Gutenberg School of Management and Economics

& Research Unit "Interdisciplinary Public Policy"

Discussion Paper Series

The Many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences

Adrian Bruhin, Ernst Fehr, and Daniel Schunk

January 04, 2016

Discussion paper number 1603

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
wiwi.uni-mainz.de

Contact details:

Adrian Bruhin
University of Lausanne
Faculty of Business and Economics (HEC Lausanne)
1015 Lausanne, Switzerland
adrian.bruhin@unil.ch

Ernst Fehr
University of Zurich
Department of Economics
8006 Zurich, Switzerland
ernst.fehr@econ.uzh.ch

Daniel Schunk
University of Mainz
Department of Economics
55099 Mainz, Germany
daniel.schunk@uni-mainz.de

The Many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences

Adrian Bruhin **Ernst Fehr** **Daniel Schunk**

February 1, 2016

Abstract

There is vast heterogeneity in the human willingness to weigh others' interests in decision making. This heterogeneity concerns the motivational intricacies as well as the strength of other-regarding behaviors, and raises the question how one can *parsimoniously* model and characterize heterogeneity across several dimensions of social preferences while still being able to predict behavior over time and across situations. We tackle this task with an experiment and a structural model of preferences that allows us to simultaneously estimate outcome-based and reciprocity-based social preferences. We find that *non-selfish preferences are the rule rather than the exception*. Neither at the level of the representative agent nor when we allow for several preference types do purely selfish types emerge. Instead, three temporally stable and qualitatively different other-regarding types emerge *endogenously*, i.e., without pre-specifying assumptions about the characteristics of types. When ahead, all three types value others' payoffs significantly more than when behind. The first type, which we denote as *strongly altruistic type*, is characterized by a relatively large weight on others' payoffs – even when behind – and moderate levels of reciprocity. The second type, denoted as *moderately altruistic type*, also puts positive weight on others' payoff, yet at a considerable lower level, and displays no positive reciprocity while the third type is *behindness averse*, i.e., puts a large negative weight on others' payoffs when behind and behaves selfishly otherwise. We also find that there is an *unambiguous and temporally stable assignment of individuals to types*. Moreover, *the three-type model substantially improves the (out-of-sample) predictions of individuals' behavior across additional games* while the information contained in subject-specific parameter estimates leads to no or only minor additional predictive power. This suggests that a parsimonious model with three types captures the bulk of the predictive power contained in the preference estimates.

JEL classification: C49, C91, D03

Keywords: Social Preferences, Heterogeneity, Stability, Finite Mixture Models

Authors' affiliations:

Adrian Bruhin: University of Lausanne, Faculty of Business and Economics (HEC Lausanne), 1015 Lausanne, Switzerland; adrian.bruhin@unil.ch

Ernst Fehr: University of Zurich, Department of Economics, 8006 Zurich, Switzerland; ernst.fehr@econ.uzh.ch

Daniel Schunk: University of Mainz, Department of Economics, 55099 Mainz, Germany; daniel.schunk@uni-mainz.de

1 Introduction

A large body of evidence suggests that social preferences can play an important role in economic and social life.[1] It is thus key to understand the motivational sources and the distribution of social preferences in a population, and to capture the prevailing preference heterogeneity in a parsimonious way. Parsimony is important because in applied contexts tractability constraints typically impose serious limits on the degree of complexity that theories can afford at the individual level. At the same time, however, favoring the most extreme form of parsimony – by relying on the assumption of a representative agent – is particularly problematic in the realm of social preferences because even minorities with particular social preferences may play an important role in strategic interactions. The reason is that social preferences are often associated with behaviors that change the incentives even for those who do not have those preferences.[2] This means that even if only a minority has social preferences they can play a disproportionately large role for aggregate outcomes. Thus, we need to be able to capture the relevant components of social preference heterogeneity while still maintaining parsimony and tractability.

It is our objective in this paper to make an important step in this direction. For this purpose we develop a structural model of social preferences that is capable of capturing both preferences for the distribution of payoffs between the players and preferences for reciprocity. These types of social preferences have played a key role in the development of this subject over the last 15 to 20 years and their relative quantitative importance is still widely debated (Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000; Charness & Rabin, 2002; Falk et al., 2008; Engelmann & Strobel, 2010). However, in the absence of an empirically estimated structural model it seems difficult to make progress on such questions. Therefore, we implement an experimental design that enables us to simultaneously estimate distribution-related preference parameters and the parameters related to (positive and negative) reciprocity preferences. The size of the estimated parameters then informs us about the relative importance of different preference components.[3] However, most importantly, our experiment provides a rich data set that allows us to characterize the distribution of social preferences at three different levels: (i) the representative agent level, (ii) the intermediate level of a small number of distinct preference types and (iii) the individual level.

From the viewpoint of achieving a compromise between tractability and parsimony, and the goal of capturing the distinct qualitative properties of important minority types the intermediate level is most interesting. We approach this level by applying a finite mixture model that endogenously identifies

[1] See, e.g., Roth, 1995; Fehr & Gächter, 2000; Charness & Rabin, 2002; Camerer, 2003; Engelmann & Strobel, 2004; Bandiera et al., 2005; Fisman et al., 2007; Dohmen et al., 2008, 2009; Erlei, 2008; Bellemare et al., 2008, 2011; Kube et al., 2012, 2013; Fisman et al., 2015. The evidence on social preferences has spurred the development of numerous models (Rabin, 1993; Levine, 1998; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000; Charness & Rabin, 2002).

[2] For example, a selfish proposer in the ultimatum game may have a reason to make fair offers even if only a (significant) minority of the responders is willing to reject unfair offers. Likewise, a selfish employer in a gift exchange game may have a reason to pay high, non-market clearing wages, although "only" a minority of employees reciprocates to high wages with higher effort. Also, in public good situations, a minority of players willing to punish free-riders can induce selfish players to contribute (see, e.g., Fehr & Schmidt 1999).

[3] This also allows us to assess the long-standing claim in the literature (see e.g., Charness & Rabin, 2002; Offerman, 2002; Al-Ubaydli & Lee, 2009) that negative reciprocity is generally more important than positive reciprocity.

different types of preferences in the population without requiring any pre-specifying assumptions about the existence and the preference properties of particular types. This means, for example, that we do not have to assume, say, a selfish or a reciprocal type of individuals. Rather, the data themselves "decide" which preference types exist and how preferences for the distribution of payoffs and for reciprocity are combined in the various types. Taken together, our finite mixture approach enables us to simultaneously identify (i) the preference characteristics of each type, (ii) the relative share of each preference type in the population and (iii) the (probabilistic) assignment of each individual to one of the preference types. The third aspect has the nice implication that our finite mixture approach provides us with the opportunity to make out-of-sample predictions *at the individual level* without the need to estimate each individual's utility function separately.

Which preference types do our finite mixture estimates yield? We find that a model with three types best characterizes the distribution of preferences at the intermediate level. A model with three types is best in the sense that it produces the most unambiguous assignment of individuals to the different preference types relative to its fit[4], that this assignment is temporally relatively stable over time and that the preference properties of the different types are stable over time. In contrast, models with two or four types produce a more ambiguous assignment of individuals to types relative to their fit and are associated with severe instabilities of the preference properties of the various types across time.

At the substantive level, what are the preference properties of the different types and how large are their shares in the population? The social preferences of the three types are best described as (i) strongly altruistic, (ii) moderately altruistic, and (iii) behindness averse. Interestingly, all three types weigh the payoff of others significantly more in the domain of advantageous inequality (i.e., when ahead) than in the domain of disadvantageous inequality (i.e., when behind), and for all of them the preference parameters that capture preferences for the distribution of payoffs are generally quantitatively more important than preferences for reciprocity.

The strong altruists, which comprise roughly 40% of our subject pool, put a relatively large positive weight on others' payoffs regardless of whether they are ahead or behind. In terms of willingness to pay to increase the other player's payoff by $1, the strong altruists are on average willing to spend 86 Cents when ahead and 19 Cents when behind. In addition, they also display moderate levels of positive and small levels of negative reciprocity, i.e., for them negative reciprocity is the weaker motivational force than positive reciprocity.

The moderate altruists, which comprise roughly 50% of our subject pool, put a significantly lower, yet still positive weight on others' payoffs. They display no positive but a small and significant level of negative reciprocity. A moderate altruist is on average willing to pay 15 Cents to increase the other player's payoff by $1 when ahead and 7 Cents when behind. It may be tempting to treat this low-cost altruism as unimportant. We believe, however, that this would be a mistake because social life is full of situations in which people can help others at low cost. Many may, for example, be willing to give directions to a stranger and help a colleague, both of which is associated with small time cost, or donate

[4] The assignment of individuals to a preference type is completely unambiguous if the individual belongs to the type either with probability one or probability zero. Intermediate probabilities mean that the assignment contains some ambiguity. See Section 3.3 for details.

some money to the victims of a hurricane although they may not be willing to engage in high-cost altruism.

Finally, the behindness averse type comprises roughly 10 percent of the subject pool and is characterized by a relatively large willingness to reduce others' income when behind – spending 78 Cents to achieve an income reduction by $1 – but no significant willingness to increase others' income when ahead or when treated kindly.

One remarkable feature of our finite mixture estimates is that no purely selfish type emerges. Instead, all three types display some form of other-regarding preference suggesting that other-regarding preferences are the rule not the exception. This conclusion is also suggested by the preference estimates for the representative agent which are characterized by intermediate levels of altruism – in between the strong and the moderately altruistic types. The absence of an independent selfish type does of course not mean that there are no circumstances – such as certain kinds of competitive markets – in which the assumption of self-interested *behavior* may well be justified.[5] However, it means that if one makes this assumption in a particular context there is a need to justify the assumption because many people may not behave selfishly in these contexts because they *are* selfish but because the institutional environment makes other-regarding behavior impossible or too costly.

Our preference estimates for the representative agent model reinforce the conclusion regarding the relative importance of distributional versus reciprocity preferences. For the representative agent distributional preferences are considerably more important than reciprocity preferences. In the absence of any kindness or hostility between the players the representative agent is, for example, willing to spend 33 Cents to increase the other player's payoff by $1 when ahead. If – in addition – the other player has previously been kind the representative agent's willingness to pay increases to 50 Cents at most. Relying on the preference estimates of the representative agent may however, be seriously misleading because according to these estimates behindness averse behaviors can only occur as a random (utility) mistake while in fact a significant minority of the subject pool – the behindness averse type – has clear preferences for income reductions when behind.

Out-of-sample predictions of individual behavior constitute the most stringent tests of a model. To study the extent to which our type-specific social preference estimates are capable of predicting individual behavior in other games, the subjects also participated in several additional games. In the first class of games they participated as second-movers in a series of ten trust games with varying costs of trustworthiness; in the second class of games they participated in two games in which they could reward and punish the previous behavior of another player. We are particularly interested in the question whether individual predictions based on our type-specific preference estimates are as good as individual predictions based on individual preference estimates. If this were the case, our type-based model would not only capture the major qualitative social preference types in a parsimonious way but there would also be no need to further disaggregate the preference estimates for predictive purposes. The results show indeed that our three-type model achieves this goal. If we predict each individual's behavior in the additional games on the basis of their types' preferences, we substantially increase the predictive

[5] One of the nice features of the various social preference models (e.g., Levine, 1998; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000) is that they show that the self-interest assumption may be unproblematic in certain environments because subjects with social preferences behave as if self-interested. Thus, by assuming self-interested subjects in these situations one does not make a mistake.

power over a model that just uses demographic and psychological personality variables as predictors. Moreover, despite its parsimony, the predictive power of the type-based model is virtually as good as the predictions that are based on estimates of each individual's preferences.

Thus, taken together the out-of-sample predictions indicate a remarkable ability of the three-type model to predict individual variation in other games. The predictive exercise also enables further insights into the strengths and the weaknesses of the type-based model. On the positive side, we find that the strength of specific behaviors such as rewarding others for a fair act is in line with the type-based model. The strong altruists reward more than the moderate altruists while the behindness averse types do not reward at all. Likewise, as predicted by the model, the behindness averse types display a considerably higher willingness to punish unfair actions (when behind) than the strong or moderate altruists. However, we also find patterns that cannot be fully reconciled with the type-based model. In particular, the behindness averse types should never reciprocate trust in the trust games because they don't put a positive value on other's payoff, but in fact we observe that they are trustworthy at moderate cost levels. These findings indicate the limits of the model and suggest certain ways to improve it – a task that we discuss in more detail in Section 4.5 of the paper.

Our paper is related to the literature on the structural estimation of social preferences at the individual level such as Andreoni and Miller (2002), Bellemare et al. (2008, 2011), Fisman et al. (2007, 2015). However, in contrast to this literature, the purpose of our paper is to provide a parsimonious classification of individuals to – endogenously determined – preference types and a characterization of the distribution of social preferences in terms of individuals' assignment to a small number of types. The results of the paper show that basically all individuals are unambiguously assigned to one of three mutually exclusive types and that individual preference estimates do not lead to superior out-of-sample predictions relative to the much more parsimonious three-type model.[6]

Our paper is also related to the literature that characterizes the latent heterogeneity in social preferences using finite mixture models. Previous studies in this literature mainly focus on distributional preferences and typically classify subjects into *predefined* preference types. For instance, Iriberri & Rey-Biel (2011, 2013) elicit distributional preferences with a series of modified three-option dictator games and apply a finite mixture model to classify subjects into four predefined types. Similarly, studies by Conte & Moffatt (2014), Conte & Levati (2014), and Bardsley & Moffatt (2007) use behavior in public good and fairness games, respectively, to classify subjects into predefined types. Such a priori assumptions may or may not be justified. For example, all of these studies assume the existence of a purely selfish type but as our analysis indicates a purely selfish type may not exist if one allows for sufficiently small costs of other-regarding behaviors. Likewise, often behindness aversion is not a feasible type by assumption

[6] There are also several other differences between these papers and our paper. First, our paper simultaneously identifies outcome-based social preferences and preferences for reciprocity while the mentioned papers – with the exception of Bellemare et al. 2011 – focus exclusively on outcome-based social preferences. Second, the structural model in Andreoni and Miller (2002) and Fisman et al. (2007, 2015) is based on a CES utility function with own and others' payoff as arguments – which rules out behindness aversion – whereas the structural model in Bellemare et al. (2008) rules out the existence of altruism and pure selfishness. This assumption in the paper by Bellemare et al. (2008) may explain why he finds a large amount of inequality aversion while we do not find evidence for a separate type that simultaneously dislikes advantageous and disadvantageous inequality. However, the large amount of inequality aversion in Bellemare et al. (2008) could also result from the fact that students are a minority in his subject pool – which is representative for the Dutch population – while our subject pool consists exclusively of students.

and therefore the structural model cannot identify such types. The only study we are aware of that identifies types endogenously instead of predefining them is by Breitmoser (2013). This study relies on existing dictator game data from Andreoni & Miller (2002) and Harrison & Johnson (2006) for testing the relative performance of different preference models with varying error specifications. However, this study as well as the others mentioned in this paragraph do not simultaneously estimate distributional *and* reciprocity-based preferences, nor do they compare the power of the type-specific and individual estimates in making out-of-sample predictions across games.

Our paper also contributes to the literature concerned with the stability of social preferences. Most studies in this literature analyze behavioral correlations. For example, Volk et al. (2012) and Carlson et al. (2014) report that contributions to public goods appear to be stable over time in the lab as well as in the field. Moreover, there is evidence that behaviors such as trust (Karlan, 2005), charitable giving (Benz & Meier, 2008), and contributions to public goods (Fehr & Leibbrandt, 2011; Laury & Taylor, 2008) seem to be correlated between the lab and field settings. Blanco et al. (2011) study the within-subject stability of inequality aversion across several games in order to understand when and why models of inequality aversion are capable of rationalizing aggregate behavior in games. However, most of these studies do not estimate a structural model of social preferences, which would be necessary for making precise quantitative behavioral predictions. As a consequence, they do not characterize the distribution and the overall characteristics of social preferences in the study population.

The remainder of the paper is organized as follows. Section 2 discusses our behavioral model and describes the experimental design. Section 3 covers our econometric strategy for estimating the behavioral model's parameters at different levels of aggregation. Section 4 presents the results and discusses their stability over time and across games. Finally, section 5 concludes.

2 Behavioral model and experimental design

2.1 Behavioral model

To characterize the distribution of social preferences at the aggregate, the type-specific and the individual level and to make out-of-sample predictions across games, we need a structural model of social preferences. To achieve our goals we apply a two-player social preference model inspired by Fehr & Schmidt (1999) and Charness & Rabin (2002) which we extended to make it also capable of capturing preferences for reciprocity. In the outcome-based part of the model, Player A's utility,

$$U^A = (1 - \alpha s - \beta r) * \Pi^A + (\alpha s + \beta r) * \Pi^B, \quad (1)$$

is piecewise-linear, where Π^A represents player A's payoff, and Π^B indicates player B's payoff.

$s = 1$ if $\Pi^A < \Pi^B$, and $s = 0$ otherwise (disadvantageous inequality);
$r = 1$ if $\Pi^A > \Pi^B$, and $r = 0$ otherwise (advantageous inequality).

Depending on the values of α and β, subjects belong to different preference types: A subject whose α and β are both zero is a purely selfish type, because she does not put any weight on the other player's

payoff. If $\alpha < 0$ the subject is behindness averse, as she weights the other's payoff negatively whenever her payoff is smaller than the other's. Analogously, if $\beta > 0$ the subject is aheadness averse, since she weights the other's payoff positively whenever her payoff is larger than the other's. Consequently, a subject who is both behindness and aheadness averse with $\alpha < 0 < \beta$ and $-\alpha < \beta$ is a difference averse type for whom disadvantageous inequality matters less than advantageous inequality. In case $\alpha < 0 < \beta$ and $-\alpha > \beta$, the subject is difference averse too, but disadvantageous inequality matters more than advantageous inequality; this is the case discussed in Fehr and Schmidt (1999). A subject with $\alpha > 0$ and $\beta > 0$ is an altruistic type, as she always weights the other's payoff positively. In contrast, a subject with $\alpha < 0$ and $\beta < 0$ is a spiteful type, since she puts a negative weight on the other's payoff, regardless of whether she is behind or ahead. Finally, a subject with $\alpha > 0 > \beta$ exhibits quite implausible preferences, since she weights the other's payoff positively when she is behind, and negatively when she is ahead. We do not expect to observe such preferences in our data.

Because we are also interested in the subjects' willingness to reciprocate kind or unkind acts, we extend model (1) to account for positive and negative reciprocity. The extension is similar to Charness & Rabin (2002) who take only negative reciprocity into account and Bellemare et al. (2011) who consider both positive and negative reciprocity. Player A's utility in the extended model is

$$U^A = (1 - \alpha s - \beta r - \gamma q - \delta v) * \Pi^A + (\alpha s + \beta r + \gamma q + \delta v) * \Pi^B, \quad (2)$$

where q and v indicate whether positive or negative reciprocity play a role. More formally,

$q = 1$ if player B behaved kindly towards A, and $q = 0$ otherwise (positive reciprocity);
$v = 1$ if player B behaved unkindly towards A, and $v = 0$ otherwise (negative reciprocity).

A positive value of γ in equation (2) means that player A exhibits a preference for positive reciprocity, i.e. a preference for rewarding a kind act of player B by increasing B's payoff. A negative value of δ represents a preference for negative reciprocity, i.e. a preference for punishing an unkind act of player B by decreasing B's payoff. In sum, the piecewise-linear model does not only nest major distributional preferences, but it also quantifies the effects of positive and negative reciprocity.

2.2 Experimental design

This subsection describes the experimental design. The experiment consists of two sessions per subject that took place three months apart from each other, one in February and one in May 2010. To test for temporal stability, both sessions included the same set of binary decision situations that allow us to estimate the subjects' preference parameters.

In each binary decision situation, the subjects had to choose one of two payoff allocations between themselves and an anonymous player B. We implemented two types of such binary decision situations: (i) dictator games for identifying the parameters α and β, and (ii) reciprocity games for identifying γ and δ. In addition to these two types of binary decision situations, the second session in May 2010 comprised a series of trust games plus two reward and punishment games for checking the stability of the estimated preferences across games.

2.2.1 Dictator games

In each dictator game, a subject in player A's role can either increase or decrease player B's payoff by choosing one of two possible payoff allocations, $X = (\Pi_X^A, \Pi_X^B)$ or $Y = (\Pi_Y^A, \Pi_Y^B)$. To identify the subject's distributional preferences, governed by α and β, we varied the cost of changing the other player's payoff systematically across the dictator games.

--- FIGURE 1 ---

Figure 1 illustrates the dictator games' design. Each of the three circles represents a set of 13 dictator games in the payoff space. In each of these dictator games, a grey line connects the two possible payoff allocations, X and Y. The slope of the grey line therefore represents A's cost of altering B's payoff. For example, consider the decision between the two options marked in red color: The slope of the grey line is -1, implying that player A has to give up one point of her own payoff for each point she wants to increase player B's payoff. Hence, if A chooses the upper-left of these two allocations, we know that A's α is greater than 0.5, since the marginal utility from increasing B's payoff, α, needs to exceed the marginal disutility of doing so, $1 - \alpha$. If, in contrast, A opts for the lower-right allocation, then A's α is lower than 0.5. Thus, by systematically varying the costs of changing the other player's payoff across all dictator games – i.e. the slope of the grey line – we can infer A's marginal rate of substitution between her own and the other player's payoff. This allows us to directly identify the corresponding parameters of the subjects' distributional preferences, α and β.

The 45° line separates the dictator games in which A's payoff is always smaller than player B's from the ones in which A's payoff is always larger than B's. Thus, the observed choices in the upper (lower) circle allow us to estimate the value of α (β) in a situation of disadvantageous (advantageous) inequality. The choices in the middle circle contribute to the identification of both α and β, as each of them involves an allocation with disadvantageous inequality as well as an allocation with advantageous inequality.

We constructed the dictator games such that the identifiable range of the parameters is between -3 and 1. The bunching of the grey lines ensures that they yield the highest resolution around parameter values of zero that separate the different preference types.

2.2.2 Reciprocity games

In addition to the dictator games, each subject played 39 positive and 39 negative reciprocity games. The reciprocity games simply add a *kind or unkind prior move* by player B to the otherwise unchanged dictator games. In this prior move, B can either implement the allocation $Z = (\Pi_Z^A, \Pi_Z^B)$ or let the subject choose between the two allocations $X = (\Pi_X^A, \Pi_X^B)$ and $Y = (\Pi_Y^A, \Pi_Y^B)$. Letting the subject choose between X and Y instead of implementing Z is either a kind or an unkind act from the subject's point of view. Hence, if player B decides not to implement Z, the subject may reward or punish B in her subsequent choice between X and Y.

