.468)	-0.27 (.136)	0.06 (.486)
After transfer	0.06 (.746)	0.08 (.661)	0.03 (.873)	0.04 (.808)	0.03 (.754)
Mood Change	L100	L150	H150	H200	All
After Lottery	0.03 (.851)	-0.15 (.408)	0.18 (.307)	-0.48 (.005)	-0.03 (.729)
After Transfer	-0.22 (.208)	-0.04 (.847)	0.02 (.898)	0.18 (.317)	-0.02 (.789)
Overall	-0.14 (.428)	-0.13 (.472)	0.20 (.277)	-0.15 (.394)	-0.00 (.961)

Table 4: Spearman's Rho and P-Values for Correlations Between Senders' Mood and the Ratio of Transfer/Endowment

After Lottery	Mood Up	Mood Down	No Change
Egoists	21	19	11
Weak Altruists	14	22	8
Strong Altruists	15	10	12
Economics	30	20	10
Non-Economics	28	23	21
After Transfer	Mood Up	Mood Down	No Change
Egoists	19	14	18
Weak Altruists	11	14	19
Strong Altruists	13	15	9
Economics	20	19	21
Non-Economics	25	22	25

Table 5: Change of Senders' Mood After the Conclusion of the Lottery and After the Transfer Decision

	All	L100	L150	H150	H200
Const	0.01	0.00	0.02	-0.01	0.05
	(0.28)	(0.03)	(0.28)	(-0.15)	(0.63)
Fair Ratio	0.43***	0.53***	0.38**	0.47***	0.29
	(5.33)	(2.94)	(2.22)	(3.28)	(1.51)
F	28.46	8.63	4.94	10.76	2.28
Prob > F	0.00	0.01	0.03	0.00	0.14
R^2	0.18	0.21	0.15	0.27	0.07

Table 6: OLS-Regression: Transfer Ratio Explained by Fair Ratio (Senders)

	ρ	p
All	0.389	0.000
L100	0.412	0.014
L150	0.289	0.114
H150	0.442	0.013
H200	0.062	0.733

Table 7: Correlation of Fair Transfer and Actual Transfer (Senders)

	Mean	StD	Median
All	.179	.141	16.7%
L100	.180	.146	20.0%
L150	.184	.125	20.0%
H150	.164	.148	13.3%
H200	.189	.147	25.0%

Table 8: Sender's Beliefs About Other Senders' Transfer Ratio

	N	Mean	StD
Goal > €7.50	28	.153	.187
Goal <= €7.50	31	.165	.165
Payoff goal	59	.159	.190
No payoff goal	73	.178	.180

Table 9: Ratio Transfer/Endowment for Subjects with Payoff Goal

	Mean	StD	Median	N
All	.463	.081	.500	258
L100	.463	.086	.500	63
L150	.465	.081	.500	63
H150	.457	.083	.500	66
H200	.467	.076	.500	66

Table 10: Fair Transfer/Endowment by Receivers

	Mean	StD	Median	N
All	.253	.170	.250	244
L100	.263	.159	.300	63
L150	.275	.153	.333	63
H150	.229	.176	.200	59
H200	.243	.176	.250	59

Table 11: Belief about Senders' Transfer/Endowment by Receivers

	Mean	StD	Median	N
All	.115	.165	.003	60
L100	.133	.055	.000	12
L150	.143	.175	.067	15
H150	.045	.088	.000	16
H200	.146	.021	.200	17

Table 12: Transfer/Endowment by Economics Students (Senders)

	Mean	StD	Median	N
All	.215	.188	.217	72
L100	.230	.200	.200	23
L150	.165	.149	.133	16
H150	.251	.220	.333	17
H200	.205	.172	.250	16

Table 13: Transfer/Endowment by Non-Economics Students (Senders)

