CERGE
Center for Economic Research and Graduate Education
Charles University Prague

Essays on Fairness, Inequality, and Uncertainty

Vítězslav Babický

Dissertation

Prague, October 2016

Dissertation Committee:

Andreas Ortmann (University of New South Wales, Sydney, Chair)
Dirk Engelmann (Humboldt University of Berlin)
Krešimir Žigić (CERGE-EI, Charles University, Prague)

Referees:

Peter Katuščák (RWTH Aachen)
Maroš Servátka (Macquarie Graduate School of Management, Sydney)

Outline

List of Tables:

List of Figures:

Acknowledgements

I want to thank my dissertation chair Andreas Ortmann for his patience and his help in finishing this dissertation.

I also want to thank my co-authors for their work on the papers, especially again to my chair Andreas Ortmann, and my colleagues Silvester van Koten and Vilém Semerák.

Thanks to my committee members who helped me with their advice and support, Andreas Ortmann, Dirk Engelmann, and Krešimir Žigić. I am also grateful to the excellent scientists I had a chance to work and study with while pursuing my PhD., including Axel Ockenfels, Martin Dufwenberg, Werner Güth, James C. Cox, and Avner Shaked.

I very much appreciate the many scientists I met, and with whom I had opportunities to discuss dissertation related topics, during workshops and seminars, at conference venues, and also those who read my drafts and provided me many useful comments.

Last but not least, I want to thank my parents for their enormous patience and support.

Prague, Czech Republic Vítězslav Babický

October, 2016

Abstract

Chapter I.

Recent theories of fairness (e.g., Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999) have typically used the assumption of ex ante known pie size. Here I explore theoretically the ramifications of pie size being unknown ex ante. Using a simple allocation problem known as dictator game, I find that attitude to fairness is systematically and intuitively related to risk and risk attitude. Results from informal experiments support the model proposed here.

Chapter II.

The relationship between risk in the environment, risk aversion and inequality aversion is not well understood. Theories of fairness have typically assumed that pie sizes are known ex-ante. Pie sizes are, however, rarely known ex ante. Using two simple allocation problems – the Dictator and Ultimatum games – we explore whether, and exactly how, unknown pie sizes with varying degrees of risk ("endowment risk") influence individual behavior. We derive theoretical predictions for these games using utility functions that capture additively separable constant relative risk aversion and inequality aversion. We experimentally test the theoretical predictions using two subject pools: students at Czech Technical University and employees of Prague City Hall. We find that: (1) Those who are more risk-averse are also more inequality-averse in the Dictator game (and also in the Ultimatum game, but not statistically significantly so) in that they give more; (2) Using the within-subject feature of our design, and in line with our theoretical prediction, varying risk does not influence behavior in the Dictator game, but does in the Ultimatum game (contradicting our theoretical prediction for that game); (3) Using the within-subject feature of our design, subjects tend to make inconsistent decisions across games; this is true on the level of individuals as well as in the aggregate. This latter finding contradicts the evidence in Blanco et al. (2011); (4) There are no subject-pool differences once we control for the elicited risk attitude and demographic variables that we collect.

Chapter III.

We analyze non-pecuniary motives (reciprocity and inequality aversion) influencing contributions to public goods (paying taxes). We achieve this by analyzing data from an experiment in which subjects first had to earn income, from which they then contributed to public good provision. We study absolute and relative taxation schemes and also tax avoidance in situations with different efficiency of public good provision. The participants in our experiment show significant deviations from pure individual income maximization; the deviations were motivated by reciprocity considerations rather than by inequality aversion.

Foreword

In the research reported here, my main goal is to explore people's behaviour under risk, uncertainty, and (their perception of) inequality. I use both theory and experiments.

In Chapter I, building on well-established outcome models of other-regarding preferences such as Bolton-Ockenfels (2000) and Fehr-Schmidt (1999), I show how exactly risk can affect decisions in the most simple context, namely that of the reward allocation game often called the Dictator game. When I wrote the paper in 2003, it was a pioneering contribution to the exploration of the interconnectedness of other-regarding behaviour under uncertainty and in risky decisions.