In the positive reciprocity games, player A is strictly better off in both allocations X and Y than in allocation Z, while B is worse off in at least one of the two allocations X and Y than in allocation Z. Consider the example with X = (1050, 270), Y = (690, 390), and Z = (550, 530). If player B forgoes allocation Z and lets A choose between the allocations X and Y, she acts kindly towards A as she

sacrifices some of her own payoff to increase A's payoff. Thus, if player A has a sufficiently strong preference for positive reciprocity, i.e. a positive and sufficiently large γ, she rewards B by choosing allocation Y instead of allocation X.

In the negative reciprocity games, player A is strictly worse off in both allocations X and Y than in allocation Z, while B is better off in at least one of the two allocations X and Y than in allocation Z. For example, consider the case where X = (450, 1020), Y = (210, 720), and Z = (590, 880). If B does not implement Z and forces A to choose between the allocations X and Y, she acts unkindly towards A as she decreases A's payoff for sure in exchange for the possibility of increasing her payoff from 880 to 1020. Hence, if A has a sufficiently strong preference for negative reciprocity, i.e. a negative and sufficiently small δ, she punishes B by opting for allocation Y instead of allocation X.

We applied the strategy method (Selten, 1967) in the reciprocity games to ask the subject how she would behave if player B gives up allocation Z, and forces her to choose between the allocations X and Y. Consequently, any behavioral differences in the choices among X and Y between the dictator games and the corresponding reciprocity games have to be due to reciprocity. Based on such behavioral differences we can identify the parameters γ and δ that reflect the subjects' preferences for positive and negative reciprocity.[7]

Taken together, we developed a design based on binary decision situations that are cognitively easy to grasp. We systematically vary the payoffs such that we are able to identify the parameters for the subjects' distributional preferences, α and β. Only small changes are necessary to extend the design such that we are additionally able to identify the reciprocity parameters, γ and δ.

2.3 Implementation in the lab

As already mentioned, we conducted two experimental sessions per subject that were three months apart. All subjects were recruited at the University of Zurich and the Swiss Federal Institute of Technology Zurich. 200 subjects participated in the first session in February 2010 (henceforth denoted Session 1) and were exposed to 117 binary decision situations involving a block of 39 dictator games (see section 2.2.1) and a block of 78 reciprocity games (see section 2.2.2) as well as a questionnaire soliciting cognitive ability, demographic data, and personality variables (i.e. the big five personality dimension). Out of these 200 subjects, 174 subjects (87%) showed up in the subsequent session that took place in May 2010 (henceforth denoted Session 2). In Session 2, the subjects completed again the 117 binary decision situations mentioned above. In addition, they played ten trust games plus the two reward and punishment games that are described in more detail in Section 4.5. We will use the preferences estimated from the dictator and reciprocity games to predict the behavior in the trust games and the reward and punishment games.

The dictator and reciprocity games were presented in blocks and appeared in random order across subjects. In the dictator games, the subjects faced a decision screen on which they had to choose between the two allocations X and Y. In the reciprocity games, the subjects initially saw allocation Z during a

[7] In this context, it is important to note that Brandts & Charness (2011) show that the strategy method typically finds qualitatively similar effects compared to the direct response method. Moreover, when we use the estimated preference parameters to predict behavior in other games we also apply the strategy method in these (other) games. Thus, we keep the mode of preference elicitation constant across the games in which reciprocity plays a role.

random interval of 3 to 5 seconds, before they had to indicate their choice between the allocation X and Y.[8]

In Session 1, after the subjects completed all dictator and reciprocity games, we additionally assessed the potential of the reciprocity games for triggering the sensation of having been treated kindly or unkindly by player B. To do so, we asked the subjects to indicate on a 5-point scale as how kind or unkind they perceived player B's action of forgoing allocation Z in a sample of 18 reciprocity games. The subjects' answers, available in table A.1 in the appendix, show that the reciprocity games have indeed succeeded in triggering the perception of having been treated kindly and unkindly by the other player B.

As payment, each subject received a show-up fee as well as an additional fixed payment for filling out the questionnaire on her personal data. After finishing the session, three of the subject's decisions as player A were randomly drawn for payment and each of them randomly matched to a partner's decision who acted as player B. Both the subject as well as her randomly matched partner received a payment according to their decisions. The experimental exchange rate was 1 CHF per 100 points displayed on the screen.[9] The average payoff in Session 1 was 52.50 CHF (std.dev. 7.47 CHF; minimum 33.30 CHF; maximum 74.10 CHF) and 55.74 CHF in Session 2 (std.dev. 7.50 CHF; minimum 28.60 CHF; maximum: 75.60 CHF). Both sessions lasted roughly 90 minutes. In Session 1 (2), the fraction of female subjects was 52% (53%) and the average age was 21.70 (21.75) years.

The subjects received detailed instructions. We examined and ensured their comprehension of the instructions with a control questionnaire. In particular, we individually looked at each subject's answers to the control questionnaire and handed it back in the (very rare) case of miscomprehension. Finally, all subjects answered the control questions correctly. They also knew that they played for real money with anonymous human interaction partners and that their decisions were treated in an anonymous way.

3 Econometric strategy

In this section, we first describe the random utility model in general which we apply for estimating the parameters of the behavioral model. Subsequently, we present three versions of the random utility model that vary in their flexibility in accounting for heterogeneity.

3.1 Random utility model

To estimate the parameters of the behavioral model, $\theta = (\alpha, \beta, \gamma, \delta)$, we apply McFadden's (1981) random utility model for discrete choices. We assume that player A's utility from choosing allocation $X_g = (\Pi^A_{Xg}, \Pi^B_{Xg}, r_{Xg}, s_{Xg}, q_{Xg}, v_{Xg})$ in game $g = 1, \dots, G$ is given by

$$\mathcal{U}^A(X_g; \theta, \sigma) = U^A(X_g; \theta) + \varepsilon_{Xg}, \tag{3}$$

[8] Screenshots of a dictator and a reciprocity game are included in figures A.1 and A.2 in the appendix.
[9] On February 1, 2010 the nominal exchange rate was 0.94 USD per CHF.

where $U^A(X_g;\theta)$ is the deterministic utility of allocation X_g, and ε_{Xg} is a random component representing noise in the utility evaluation. The random component ε_{Xg} follows a type 1 extreme value distribution with scale parameter $1/\sigma$. According to this model player A chooses allocation X_g over allocation Y_g if $\mathcal{U}^A(X_g;\theta,\sigma) \geq \mathcal{U}^A(Y_g;\theta,\sigma)$. Since utility has a random component, the probability that payer A's choice in game g, C_g, equals X_g is given by

$$\begin{aligned}\Pr\left(C_g = X_g;\ \theta,\sigma,X_g,Y_g\right) &= \Pr\left(U^A\left(X_g;\theta\right) - U^A\left(Y_g;\theta\right) \geq \varepsilon_{Yg} - \varepsilon_{Xg}\right) \\ &= \frac{\exp\left(\sigma\ U^A\left(X_g;\theta\right)\right)}{\exp\left(\sigma\ U^A\left(X_g;\theta\right)\right) + \exp\left(\sigma\ U^A\left(Y_g;\theta\right)\right)}.\end{aligned} \tag{4}$$

Note that the parameter σ governs the choice sensitivity towards differences in deterministic utility. If σ is 0 player A chooses each option with the same probability of 50% regardless of its deterministic utility. If σ is arbitrarily large the probability of choosing the option with the higher deterministic utility approaches 1.

A subject i's individual contribution to the conditional density of the model follows directly from the product of the above probabilities over all G games:

$$f(\theta,\sigma;X,Y,C_i) = \prod_{g=1}^{G} \Pr\left(C_{ig} = X_g;\ \theta,\sigma,X_g,Y_g\right)^{I(C_{ig}=X_g)}\ \Pr\left(C_{ig} = Y_g;\ \theta,\sigma,X_g,Y_g\right)^{1-I(C_{ig}=X_g)}, \tag{5}$$

where the indicator $I(C_{ig} = X_g)$ equals 1 if the subject chooses allocation X_g and 0 otherwise.

3.2 Aggregate estimation

The first version of the random utility model pools the data and estimates aggregate parameters, (θ,σ), that are representative for all subjects. These aggregate estimates represent the most parsimonious characterization of social preferences. They are useful mainly for comparisons with the existing literature, such as Charness & Rabin (2002) or Engelmann & Strobel (2004). However, since the aggregate estimates completely neglect heterogeneity they may fit the data only poorly.

3.3 Finite mixture estimation

The second version takes individual heterogeneity into account and estimates finite mixture models. Finite mixture models r enough to take the most important aspects of heterogeneity into account, namely the existence of distinct preference types. But on the other hand, they remain relatively parsimonious, as they require much less parameters than estimations at the individual level.

Finite mixture models assume that the population is made up by a finite number of K distinct preference types, each characterized by its own set of parameters, (θ_k,σ_k). This assumption of distinctly different preference types implies latent heterogeneity in the data, since each subject belongs to one of the K

types, but individual type-membership is not directly observable. Consequently, a given subject i's likelihood contribution depends on the whole parameter vector of the finite mixture model, $\Psi = (\theta_1, \dots, \theta_K, \sigma_1, \dots, \sigma_K, \pi_1, \dots, \pi_{K-1})$, and corresponds to

$$\ell(\Psi; X, Y, C_i) = \sum_{k=1}^{K} \pi_k \, f(\theta_k, \sigma_k; X, Y, C_i). \tag{6}$$

It equals the sum of all type-specific conditional densities, $f(\theta_k, \sigma_k; X, Y, C_i)$, weighted by the ex-ante probability, π_k, that subject i belongs to the corresponding preference type k. Since individual type-membership cannot be observed directly, the unknown probabilities π_k are ex-ante the same for all subjects and equal to the preference types' relative sizes. The parameter vector $\Psi = (\theta_1, \dots, \theta_K, \sigma_1, \dots, \sigma_K, \pi_1, \dots, \pi_{K-1})$ consists of K type-specific sets of parameters reflecting the types' preferences and choice sensitivities as well as $K - 1$ parameters governing the types' relative sizes. Thus, estimating a finite mixture model results in a parsimonious characterization of the K types by their type-specific behaviors and relative sizes.

Once we estimated the parameters of the finite mixture model, we can endogenously classify each subject into the preference type that best describes her behavior. Given the fitted parameters, $\widehat{\Psi}$, any subject i's ex-post probabilities of individual type-membership,

$$\tau_{ik} = \frac{\hat{\pi}_k f(\hat{\theta}_k, \hat{\sigma}_k; X, Y, C_i)}{\sum_{m=1}^{K} \hat{\pi}_m f(\hat{\theta}_m, \hat{\sigma}_m; X, Y, C_i)}, \tag{7}$$

follows from Bayes' rule. These ex-post probabilities of individual type-membership directly yield the preference type the subject most likely stems from.

An important aspect of estimating a finite mixture model is to find the appropriate number of preference types K that represent a compromise between flexibility and parsimony. If K is too small, the model lacks the flexibility to cope with the heterogeneity in the data and may disregard minority types. If K is too large, on the other hand, the model is overspecified and tries to capture types that do not exist. Such an overspecified model results in considerable overlap between the estimated preference types and an ambiguous classification of subjects into types. In either case, the stability and predictive power of the model's estimates are likely compromised.

Unfortunately, there is no general single best strategy for determining the optimal number of types in a finite mixture model. Due to the non-linearity of any finite mixture model's likelihood function there exists no statistical test for determining K that exhibits a test statistic with a known distribution (McLachlan, 2000)[10]. Furthermore, classical model selection criteria, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), are known to perform badly in the context of finite mixture models. The AIC is order inconsistent and therefore tends to overestimate the optimal

[10] Lo et al. (2001) proposed a statistical test (LMR-test) to select among finite mixture models with varying numbers of types, which is based on Vuong (1989)'s test for non-nested models. However, the LMR-test is unlikely to be suitable when the alternative model has non-normal outcomes Muthen (2003).

number of types (Atkinson, 1981; Geweke & Meese, 1981; Celeux & Soromenho, 1996). The BIC is consistent under suitable regularity conditions, but still shows weak performance in simulations when being applied as a tool for determining K (Biernacki et al., 2000).

But in any case, the classification of subjects into preference types should be unambiguous in the sense that τ_{ik} is either close to zero or close to 1, and the estimated type-specific parameters should be stable over time. We apply the normalized entropy criterion (NEC) to summarize the ambiguity in the individual classification of subjects into preference types (Celeux & Soromenho, 1996; Biernacki et al., 1999). The NEC allows us to select the finite mixture model with $K > 1$ types that yields the cleanest possible classification of subjects into types relative to its fit. The NEC for K preference types,

$$NEC(K) = \frac{E(K)}{L(K) - L(1)}, \tag{8}$$

is based on the entropy,

$$E(K) = -\sum_{k=1}^{K}\sum_{i=1}^{N} \tau_{ik} \ln \tau_{ik} \geq 0, \tag{9}$$

normalized by the difference in the log likelihood between the finite mixture model with K types, $L(K)$, and the aggregate model, $L(1)$. The entropy, $E(K)$, quantifies the ambiguity in the ex-post probabilities of type-membership, τ_{ik}. If all τ_{ik} are either close to 1 or close to 0, meaning that each subject is classified unambiguously into exactly one behavioral type, $E(K)$ is close to 0. But if many τ_{ik} are close to $1/K$, indicating that many subjects cannot be cleanly assigned to one type, $E(K)$ is large.

Consequently, we opt for the number of preference types K that minimizes the NEC and yields the cleanest segregation of subjects into types relative to the finite mixture model's fit. Subsequently, we examine whether the estimated type-specific parameters θ_k are stable over time.

3.4 Individual estimations

Finally, the third version of the random utility model estimates the parameters, (θ_i, σ_i), separately for each individual. The resulting individual estimates reveal the full extent of behavioral heterogeneity in the data. However, they lack parsimony and likely suffer from small sample bias. Thus, we expect them to be less stable over time than the aggregate estimates and the finite mixture models' type-specific estimates. Moreover, a researcher interested in developing a parsimonious theoretical model with different social preference types may find it hard to infer the general behavioral patterns from a plethora of individual estimates.

4 Results

A key purpose of our study is to provide a characterization of the distribution of social preferences that is (i) parsimonious, (ii) captures the major qualitative regularities of the data, (iii) displays reasonable levels of stability over time, and (iv) is capable of predicting behavior out-of-sample in other games. To achieve this purpose we proceed as follows. First, we estimate the preference parameters of a representative agent and examine how well these parameters capture the various aspects of our data. Clearly, the representative agent model is the most parsimonious one but – as we will see below – it misses important characteristics of the behavioral patterns. Second, we estimate the parameters of a model that allows for a small number of types without imposing ex-ante restrictions on the qualitative properties of the types. Third, we estimate the preference parameters for each individual separately thus allowing that each individual is its own preference type.

We had to exclude 14 of the 174 subjects from the sample because they behaved very inconsistently. These 14 subjects switched several times between the allocations X and Y within a given circle of the experimental design. In other words, they reversed their preferences for the other player's payoff several times when the cost of doing so rose monotonically. Consequently, it is not possible to estimate the individual preferences of these 14 subjects. Their estimated choice sensitivity $\hat{\sigma}$ is close to 0, indicating an abysmal fit of the empirical model. With $\hat{\sigma}$ almost 0, the preference parameters are no longer identified and at least one of their estimates lies outside the identifiable range of -3 to 1. Hence, we dropped these 14 subjects and report all following results for the remaining 160 subjects.

4.1 Preferences of the representative agent

Table 1 presents the parameter estimates $(\widehat{\theta}, \hat{\sigma})$ of the aggregate model that are representative for all subjects. The estimates indicate that the distributional preference parameters (α and β) are important for aggregate behavior. Both in Sessions 1 and 2 the representative agent values the payoff of others positively ($\hat{\alpha} > 0$ and $\hat{\beta} > 0$) regardless of whether the other player is better or worse off. However, the valuation of the other player's payoff is much higher when ahead than when behind, implying that the representative agent displays asymmetric altruism. More, specifically, in Session 1 (2), the estimate of α equals 0.083 (0.098) while the estimated value of β is much bigger and amounts to 0.261 (0.245). Thus, the weight of the other player's payoff is almost three times as high in situations of advantageous inequality than in situations of disadvantageous inequality (z-tests with $H_0: \alpha = \beta$ yield a p-value < 0.001 in both sessions). In terms of the willingness to pay, these numbers imply that the representative agent is willing to pay approximately 33 Cents to increase the other player's payoff by \$1 when ahead while when behind he is only willing to pay approximately 10.5 Cents.[11]

--- TABLE 1 ---

[11] The willingness to pay for a \$1 increase in the other player's payoff when ahead is given by $\beta/(1-\beta)$; when behind this willingness is given by $\alpha/(1-\alpha)$. With a value of $\beta = 0.25$, which is in the confidence interval of the preference estimates for both sessions, a subject is willing to pay 33 Cents to increase the other's payoff by \$1 when ahead. With a value of $\alpha = 0.095$, which is contained in the confidence interval for both sessions, a subject is willing to pay 10.5 Cents to increase the other's payoff by \$1 when behind.

The estimates of the reciprocity parameters γ and δ imply that the subjects' preferences are on average somewhat reciprocal. Kind acts increase the weight of the other player's payoff ($\hat{\gamma} > 0$), while unkind acts decrease the weight of the other player's payoff ($\hat{\delta} < 0$). However, the magnitude of the estimated reciprocity parameters is small, suggesting that both positive and negative reciprocity play a less important role than distributional preferences. Moreover, although there seems to be a consensus in the literature that negative reciprocity is more important than positive reciprocity[12] the preference estimates of the representative agent model do not support this. In fact, the parameter for positive reciprocity is even higher than the one for negative reciprocity (z-tests with $H_0: \gamma = \delta$ yield a p-value < 0.001 in both sessions).

The last column of table 1 shows that aggregate behavior is rather stable over time. The parameter estimates of α, β and δ are clearly not significantly different between the Sessions 1 and 2. Only the estimates for positive reciprocity, $\hat{\gamma}$, and the choice sensitivity, $\hat{\sigma}$, differ significantly between the two sessions. The significant decline in $\hat{\gamma}$ across sessions suggests that positive reciprocity is a more fragile preference component compared to the other components.

Note that this instability of the reciprocity parameter cannot be attributed to attrition bias because the estimates in Table 1 are based on the behavior of the *same* subjects in the two sessions. In addition, we find no evidence for attrition bias. The Session 1 estimates in the sample of all subjects who participated in that session are statistically indistinguishable from the Session 1 estimates in the subsample of the 160 subjects who participated in both sessions (see Table A.2 in the appendix).

How well do the preference parameters of the representative agent predict the aggregate data? In Figure 2, the solid lines represent the subjects' empirical behavior in Session 1 while the dashed lines correspond to their predicted behavior.[13] The panels on the left and right show the share of subjects willing to increase and decrease the other player's payoff, respectively. The upper panels describe the behavior of the subjects when behind, which is predicted (jointly with $\hat{\sigma}$) by the estimated preference parameter $\hat{\alpha}$, while the lower panels describe the behavior of subjects when ahead, which is predicted (jointly with $\hat{\sigma}$) by the estimated parameter $\hat{\beta}$.

--- FIGURE 2 ---

At first glance the aggregate model fits the data well, as the empirical and predicted shares of subjects willing to change the other's payoff almost coincide. In particular, the lower-left and upper-left panels show that the share of subjects increasing the other's payoff is higher in situations of advantageous than disadvantageous inequality. For example, at a cost of 0.39, more than 40 % of the subjects are willing to increase the other player's payoff when ahead but less than 20% are willing to do so when behind. Hence, $\hat{\beta} > \hat{\alpha}$ in the aggregate model.

There are, however, important behavioral regularities that the representative agent model fails to explain. The right panels of Figure 2 indicate that there exists a minority of subjects who decrease the other player's payoff even at a cost, especially when they are behind. If all individuals would have qualitatively similar preferences as the representative agent, i.e., if all of them had a positive valuation

[12] See, e.g., Charness & Rabin (2002), Offerman (2002) and Al-Ubaydli & Lee (2009).
[13] Figure A.3 in the appendix depicts Figure 2's analogue for session 2, which is very similar.

of others' payoff ($\alpha > 0$ and $\beta > 0$) there should be nobody who decreases the other player's payoff. In fact, however, the right panels show that up to 20% of the subjects decrease other's payoff. The aggregate model "neglects" these subjects in the sense that it assigns a positive α and a positive β to the representative agent because the share of subjects who increase the other's payoff at a given cost level is larger than the share of subjects that decreases the other's payoff (compare right to left panels).

However, understanding the behavior of subjects that decrease the other player's payoff can be crucial for predicting aggregate outcomes even if these subjects constitute only a minority: For example, in ultimatum games or public goods games with punishment, even a minority of subjects who are willing to reject unfair offers or punish freeriding can discipline a majority of selfish players and entirely determine the aggregate outcome (Fehr & Schmidt, 1999). But the aggregate model absorbs the behavior of these subjects in the random utility component, as it is not flexible enough to take minorities of subjects into account whose preference parameters systematically differ from those of the majority.

4.2 A parsimonious model of preference types

In view of the relevance of the existence of minority types for aggregate outcomes it is important to be able to characterize the heterogeneity of preferences of suitably defined sub-populations. In addition, we need to be able to characterize the preferences of these subgroups because this provides insights into their potential role in social interactions. For example, it is important to know whether a subgroup values the payoffs of others generally negatively – which would define them as spiteful types – or whether they only value the payoffs of others negatively when behind or treated unkindly.[14] However, the a priori definition of subgroups or preference types is always associated with some arbitrariness and the danger that the pre-defined groups or preferences characteristics of the group do not do justice to the data. Therefore, we apply an approach that *simultaneously* identifies (i) the preference characteristics of each type, (ii) the relative share of each type in the population, and (iii) the assignment of each individual to one of the preference types.

The finite mixture approach we use in this section fits this bill. To apply this approach we need to specify a priori the number of distinct preference types we consider. To obtain a compromise between flexibility and parsimony, we choose the number of preference types, K, based on the normalized entropy criterion (NEC, see section 3.3). Figure 3 shows the NEC's value for $K = 2$, $K = 3$, and $K = 4$ preference types. In both sessions, the NEC favors a finite mixture model with $K = 3$ preference types providing the cleanest assignment of subjects to types relative to its fit.

--- FIGURE 3 ---

Furthermore, when judging the appropriateness of the assumed number of types, we also examine below whether qualitatively new types emerge if one increases K or whether an increase in K is just associated with splitting up a given type while maintaining the sign of the various preference parameters. Finally, a further desirable feature when judging the appropriateness of the assumed number of types is that the preference characteristics of the different types should be relatively stable across time.

[14] Fehr et al. (2008) provide, for example, evidence that members of higher castes in India seem to have more frequently spiteful preferences. A generally negative valuation of others' payoff may have very different implications for, e.g., contract design and other institutional design questions compared to negative reciprocity.