	Logit1	*Logit2*	*Logit3*	*Logit4*	*Logit5*	*Logit6*	*Logit7*
Const	-0.06	0.08	0.69	-0.59	-0.35	0.86	1.28**
	(-0.17)	(0.18)	(0.68)	(-1.32)	(-0.84)	(1.49)	(2.36)
Won^a	-0.83*	-0.83*	-0.88**	-0.89*	-0.85*	-0.74*	-0.78*
	(-1.58)	(-1.57)	(-1.66)	(-1.63)	(-1.59)	(-1.34)	(-1.31)
Female		-0.26					
		(-0.50)					
Initial Mood^b			-0.11				
			(-0.79)				
Econ. Student				1.09**			
				(2.00)			
Monetary Goal					0.68		
					(1.28)		
Fair Ratio						-2.94**	
						(-2.05)	
Belief Ratio							-8.81***
							(-3.32)
Log likelihood	-41.53	-41.41	-40.96	-39.44	-40.70	-39.21	-34.08
LR chi^2(2)	2.57	2.82	3.71	6.76	4.23	7.22	17.48
Prob > chi^2	0.11	0.24	0.16	0.03	0.12	0.03	0.00

a) one-sided, b) all other mood variables are insignificant, too

Table 14: Logit-Regressions for *L150* and *H150*

Curriculum Vitae – Björn Hartig

EDUCATION AND QUALIFICATIONS

OCTOBER 2008 – SEPTEMBER 2011
PH.D. scholarship at the Cologne Graduate School of Management, Economics, and Social Sciences at the University of Cologne
Supervisor: Prof. Dr. Axel Ockenfels

MAY 2003 – SEPTEMBER 2008
Diploma in economics at the University of Bonn (Grade: 1.7)
Thesis: Does culture affect reward related brain activity in social comparison? (Supervisor: Prof. Dr. Armin Falk; Grade: 1.0)

OCTOBER 1999 – JULY 2006
Magister Artium (Master of Arts) in Sinology at the University of Bonn (Grade: 1.6), Minors: Tibetology, Mongolian Studies
Thesis: Minority Rights in the Chinese Constitution (Supervisor: Prof. Dr. Wolfgang Kubin, Grade: 1.3)

JUNE 1997
Abitur at the Gymnasium Carolinum in Osnabrück (Grade: 1.2)

AWARDS

2010
Heinz Sauermann-Grant for Experimental Economics of the German Society for Experimental Economics (together with Felix Kölle)

WORK EXPERIENCE

OCTOBER 2011 – CURRENT
Experimental Laboratory Officer at Royal Holloway, University of London

OCTOBER 2009 – APRIL 2011
Development and Lecture of undergraduate course "Experiments in Economics and Social Sciences"

APRIL 2006 – SEP. 2008
Teacher for Chinese language, culture and society at the Federal Language Office of the German Federal Armed Forces

APRIL – JULY 2005 & 2006
Assistant teacher (tutor) for the course BWL III (Decision theory, organization and marketing) by Prof. Dr. Kräkel, University of Bonn

OCTOBER 2002 – APRIL 2003
Teacher for German language at the Dehua Xinri Education Consultant Center in Beijing, China

FEBRUARY 2003
Teacher for English language at the Capital Normal University Winter Camp in Harbin, China

JULY 2002
Teacher for English language at the Capital Normal University Summer Camp in Beijing, China

MAY 1999 – AUGUST 2002
Coder (Visual Basic) for CJ Games in Aachen

AUGUST 1997 – AUGUST 1998
Civil Service

ADDITIONAL EXPERIENCES

2000 – 2006
Member of the students council of Asian and Oriental Language and Cultural Studies, University of Bonn

JULY 1996
Participation in the course “Paradoxien” (Paradox) at the Deutsche Schüler Akademie of the Verein für Bildung und Begabung e.V. in Annweiler, Germany

C Acknowledgments

This dissertation would have never come into being had it not been for the support of many different people.

I would like to take the chance and thank all of them from the bottom of my heart.

First, I want to thank Prof. Dr. Dirk Sliwka and the board of the Cologne Graduate School for taking a chance on me and giving me the opportunity to write my dissertation and pursue my dream.

Many special thanks go to Prof. Dr. Axel Ockenfels for agreeing to supervise my dissertation, for giving me the freedom to explore my own ideas, for all his valuable feedback, and for teaching me not to be contend with mediocrity.

I am heartily thankful to my wife Lan for being patient with me and for grounding me, to my parents for always supporting me no matter what, to my grandmothers for giving me perspective, and to my son Jonathan for having the courtesy to wait long enough to come into this world until his old man was pretty much finished with his dissertation.

I also want to thank all my colleagues at the CGS and everyone from the WISO faculty for always being kind, helpful and fun to be around, making the last three years a very special time in my life that I will cherish forever.

I would also like to extend a thank you to the Turks for introducing coffee to Europe and to everybody involved in the invention of personal computers, typewriters, and the Internet. Without those milestones in the history of humankind, I would most likely never have graduated from University in the first place.

Thank you all! :)