In Chapter II, building on my theoretical contribution, my co-authors and I report the results of experimental tests of theoretical predictions for the Dictator game as well as predictions for the closely related ultimatum and simple investment games, with the latter's second stage being a Dictator game. We describe the experimental setting and the procedures used, and also the two subject pools (standard student subjects and employees at Prague City Hall), and the demographic characteristic that we elicited.

Finally, in Chapter III, I report experimental results from joint work on public-goods financing, tax avoidance, reciprocity and inequality aversion with earned wealth and under absolute and relative contribution schemes. The experiment was designed and implemented in 2005–2006. It provides a rather complex setting that was constructed from scratch and addressed several experimental design and implementation challenges, the issue of earned wealth and wealth-based stratification of subjects among them.

However, the lessons learned from the lengthy experiments conducted in accordance with all contemporary methodology were, at least for me at this point of time, that the preferences of different subjects are also in many respects very different from each other, and it is extremely difficult to create a single model that can predict every type of recognized behaviour well. Moreover, looking at all data more thoroughly, it is also clear that not all subjects in experiments decide consistently over time, that their decisions can be quite noisy and are sometimes even contradictory, as we can observe not only in laboratory controlled environments but also in every day of our real lives.

Chapter I. Fairness under Risk: Insights from Dictator Games

(published in CERGE-EI Working Papers series as WP No. 217)

1. Introduction

Important papers (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999) have tried to explain the results of pie distribution experiments which suggest that many subjects do not behave in the purely selfish manner postulated by standard economic theory[1]. Both models incorporate other-regarding behavior in the form of inequity aversion as their key explanatory component. They are also constructed under the assumption of ex ante known pie sizes. The world, however, is not always fully known ex ante.

Take the situation of a couple who want to be together for the rest of their lives. While deeply in love, she is rational enough to know that there is a – ever so slightly – chance that things will not work out as planned. She is the better prospect commercially (being as it is, a hot-shot lawyer, fresh out of a top-notch law school) while he is a sensitive guy who writes wonderful poems but is unlikely to eke out more than a meager living from his profession. Hence, she wants a prenuptial agreement. He has no choice but to accept in its entirety whatever it is that she wants.

Clearly, this is a one-shot dictator game. It is also a dictator game under uncertainty or risk (dependent on whether we assume the range of possible outcomes to be known or not) because the dictator does not know what the size of the pie will be if, contrary to today's blissful expectations of living happily ever after, push would come to shove. What will the dictator do in such a situation? In this paper, based on results from informal experiments[2], I assume that she has a preference for relative rather than for ”absolute” giving, and I investigate how the variance of possible pie sizes, i.e. the risk associated with the distribution, will affect her offers. I also explore how this decision is related to her risk attitude.

[1] For example, the game-theoretic prediction in dictator and ultimatum games suggests zero giving using standard selfish preferences. Experimental studies, however, provide clear evidence on positive giving for both games; the transfer to the recipient amounts to about 20% of the pie size for the dictator game and more than twice that for the ultimatum game.
[2] And also based on the key behavioral assumption of most models of reciprocity and fairness.

2. ERC[3] analysis of the game

Both the Fehr-Schmidt model and the Bolton-Ockenfels ERC model study interactions of n people. In both models, people care about their own payoffs. The difference lies in the modelling of inequity aversion. In Fehr & Schmidt (1999), it is expressed as some linear function of the difference of one's own payoff and the various payoffs of other actors, while in the ERC model it is expressed as some function of the relative payoff, i.e. the ratio of one's own payoff to the sum of all payoffs[4].

For two-person games the distribution of payoffs is fully, and conveniently, determined by either the absolute or the relative payoff of a single agent. Consequently, with any sum of total payoffs, the Fehr-Schmidt utility can be viewed as a special case of the ERC motivation function (the difference in absolute payoffs is equal to the difference in relative payoffs times a constant representing the size of the pie to be distributed[5]). Hence, the following ERC analysis of the game can be easily translated into the corresponding Fehr-Schmidt analysis.