Table 2 reports the results of our finite mixture estimates for both sessions. As in the case of the representative agent, the estimates of the parameters that capture outcome-based distributional preferences ($\hat{\alpha}$ and $\hat{\beta}$) are generally much higher than the reciprocity parameters ($\hat{\gamma}$ and $\hat{\delta}$). For this reason, we characterize the different types according to their distributional preference parameters. The table shows the existence of (i) a *Moderately Altruistic* type (MA), (ii) a *Strongly Altruistic type* (SA) and (iii) of a *Behindness Averse* (BA) type. A remarkable feature of all three types is that they value the payoff of others' much more when they are ahead than when behind. For this reason, one may also speak of Moderate (asymmetric) Altruists and Strong (asymmetric) Altruists. Another remarkable feature of Table 2 is that a purely selfish type, that puts zero value on others' payoffs, does not exist. All types display positive or negative valuations of others' payoffs.[15]

--- TABLE 2 ---

The MA-type makes up roughly 50% of the population and puts positive but modest weight on the other player's payoff, regardless of whether they are ahead or behind. This type also displays basically no positive reciprocity but moderate levels of negative reciprocity. The distributional preferences of the MA-type (inferred from Session 1) implies that members of this group are on average willing to spend 15 Cents to increase the other player's payoffs by $1 when ahead and 7 Cents when behind. Thus, the MA-types are willing to behave altruistically when the cost is relatively low.

The SA-type roughly comprises between 35% and 40% of the population. Subjects in this group display a valuation of the other player's payoff that is two to three times larger than that of the MA-type. The Strong Altruists also show relatively high levels of positive reciprocity and somewhat lower levels of negative reciprocity. Based on their distributional preferences (in Session 1) the SA-type is willing to spend 86 Cents to increase other player's payoff by $1 when ahead and 19 Cents when behind. Moreover, if a strong altruist has been treated kindly, such that the positive reciprocity parameter becomes relevant, the willingness to increase the other's payoff increases to 159 Cents when ahead and 45 Cents when behind.16 Thus, this group indeed displays rather strong social preferences.

Finally, the BA-type comprises roughly 10% of the population and weighs the other player's payoff negatively in situations of disadvantageous inequality. Interestingly, this type also tends to value others' payoffs negatively when ahead but the relevant preference parameter (β) is not significantly different from zero. This type also displays no significant preferences for positive or negative reciprocity. However, the behindness averse component of the BA-type is rather strong: they are on average willing to spend 78 Cents to decrease the other player's payoff by $1 when behind.

4.3 Stability and fit of the preference types

One desirable characteristic of a parsimonious distribution of types is that the preferences of the different types as well as their shares remain stable over time. We can address this issue by comparing the relevant parameter estimates between Session 1 and 2. Table 3 shows the p-values of z-tests for differences in the types' relative sizes, π_k, and preference parameters, θ_k, between Sessions 1 and 2. The estimates of

[15] Separate selfish types also do not emerge if we increase the number of types to $K = 4$ (see Table A.4) in the appendix). With four distinct types we further disaggregate the group of MA-types.
[16] These numbers are based on the preference parameters of Session 1. The numbers for Session 2 would differ slightly but the general thrust of the argument remains the same.

the finite mixture model with $K = 3$ types are remarkably stable over time, as none of the differences between Sessions 1 and 2 are statistically significant at the 5% level. In contrast, the parameter estimates of the finite mixture models with $K = 2$ and $K = 4$ types vary significantly over time, indicating that these models are misspecified. For example, in the two-type model all four preference parameters of the strongly altruistic type are unstable. Likewise, in the four-type model most preference parameters of the (strongly and moderately) altruistic types are unstable.

--- TABLE 3 ---

In addition, the model with $K = 2$ types lacks the flexibility to capture the minority of BA-types (see also Table A.3 in the appendix), while the model with $K = 4$ types overfits the data as it tries to isolate a second moderately altruistic type that is not stable over time (see also Table A.4 in the appendix). Hence, the finite mixture model with $K = 3$ preference types not only represents the best compromise between flexibility and parsimony but also yields the most temporally stable characterization of social preferences.

Figure 4 illustrates the temporal stability of the type-specific preference parameters in the (α, β)-space. The MA-type's parameter estimates are shown in green, the SA-type's in blue, and the BA-type's in red. Note that for all three types, even for the very precisely estimated MA- and BA-types, the 95% confidence intervals for the estimates of Sessions 1 and 2 overlap, indicating preference stability at the type level.

--- FIGURE 4 ---

The finite mixture method we used not only provides a characterization of the preferences of each type but also provides for each individual posterior probabilities of type-membership (see equation (7)). A good model assigns individuals unambiguously to one of the preference types in the sense that the probability of belonging to that type is close to 1. The upper panel in Figure 5 shows the distribution of the posterior probabilities of individual type-membership in Session 1. The histograms reveal that there are almost no interior probabilities of individual type-membership suggesting that almost all subjects are unambiguously assigned to one of the three types. Moreover, by comparing the lower panel – which shows the distribution of the posterior probabilities of individual type-membership in Session 2 – with the upper panel we observe that the distribution is very stable across time.

--- FIGURE 5 ---

To what extent do the preference parameters estimated in the $K = 3$ model predict the empirical behavior of each of the three types? Figure 6 provides the answer to this question. It displays the empirical and predicted type-specific share of subjects willing to change the other player's payoff at different cost levels in Session 1[17]. The empirical and predicted shares of subjects follow each other closely and pick up behavioral differences between the preference types which the aggregate model cannot explain due to its rigidity. The SA-types exhibit the highest willingness for increasing the other player's payoff, regardless of whether they are ahead or behind, and they are almost never willing reduce the other's payoff. The MA-types, on the other hand, only increase the other player's payoff if such an

[17] Figure A.4 in the appendix depicts Figure 6's analogue for Session 2 which is very similar.

increase is relatively cheap, and they are also almost never willing to decrease the other's payoff. Finally, a substantial share of BA-types opts for decreasing the other's payoff, while almost no BA-types are willing to increase the other's payoff.

--- FIGURE 6 ---

4.4 Individual preference estimates

In this subsection, we provide an overview of the individual-specific estimates of the random utility model. These estimates capture the full extent of behavioral heterogeneity as they characterize each subject by her own vector of parameters, $(\hat{\theta}_i, \hat{\sigma}_i)$. However, they also consume a lot of degrees of freedom and thus may suffer from small sample bias. Moreover, the plethora of individual parameter estimates is ill suited for developing parsimonious theoretical models of heterogeneous social preferences.

Table 4 summarizes the individual-specific estimates of the 160 subjects participating in both sessions. The summary statistics confirm that, on average, distributional preferences play a more important role than motives for reciprocity. In particular, the means and medians of $\hat{\alpha}_i$ and $\hat{\beta}_i$ are close to the aggregate estimates and indicate that subjects display on average asymmetric altruism. The means and medians of the reciprocity parameters, $\hat{\gamma}_i$ and $\hat{\delta}_i$, exhibit the same signs as their aggregate counterparts but tend to be smaller in absolute values. Moreover, the large standard deviations and ranges between the minima and maxima reveal that the individual estimates are highly dispersed.

--- TABLE 4 ---

4.5 The predictive power of preference estimates across games

In this section, we examine the overall predictive power of the types-specific preferences and the individual preferences estimated in Session 2 for two types of games – trust games as well as reward and punishment games. Both games are described in more detail below. We compare, in particular, the predictions that follow from the type-specific preference estimates of the finite mixture model with (i) predictions that are based exclusively on psychological and demographic variables such as personality traits, cognitive skills, age, gender, income, and field of study, and (ii) with predictions that are – in addition – based on individual-specific preference estimates. Because the finite mixture model also provides an assignment of each individual to one of the three types, and because we know the preference parameters of each type, the type-specific model also gives us predictions for each individual. Thus, the first comparison informs us whether and how much the preferences estimates of the three type-model (together with each individuals' assignment to one type) increases the power to predict individual behavior in other games. The second comparison tells us to what extent the inclusion of further individual-specific preference information improves the predictions over the finite mixture model's predictions.

In addition to predicting the behavioral variation across individuals we are in this section also interested in the extent to which the predicted behavioral variation across types is qualitatively similar to the actual variation. For example, in the trust game discussed below the SW-types are predicted to be more trustworthy than the WW-types. In the reward and punishment games both the WW-types and the SW-

types should only reward but never punish other players because their estimated negative reciprocity parameters are too small to overturn the positive weight they put on the other player's payoffs that follows from the outcome-based social preferences. Likewise, the estimated preferences of the behindness averse type imply that this type should not make any positive back-transfers in the trust game and should never reward the other player in the reward and punishment games because only her (envious) other-regarding preferences in the domain of *disadvantageous* inequality are significantly different from zero. Yet, rewarding a fair action in the reward and punishment games as well as reciprocating trust in the trust game requires putting a positive weight on other's payoff in the domain of *advantageous* inequality.

Examining the validity of these qualitative predictions is important. In particular, if the qualitative predictions are violated they inform us about potentially relevant behavioral factors that are not yet captured by our model. In addition, deviations from the qualitative predictions may provide hints about the instability of certain preference components or certain preference types which is also an important piece of information.

4.5.1 Predicting behavior in trust games

In the ten trust games, shown in Figure 7, player B can refrain from trusting, which yields a payoff of $(600, 600)$ or B can trust. In case that player B trusts, player A chooses whether she is trustworthy, yielding the payoffs $(1200 - c, 900)$, or not trustworthy, resulting in $(1200, 0)$, where c denotes the cost of being trustworthy. The cost of being trustworthy, c, increase over the ten trust games from 0 to 900 in equally sized steps. Player A was asked to indicate his choice for the case that player B chooses to trust. Because a trusting move by player B unambiguously increases A's payoff opportunities and thus constitutes an act of kindness, positive reciprocity, γ, can play a role. In addition, depending on the cost of trustworthiness, distributional preferences, α or β, can play a role as well. Therefore, to predict player A's choice for a given cost of being trustworthy, we apply the following behavioral model that captures the deterministic utilities of the two options, i.e.

$$U^A(trustworthy; \theta, c) = (1 - \alpha s - \beta r - \gamma) * (1200 - c) + (\alpha s + \beta r + \gamma) * 900, \tag{10}$$

and

$$U^A(not\ trustworthy; \theta) = (1 - \beta - \gamma) * 1200. \tag{11}$$

Next, we use the random utility model (3) to predict the probability that the subject is trustworthy,

$$Pr(trustworthy; \theta, \sigma, c) = \frac{\exp\left(\sigma\ U^A(trustworthy; \theta, c)\right)}{\exp(\sigma\ U^A(trustworthy; \theta, c)) + \exp\left(\sigma\ U^A(not\ trustworthy; \theta)\right)}. \tag{12}$$

Finally, for the predictions based on the finite mixture model, we evaluate (12) using the type-specific estimates and each individual's assignment to a type, $Pr\left(trustworthy; \hat{\theta}_k, \hat{\sigma}_k, c\right)$, while for the

individual predictions, we evaluate (12) using the individual-specific estimates, $Pr(trustworthy; \hat{\theta}_i, \hat{\sigma}_i, c)$.

--- FIGURE 7 ---

Table 5 shows the results of four OLS regressions of the subjects' empirical trustworthiness on their predicted probability of being trustworthy. In all regressions we control for the above mentioned psychological and demographic measures. The first two regressions show that the psychological and demographic measures alone explain only 6% of the empirical variance in trustworthiness while the type-specific estimates increase the explained variance to 35%. Moreover, the third regression indicates that only using the individual-specific estimates of all 160 subjects does not increase the explained variance. Finally, if we use both the type-specific estimates and the additional information contained in the individual-specific estimates – as indicated by the difference between $Pr(trustworthy; \hat{\theta}_i, \hat{\sigma}_i, c)$ and $Pr(trustworthy; \hat{\theta}_k, \hat{\sigma}_k, c)$ – we are able to explain 37% of the variance in trustworthiness (see column 4).[18] In this regression, the coefficient on the type-specific predictions implies that a 100% increase in the predicted trustworthiness increases the actual probability of trustworthiness by 65%.

--- TABLE 5 ---

The remarkable implication of Table 5 is that the individual-specific preference estimates do not lead to any major improvements in predictive power (in terms of explained variance), suggesting that the bulk of the relevant preference information is already contained in the type-specific estimates of the finite mixture model. In other words, for predictive purposes a parsimonious model with only 3 types is basically as good as a model with 160 types. One reason for this result could be that while the type-specific estimates tend to average out noise the individual-specific estimates may have the tendency to fit noise.

To what extent do the type-specific estimates of the finite mixture model not only predict the variation across individuals but also the levels of trustworthiness? Figure 8 provides an answer to this question. It documents the levels (including the 95% confidence intervals) and the variation in trustworthiness of the different types across the various cost levels. The figure shows that, as predicted by the type-specific estimates in Table 2, the SA-type is by far the most trustworthy type. This type reduces trustworthy responses only if their cost become so high that they imply the acceptance of relatively high levels of disadvantageous inequality. In addition, the figure shows a surprising behavior of the BA-type. According to the point estimates of this type's preference parameters, subjects in this category should never behave trustworthily when ahead but in fact they behave in the trust game almost like the MA-type: they are frequently making trustworthy choices when the costs are low to medium, indicating a substantial willingness to reciprocate trust. Thus, while the behavioral pattern of the SA- and the MA-

[18] The reader may wonder why it is possible that the subject-specific estimates used in regression 3 lead to a decrease in R^2 relative to regression 2 while if one uses both the subject-specific and the type-specific information (regression 4) there is an increase in R^2. Denote the total variance in the dependent variable by V, the variance explained by the type-specific prediction by V^k and the variance explained by the subject-specific estimates by V^i, with $V^k \subset V$, $V^i \subset V$, $V^k \neq V^i$. Then the variance explained by combining the subject-specific and the type-specific predictions, $V^k \cup V^i$, is larger than both V^k and V^i.

type is relatively well captured by the type-specific estimates, the behindness averse type's behavior is puzzling.

--- FIGURE 8 ---

In our view, this finding is interesting for the following two reasons. First, it highlights the instability of the behindness averse type's preferences and second it points towards a potentially relevant, yet omitted, factor in our structural model. The instability of this type's preferences is well captured by our estimated empirical model. Table 2 and Figure 4 show that the standard errors of the BA-type's preference parameters are much higher than those of the other types. A particularly striking example of this instability are the parameter estimates for positive reciprocity in Session 1: the BA-type displays the highest estimate of the reciprocity parameter among all three types but with a standard error that is almost 10 times larger than the SA-type's and almost 5 times larger than the MA-type's. The instability of the BA-type is also reflected in the assignment of individuals to this type. In Session 2, 84% of the individuals assigned to the SA-type have already been assigned to this type in Session 1; in contrast, only 56.5% of the individuals assigned to the BA-type in Session 2 have already been assigned to this type in Session 1.

This instability in preferences also suggests that the BA-type's behavior reacts sensitively to even relatively small contextual changes. One such contextual effect may be the extent to which the positive or negative intentions of the first-mover become salient across different games and situations. In the case of the trust game, for example, it is very transparent and clear that a trusting move by player B signals kind intentions and this clarity may have tilted behindness averse players towards a relatively strong reciprocation of trust. One way to include this sensitivity to contextual changes into a structural model of social preferences would be to have an extra parameter that explicitly captures and measures the perceived kindness of actions such that one is capable of estimating this parameter.

4.5.2 Predicting behavior in reward and punishment games

We conducted two reward and punishment games, RP1 and RP2. In these games, a subject in the role of player A decides on whether she wants to reward or punish player B for her previous choice. In RP1 player B can choose between $X = (X^A, X^B) = (600, 600)$ and $Y = (Y^A, Y^B) = (300, 900)$, i.e., in the first allocation B sacrifices money to increase A's payoff while the choice of the second allocation can be viewed as an unkind act that favors player B. In RP2 player B can choose between $X = (700, 500)$ and $Y = (500, 700)$; in this case the choice of the first allocation is clearly a kind act while choosing the second allocation constitutes a selfish (unkind) act by B. In both games we elicit player A's willingness to reward or punish player B for both of B's choices. Player A can pay 10, 20 or 30 to reward B, which increases B's payoff by 100, 200 or 300, respectively. But A can also pay 10, 20 or 30 to punish B, which decreases B's payoff by 100, 200 or 300, respectively. Finally, A may also decide to neither reward or punish B.[19]

To assess the quantitative predictive power of the type-specific and individual preference estimates in RP1 and RP2 we calculate how much a subject is willing to spend on rewarding or punishing player B in response to the kind choice, X, and the unkind choice, Y. Note that in both games the parameters for

[19] The experimental instructions used neutral wording, i.e., terms such as "punishment" or "reward" were avoided.

outcome-based and the reciprocity-based social preferences can play a role. Therefore, if the subject spends w on rewarding player B for choosing allocation $C \in \{X, Y\}$ her utility is

$$U^A(w;\theta) = (1 - \alpha s - \beta r - \gamma q - \delta v) * (C^A - w) + (\alpha s + \beta r + \gamma q + \delta v) * (C^B + 10w). \quad (13)$$

On the other hand, if the subject spends p on punishing player B her utility is

$$U^A(p;\theta) = (1 - \alpha s - \beta r - \gamma q - \delta v) * (C^A - p) + (\alpha s + \beta r + \gamma q + \delta v) * (C^B - 10p). \quad (14)$$

Finally, if the subject neither rewards nor punishes player B she obtains a utility of

$$U^A(0;\theta) = (1 - \alpha s - \beta r - \gamma q - \delta v) * C^A + (\alpha s + \beta r + \gamma q + \delta v) * C^B. \quad (15)$$

We apply the random utility model (3) to predict the probability of each reward and punishment level the subject can choose from based on the finite mixture model's assignment to a type and the type-specific estimates, $(\hat{\theta}_k, \hat{\sigma}_k)$, as well as the individual-specific estimates, $(\hat{\theta}_i, \hat{\sigma}_i)$. Subsequently, we use these probabilities for computing the expected reward/punishment levels which we then use as regressors to explain the actual reward/punishment levels.

Table 6 shows four OLS regressions of the actual on the expected reward and punishment levels. The first regression shows that psychological and demographic measures explain only 3.5% of the variance while the second regression indicates that our type-specific preference estimates increase the explained variance to 26.7%. Moreover, the third regression indicates that using the individual-specific estimates of the preference parameters of all 160 individuals even decreases the R^2 slightly. Finally, the explained variance rises to roughly 30% when we use both the type-specific prediction and the deviation of the individual-specific from the type-specific prediction, and both predictors are significant.

--- TABLE 6 ---

The results of Table 6 reinforce one of the main conclusions from Table 5 – the bulk of the relevant preference information is already contained in the type-specific estimates of the finite mixture model because the individual-specific preference estimates do not lead to any major improvements in predictive power (in terms of explained variance). In fact, the explained variance of prediction based on the individual-specific estimates (column 3) is even lower compared to the one based on the type-specific estimates.

To what extent is the predicted behavioral variation across types qualitatively similar to the actual variation? We can answer this question with the help of Figure 9 which shows player A's average reward and punishment behavior in RP1 and RP2. According to the type-specific parameter estimates the BA-type should only punish B's behavior if behind but never reward B. The red bars in Figure 9 show that this prediction is neatly born out by the data. In fact, the BA-type is clearly the most punitive among the three types. According to our estimates, the SA-type should be more willing to reward than the MA-type – a prediction that is also qualitatively met by the data. However, our type-specific estimates also imply that the SA- and the MA-types never punish the other player because their coefficient for negative

reciprocity is far too small to compensate the positive weight they put on the other player's payoff in the domain of disadvantageous inequality.

--- FIGURE 9 ---

In contrast to this prediction, we observe that the SA- and MA-type both punish player B for choosing the unkind allocation Y. In our view this behavior again points towards the instability of reciprocal behaviors across various games and situations. Because reciprocity is critically dependent on inferences about the kindness or hostility of the other players' actions, instability in kindness/hostility perceptions across games may give rise to instability in reciprocal behaviors. For example, in RP1 an offer of $(300, 900)$ is saliently unfair because the equal split of $(600, 600)$ is available; it seems plausible that such saliently unfair actions may magnify negative reciprocity concerns and lead to higher than predicted punishment relative to other games in which unfairness is less salient. From an empirical perspective, this points towards the necessity to identify the levels and the determinants of subjects' kindness and hostility perceptions, because if we get an empirical grip on these perceptions, we may be able to explain the variation in the strengths of reciprocity across games.

Taken together, this section yields several key insights. First, a parsimonious model with a few other-regarding types contains the bulk of the preference information that helps explain the behavioral variation across individuals. In fact, the predictions based on the type-specific estimates explain a substantial fraction of the behavioral variation across individuals, while the predictions based on the individual-specific estimates of all 160 subjects do not further improve the explained fraction of behavioral variation.

Second, the predicted behavioral variation across types is by and large qualitative met by the types' actual behavioral variation. In particular, in the reward and punishment games, the strength of reward is highest among individuals in the SA-type and lowest among those in the BA-type, while the willingness to punish is highest among individuals in the BA-type and lowest among those in the SA-type.

Third, however, in situations that likely render positive and negative reciprocity very salient our type-specific estimates fail to predict actually occurring rewarding and punishing behavior which points – in our view – towards a key problem in the empirical application of reciprocity models. Unless these models are accompanied with an explicit, *empirically implementable*, theory of kindness and hostility perceptions such that these perceptions can be quantitatively predicted, they will not be able to accurately predict the strength of reciprocity motives across games.

Finally, the findings illustrated in Figures 8 and 9 reinforce the conclusion that purely selfish behavior is sufficiently rare such that no independent selfish type emerges in a parsimonious model of social preference types. If the cost of other-regarding behavior becomes low most people generally seem to be willing to increase or decrease the other player's payoff to some degree. In Figure 8, for example, we find very high levels of trustworthiness even among the individuals in the MA- and the BA-type when costs are low. Likewise, there are significant levels of average punishment among all three preference types in the reward and punishment games.

5 Conclusions

The analysis in this paper combines a systematic experimental design with a flexible structural model to uncover the distribution and stability of social preferences. It yields several main conclusions. First, purely selfish behavior seems to be the exception rather than the rule, i.e. once the costs of altering the other player's payoff are sufficiently low, selfish behavior is very rare and the vast majority of individuals exhibits some sort of social preferences. Second, the distribution of social preferences can be characterized in a parsimonious way by three temporally stable preference types: a moderately altruistic type that makes up roughly 50% of the population, a strongly altruistic type that constitutes 40% of the population, and a behindness averse type that accounts for 10% of the population. Third, this parsimonious characterization of the distribution of social preferences is not only stable over time but also exhibits considerable power in making out-of-sample predictions across games. In particular, the type-specific preference estimates combined with the classification of subjects into types explain a substantial fraction of behavioral variation in additional games. In fact, they are virtually as good in making out-of-sample predictions across games as the individual preference estimates, suggesting that the individual preference estimates are noisy and may suffer from small sample bias. Finally, preferences for reciprocity seem to be less important and less stable than distributional preferences, and there is little evidence that preferences for negative reciprocity are stronger than preferences for positive reciprocity.

However, some caveats are in order here. The notion of preference types must be understood with respect to the specific subject pool and the piecewise-linear model we use in this paper. This has two important implications. First, the finding that reciprocal motives turn out to be relatively minor compared to distributional ones should not be generalized to other framings and institutional arrangements where reciprocity may play a much more prominent or even the dominant role. Second, the result that there are essentially no inequality averse types may be specific to the student subject pool and the dictator games we used here. For instance, Bellemare et al. (2008) found in a representative sample of the Dutch population that young and well educated subjects tend to be less inequality averse than the average. Moreover, in the dictator games, the subjects always had to choose between two unequal allocations and could never implement an equal payoff distribution between themselves and the other player. Thus, in a representative sample we may find more evidence for inequality aversion, especially if the subjects can also implement equal payoff distributions.

There are mainly three main avenues for future research. First, the methodology presented in this paper could be applied to representative subject pools to learn more about the distribution of social preferences in the general population or in specific institutional contexts. Second, as already outlined, our results may prove helpful for both developing theoretical models that explicitly incorporate social preference heterogeneity in a parsimonious way and setting up agent-based simulations that illuminate behavioral dynamics in various contexts. Finally, the experimental design and preference model could be extended to capture eventual nonlinearities in social preferences.

Acknowledgements

We are grateful for insightful comments from Charles Bellemare, Anna Conte, David Gill, Daniel Houser, Peter Moffatt, and from the participants of the research seminars at Universities of Lausanne, Auckland, Otago, Waikato, New South Wales, Monash University and the Queensland University of Technology, the EEA|ESEM meeting 2014 in Toulouse, the Experimetrix Workshop 2015 in Alicante, the European ESA-meeting 2015 in Heidelberg, and the CESifo Area Conference on Behavioral Economics 2015 in Munich. Any errors and/or omissions are solely our own. This study is part of the grant #152937 of the Swiss National Science Foundation (SNSF).

References

Al-Ubaydli, O. and M. Lee (2009): An experimental study of asymmetric reciprocity. Journal of Economic Behavior & Organization, 72, 738-749.

Andreoni, J. and J. H. Miller (2002): Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. Econometrica, 70, 2, 737-753.

Atkinson, A. (1981): Likelihood Ratios, Posterior Odds and Information Criteria. Journal of Econometrics, 16, 15–20.

Bandiera, O., I. Barankay and I. Rasul (2005): Social Preferences and the Response to Incentives: Evidence from Personnel Data. Quarterly Journal of Economics, 120, 917-962.

Bardsley, N. and P. Moffatt (2007): The experimetrics of public goods: inferring motivations from contributions. Theory and Decision, 62, 161-193.

Bellemare, C., S. Kröger and A. van Soest (2008): Measuring Inequity Aversion in a Heterogeneous Population using Experimental Decisions and Subjective Probabilities. Econometrica, 76, 815-839.

Bellemare, C., S. Kröger and A. van Soest (2011): Preferences, Intentions, and Expectations Violations: a Large-Scale Experiment with a Representative Subject Pool. Journal of Economic Behavior and Organization, 78, 349-365.

Benz, M. and S. Meier (2008): Do people behave in experiments as in the field? - evidence from donations. Experimental Economics, 11, 268-281.

Biernacki, C., G. Celeux and G. Govaert (1999): An Improvement of the NEC Criterion for Assessing the Number of Clusters in a Mixture Model. Pattern Recognition Letters, 20, 267–272.

Biernacki, C., G. Celeux and G. Govaert (2000): Assessing a Mixture Model for Clustering With the Integrated Completed Likelihood. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 719–725.

Blanco, M., D. Engelmann and H. Normann (2011): A within-subject analysis of other-regarding preferences. Games and Economic Behavior, 72, 321-338.

Bolton, G. and A. Ockenfels (2000): ERC: A theory of equity, reciprocity, and competition. American Economic Review, 90, 166-193.

Brandts, J. and G. Charness (2011): The strategy versus the direct-response method: a first survey of experimental comparisons. Experimental Economics, 14, 375-398.

Breitmoser, Y. (2013): Estimating social preferences in generalized dictator games. Economics Letters, 121, 192-197.

Bruhin, A., H. Fehr-Duda and T. Epper (2010): Risk and Rationality: Uncovering Heterogeneity in Probability Distortion. Econometrica, 78, 1375-1412.

Camerer, C. (2003): Behavioral Game Theory: Experiments on Strategic Interaction, Princeton: Princeton University Press.

Carlson, F., O. Johansson-Stenman and P. Nam (2014): Social preferences are stable over long periods of time. Journal of Public Economics, 117, 104-114.

Celeux, G. and G. Soromenho (1996): An Entropy Criterion for Assessing the Number of Clusters in a Mixture Model. Journal of Classification, 13, 195–212.

Charness, G. and M. Rabin (2002): Understanding Social Preferences with Simple Tests. Quarterly Journal of Economics, 117, 817-869.

Conte, A. and M. Levati (2014): Use of Data on Planned Contributions and Stated Beliefs in the Measurement of Social Preferences. Theory and Decision, 76, 201–223.

Conte, A. and P. Moffatt (2014): The Econometric Modelling of Social Preferences. Theory and Decision, 76, 119–145.

Dohmen, T., A. Falk, D. Huffman and U. Sunde (2008): Representative trust and reciprocity: prevalence and determinants. Economic Inquiry, 46, 1, 84-90.

Dohmen, T., A. Falk, D. Huffman and U. Sunde (2009): Homo Reciprocans: Survey Evidence on Behavioural Outcomes. The Economic Journal, 119, 592-612.

Dufwenberg, M. and G. Kirchsteiger (2004): A Theory of Sequential Reciprocity. Games and Economic Behavior, 47, 268-98.

Engelmann, D. and M. Strobel (2004): Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. American Economic Review, 94, 4, 857-869.

Engelmann, D. and M. Strobel (2010): Inequality Aversion and Reciprocity in Moonlighting Games. Games, 1, 459-477.

Erlei, M. (2008): Heterogeneous Social Preferences. Journal of Economic Behavior and Organization, 65, 436-457.

Falk, A., E. Fehr and U. Fischbacher (2008): Testing theories of fairness – Intentions matter. Games and Economic Behavior, 62, 287-303.

Falk, A. and U. Fischbacher (2006): A theory of reciprocity. Games and Economic Behavior, 54, 293-315.

Fehr, E. and U. Fischbacher (2002): Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. The Economic Journal, 112, C1-C33.

Fehr, E. and S. Gächter (2000): Fairness and retaliation: The economics of reciprocity. The Journal of Economic Perspectives, 14, 159‐181.

Fehr, E., K. Hoff and M. Kshetramade (2008): Spite and Development. American Economic Review, 98, 494-499.

Fehr, E. and A. Leibbrandt (2011): A Field Study on Cooperativeness and Impatience in the Tragedy of the Commons. Journal of Public Economics, 95, 1144-55.

Fehr, E. and K. Schmidt (1999): A theory of fairness, competition, and cooperation. Quarterly Journal of Economics, 114, 817-868.

Fisman, R., P. Jakiela, S. Kariv and D. Markovits (2015): The distributional preferences of an elite. Science, 349, aab0096.

Fisman, R., S. Kariv and D. Markovits (2007): Individual Preferences for Giving. American Economic Review, 97, 1858-1876.

Geweke, J. and R. Meese (1981): Estimating Regression Models of Finite but Unknown Order. International Economic Review, 22, 55–70.

Harrison, G. and L. Johnson (2006) Identifying altruism in the laboratory. In: R. Isaac, D. Davis (Eds.), Experiments Investigating Fundraising and Charitable Contributors, Research in Experimental Economics, 11, Emerald Group Publishing Limited, 177–223

Iriberri, N. and P. Rey-Biel (2011): The role of role uncertainty in modified dictator games, Experimental Economics, 14, 160-180.

Iriberri, N. and P. Rey-Biel (2013): Elicited Beliefs and Social Information in Modified Dictator Games: What Do Dictators Believe Other Dictators Do? Quantitative Economics, 4, 515-547.

Karlan, D. (2005): Using experimental economics to measure social capital and predict financial decisions. American Economic Review, 95, 1688-1699.

Kube, S., M. Marechal and C. Puppe (2012): The Currency of Reciprocity: Gift-Exchange at the Workplace. American Economic Review, 102, 1644-1662.

Kube, S., M. Marechal and C. Puppe (2013): Do Wage Cuts Damage Work Morale: Evidence From a Natural Field Experiment. Journal of the European Economic Association, 11, 853-870.

Laury, S. and L. Taylor (2008): Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good. Journal of Economic Behavior & Organization, 65, 9-29.

Levine, D. (1998): Modeling Altruism and Spitefulness in Experiments. Review of Economic Dynamics 1, 593-622.

Lo, Y., N. Mendell and D. Rubin (2001): Testing the Number of Components in a Normal Mixture. Biometrika, 88, 767-778.

McLachlan, G. and D. Peel (2000): Finite Mixture Models. Wiley Series in Probabilities and Statistics. New York: Wiley.

McFadden, D. (1981): Econometric Models for Probabilistic Choice. In: C. Manski, D. McFadden (eds.), Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge.

Muthen, B. (2003): Statistical and Substantive Checking in Growth Mixture Modeling: Comment on Bauer and Curran (2003). Psychological Methods, 8, 369-377.

Offerman, T. (2002): Hurting Hurts More Than Helping Helps. European Economic Review, 46, 1423-1437.

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

Roth, A. E. (1995): Bargaining Experiments, In: Handbook of Experimental Economics, John Kagel and Alvin E. Roth, editors, Princeton University Press, 253-348.

Selten, R. (1967): Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In: H. Sauermann (ed.), Beiträge zur experimentellen Wirtschaftsforschung, Tübingen: Mohr, 136-168.

Volk, S., C. Thöni and W. Ruigrok (2012): Temporal stability and psychological foundations of cooperation preferences. Journal of Economic Behavior and Organization, 81, 664–676.

Vuong, Q. (1989): Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses. Econometrica, 57, 307-333.

Tables

	Estimates of Session 1	*Estimates of Session 2*	*p-value of z-test with H_0: Session1=Session2*
α	0.083***	0.098***	0.468
	(0.015)	(0.013)	
β	0.261***	0.245***	0.551
	(0.019)	(0.019)	
γ	0.072***	0.029***	0.010
	(0.014)	(0.010)	
δ	-0.042***	-0.043***	0.918
	(0.011)	(0.008)	
σ	0.016***	0.019***	0.006
	(0.001)	(0.001)	
# of observations	18,720	18,720	
# of subjects	160	160	
Log Likelihood	-5,472.31	-4,540.74	

Individual cluster robust standard errors in parentheses.
*** significant at 1%; ** significant at 5%; * significant at 10%

Table 1: Preferences of the representative agent ($K = 1$) in Sessions 1 and 2.

	Strongly Altruistic Type	Moderately Altruistic Type	Behindness Averse Type
Session 1			
π	0.405***	0.474***	0.121***
	(0.047)	(0.042)	(0.039)
α	0.159***	0.065***	-0.437***
	(0.036)	(0.013)	(0.130)
β	0.463***	0.130***	-0.147
	(0.028)	(0.017)	(0.147)
γ	0.151***	-0.001	0.170
	(0.026)	(0.012)	(0.119)
δ	-0.053**	-0.027**	-0.077
	(0.025)	(0.012)	(0.162)
σ	0.018***	0.032***	0.008***
	(0.001)	(0.002)	(0.002)
Session 2			
π	0.356***	0.544***	0.100***
	(0.039)	(0.041)	(0.024)
α	0.193***	0.061***	-0.328***
	(0.019)	(0.009)	(0.073)
β	0.494***	0.095***	-0.048
	(0.020)	(0.012)	(0.053)
γ	0.099***	-0.005	-0.028
	(0.024)	(0.006)	(0.030)
δ	-0.082***	-0.019***	-0.015
	(0.018)	(0.007)	(0.035)
σ	0.019***	0.049***	0.015***
	(0.001)	(0.004)	(0.002)
# of observations (both sessions)		18,720	
# of subjects (both sessions)		160	
Log Likelihood in Session 1		-4,202.17	
Log Likelihood in Session 2		-3,166.32	

Individual cluster robust standard errors in parentheses.
*** significant at 1%; ** significant at 5%; significant at 10%

Table 2: Finite mixture estimations ($K = 3$) in Sessions 1 and 2.

	π	α	β	γ	δ
Finite Mixture Model with $K = 2$ Types					
Strongly Altruistic Type	0.067	0.000	0.000	0.000	0.011
Moderately Altruistic Type	0.067	0.109	0.000	0.000	0.092
Finite Mixture Model with $K = 3$ Types					
Strongly Altruistic Type	0.423	0.397	0.365	0.139	0.362
Moderately Altruistic Type	0.238	0.780	0.100	0.776	0.538
Behindness Averse Type	0.657	0.464	0.528	0.106	0.708
Finite Mixture Model with $K = 4$ Types					
Strongly Altruistic Type	0.472	0.002	0.000	0.000	0.288
Moderately Altruistic Type I	0.668	0.000	0.000	0.012	0.069
Moderately Altruistic Type II	0.133	0.088	0.011	0.984	0.559
Behindness Averse Type	0.948	0.329	0.285	0.096	0.658

Table 3: p-values of z-tests for differences in the types' relative sizes and preference parameters between Sessions 1 and 2.

	Median	Mean	Std. Dev.	Min.	Max.
Session 1					
α	0.054	0.018	0.285	-1.394	0.471
β	0.211	0.216	0.328	-1.977	0.998
γ	0.043	0.082	0.205	-0.366	0.783
δ	-0.010	-0.056	0.196	-1.106	0.598
σ	0.035	0.168	0.248	0.004	0.847
Session 2					
α	0.060	0.048	0.236	-1.636	0.401
β	0.169	0.225	0.248	-0.405	0.905
γ	0.000	0.030	0.166	-1.087	0.679
δ	-0.010	-0.045	0.119	-0.553	0.229
σ	0.069	0.269	0.278	0.007	0.886

Table 4: Summary statistics of individual estimations in Sessions 1 and 2.

OLS regression with dependent variable: trustworthy [0/1]				
Intercept	0.392	0.233	0.228	0.215
	(0.294)	(0.210)	(0.191)	(0.196)
Prediction based on type-specific estimates		0.607***		0.650***
		(0.033)		(0.033)
Prediction based on individual-specific estimates			0.577***	
			(0.031)	
Difference between predictions based on individual- and type-specific estimates				0.300***
				(0.053)
Additional control variables	yes	yes	yes	yes
# of observations	1,600	1,600	1,600	1,600
# of subjects	160	160	160	160
R^2	0.059	0.349	0.343	0.374

Additional control variables include: Big 5 personality traits, cognitive ability, age, gender, monthly income, and field of study. Individual cluster robust standard errors in parentheses.
*** significant at 1%; ** significant at 5%; * significant at 10%

Table 5: Predictive power of preference estimates in the trust games.

OLS regression with dependent variable: Reward / Punishment				
Intercept	59.976	-13.644	-31.029	-36.756
	(87.566)	(68.128)	(63.755)	(62.194)
Prediction based on type-specific estimates		1.123***		1.065***
		(0.089)		(0.084)
Prediction based on individual-specific estimates			0.637***	
			(0.052)	
Difference between predictions based on individual- and type-specific estimates				-0.348***
				(0.074)
Additional control variables	yes	yes	yes	yes
# of observations	640	640	640	640
# of subjects	160	160	160	160
R^2	0.035	0.267	0.251	0.302

Additional control variables include: Big 5 personality traits, cognitive ability, age, gender, monthly income, and field of study. Individual cluster robust standard errors in parentheses.
*** significant at 1%; ** significant at 5%; * significant at 10%.

Table 6: Predictive power of preference estimates in the reward and punishment games.

Figures

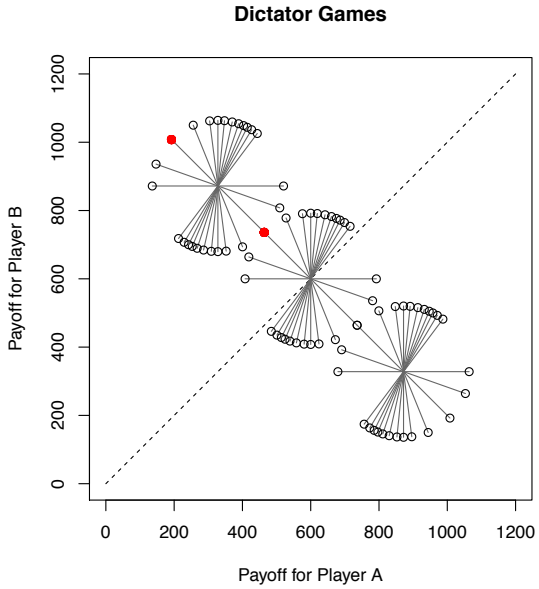


Figure 1: The dictator games.
(Upper circle: Disadvantageous inequality. Lower circle: Advantageous inequality.)

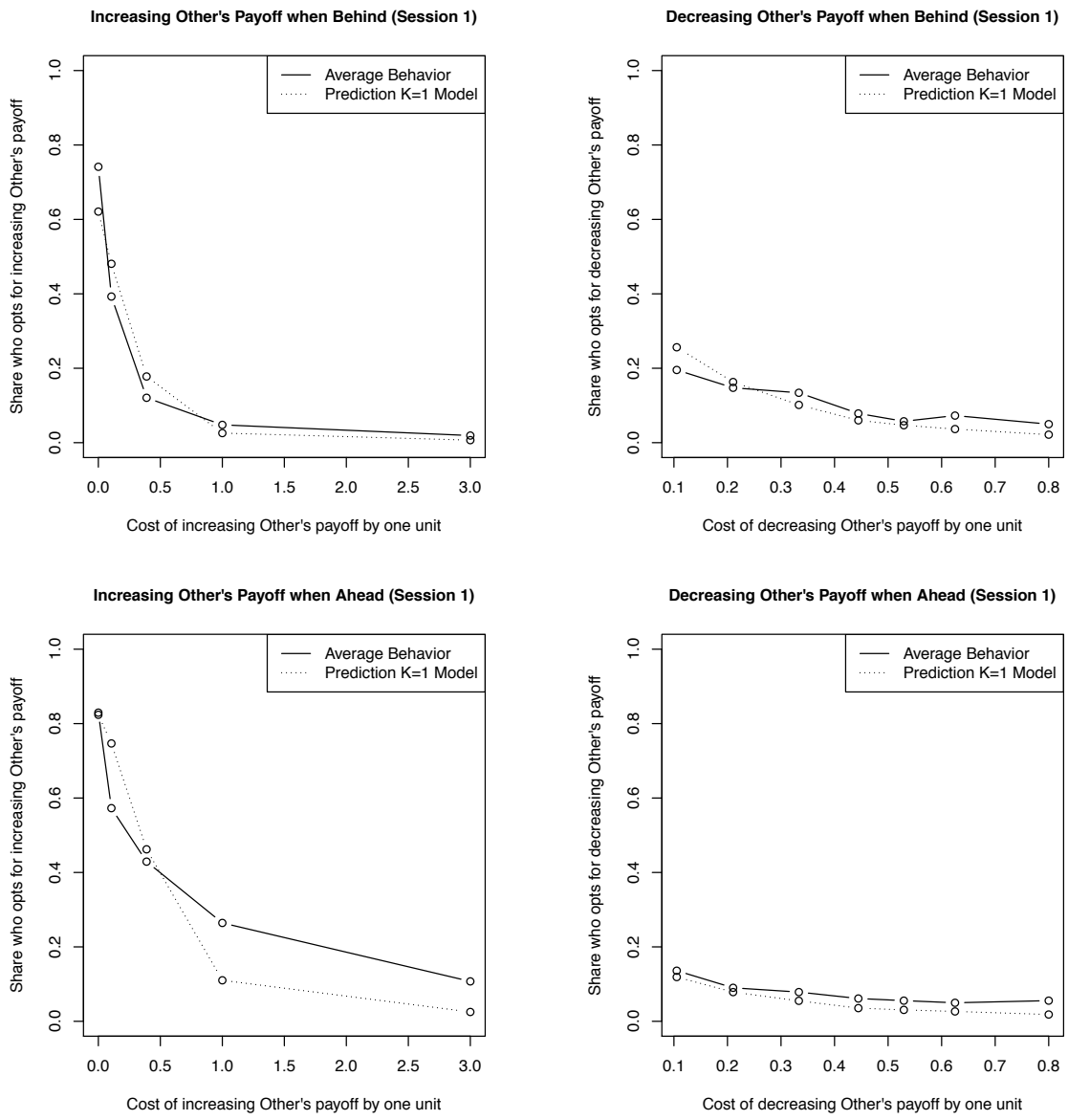


Figure 2: Empirical and predicted share of subjects willing to change the other player's payoff in Session 1.

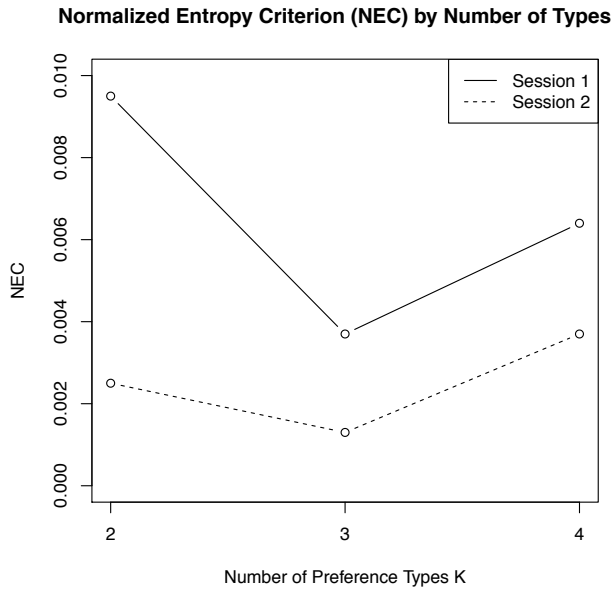


Figure 3: Normalized entropy criterion (NEC) for different numbers of preference types.

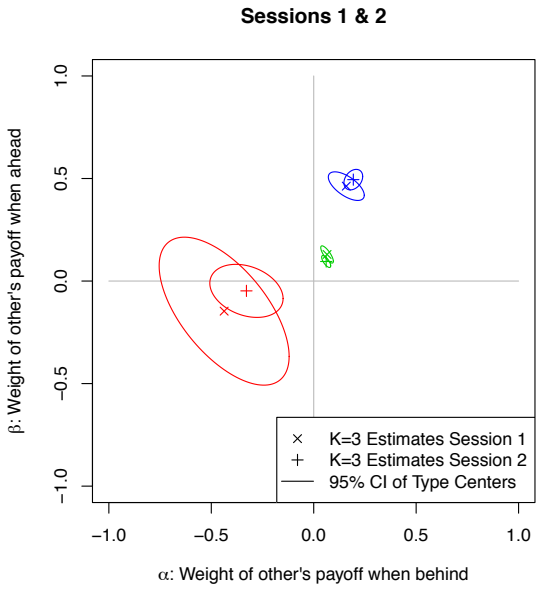


Figure 4: Temporal stability of preference estimates in the three-type model ($K = 3$).