Let the motivation function be additively separable:

$$v(\mathrm{y}, \sigma) = u(\mathrm{y}) - \mathrm{k}\, f(\sigma)$$

where y is the absolute payoff of the player we are interested in (the dictator) and σ is her relative payoff (i.e. the ratio of her absolute payoff to the sum of all payoffs). To fulfill the assumptions of ERC theory, let u be a continuous increasing concave function (i.e. the marginal utility of one's own payoff is decreasing), f be a continuous strictly convex function attaining its minimum at $\sigma = 0.5$ (the disutility which a player experiences from her relative position in the game is minimized when her payoff equals that of the other player), and $k > 0$ be a constant (the coefficient k quantifies how much she cares about her relative payoff). As k approaches 0, she cares less and less about her relative standing and becomes, in the limit, a selfish actor with utility function u postulated by standard economic theory.

Let C be a random variable which determines the size of the pie to be distributed. Let p be the proportion of the pie that the dictator is willing to transfer to the recipient. Then, the dictator's maximization problem is[6]:

[3] ERC = Equity, Reciprocity, and Competition

[4] For example, if the payoffs are 6, 3 and 1 for the other players and 5 for oneself, then inequity aversion according to Fehr & Schmidt is measured as 6-5=1 on one side and (5-3)+(5-1)=6 on the other side; for the final inequity aversion both of these values (weighted by possibly different fairness sensitivity parameters) are used. For the ERC model, the difference of the relative payoff to the equal division matters, i. e. 5/15 – 1/4, where the second number normalizes the relative payoff with respect to the equal standing.

[5] Take for example payoffs 5 and 3. For the Fehr-Schmidt model, inequity aversion is evaluated as 5-3=2, and the difference of relative payoffs from ERC is also 5/8 – 3/8 = 2/8, 8 is the total size of the pie in the game.

[6] I assume for now that agents are expected utility maximizers. I am aware that this assumption is a topic of ongoing disputes on which I remain agnostic. My interest here is, within the framework of previous studies, to analyze the ramifications of decision making under risk. I am in the process of extending my analysis to prospect theory.

$$\max_p \mathrm{E}v((1-p)C, 1-p) = \max_p [\mathrm{E}u((1-p)C) - k\, f(1-p)].$$

Note that for the utility maximizing decision holds $p \in [0, 0.5]$ since the allocation $p = 0.5$ is always preferred to any allocation $p' > 0.5$. Note also that the level of inequity aversion is the same for all realizations of the random variable *C* since the dictator's decision determines the relative payoffs no matter what the actual size of the pie will be.

Since under our assumptions above (concavity of both components of the motivation function, and hence of the motivation function itself), the second-order condition is automatically fulfilled, the optimal dictator giving *p* follows from the first-order condition:

$$d(\mathrm{E}u((1-p)C))/dp + k\, f'(1-p) = 0.$$

To be able to compute comparative static results, and for the sake of computational convenience, I assume that *u* is a function of the constant relative risk aversion variety, namely $u(x) = \mathrm{sgn}(r)x^r$, with $r \leq 1$, $r \neq 0$[7]. (Recall that the coefficient of relative risk aversion is equal to $r - 1$ for such functions). I can then rewrite the first-order condition as

$$(1-p)^{1-r} f'(1-p) = r\ \mathrm{sgn}(r)\ \mathrm{E}(C^r)/k. \qquad (1)$$

Note that if the right hand side does not belong to the interval $(0, f'(1))$ then border dictator giving occurs - either none or half of the pie will be transferred.