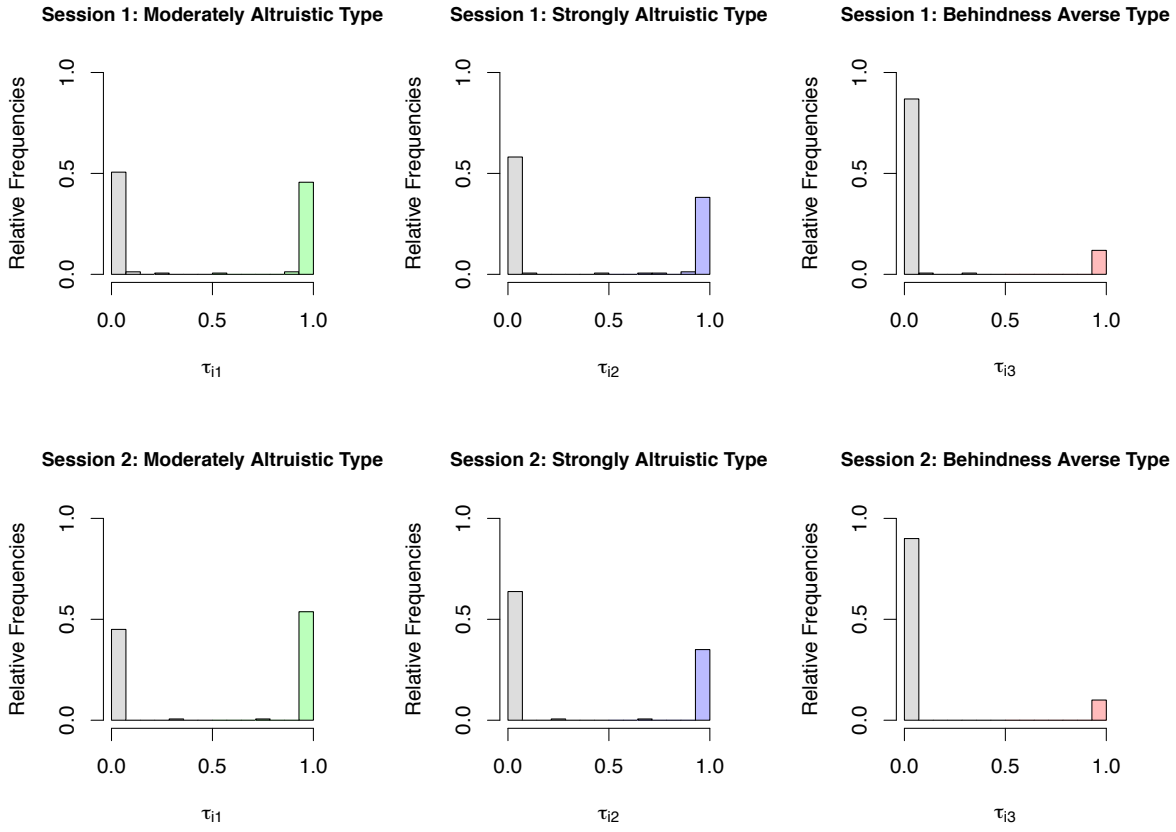


Figure 5: Distribution of posterior probabilities of individual type-membership in Sessions 1 (upper row) and 2 (lower row).

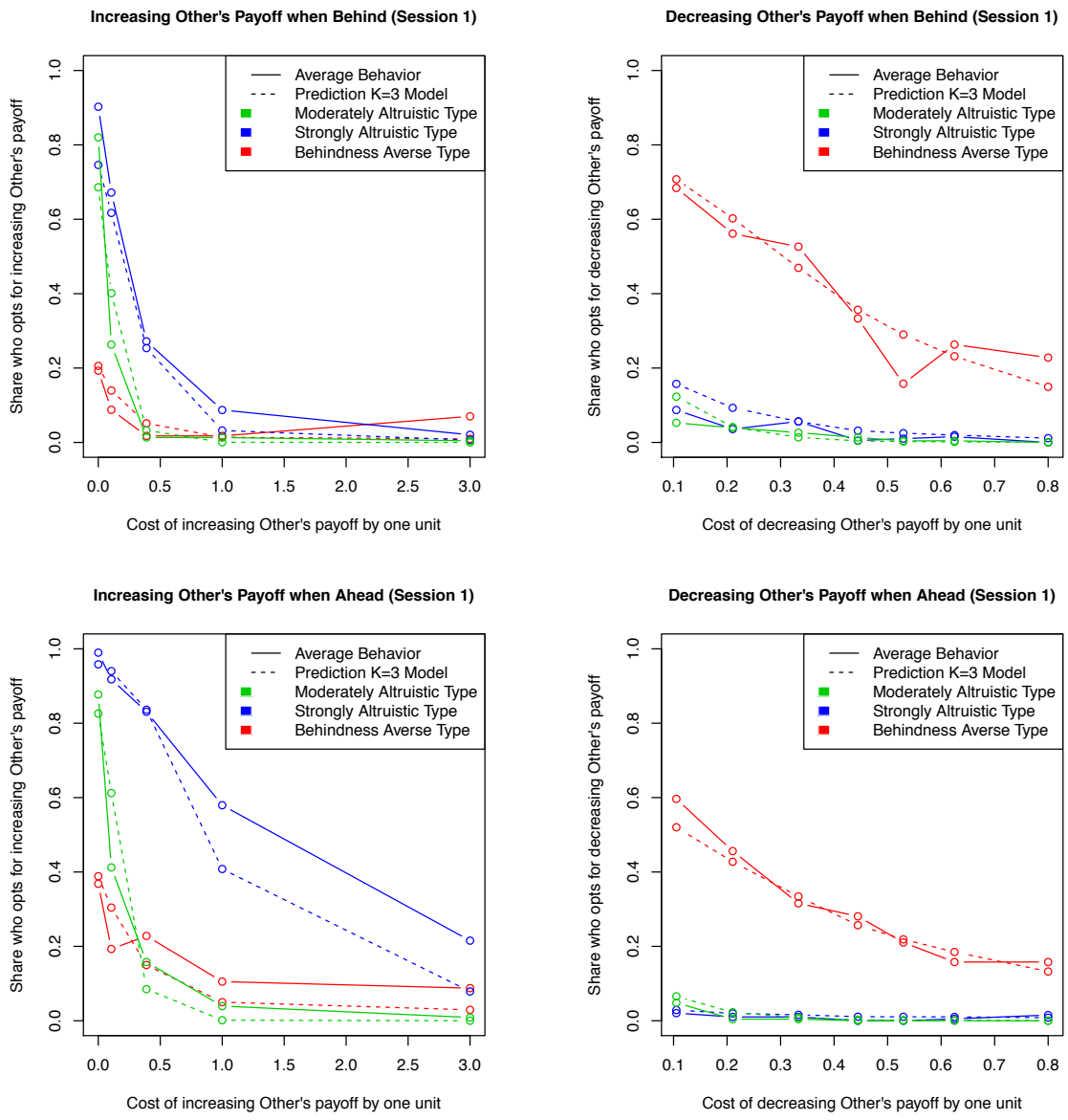
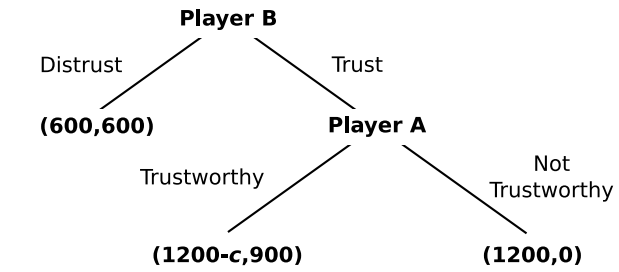


Figure 6: Empirical and predicted fraction of subjects willing to alter the other's payoff in Session 1.



Payoffs: (Π^A,Π^B)

Cost of Being Trustworthy: $c \in \{0, 100, 200, ..., 900\}$

Figure 7: The ten trust games with varying costs of being trustworthy.

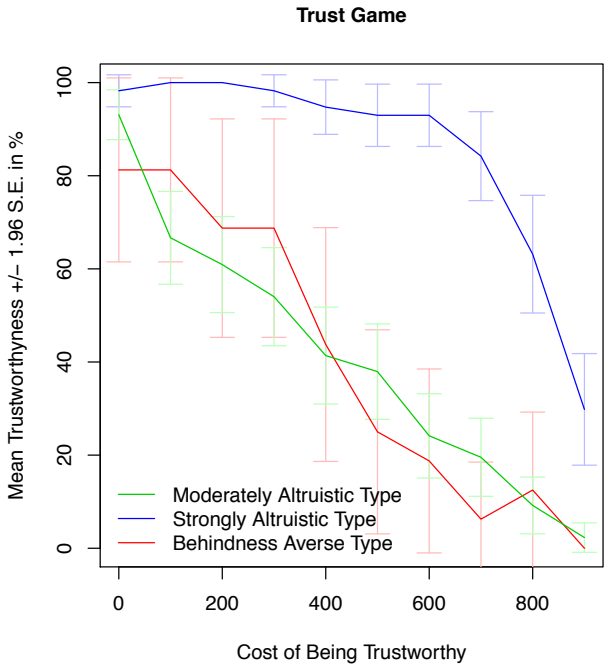


Figure 8: Different types' mean trustworthiness across cost levels (with 95% confidence intervals)

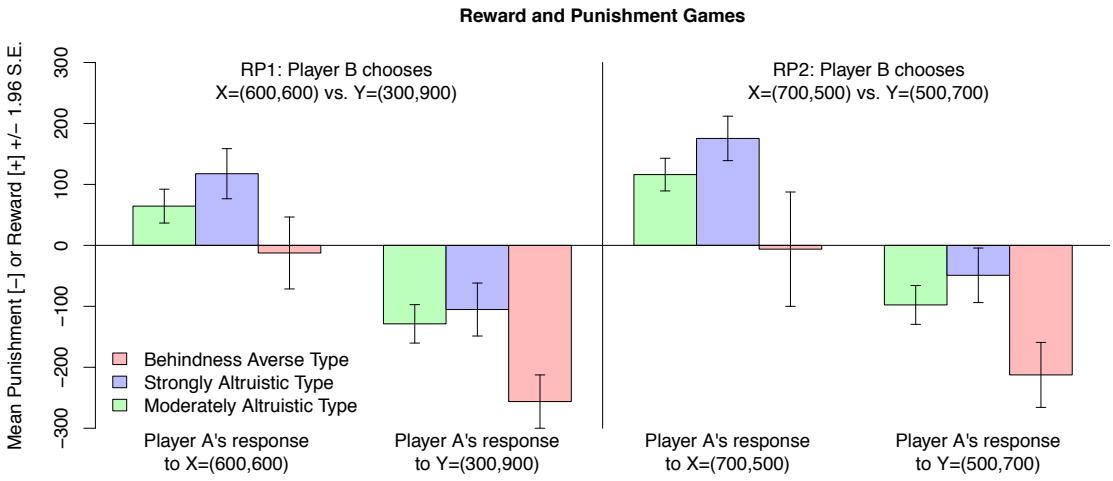


Figure 9: Player A's mean reward and punishment in response to player B's choice (with 95% confidence intervals).

Online Appendix

A1 Screenshots

Dictator Game (translated from German):

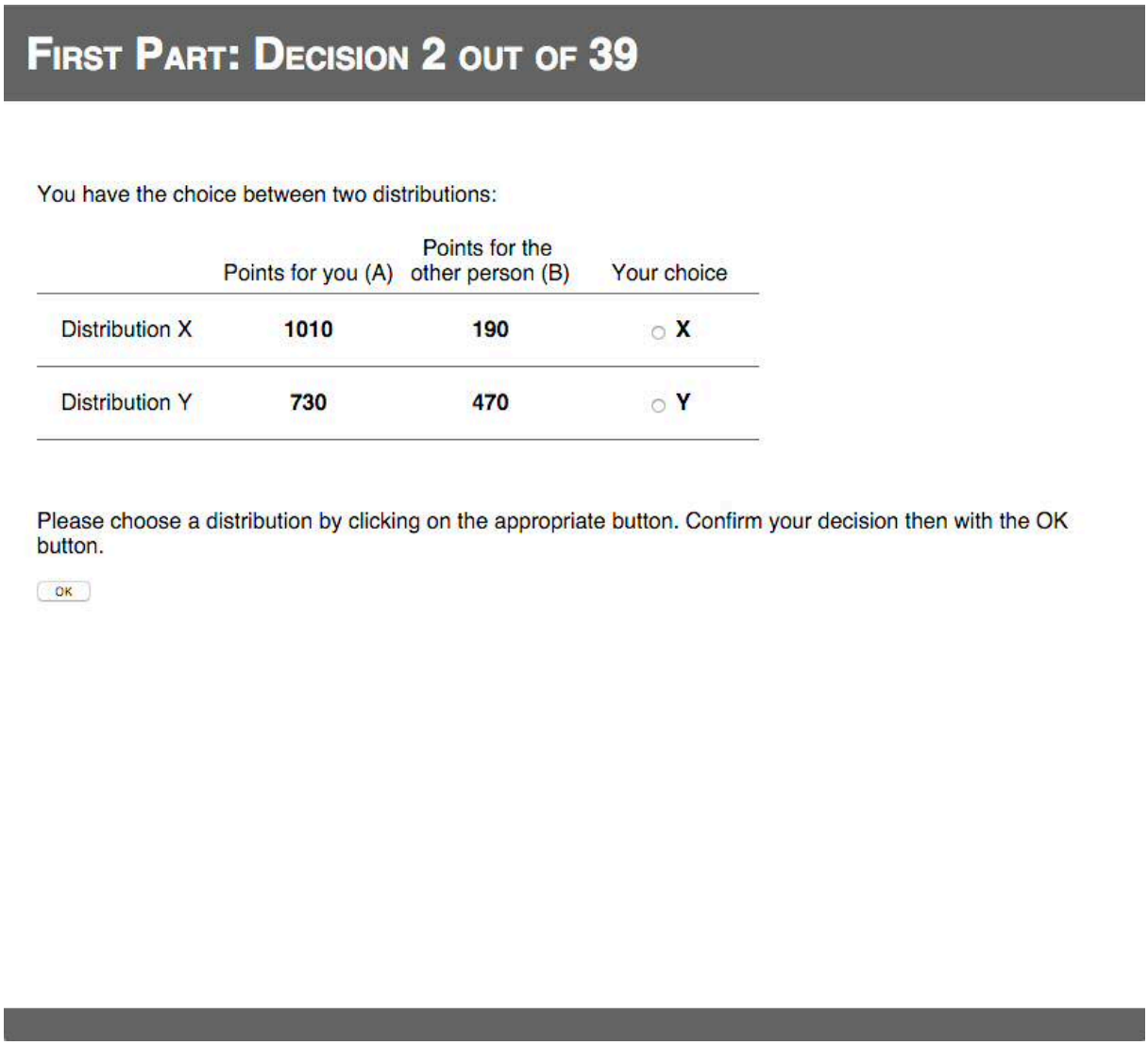


Figure A.1: Screenshot of a dictator game. Choice by player A.

Reciprocity Game (translated from German):

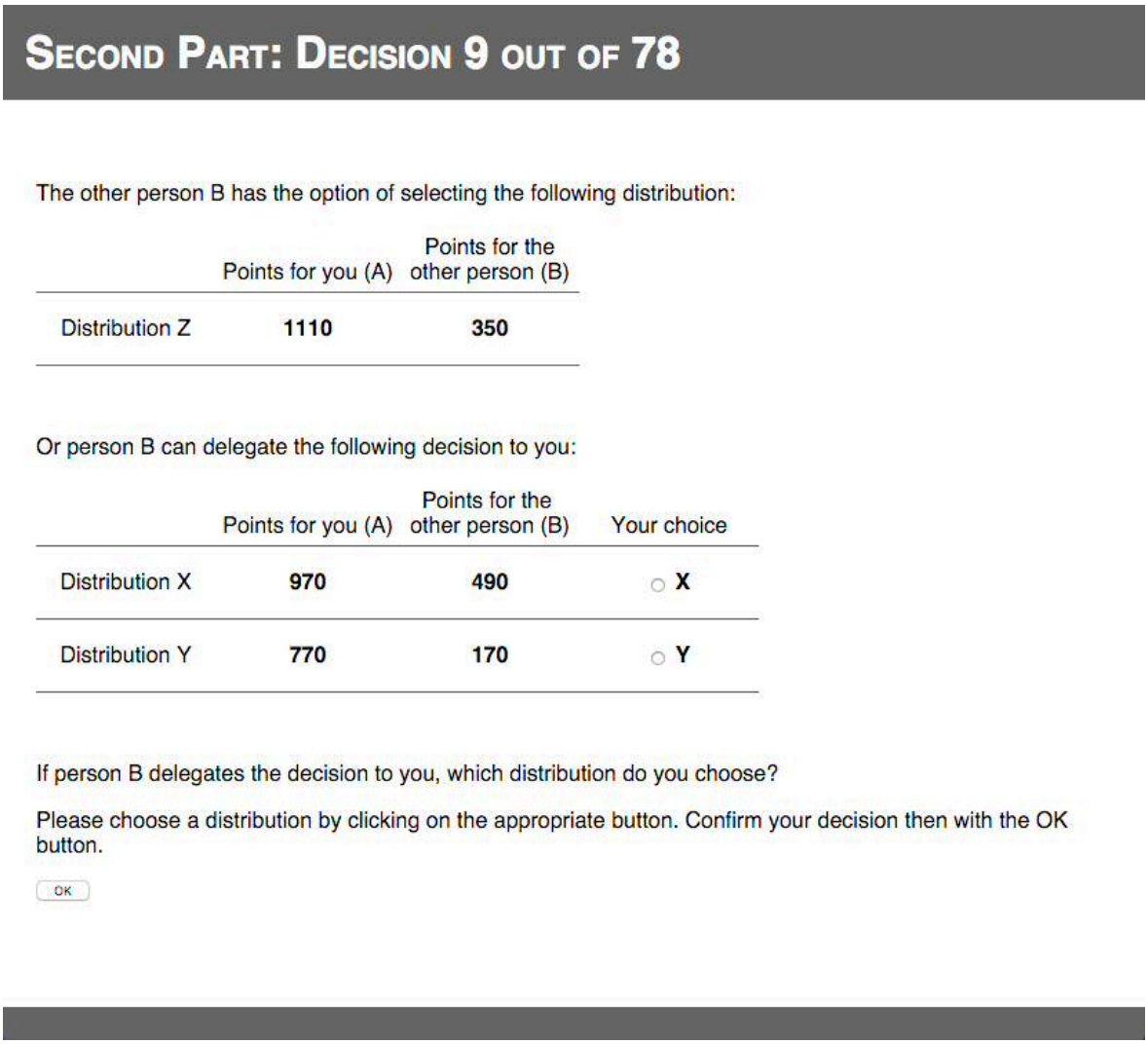


Figure A.2: Screenshot of a reciprocity game. Choice by player A.

A2 Potential of reciprocity games to trigger the sensation of having been treated kindly or or unkindly by the other player

Allocation X (Π_X^A, Π_X^B)	Allocation Y (Π_Y^A, Π_Y^B)	Allocation Z (Π_Z^A, Π_Z^B)	Average Kindness Rating	Std. Err.
(470, 730)	(190, 1010)	(610, 590)	2.087	0.065
(520, 870)	(140, 870)	(660, 730)	2.294	0.063
(450, 1020)	(210, 720)	(590, 880)	2.306	0.058
(790, 600)	(410, 600)	(930, 460)	2.375	0.061
(740, 460)	(460, 740)	(880, 320)	2.487	0.059
(720, 750)	(480, 450)	(860, 610)	2.669	0.061
(1060, 330)	(680, 330)	(1200, 190)	2.712	0.064
(990, 480)	(750, 180)	(1130, 340)	2.725	0.061
(1010, 190)	(730, 470)	(1150, 50)	2.831	0.060
(450, 1020)	(210, 720)	(70, 860)	3.825	0.063
(720, 750)	(480, 450)	(340, 590)	3.888	0.061
(990, 480)	(750, 180)	(610, 320)	3.888	0.059
(790, 600)	(410, 600)	(270, 740)	4.725	0.045
(1060, 330)	(680, 330)	(540, 470)	4.763	0.051
(470, 730)	(190, 1010)	(50, 1150)	4.763	0.044
(520, 870)	(140, 870)	(0, 1010)	4.794	0.050
(740, 460)	(460, 740)	(320, 880)	4.800	0.046
(1010, 190)	(730, 470)	(590, 610)	4.831	0.039

Table A.1: Kindness rating if player B forgoes allocation Z and leaves player A the choice between allocations X and Y. (1=very unkind; 5=very kind)

Table A.1 allows to check the potential of the reciprocity games for triggering reciprocal actions. It shows for a sample of 18 reciprocity games, how subjects in the role of player A on average rated player B's kindness when player B forgoes allocation Z and gives them the choice between allocations X and Y. Subjects had to rate player B's kindness on a 5-point scale form 1 (very unkind) to 5 (very kind

A3 No evidence of attrition bias

	Subjects participating in Session 1 and Session 2 (N=160)	*All Subjects participating in Session 1 (N=183)*	*p-value of z-test with H_0: Equal estimates in columns 1 and 2*
α	0.083***	0.076***	0.699
	(0.015)	(0.014)	
β	0.261***	0.261***	0.984
	(0.019)	(0.018)	
γ	0.072***	0.074***	0.916
	(0.014)	(0.012)	
δ	-0.042***	-0.036***	0.723
	(0.011)	(0.010)	
σ	0.016***	0.015***	0.862
	(0.001)	(0.001)	
# of obs.	18,720	21,411	
# of subj.	160	183	
Log Lik	-5,472.31	-6,332.84	

Individual cluster robust standard errors in parentheses.

*** significant at 1%; ** significant at 5%; * significant at 10%

Subjects with inconsistent choices and at least one estimated preference parameter outside the identifiable range of -3 to 1 are dropped.

Table A.2: No evidence of attrition bias.