For any distribution (here, of pie sizes), the value of $\mathrm{E}(C^r)$ represents a risk associated with a given distribution. For example, in the case of $\mathrm{E}C_1 = \mathrm{E}C_2$ it is easy to see that $\mathrm{Var}\ C_1 < \mathrm{Var}\ C_2$ if and only if $\mathrm{E}(C_1^2) < \mathrm{E}(C_2^2)$, since $\mathrm{Var}\ C_i = \mathrm{E}(C_i^2) - (\mathrm{E}C_i)^2$. Similarly, for symmetric distributions it is always true (and in fact it is typically true for arbitrary distributions) that if $\mathrm{E}C_1 = \mathrm{E}C_2$ and $\mathrm{Var}\ C_1 < \mathrm{Var}\ C_2$, then also $\mathrm{E}(C_1^r) < \mathrm{E}(C_2^r)$ for all $r < 0$ and $\mathrm{E}(C_1^r) > \mathrm{E}(C_2^r)$ for all $r \in (0, 1)$. This is due to the convexity/concavity of function x^r. Thus, the right-hand side of equality (1) is increasing or decreasing with the increasing "risk" of a given distribution depending on the relative risk-aversion of the economic agent, i.e., whether $r < 0$ or $r \in (0, 1)$, respectively. Note that for $r = 1$ (i.e., risk-neutral players) the decision *p* depends only on the expected size of the pie.

An analogous analysis for the constant relative risk-aversion function corresponding to $r = 0$, i.e. for $u(x) = \log x$, yields the following first-order condition:

$$k\, f'(1-p) = 1/(1-p).$$

[7] See for example experimental results by Holt & Laury, 2002.

Note that the dictator's decision under this functional specification does not depend on the size of the pie distributed in the dictator game.

It is possible to refine the above analysis further if we also assume f to be of the constant relative risk-aversion variety (although this term is not about risk aversion but reflects rather inequity aversion), i.e.

$$f(\sigma) = (\sigma - 0.5)^{\gamma}$$

where $\gamma > 1$ to assure its strict convexity. Consequently, $f'(1-p) = \gamma(1-2p)^{\gamma-1}/2^{\gamma-1}$ and dictator giving, using (1), now satisfies the equality

$$(1 - p)^{1-r} (1 - 2p)^{\gamma-1} = r \operatorname{sgn}(r)E(C^r)/l \qquad (2)$$

where $l = \gamma k/2^{\gamma-1}$ is a constant.

Under the given parameter assumptions ($r < 1$, $\gamma > 1$), it is easy to see that the left-hand side of (2) is a decreasing function of p. This means that dictator giving is lower when the right-hand side is higher and vice-versa. Together with my analysis of the effects of risk attitude on the right-hand side of (1) above, I prove the following proposition:

Proposition 1. Within the ERC framework, people characterized by a coefficient of relative risk aversion below -1 will decrease their dictator giving for any given pie size as risk increases, and people with a coefficient of relative risk aversion above -1 will increase their dictator giving in such a situation. Decisions of people with relative risk aversion equal to -1, as well as decisions of risk-neutral agents, will be unaffected by risk when pie size is unknown ex ante.

What is the intuition behind this theoretical result? As the coefficient of relative risk aversion decreases from 1, people are willing to be more altruistic up to a certain level; they substitute risk aversion for fairness. However, after that level, risk aversion prevails and people start to treat risk and fairness attributes as complements, decreasing the giving with higher risk. When the coefficient of relative risk aversion is equal to -1, the behavior crosses the neutral point as it was in the starting coefficient of 1.

Remark 1: A similar comparative statics analysis can be done with respect to changes in the size of the pie if it is not uncertain. In that case, all expected value operators will disappear and, as a result, the ERC model suggests higher (lower) dictator giving for larger sizes of the pie for coefficients of relative risk aversion below (above) -1, and no influence of the pie size in the case of logarithmic utility. Similarly, risk-neutral people would decrease their offers with increasing pie.