A4 Share of subjects willing to change the other's payoff in Session 2 ($K = 1$ model)

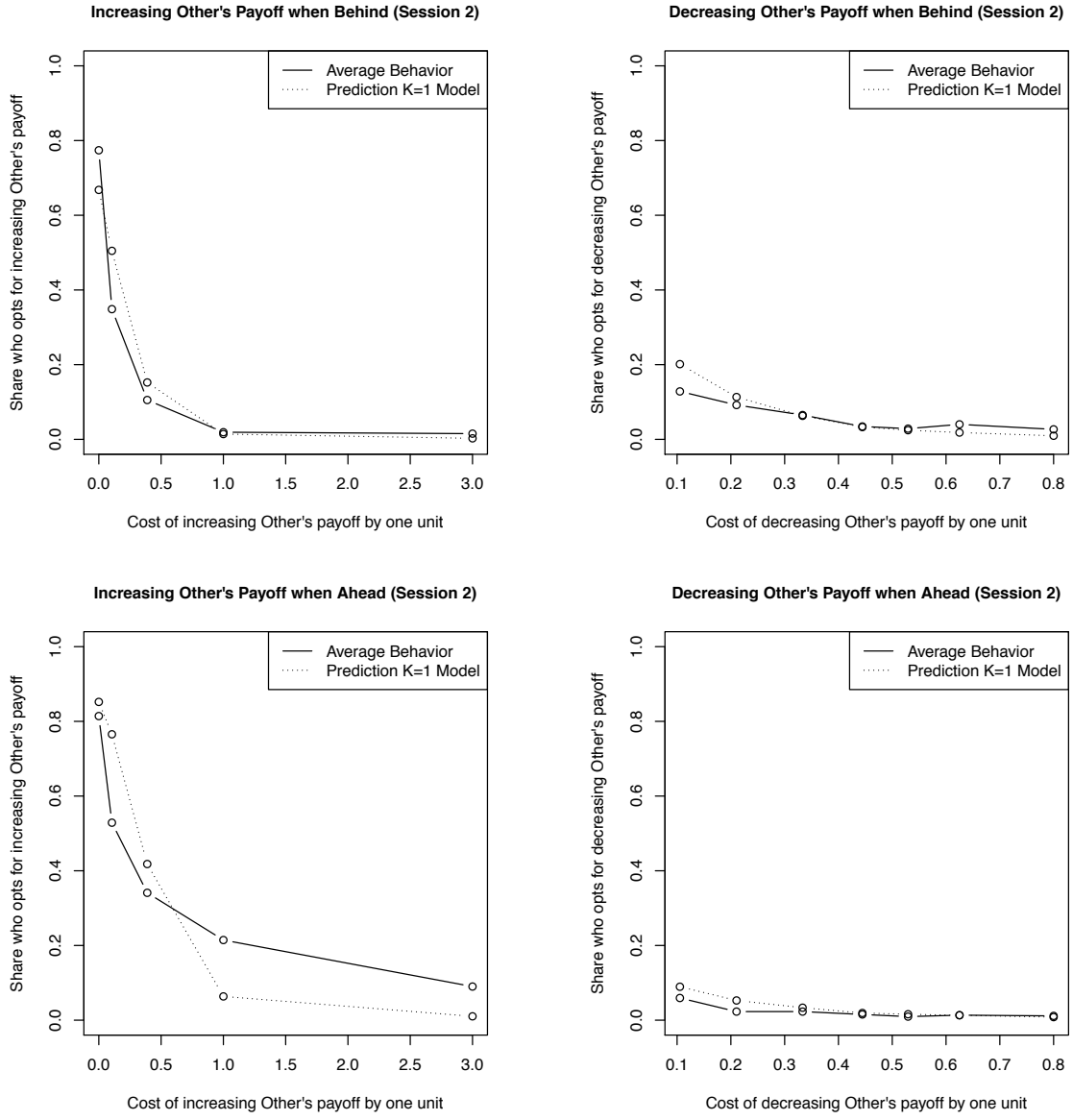


Figure A.3: Empirical and predicted share of subjects willing to change the other player's payoff in Session 2.

A5 Finite mixture model with $K = 2$ preference types

	Social Welfare Type I	Social Welfare Type II
Session 1		
π	0.477***	0.523***
	(0.049)	(0.049)
α	0.061***	0.085***
	(0.015)	(0.030)
β	0.122***	0.370***
	(0.023)	(0.027)
γ	0.000	0.141***
	(0.011)	(0.023)
δ	-0.026**	-0.055***
	(0.012)	(0.019)
σ	0.032***	0.012***
	(0.003)	(0.001)
Session 2		
π	0.638***	0.362***
	(0.039)	(0.039)
α	0.033**	0.188***
	(0.013)	(0.019)
β	0.089***	0.493***
	(0.012)	(0.020)
γ	-0.008	0.098***
	(0.007)	(0.023)
δ	-0.021***	-0.080***
	(0.007)	(0.018)
σ	0.028***	0.018***
	(0.003)	(0.001)
# of observations (both sessions)		18,720
# of subjects (both sessions)		160
Log Likelihood in Session 1		-4,920.77
Log Likelihood in Session 2		-3,689.26

Table A.3: Finite mixture estimations ($K = 2$) in Sessions 1 and 2.

A6 Finite mixture model with $K = 4$ preference types

	Strongly Altruistic Type	Moderately Altruistic Type I	Moderately Altruistic Type II	Behindness Averse Type
Session 1				
π	0.360***	0.367***	0.170***	0.103***
	(0.054)	(0.045)	(0.035)	(0.033)
α	0.180***	0.056*	0.057***	-0.558**
	(0.023)	(0.028)	(0.017)	(0.223)
β	0.484***	0.178***	0.072***	-0.207
	(0.031)	(0.040)	(0.012)	(0.139)
γ	0.150***	0.022	-0.003	0.211
	(0.026)	(0.023)	(0.011)	(0.140)
δ	-0.060***	-0.032	-0.020	-0.071
	(0.021)	(0.020)	(0.017)	(0.123)
σ	0.018***	0.023***	0.139***	0.008***
	(0.001)	(0.002)	(0.034)	(0.002)
Session 2				
π	0.342***	0.313***	0.245***	0.100***
	(0.038)	(0.039)	(0.037)	(0.024)
α	0.193***	0.097***	0.026***	-0.329***
	(0.019)	(0.013)	(0.007)	(0.073)
β	0.503***	0.168***	0.033***	-0.048
	(0.019)	(0.015)	(0.010)	(0.053)
γ	0.103***	-0.005	-0.004	-0.028
	(0.023)	(0.012)	(0.005)	(0.030)
δ	-0.081***	-0.034***	-0.009	-0.015
	(0.019)	(0.012)	(0.006)	(0.035)
σ	0.019***	0.039***	0.109***	0.015***
	(0.001)	(0.003)	(0.009)	(0.002)
# of observation (both sessions)			18,720	
# of subjects (both sessions)			160	
Log Likelihood in Session 1			-4,039.43	
Log Likelihood in Session 2			-3,016.26	

Individual cluster robust standard errors in parentheses.
*** significant at 1%; ** significant at 5%; * significant at 10%

Table A.4: Finite mixture estimations ($K = 4$) in Sessions 1 and 2.

A7 Share of subjects willing to change the other's payoff in Session 2 ($K = 3$ model)

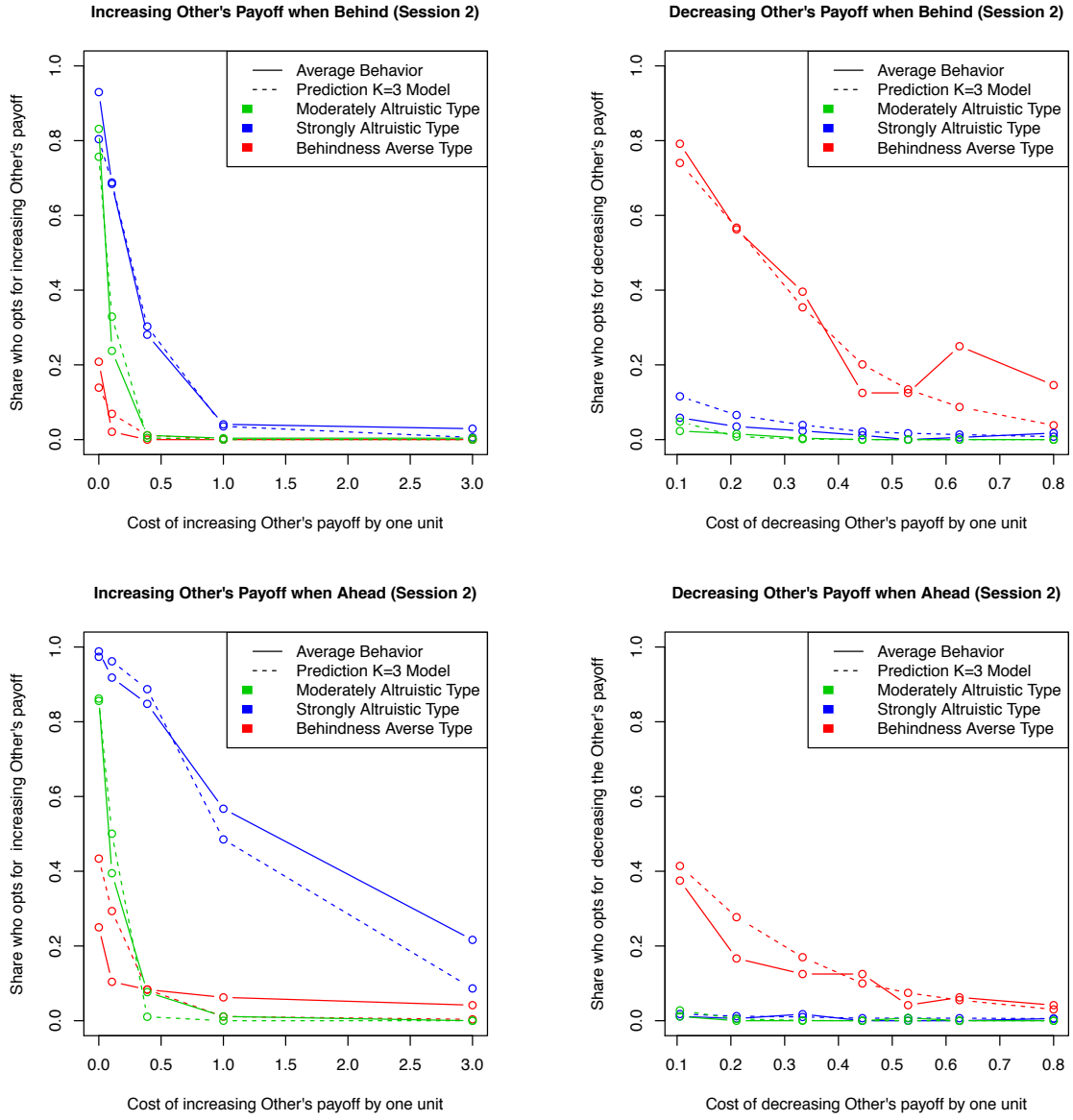


Figure A.4: Empirical and predicted share of subjects willing to alter the other's payoff in Session 2.

The Many Faces of Human Sociality:
Uncovering the Distribution and Stability of Social Preferences

Experimental Instructions

Translated from German to English

Comments not shown to the subjects are marked in green.

Instructions

Welcome to the Institute for Empirical Research in Economics at the University of Zurich. We thank you for again participating in our economic study. You can again earn money by participating. The amount you earn depends on your decisions in the study.
Please note that you may not communicate with one another during the study. If you have questions, please raise your hand. A study administrator will come to your seat and you can discuss the question. The violation of the rule against communication will result in exclusion from the study and from all payments.

This experiment consists of a total of ***9 parts***. At the beginning of each part you will get the corresponding instructions.

- **The first 3 parts of the experiment will take the most time. In these 3 parts,** you must decide how certain monetary payments between you *(Person A)* and another specific participant in the experiment *(Person B)* should be distributed.

- **In parts 4 and 5,** we ask you to estimate how other persons who have to make similar decisions would decide on average.

- **In part 6 of the experiment,** we ask you to complete a questionnaire.

How are the payments in this experiment determined?

1. You receive a fixed payment of CHF 5 for participating in the study. Additionally, you will also receive the payments described below.

2. 3 decision situations from the parts 1 to 3 will be randomly selected for payment. The distribution of payoffs in these 3 decision situations will be paid out to you (*Person A*) and a randomly selected other individual in the role of the receiver (*Person B*).

3. One of the decision situations in part 4 will be randomly selected. Your payment depends on the accuracy of your estimate.

4. Finally, you will get 5 CHF for completing the questionnaire in part 6.

The determination of the random selections determining the payments will first be made *after the conclusion of the entire experiment*. The money will be paid in cash to you.

The entire experiment is completely anonymous, i.e. you will not be informed of the identity of the participant paired with you, and your identity remains unknown to the other participants.

The order of the first and second part was randomized across subjects to avoid order effects.

Instructions for the first part

In this part of the experiment, you will make 39 decisions that concern you and another person participating in this experiment. The other person will be randomly paired with you in each decision situation. You will never learn who this person is, and the other person will also not learn of your identity.

In each of the 39 decision situations, you have exactly two options, an option X and an option Y. Each option involves a monetary amount for you *(Person A)* and a monetary amount for the other person *(Person B)* who is paired with you. You determine the distribution of the payment definitely with your decision. The other person *(Person B)* thus cannot change the income.

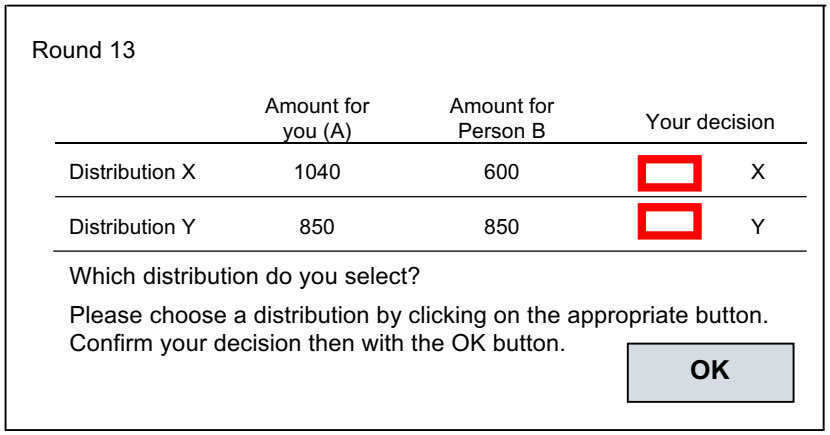
Please note: We present monetary amounts as points on the computer screen. 100 points are worth 1 CHF.

The amounts can also be negative. If you choose an option with a negative amount, the corresponding number of points will be deducted from you or the other person, respectively.

The procedure on the computer

The 39 different situations will be presented successively on a computer screen. You will see the options in the rows, and the columns show the amounts for you and the other person.

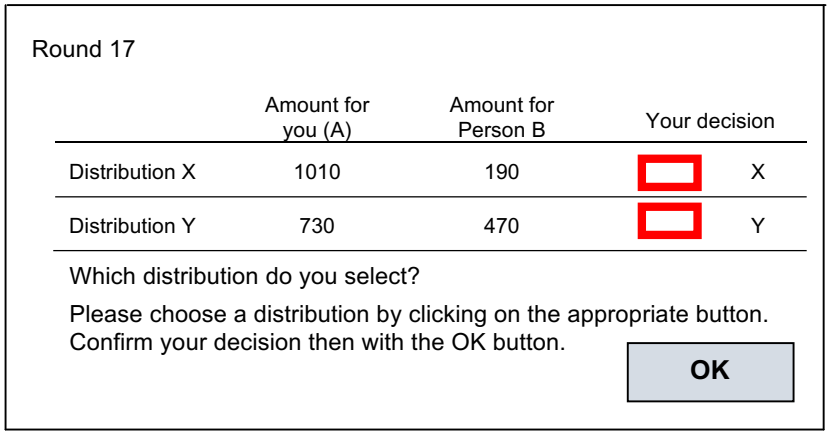
In the screen shown below, for example, you receive 1040 points while the other person only gets 600 points if you select option X. If you choose option Y, then both you and the other person receive 850 points each.



Control questions

Please answer the questions below. The objective is to have complete clarity about the rules in the experiment. When you are done, raise your hand. We will check if the control questions are answered correctly and start the experiment.

1. Please look at the screen below:



(a) How large is the income gap (in points) between you and Person B?

Income gap (in points) in case of selection of distribution X:_____.

Income gap (in points) in case of selection of distribution Y:_____.

(b) How large is the aggregate income (in points) of you and Person B for distributions X and Y?

Aggregate income (in points) for distribution X:_____.

Aggregate income (in points) for distribution Y:_____.

(c) If you select distribution X…

What is your income in CHF? _____.

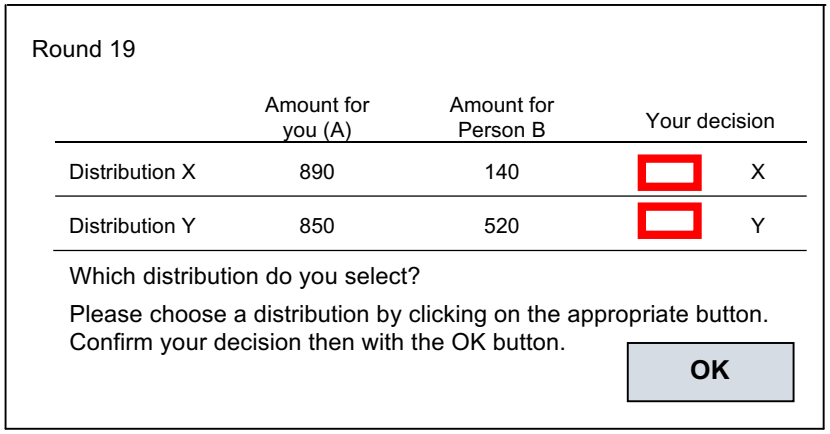
What is Person B's income in CHF? _____.

(d) If you select distribution Y …

What is your income in CHF? _____.

What is Person B's income in CHF? _____.

2. Please look at the screen below:



(a) How large is the income gap (in points) between you and Person B?

Income gap (in points) in case of selection of distribution X:_____.

Income gap (in points) in case of selection of distribution Y:_____.

(b) How large is the aggregate income (in points) of you and Person B for distributions X and Y?

Aggregate income (in points) for distribution X:_____.

Aggregate income (in points) for distribution Y:_____.

(c) If you select distribution X…

What is your income in CHF? _____.

What is Person B's income in CHF? _____.

(d) If you select distribution Y …

What is your income in CHF? _____.

What is Person B's income in CHF? _____.

3. Will the other person who is matched with you (*Person B*) learn about your identity?
_____ Yes
_____ No

Instructions for the second part

In this part of the experiment, you will make 78 decisions that affect you and another person participating in this experiment. The other person will be randomly assigned to you in each decision situation. You will not learn, however, who the other person is nor will the other person learn of your identity.

You have two options in each of the 78 decision situations, an option X and an option Y. Each decision concerns a monetary amount for you (*Person A*) and a monetary amount for another person (*Person B*) who is paired with you. In this experiment, you will determine the final distribution of the payment with your decision. The other person (*Person B*) thus can no longer change his or her income after you have made your decision.

The difference to the first part of the experiment is as follows: Now, Person B can determine **before your decision** whether he or she would like to fix a certain final distribution Z of the payments. If Person B determines this final distribution, you no longer can influence the distribution of the payments. As an alternative, Person B can delegate the decision about the distribution to you. In this case, you must select between options X and Y as described above, meaning that option Z is not available to you.

Please take exact note of the decisions *Person B* must make before you make your decision. We will show you some examples here:

- *Example 1:*
  Person B has the option of selecting the following distributions:

	Amount for you (A)	Amount for Person B
Distribution Z	170	1200

  Or Person B can delegate the following decision to you:

	Amount for you (A)	Amount for Person B
Distribution X	310	1060
Distribution Y	350	680

  If Person B delegates the decision to you in this case, it has the following consequences:
  (1) Comparison with Z:
  - You receive a higher payment in both distributions X and Y, which Person B left for your selection, than you would in distribution Z.
  - Person B receives a lower payment in both distributions X and Y, which Person B left to you, than he or she would in distribution Z.
  - The aggregate income of A and B in distribution X amounts to 1370 and is thus exactly the same size as in distribution Z. The aggregate income of A and B is smaller in distribution Y, however; it amounts to 1030.
  - The income gap is lower in the distributions X and Y, from which you can choose, than in distribution Z, where the income gap amounts to 1030 points. The income gap is only 750 in X and just 330 points in Y.

(2) Comparison between X and Y:

- The following applies to distribution X, which is an option you can choose:
  - You receive a lower payment than in distribution Y.
  - Person B receives a higher payment in X than in Y.
  - The aggregate income is higher in X than in Y.
  - The income gap is less in X than in Y.

- *Example 2:*
  Person B has the option of choosing the following distribution:

	Amount for you (A)	Amount for Person B
Distribution Z	1070	370

Or person B can delegate the following decision to you:

	Amount for you (A)	Amount for Person B
Distribution X	930	510
Distribution Y	810	150

If person B delegates the decision to you in this situation, this decision will have the following consequences:

(1) Comparison with Z:

- You will receive a lower payment in both distributions X and Y, which Person B left for your selection, than you would in payment Z.
- Person B receives a larger payment in X than in Z, but B receives a smaller payment in Y than in Z.
- The aggregate income of A and B in distribution X amounts to 1440 and is thus exactly the same size as in distribution Z. The aggregate income of A and B is smaller in distribution Y, however, and amounts to 960.
- The income gap is lower in the distributions X and Y, from which you can choose, than in distribution Z, where the income gap amounts to 700 points. The income gap is only 420 in X and 660 points in Y

(2) Comparison between X and Y:

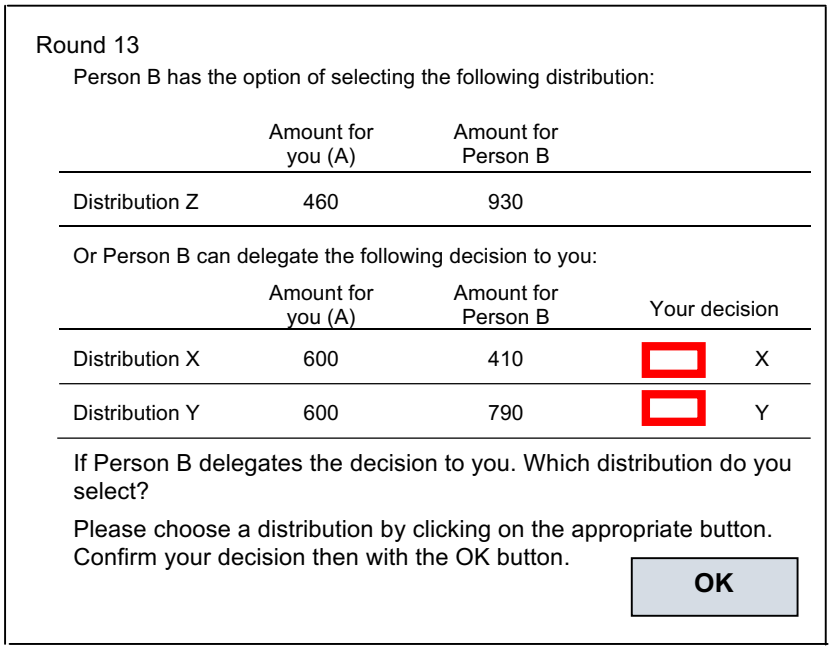
- The following applies to distribution X, which is an option you can choose:
  - You receive a higher payment than in distribution Y.
  - Person B receives a higher payment than in Y.
  - The aggregate income is higher than in Y.
  - The income gap is less than in Y.