Remark 2: A similar analysis can also be done for a generalized model of the Fehr-Schmidt type where the argument in function f is now the difference of the payoffs of the players, i.e. $(1 - 2p)C$. The situation is then somewhat more complicated since $E(C^{\gamma})$ also enters into the denominator of the right-hand side in (1) and hence plays a role in the behavior of dictator giving in such a model. But then, since the term $E(C^{\gamma})$ in the denominator decreases the right-hand side in equations (1) and (2) when the risk increases, the only difference in the new specification of the model is that the critical value of parameter *r* (i.e. the value when the dependence of behavior on risk switching is similar to that in the proposition above) is lower than 0 (or, equivalently, the critical coefficient of relative risk aversion is lower than -1); it decreases even more with increasing parameter γ and such a change may also differ for different types of probability distributions.

Remark 3: The analysis is applicable also to risk-loving agents, i.e. those with convex selfish utility *u*. However, the second-order condition can then be invalid and in such a case, the dictator giving $p \in [0, 1/2]$ will not be the interior point. Typically, in such a situation the model predicts zero dictator giving in riskier conditions for these kinds of agents.

In fact, such a result conforms to the intuition that very high payoffs are really attractive for risk-loving agents and, thus, these agents do not like to share such payoffs with others, at least compared to lower payoffs which are more likely to happen.

3. Discussion

I chose to analyze the dictator game because giving behavior in this game depends only on a single person's preferences. I thus could study preferences in their purest form. The results of the informal experiments I conducted in Prague and Jena demonstrate, quite intuitively, that risk aversion matters and hence ought to be incorporated into models explaining other-regarding behavior. The pilot experiments suggest that people prefer relative over absolute offers under risky conditions. This fact conforms to the intuition that, if there is a choice, risk-averse agents prefer to share risk over bearing it themselves. The pilot experiments also suggest that on average, again quite in line with intuition, decision makers want to keep a certain risk premium, and that subjects decrease their giving (both absolute and relative) with increasing risk, suggesting that for the average subject the coefficient of relative risk aversion is less than -1[8]. Of course, people are heterogeneous so the actual giving behavior is different

[8] For example, an econometric analysis of auction data suggests this coefficient to be around 0.5. However, the difference here can be affected by different settings of the game or other assumptions.

for some people. That said, for almost all subjects both risk and fairness attitudes factor into their giving behavior. My model above formalizes this result.

The present research can be expanded in various ways. First, the expected utility approach can be replaced by an approach based on prospect theory (Kahneman & Tversky, 1979). Such an extension allows for modeling divergent perceptions of gains and losses and results in different predictions for different types of people, which also conforms to different behavioral patterns observed in various experiments.

Second, an extension of my analysis to more complicated games such as the ultimatum game seems desirable even though the experimental results for such a setting are going to be noisier due to beliefs playing a role in the decision making. Also desirable seems experimental work that tries to assess empirically the correlation between risk and fairness attitudes.

Acknowledgement

I am grateful to Dirk Engelmann, Werner Güth, Axel Ockenfels and Avner Shaked for fruitful discussions and to Anthony Ziegelmeyer and other seminar participants at the Max Planck Institute for Research into Economic Systems in Jena, Germany, for their comments. Special thanks go to Andreas Ortmann for numerous comments on earlier versions of this paper.

References

1. **Bolton G., Ockenfels A., 2000.** ERC: A Theory of Equity, Reciprocity and Competition. *American Economic Review 90*, 166-193.
2. **Fehr E., Schmidt K., 1999.** A Theory of Fairness, Competition and Cooperation. *Quarterly Journal of Economics 114*, 817-868.
3. **Holt C., Laury S., 2002.** Risk Aversion and Incentive Effects. *American Economic Review 92*, 1644-1655.
4. **Kahneman D., Tversky A., 1979.** Prospect Theory: An Analysis of Decision Under Risk. *Econometrica 47*, 263-291.