Please note again: We present monetary amounts again as points on the computer screen. 100 points are worth 1 CHF.

The procedure on the computer

The 78 different situations will be presented successively on a computer screen. On the screen presented below, Person B can decide whether he/she will give you 460 points and retain 930 points, or if he or she will leave the following decision to you:

- Option X: 600 for you (Person A) und 410 for Person B.
- Option Y: 600 for you (Person A) und 790 for Person B.

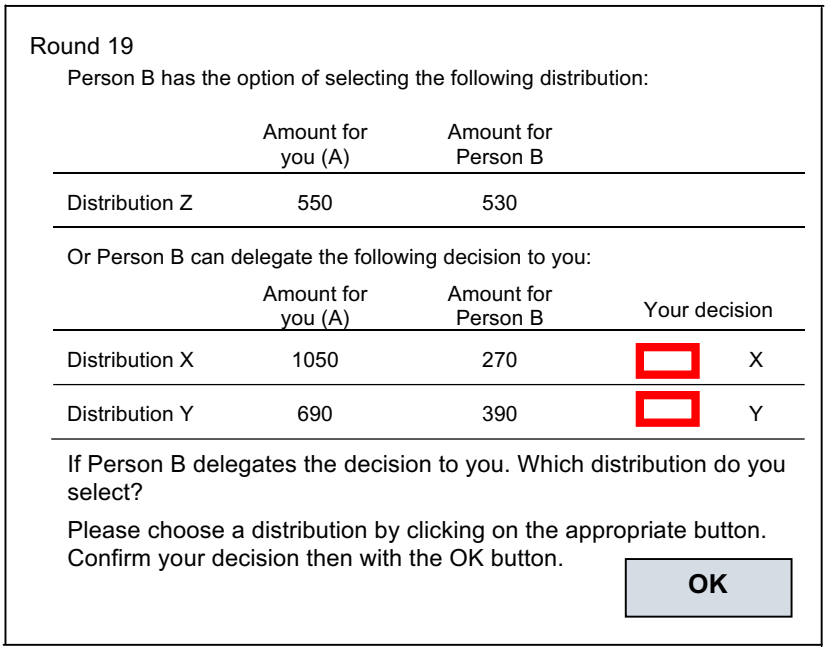


This is the longest part of the experiment, as it contains 78 questions. We kindly ask you to answer all 78 questions conscientiously, despite the number of questions.

Control questions

Please answer the questions below. The objective is to have complete clarity about the rules in the experiment. When you are done, raise your hand. We will check if the control questions are answered correctly and start the experiment.

1. Please look at the screen below:



Assuming that Person B delegated the decision to you. You thus must make a decision between the distributions X and Y.

(a) Did Person B improve his or her situation by delegating the decision?
- ☐ Yes, B is in a better position than in distribution Z.
- ☐ No, B is in a worse position than in distribution Z.
- ☐ Whether B is better or worse off depends on whether I select X or Y.

(b) Did Person B improve your (Person A's) situation by delegating the decision?
- ☐ Yes, I am in a better situation than in distribution Z.
- ☐ No, I am in a worse position than in distribution Z.
- ☐ Whether I am better or worse off depends on whether I select X or Y.

(c) How great is the income gap (in points) between you and Person B?

Income gap (in points) in distribution X:_____.
Income gap (in points) in distribution Y:_____.

(d) Is the income gap between you and Person B in distributions X and Y less than, greater than, or equal to the distribution Z?
- ☐ Less than in distribution Z.
- ☐ Greater than in distribution Z.
- ☐ Equal to distribution Z.

(e) What is the aggregate income (in points) of you and Person B in the distributions X, Y, and Z?

Aggregate income (in points) in distribution X:_____.
Aggregate income (in points) in distribution Y:_____.
Aggregate income (in points) in distribution Z:_____.

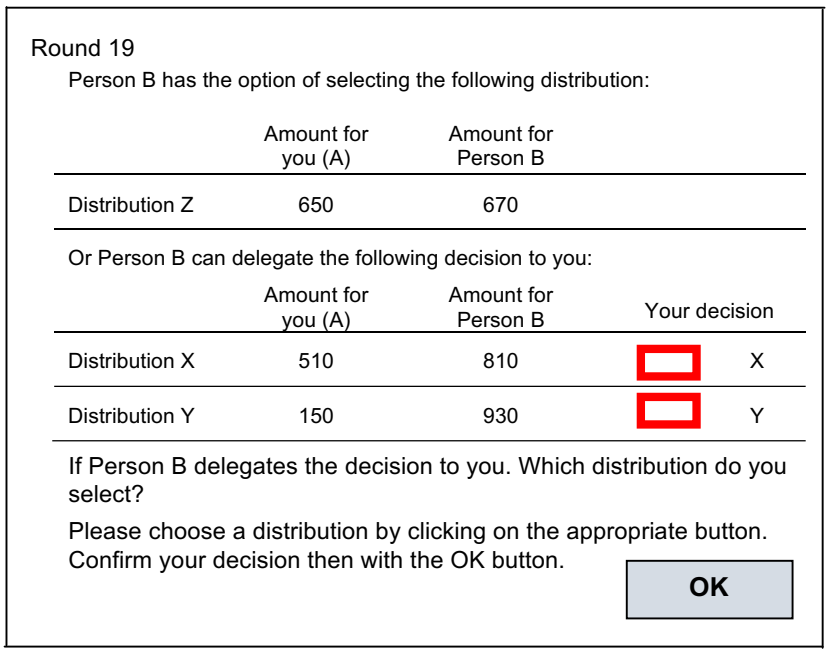
(f) Do you have a smaller, a larger, or an equal payment in distributions X and Y compared to distribution Z?
- ☐ My payment in distributions X and Y is less than that in distribution Z.
- ☐ My payment in distributions X and Y is greater than that in distribution Z.
- ☐ My payment in distributions X and Y is equal to that in distribution Z.

(g) Does the other person (B) have a smaller, a larger, or an equal payment in distributions X and Y compared to distribution Z?
- ☐ Person B's payment in distributions X and Y is smaller than in distribution Z.
- ☐ Person B's payment in distributions X and Y is greater than in distribution Z.
- ☐ Person B's payment in distributions X and Y is equal to that in distribution Z.

(h) If you select payment X, how large is your income in CHF? _____

(i) Will the person paired with you (Person B) learn of your identity?
- ☐ Yes.
- ☐ No.

1. Please look at the screen below:



Round 19

Person B has the option of selecting the following distribution:

	Amount for you (A)	Amount for Person B
Distribution Z	650	670

Or Person B can delegate the following decision to you:

	Amount for you (A)	Amount for Person B	Your decision	
Distribution X	510	810	☐	X
Distribution Y	150	930	☐	Y

If Person B delegates the decision to you. Which distribution do you select?

Please choose a distribution by clicking on the appropriate button.
Confirm your decision then with the OK button.

OK

Assuming that Person B delegated the decision to you. You thus must make a decision between the distributions X and Y.

(a) Did Person B improve his or her situation by delegating the decision?
- ☐ Yes, B is in a better position than in distribution Z.
- ☐ No, B is in a worse position than in distribution Z.
- ☐ Whether B is better or worse off depends on whether I select X or Y.

(b) Did Person B improve your (Person A's) situation by delegating the decision?
- ☐ Yes, I am in a better situation than in distribution Z.
- ☐ No, I am in a worse position than in distribution Z.
- ☐ Whether I am better or worse off depends on whether I select X or Y.

(c) How great is the income gap (in points) between you and Person B?

Income gap (in points) in distribution X:_____.
Income gap (in points) in distribution Y:_____.

(d) Is the income gap between you and Person B in distributions X and Y less than, greater than, or equal to the distribution Z?
- ☐ Less than in distribution Z.
- ☐ Greater than in distribution Z.
- ☐ Equal to distribution Z.

(e) What is the aggregate income (in points) of you and Person B in the distributions X, Y, and Z?

Aggregate income (in points) in distribution X:_____.
Aggregate income (in points) in distribution Y:_____.
Aggregate income (in points) in distribution Z:_____.

(f) Do you have a smaller, a larger, or an equal payment in distributions X and Y compared to distribution Z?

☐ My payment in distributions X and Y is less than that in distribution Z.
☐ My payment in distributions X and Y is greater than that in distribution Z.
☐ My payment in distributions X and Y is equal to that in distribution Z.

(g) Does the other person (B) have a smaller, a larger, or an equal payment in distributions X and Y compared to distribution Z?

☐ Person B's payment in distributions X and Y is smaller than in distribution Z.
☐ Person B's payment in distributions X and Y is greater than in distribution Z.
☐ Person B's payment in distributions X and Y is equal to that in distribution Z.

(h) If you select payment X, how large is your income in CHF? _____

(i) Will the person paired with you (Person B) learn of your identity?

☐ Yes.
☐ No.

Instructions for the third part

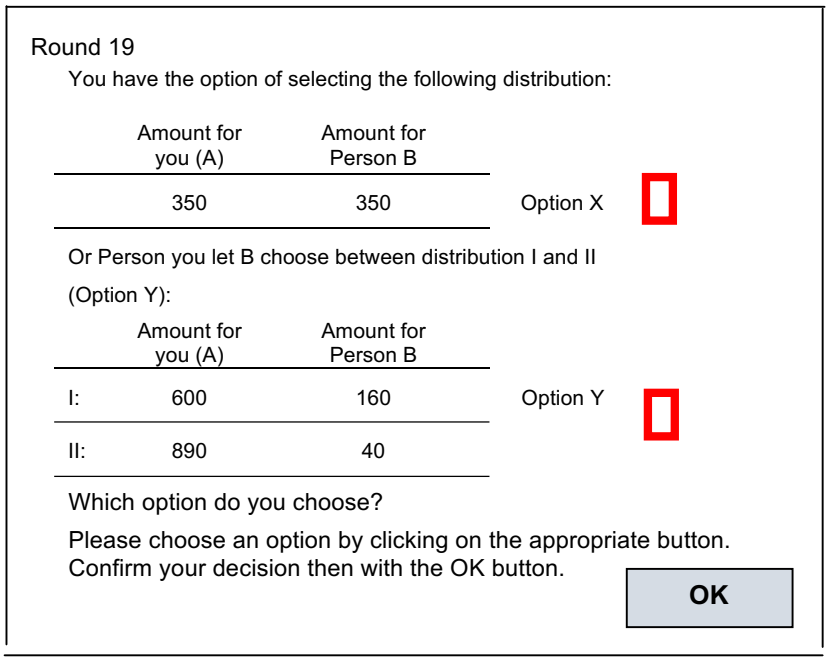
In this part of the experiment, you will make 6 decisions. This part is substantially shorter than the previous parts. Your decisions will again affect you and another person participating in the experiment. The other person will be randomly assigned to you in each decision situation. You will not learn, however, who the other person is nor will the other person learn of your identity.

In each of the 6 decision situations, you have exactly two options, option X and option Y, and you will be randomly matched again with another person (*Person B*). However, the difference to the previous two parts is as follows: If you choose option X, you determine a distribution of monetary amounts between yourself (*Person A*) and another person (*Person B*). Thus, the other person (*Person B*) cannot change the income. If you choose option Y, you let *Person B* choose between two distributions. Thus, the other person (*Person B*) can choose the definitive distributions of incomes.

Please note again: We present monetary amounts again as points on the computer screen. 100 points are worth 1 CHF.

The procedure on the computer

The 6 different decision situations will be presented successively on a computer screen. On the computer screen below, you have the option to give yourself and the other person 350 points each (Option X). If you choose option Y, however, you give person B the choice between two distributions: (I) 600 for you and 160 for Person B. (II) 890 for you and 40 for Person B.



The following instructions for the parts 4, 5, and 6 were only shown on the computer screen.

Part 4 (belief questions):

In the following fourth part oft the experiment, we show you again 12 decision situations in which there is a choice between two distributions of monetary amounts. You already know the 12 decisions from the previous parts of the experiment. However, this time, you do not have to choose between the two distributions of monetary amounts.
We ask you to estimate, how 100 randomly selected persons would choose in such decision situations. In particular, we will ask you how many out of 100 persons would choose one of the two options.

Of these 12 decision situations, one will be randomly selected. If your estimate deviates by less than 10 percentage points from the true fraction of persons, you gain an additional 5 CHF. For example: If you estimate that out of 100 persons 72 persons would choose option X, and the true number is 75 persons, you get 5 CHF. However, if you estimate that 64 persons choose option X, you do not get the additional 5 CHF (since your estimate deviates by 75-64 = 11 percentage points from the true fraction).

Part 5 (reciprocity perception):

In the following fifth part of the experiment, we show you again 18 decision situations in which there is a choice between two distributions of monetary amounts. You already know the 18 decisions from the previous parts of the experiment. However, this time, you do not have to choose between the two distributions of monetary amounts.

We ask you to give us an indication whether you perceive the other person's behavior as kind or unkind.

Part 6 (Questionnaire):

The final part of the experiment consists of a questionnaire. It is important for us that you answer the questions as good as possible. Many questions are about views or values. Thus, most of the time there are no right or wrong answers. Your answers best fulfill the purpose of the questionnaire if they are as truthful as possible. Sometimes there are some very personal questions. Please answer them as well as truthfully as possible. Your answers will be treated confidentially and analyzed in an anonymous way.

Before the cognitive ability test (end of part 6):

Finally, we would like to ask you to complete a series of patterns. You will each time see a pattern in which a part is cut out. Please look at the pattern and think which of the cut out parts best complete the pattern, both in horizontal and vertical direction. Your task is to identify the correct cut out part out of 8 possibilities.

We first show you 2 examples to practice.
[2 examples in which the subject gets a feedback on whether she selected the correct pattern]

Now, we show you 12 patterns. In total you have 12 minutes to complete the 12 patterns.
[12 Raven's matrices]

Instructions

Welcome to the Institute for Empirical Research in Economics at the University of Zurich. We thank you for again participating in our economic study. You can again earn money by participating. The amount you earn depends on your decisions in the study.
Please note that you may not communicate with one another during the study. If you have questions, please raise your hand. A study administrator will come to your seat and you can discuss the question. The violation of the rule against communication will result in exclusion from the study and from all payments.

This experiment consists of a total of ***9 parts***. The parts are of different lengths; they might last for more than 10 minutes or just for a few minutes. We expect a total work time of approximately one and a half hours.

- **In parts 1 to 8**, you must decide how certain monetary payments between you *(Person A)* and another specific participant in the experiment *(Person B)* should be distributed.

- We ask that you complete a questionnaire **in part 9** of the experiment.

How are the payments in this experiment determined?

1. You receive a fixed payment of CHF 20 for participating in the study. You will also receive the payments described below.

2. Precisely three situations will be randomly selected from the decision situations in parts 1 to 8. The monetary distributions in these three rounds will be paid to you and to a randomly chosen individual in the role of the recipient.

The determination of the random selections determining the payments will first be made *after the conclusion of the entire experiment.* The money will be paid in cash to you.

Please note that we present monetary amounts as points on the computer screens. In this case, 100 points are worth one Swiss franc.

The entire experiment is completely anonymous, i.e. you will not be informed of the identity of the participant paired with you, and your identity remains unknown to the other participants.

Instructions for the first part

In this part of the experiment, you will make 78 decisions that affect you and another person participating in this experiment. The other person will be randomly assigned to you in each decision situation. You will not learn, however, who the other person is nor will the other person learn of your identity.

You have two options in each of the 78 decision situations, an option X and an option Y. Each decision concerns a monetary amount for you (*Person A*) and a monetary amount for another person (*Person B*) who is paired with you. In this experiment, you will determine the final distribution of the payment with your decision. The other person (*Person B*) thus can no longer change his or her income after you have made your decision.

Please take note of the following: Person B can determine **before your decision** whether he or she would like to fix a certain final distribution Z of the payments. If Person B determines this final distribution, you no longer can influence the distribution of the payments. As an alternative, Person B can delegate the decision about the distribution to you. In this case, you must select between options X and Y as described above, meaning that option Z is not available to you.

Please take exact note of the decisions *Person B* must make before you make your decision. We will show you some examples here:

- *Example 1:*
  Person B has the option of selecting the following distributions:

	Amount for you (A)	Amount for Person B
Distribution Z	170	1200

Or Person B can delegate the following decision to you:

	Amount for you (A)	Amount for Person B
Distribution X	310	1060
Distribution Y	350	680

If Person B delegates the decision to you in this case, it has the following consequences:
(1) Comparison with Z:
- You receive a higher payment in both distributions X and Y, which Person B left for your selection, than you would in distribution Z.
- Person B receives a lower payment in both distributions X and Y, which Person B left to you, than he or she would in distribution Z.
- The aggregate income of A and B in distribution X amounts to 1370 and is thus exactly the same size as in distribution Z. The aggregate income of A and B is smaller in distribution Y, however; it amounts to 1030.
- The income gap is lower in the distributions X and Y, from which you can choose, than in distribution Z, where the income gap amounts to 1030 points. The income gap is only 750 in X and just 330 points in Y.

(2) Comparison between X and Y:
- The following applies to distribution X, which is an option you can choose:

- You receive a lower payment than in distribution Y.
- Person B receives a higher payment in X than in Y.
- The aggregate income is higher in X than in Y.
- The income gap is less in X than in Y.

- *Example 2:*
  Person B has the option of choosing the following distribution:

	Amount for you (A)	Amount for Person B
Distribution Z	1070	370

Or person B can delegate the following decision to you:

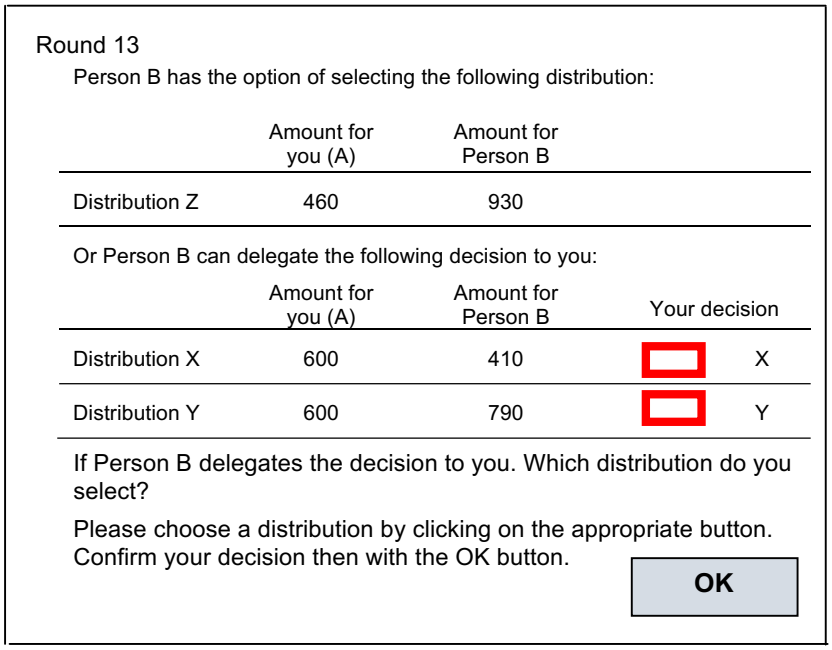
	Amount for you (A)	Amount for Person B
Distribution X	930	510
Distribution Y	810	150

If person B delegates the decision to you in this situation, this decision will have the following consequences:
(1) Comparison with Z:
- You will receive a lower payment in both distributions X and Y, which Person B left for your selection, than you would in payment Z.
- Person B receives a larger payment in X than in Z, but B receives a smaller payment in Y than in Z.
- The aggregate income of A and B in distribution X amounts to 1440 and is thus exactly the same size as in distribution Z. The aggregate income of A and B is smaller in distribution Y, however, and amounts to 960.
- The income gap is lower in the distributions X and Y, from which you can choose, than in distribution Z, where the income gap amounts to 700 points. The income gap is only 420 in X and 660 points in Y

(2) Comparison between X and Y:
- The following applies to distribution X, which is an option you can choose:
  - You receive a higher payment than in distribution Y.
  - Person B receives a higher payment than in Y.
  - The aggregate income is higher than in Y.
  - The income gap is less than in Y.

The procedure on the computer

The 78 different situations will be presented successively on a computer screen. On the screen presented below, Person B can decide whether he/she will give you 460 points and retain 930 points, or if he or she will leave the following decision to you:
- Option X: 600 for you (Person A) und 410 for Person B.
- Option Y: 600 for you (Person A) und 790 for Person B.

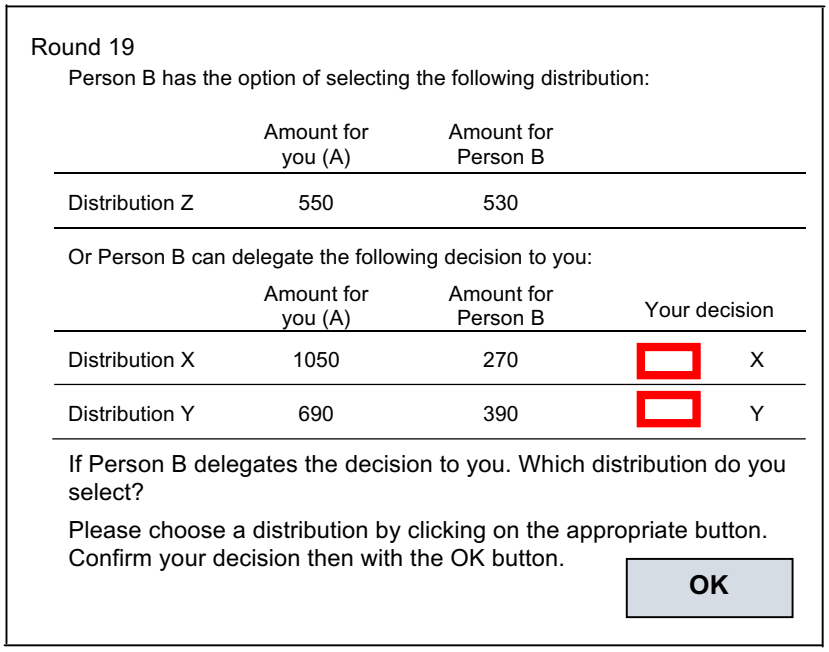


This is the longest part of the experiment, as it contains 78 questions. We kindly ask you to answer all 78 questions conscientiously, despite the number of questions.

Control questions

Please answer the questions below. The objective is to have complete clarity about the rules in the experiment. When you are done, raise your hand. We will check if the control questions are answered correctly and start the experiment.

1. Please look at the screen below:



Assuming that Person B delegated the decision to you. You thus must make a decision between the distributions X and Y.

(d) Did Person B improve his or her situation by delegating the decision?
- ☐ Yes, B is in a better position than in distribution Z.
- ☐ No, B is in a worse position than in distribution Z.
- ☐ Whether B is better or worse off depends on whether I select X or Y.

(e) Did Person B improve your (Person A's) situation by delegating the decision?
- ☐ Yes, I am in a better situation than in distribution Z.
- ☐ No, I am in a worse position than in distribution Z.
- ☐ Whether I am better or worse off depends on whether I select X or Y.