Chapter II. Fairness in Risky Environments: Theory and Evidence

(joint work with S. van Koten and A. Ortmann, published in *Games*, 2013[9])

1. Introduction

The relationship between risk aversion and inequality aversion is not well understood. It has been noted, however, that they are closely related in certain choice situations.[10] For example, risk-averse and inequality-averse choices are indistinguishable in situations where a decision maker picks one realization from a set of income distributions, but does not know his position in the distribution (see Rawls [1]). This does not, however, imply that they are related in more general choice situations (see Chambers [2] for a counter-example). The empirical evidence, though limited, suggests that more risk-averse people are also more inequality-averse. In particular, Ferrer-i-Carbonell & Ramos [3] use German (non-incentivized) survey data and Carlsson *et al.* [4] use non-incentivized choices between imagined societies and lotteries to show that risk aversion and inequality aversion are positively related. The results, however, are most likely affected by the hypothetical nature of the tasks, as it is not difficult to be inequality-averse when there are no monetary consequences to making decisions (a problem Ferrer-i-Carbonell & Ramos [3] discuss). Experimental evidence also suggests that the degree of risk aversion is heavily dependent on the non-hypothetical nature of the task(s), and the scale of financial stakes used (e.g., Holt & Laury [5]; Harrison *et al.* [6]). The relationship between risk aversion and inequality aversion would thus be better understood with a properly incentivized experiment, where choices explicitly contain tradeoffs between (selfish) utility and fairness (*i.e.*, inequality aversion).

The Dictator game and the Ultimatum game are standard decision situations that feature this tradeoff. The theoretic prediction for the Dictator game, under the standard assumption of selfish preferences, is zero giving. For the Ultimatum game, the theoretic prediction is zero giving and an acceptance threshold of zero or the smallest monetary unit above zero. The experimental evidence, however, shows that there is significant giving in both games, and that thresholds are greater than the smallest monetary unit in the Ultimatum game. Specifically, transfers to the recipient are about twenty percent of the pie size for Dictator games (but see

[9] Van Koten, S.; Ortmann A.; Babicky V., 2013. Fairness in Risky Environments: Theory and Evidence. *Games* 4, 208-242.

[10] A reasonable intuition for that finding could be this: encountering states of inequality in the world makes one aware of the danger that one self may end up in some such state. This is likely to be the more threatening for a person the more s/he is risk-averse. This threat might activate, or intensify, inequality aversion.

Cherry *et al.* [7][11]), and more than twice of that for Ultimatum games (see Camerer [8] and Güth & Ortmann [9]).

Bolton & Ockenfels [10] and Fehr & Schmidt [11] claim that positive giving and thresholds in these kind of game experiments can be explained by incorporating other-regarding preferences such as "inequality aversion" (a form of fairness) to subjects' utility function.[12] Greater aversion to inequality corresponds to higher giving in both games, and higher thresholds in Ultimatum games. We can thus test the relationship between risk aversion and inequality aversion by analyzing subjects' choices in Dictator and Ultimatum games.

Blanco *et al.* [20] show that Fehr & Schmidt's [11] model is relatively accurate in predicting choices across different games on the aggregate level, but does poorly on the individual level.[13] They thus suggest that people's choices are not consistent across different games. Their study, however, does not look at the effects that risk in the environment and risk aversion have on behavior, as Fehr & Schmidt's (and Bolton-Ockenfels's) [10,11] models were constructed under the assumption that pie sizes are known ex ante. The world, however, is rarely known ex ante, and so risk in the environment and risk preferences may play an important role in influencing fairness and reciprocity. It therefore remains an open question if choices in different games, and for each game under different degrees of risk, are consistent. We use a within-subject design to test whether subjects behave consistently under different degrees of risk for each game and across games.

This paper contains three contributions. First, we study the relationship between risk aversion and inequality aversion using properly incentivized Dictator and Ultimatum games,

[11] Cherry *et al.* [7] have demonstrated persuasively that the experimental results reported in the literature are dependent on both the degree of asset legitimacy and social distance. When the pie was not provided by the experimenter but had to be earned by their student dictators first ("asset legitimacy"), and when the game was also played under double anonymity (maximal "social distance"; see Hoffman *et al.* [12]), giving was as predicted by the standard economic assumption of selfishness. Bekkers [13] provides similar evidence through a field experiment. Smith [14] argues that asset legitimacy is an important challenge that experimental economists need to address.