(f) How great is the income gap (in points) between you and Person B?

Income gap (in points) in distribution X:_____.
Income gap (in points) in distribution Y:_____.

(d) Is the income gap between you and Person B in distributions X and Y less than, greater than, or equal to the distribution Z?
- ☐ Less than in distribution Z.
- ☐ Greater than in distribution Z.
- ☐ Equal to distribution Z.

(e) What is the aggregate income (in points) of you and Person B in the distributions X, Y, and Z?

Aggregate income (in points) in distribution X:_____.
Aggregate income (in points) in distribution Y:_____.
Aggregate income (in points) in distribution Z:_____.

(j) Do you have a smaller, a larger, or an equal payment in distributions X and Y compared to distribution Z?
- ☐ My payment in distributions X and Y is less than that in distribution Z.
- ☐ My payment in distributions X and Y is greater than that in distribution Z.
- ☐ My payment in distributions X and Y is equal to that in distribution Z.

(k) Does the other person (B) have a smaller, a larger, or an equal payment in distributions X and Y compared to distribution Z?
- ☐ Person B's payment in distributions X and Y is smaller than in distribution Z.
- ☐ Person B's payment in distributions X and Y is greater than in distribution Z.
- ☐ Person B's payment in distributions X and Y is equal to that in distribution Z.

(l) If you select payment X, how large is your income in CHF? _____

(m) Will the person paired with you (Person B) learn of your identity?
- ☐ Yes.
- ☐ No.

Please raise your hand when you have completed the control questions up to this point. A study administrator will look at your answers; after this, the study will begin on your screen.

Instructions for the second part

In this part of the experiment, you will see eight decision situations concerning a monetary distribution between you and another person participating in this experiment. The other person will be randomly paired with you in each decision situation. You will never learn who this person is, and the other person will also not learn of your identity. As in the last part, the experiment now concerns a monetary amount for you *(Person A)* and the other person *(Person B)*.

The difference to the first part of the experiment is the following: This time, you cannot choose between two options X and Y. Instead, Person B has this choice, and you can **react to Person B's choice** by either **crediting** Person B with points or **deducting** points from him/her.

Please again take exact note of which decision *Person B* must make. We will show you an example below:

- Person B must choose between the following distributions:

	Amount for you (A)	Amount for Person B
Distribution X	600	600
Distribution Y	200	1000

The aggregate income of you and Person B is the same in both distributions X and Y and amounts to 1200 points. You and Person B receive the same amount in distribution X, but in distribution Y, Person B receives 800 more points than you do. You must now decide for both of Person B's selection options (X and Y) whether you want to credit points to or deduct points from Person B.

- If Person B selects distribution **X**, you have three options:
  1st option: You can **credit** Person B with **positive points** using the following table:

Cost for you (in points)	10	20	30
Positive points for the other person	50	80	100

The table shows that you have to pay 10 points (i.e. relinquish) if you want to credit Person B with 50 positive points. If you want to credit Person B with 80 points, you have to pay 20, and if you want to credit Person B with 100 points, you have to pay 30.

2nd option: You can **deduct points** from Person B using the following table:

Cost for you (in points)	10	20	30
Point deduction for the other person	50	80	100

The table shows that you have to pay (i.e. relinquish) 10 points if you want to deduct 50 points from Person B. If you want to deduct 80 points from Person B, you have to pay 20, and if you want to deduct 100 points from Person B, you have to pay 30.

3rd option: You can **neither credit points to nor deduct points** from Person B.

- If Person B selects distribution **Y**, you again have three options:
  1st option: You can **credit** Person B with **positive points** using the following table:

Cost for you (in points)	10	20	30

Positive points for the other person	50	80	100

The table shows that you have to pay (i.e. relinquish) 10 points if you want to credit Person B with 50 positive points. If you want to credit Person B with 80 points, you have to pay 20, and if you want to credit Person B with 100 points, you have to pay 30.

2nd option: You can **deduct points** from Person B using the following table:

Cost for you (in points)	10	20	30
Point deduction for the other person B	50	80	100

The table shows that you have to pay (i.e. relinquish) 10 points if you want to deduct 50 points from Person B. If you want to deduct 80 points from Person B, you have to pay 20, and if you want to deduct 100 points from Person B, you have to pay 30.

3rd option: You can **neither credit points to nor deduct points** from Person B.

Control questions

Please answer the questions below.

1. Please look at the following screen shots. These two screen shots will always be shown directly after one another, as they concern the **same decision situation** between distributions X and Y **for Person B**. The only difference between the two screens is the following: on the first, you may indicate how you will react to Person B's behavior if B decides to select **distribution X**. In contrast, you may indicate on the second screen how you will react to Person B's behavior if B decides to select **distribution Y**.

Screen 1

Person B has the option of choosing between the following distributions:

	Amount for you (A)	Amount for Person B	Total points
Distribution X	600	600	1200
Distribution Y	500	700	1200

If Person B selects distribution X…

… would you like to credit Person B with points?

Cost for you in points	10	20	30
Point increase for the other Person B	35	50	60
Your choice	□	□	□

… would you like to deduct points from Person B?

Cost for you in points	10	20	30
Point deduction for the other Person B	35	50	60
Your choice	□	□	□

… would you neither like to credit points to nor deduct points from Person B? Then please click here: □

OK

Screen 2

Person B has the option of choosing between the following distributions:

	Amount for you (A)	Amount for Person B		Total points
Distribution X	600	600		1200
Distribution Y	500	700		1200
If Person B selects distribution Y…				
… would you like to credit Person B with points?				
Cost for you in points		10	20	30
Point increase for the other Person B		35	50	60
Your choice		□	□	□
… would you like to deduct points from Person B?				
Cost for you in points		10	20	30
Point deduction for the other Person B		35	50	60
Your choice		□	□	□
… would you neither like to credit points to nor deduct points from Person B? Then please click here: □				

OK

(a) How large is the income gap (in points) between you and Person B?

Income gap (in points) in case of selection of distribution X:_____.

Income gap (in points) in case of selection of distribution Y:_____.

(b) How large is the aggregate income (in points) of you and Person B for distributions X and Y?

Aggregate income in points for distribution X:_____.

Aggregate income in points for distribution Y:_____.

(c) If Person B selects distribution X and you would like to credit 60 points to Person B:

(c.1.) Which screen can you use for this decision? _____

(c.2) How many points must you pay to credit 60 points to Person B? _____

(c.3) How large is *your* income (in CHF) in this case? _____

(c.4) How large is *Person B's* income (in CHF) in this case? _____

(d) If Person B selects distribution Y and you would like to credit 80 points to Person B:

(d.1.) Which screen can you use for this decision? _____

(d.2.) How many points must you pay to credit 80 points to Person B? _____

(d.3) How large is *your* income (in CHF) in this case? _____

(d.4) How large is *Person B's* income (in CHF) in this case? _____

(e) If Person B selects distribution Y and you would like to deduct 80 points from Person B:

(e.1.) Which screen can you use for this decision? _____

(e.2.) How many points must you pay to deduct 100 points from Person B?? _____

(e.3) How large is *your* income (in CHF) in this case? _____

(e.4) How large is *Person B's* income (in CHF) in this case? _____

Please raise your hand when you have completed the control questions up to this point. A study administrator will look at your answers; after this, the study will begin on your screen.

Instructions for the third part

The third part of the experiment consists of only one decision.[1] The task is to divide the total amount of 1200 points between you (Person A) and another person (Person B). The following rules apply:

The other Person (B) may **first** determine how much of the total amount of 1200 points he or she would like to keep for him/herself. **Afterwards**, you (Person A) may determine how many points you would like to have. **If the total sum (i.e. the amounts that both persons have determined) exceeds 1200 points, neither person, i.e. neither Person A nor Person B, will receive points.** If the sum the two persons determine **does not exceed** 1200, each receives the points he or she determined.

Control questions

1. Person B determines that he or she would like 900 points of the total amount for him/herself, and you (Person A) claim 300 points for yourself. How large is the payment…

   …for you (in CHF)? _____

   …for Person B (in CHF)? _____

2. Person B determines that he or she would like 700 points of the total amount for him/herself, and you (Person A) claim 600 points for yourself. How large is the payment …

   …for you (in CHF)? _____

   …for Person B (in CHF)? _____

3. Person B determines that he or she would like 600 points of the total amount for him/herself, and you (Person A) claim 300 points for yourself. How large is the payment …

   …for you (in CHF)? _____

   …for Person B (in CHF)? _____

Please raise your hand when you have completed the control questions up to this point. A study administrator will look at your answers; after this, the study will begin on your screen.

[1] In addition to the games analyzed in our paper "The Many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences", session 2 also contained four public good games, an ultimatum game in the design of Blount and Larrick (2000), as well as a multiple price list dictator game.
We did not analyze these games in the paper mentioned above, because our estimated behavioral model used in this paper does not allow to make quantitative predictions of behavior for these games: In the public good and the Blount-type ultimatum games, the decision space is (quasi-)continuous, but our econometric model assumes discrete decisions. In the multiple price list dictator game, the decisions are not independent, while our econometric model assumes independence.

Instructions for the fourth part

In this part of the experiment, you will make 39 decisions that concern you and another person participating in this experiment. The other person will be randomly paired with you in each decision situation. You will never learn who this person is, and the other person will also not learn of your identity.

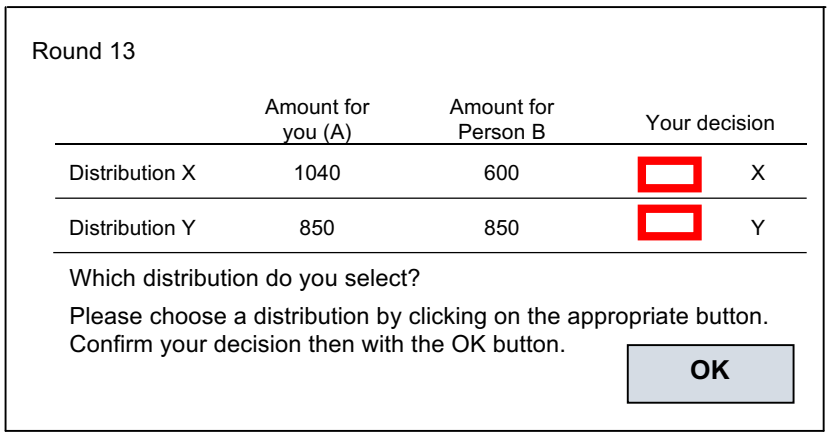
As in the first part of the experiment, you have exactly two options, an option X and an option Y, in each of the 39 decision situations. Each option involves a monetary amount for you *(Person A)* and a monetary amount for the other person *(Person B)* who is paired with you. You determine the distribution of the payment definitely with your decision. The other person *(Person B)* thus cannot change the income.

The difference to the first three parts of the experiment is the fact that Person B **cannot** make any decision before your decision this time. This time, you decide completely uninfluenced by any preliminary decision that B makes about the definite distribution of the payments.

The procedure on the computer

The 39 different situations will be presented successively on a computer screen. You will see the options in the rows, and the columns show the amounts for you and the other person.

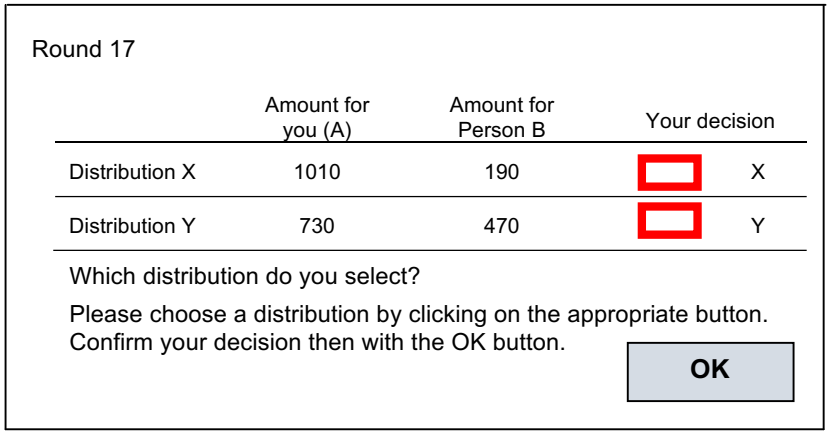
In the screen shown below, for example, you receive 1040 points while the other person only gets 600 points if you select option X. If you choose option Y, then both you and the other person receive 850 points each.



<u>Control questions</u>

Please answer the questions below. The objective is to have complete clarity about the rules in the experiment. When you are done, raise your hand. We will check if the control questions are answered correctly and start the experiment.

1. Please look at the screen below:



(a) How large is the income gap (in points) between you and Person B?

Income gap (in points) in case of selection of distribution X:_____.

Income gap (in points) in case of selection of distribution Y:_____.

(b) How large is the aggregate income (in points) of you and Person B for distributions X and Y?

Aggregate income (in points) for distribution X:_____.

Aggregate income (in points) for distribution Y:_____.

(c) If you select distribution X…

What is your income in CHF? _____.

What is Person B's income in CHF? _____.

(d) If you select distribution Y …

What is your income in CHF? _____.

What is Person B's income in CHF? _____.

Please raise your hand when you have completed the control questions up to this point. A study administrator will look at your answers; after this, the study will begin on your screen.

Instructions for the fifth part

In this part of the experiment, you will again make decisions that concern you and another person participating in this experiment. The other person will be, as always, randomly paired with you, and his or her identity will remain unknown.

You have two options in each decision situation, an option X and an option Y. Each decision concerns a monetary amount for you (*Person A*) and a monetary amount for another person (*Person B*) who is paired with you. In this experiment, you will determine the final distribution of the payment with your decision. The other person (Person B) thus can no longer change his or her income after you have made your decision.

The decisions will be presented to you in a table as shown below:

Distribution X		Your choice		**Distribution Y**	
Points for you (A)	Points for Person B			Points for you (A)	Points for Person B
600	800	o	o	590	1200
600	800	o	o	590	1100
600	800	o	o	590	1000
600	800	o	o	590	900
600	800	o	o	590	800
600	800	o	o	590	700
600	800	o	o	590	600
600	800	o	o	590	500
600	800	o	o	590	400
600	800	o	o	590	300

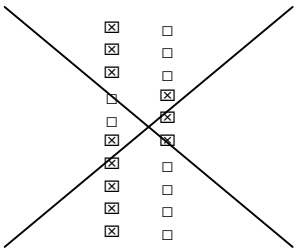
The individual decisions are each presented as rows in the table. You must decide for each row if you prefer distribution X or distribution Y.
Let us look, for example, at the first row of the table above.

- If you select distribution X, you will receive 600 points and Person B will receive 800 points. The aggregate income of you and Person B thus amounts to 1400 and the income gap between you and Person B is 200 points.
- If you select distribution Y, you will receive 590 points and person B will receive 1200 points. The aggregate income of you and Person B thus amounts to 1790 and the income gap between you and Person B is 610 points

If you look at the table carefully, you will see that it is constructed in such a way that there is only one spot where a change from option X to option Y is possible.

This means that the two selection models shown below to the left are possible, but that on the right is not.

possible



not possible

Control questions

These control questions refer to the table presented above. We will now look at the fifth **row** of this table.

(a) How large is the income gap (in points) between you and Person B?

Income gap (in points) in case of selection of distribution X:_____.

Income gap (in points) in case of selection of distribution Y:_____.

(b) How large is the aggregate income (in points) of you and Person B for distributions X and Y?

Aggregate income in points for distribution X:_____.

Aggregate income in points for distribution Y:_____.

(c) If you select distribution X…

What is your income in CHF? _____.

What is Person B's income in CHF? _____.

(d) If you select distribution Y …

What is your income in CHF? _____.

What is Person B's income in CHF? _____.

Please raise your hand when you have completed the control questions up to this point. A study administrator will look at your answers; after this, the study will begin on your screen.

Instructions for the sixth part

In this part of the experiment, you will make ten decisions that concern you and another person participating in this experiment. As always, you will not learn of the identity of the other participant paired with you, and your identity also remains unknown to the other person.

You already know the form of these decisions. You have two options in each decision situation, an option X and an option Y. Each decision concerns a monetary amount for you (*Person A*) and a monetary amount for another person (*Person B*) who is paired with you. Person B can determine **before your decision** whether he or she would like to fix a certain final distribution Z of the payments. If Person B determines this final distribution, you no longer can influence the distribution of the payments. As an alternative, Person B can delegate the decision about the distribution to you. In this case, you must select between options X and Y as described above, meaning that option Z is not available to you.

Please take careful note of the decision that *Person B* must make before you decide. We show you an example here:

Round 2

Person B has the option of selecting the following distribution:

	Amount for you (A)	Amount for Person B
Distribution Z	600	600

Or Person B can delegate the following decision to you:

	Amount for you (A)	Amount for Person B	Your decision	
Distribution X	800	900	☐	X
Distribution Y	1200	0	☐	Y

If Person B delegates the decision to you. Which distribution do you select?

Please confirm your decision with the OK button.

OK

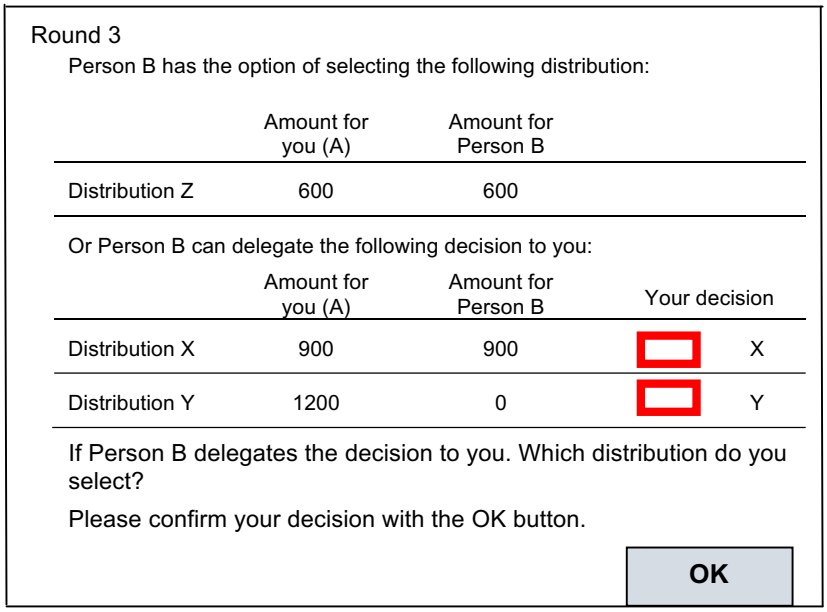
If Person B selects distribution Z, then both you and Person B receive 600 points. Person B can also entrust you with the decision between X and Y. If Person B delegates the decision to you, you have two possibilities, distribution X and distribution Y:

- If you select option X, the aggregate income of you and Person B amounts to 1700 (i.e. it is higher than in option Z, where the aggregate income is 1200 points). You will then receive 800 points and Person B will receive 900 points.
- The aggregate income of you and Person B amounts to 1200 points in option Y (i.e. it is equal to option Z, where the aggregate income is also 1200 points, but less than in option X). You receive 1200 points in option Y and Person B receives nothing.

If you select option Y, you would then abuse the trust that Person B placed in you in order to gain an advantage for yourself. If you select option X you will have 400 fewer points (800 instead of 1200 points), and reward Person B for trusting you.

Control questions

1. Please look at the screen below:



Assume that Person B delegates the decision to you. You now have to decide between distributions X and Y.

(a) How large is the income gap (in points) between you and Person B?

Income gap (in points) in case of selection of distribution X:_____.

Income gap (in points) in case of selection of distribution Y:_____.

(b) How large is the aggregate income (in points) of you and Person B for distributions X and Y?

Aggregate income (in points) for distribution X:_____.

Aggregate income (in points) for distribution Y:_____.

(c) Do you have a smaller, a larger, or an equal payment in distributions X and Y compared to distribution Z?

☐ My payment in distributions X and Y is less than that in distribution Z.
☐ My payment in distributions X and Y is greater than that in distribution Z.
☐ My payment in distributions X and Y is equal to that in distribution Z.

(d) If you select distribution X, what is your income in CHF? _____

(e) If you select distribution Y, what is your income in CHF? _____

Please raise your hand when you have completed the control questions up to this point. A study administrator will look at your answers; after this, the study will begin on your screen.

Instructions for the seventh part

You will again be randomly paired with another person in this part of the experiment. Both you and Person B must decide on the use of 600 points. You can leave the 600 points on a **private account** or you can invest them **partially or completely** in a shared project. You automatically deposit each point that you do not invest in the project in the private account.

Income from the private account:
You can retain with certainty every point that you deposit on the private account. For example, if you deposit 600 points in the private account (and thus invest nothing in the project), you will earn exactly 600 points from the private account. If, for example, 300 points are deposited in the private account, you will receive 300 points from the private account. *No one other than you can draw an income from your private account.*

Income from the shared project
Both you and Person B earn something in the same way from the amount that you invest in the project. Likewise, you earn something on Person B's investment. **Both you and Person B earn 0.6 point** for every point that you (or Person B) uses for the shared project. **A total of 1.2 points are distributed equally to both persons for every point invested in the project.**
If, for example, you and Person B invest 1000 points in the project together, you and Person B will each receive 1000 x 0.6 = 600 points from the project.

Total income:
Your total income consists of the sum of your income from the private account and your income from the project. If both you and Person B each contribute 500 points to the project and retain 100 points each for the private account, then both you and Person B will receive 1000 x 0.6 = 600 points from the project plus an additional 100 points from the private account. In total, both you and Person B receive 700 points each.

Control questions

1. Assume that neither you nor Person B contribute to the project.
What is *your* total income? _____
What is *Person B's total income*? _____

2. Assume that both you and Person B contribute the entire available amount of 600 points to the project.
What is *your* total income? _____
What is *Person B's total income*? _____

3. Assume that you contribute 500 points to the project but that Person B contributes no points.
What is *your* total income? _____
What is *Person B's total income*? _____

4. Assume that you contribute nothing to the project and Person B contributes the entire available amount of 600.

What is *your* total income? _____

What is *Person B's total income*? _____

5. This part will not just be completed on the computer with the **multiplication factor of 0.6**; instead, we will present you this task four times with various multiplication factors. The multiplication factors are chosen in such a way that each point invested in the project yields **more than one point** for the entire project. How large must the multiplication factor be at least so that each point invested in the project yields **more than one point** for the entire project?

☐ Greater than 0.3
☐ Greater than 0.4
☐ Greater than 0.5
☐ Greater than 0.6

Please raise your hand when you have completed the control questions up to this point. A study administrator will look at your answers; after this, the study will begin on your screen.

The eighth part of the experiment was to elicit proposer decisions which were needed to calculate the payments. In this part of the experiment, the subject played in the role of player B, the proposer. The instructions were directly given on the screen.

The ninth of the experiment consisted of a questionnaire about the subject's personal details to match her to Session 1.