[12] Engelmann & Strobel [15] presented an experimental test of the Bolton & Ockenfels and Fehr & Schmidt models which suggests that people, following Rawls's theory of justice (Rawls [1]), want to maximize the welfare of the person who is the worst off (a form of other-regarding behaviour); these authors (see also Engelmann & Strobel [16] for a balanced review of the literature and Engelmann [17] for an important caveat regarding the appropriate modelling of welfare maximization) identify the importance of efficiency concerns (defined as the sum of all payoffs in the game) as an explanatory variable. In their theory section, Charness & Rabin [18] use a social welfare criterion, which is defined as a weighted combination of minimal payoff (again, a form of other-regarding behaviour) and the sum of all payoffs in the game. Cox & Sadiraj [19] provide a non-linear generalization of that model.

[13] Recent experimental work conducted in parallel by Güth *et al.* [21], Cappelen *et al.* [22], and Krawczyk & Le Lec [23], introduce theories or experimental results of distributive choice under risk. Güth *et al.* [21], continuing the work of Brennan *et al.* [24] and, using a within-subject design and incentive-compatible elicitation of valuations of 16 different prospects, find that individuals are self-oriented towards the social allocation of risk and delay and other-regarding with respect to expected payoffs. Cappelen *et al.* [22] use a two-stage dictator game to study to what extent, and under what circumstances, ex-ante fairness under risk is robust to ex-post redistribution. Krawczyk & Le Lec [23] use a within-subject design and probabilistic versions of the dictator game that are manipulated along two dimensions (the relative cost of giving and the nature of the lottery) to try to tease out the relative merits of outcome-based and intention-based models of fairness under risk. All the models mentioned above were constructed under the assumption of pie sizes that are known ex ante.

with varying degrees of risk ("endowment risk").[14] We use a variant of the procedure from Holt & Laury [5] to elicit risk preferences, which allow us to compare risk preferences and choices made in the two games. Second, we use a within-subject design and the elicited risk preferences to explore whether choices under different degrees of risk (within the same game) are consistent. Third, we use the within-subject design feature to explore whether choices are consistent across games, conditional on the same degree of risk. We derive theoretical predictions for our second and third contribution using Bolton & Ockenfels's [10] ERC model. Assuming constant relative risk aversion, we extend their model and test our theoretical predictions for risky environments.

Our predictions are presented in Section 2. Section 3 contains our experimental design and implementation. In Section 4, we summarize our data and results. In Section 5, we discuss our findings and enumerate questions regarding our study. The Appendix contains simulations, an overview of the socio-demographic characteristics of our subjects, and a copy of the (translated) instructions and the precise sequencing of the scenarios used in the experiment.

2. Theoretical Predictions for Fairness in Risky Environments

2.1. Risk Attitude and Inequality Aversion

In the Dictator game, a dictator decides on the proportion p of an uncertain endowment S to give to a (anonymous) recipient.[15] To disambiguate, we call the generic (interactive) decision situation "game", and the different realizations of the endowment S "scenarios". To determine the relationship between risk aversion and inequality aversion, we measure participants' risk preferences using a task inspired by Holt & Laury [5]. The assertion that risk aversion and inequality aversion are positively related leads to the following hypothesis:

- ***Hypothesis 1D:*** *In the Dictator game, those with higher risk aversion give more than those with lower risk aversion, independent of the degree of risk in the environment.*

The relationship between risk aversion and inequality aversion can also be tested with an Ultimatum game. As in the Dictator game, a proposer decides on the proportion p of an

[14] In addition, we also studied the so-called Trust game in such an environment (see Appendices A.5, A.6, and A.8). Assuming self-regarding preferences, the situation for responders in the Trust game is theoretically equivalent to that of dictators in a standard Dictator scenario. However, responders in our design and implementation of the Trust game cannot infer precisely the proposer's initial decision (because the amount sent is multiplied by an unknown random factor X), and proposers cannot foresee the responders' reaction. Proposers probably developed beliefs about responders' behaviour, but we were not able, given time constraints, to control for these during the experiment. We are therefore not able to theoretically derive the effects of risk preferences for proposing and responding in the Trust game and therefore decided not to include the Trust game in our analysis. We note that none of the relationships turned out significant for the Trust game.

[15] In principle, the recipient's proposed share can also be determined in absolute terms. There are three reasons why we did not use absolute terms. First, in theories of fairness and reciprocity only relative terms matter. Second, an ex-ante allocation in absolute terms could result in a negative outcome for the decision maker (when the realized pie is small), which might trigger loss aversion and confound our results. Third, in the present paper we are not interested in optimal contract design (this could solve the preceding problem, but would also complicate our design beyond what seems feasible to implement.)

uncertain endowment S to give to an (anonymous) recipient. Unlike the Dictator game, the recipient is able to reject the proposal (she is thus called the responder). If the responder rejects the proposal, both proposer and responder receive zero. Under the assumption of selfishness, standard theory predicts that the responder rejects the proposal when her expected utility is lower than the disutility from having an unequal distribution. The responder thus sets a threshold for offers (the acceptance threshold), below which she rejects them. Responders that are more inequality-averse set higher thresholds, independent of the degree of risk in the environment. The assertion that risk aversion and inequality aversion are related leads to the following hypothesis:[16]

- ***Hypothesis 1U:*** *In the Ultimatum game, those with higher risk aversion set higher thresholds than those with lower risk aversion, independent of the degree of risk in the environment.*

For Ultimatum giving decisions we cannot derive predictions from theory as we do not know the expectations that givers had about the inequality aversion of the receiver they would be paired with.

2.2. Consistency within Games

We extend our analysis by exploiting the within-subject feature of our design (which is explained in more detail in Section 3). We derive predictions for within-subject consistency across different degrees of risk. Risk is added to the Dictator and Ultimatum games by providing endowments in the form of lottery tickets, which represent a mean-preserving spread. This requires the following extension of the standard theory of ERC[17] [10].

Let the motivation function ($v[y]$) be additively separable in utility ($u[y]$) and inequality aversion ($k \cdot f[\sigma]$):

$$v[y] = u[y] - k \cdot f[\sigma] \tag{1}$$

In Equation (1), y is the absolute payoff of the decision maker, and σ is his relative payoff (*i.e.*, the ratio of his absolute payoff to the sum of all payoffs in the interaction). To fulfill the assumptions of ERC theory, let u be a continuously increasing concave utility function (*i.e.*, the marginal utility of his payoff is decreasing), f be a continuous strictly convex inequality

[16] Note that we have not used the ERC formulation to derive our hypotheses regarding the relation between risk aversion and inequality aversion in the Dictator and Ultimatum game, as this formulation is moot on the possible effects of risk aversion on inequality aversion. Predictions could be derived for people with different risk preferences, if the inequality aversion parameter k could be assumed to remain constant between different people with different risk preferences but with equal inequality aversion. This assumption does not hold as the inequality aversion parameter k conveys both an inequality aversion and an (unknown) rescaling effect. An increase in risk aversion is modelled by a transformation that results in a more concave utility function. This transformation also rescales the magnitude and the slope of the utility function and the parameter k needs to change to correct for this rescaling effect, making it impossible to deduce the individual inequality aversion effect. For example, using formula (4) below to estimate the parameter k shows that in our sample it ranges from an average of 0.2 (for highly risk-averse subjects) to 100 (for somewhat risk-averse subjects) to 5500 (for risk-loving subjects).

[17] This can easily be reformulated in terms of Fehr & Schmidt [11]; see Babicky [25].

aversion function attaining its minimum equal to 0 at $\sigma = 0.5$ if n = 2 (*i.e.*, the decision maker's disutility from his relative payoff is minimized when his payoff equals the other player's payoff), and $k > 0$. The parameter k quantifies how much he cares about his relative payoff. As k approaches 0, he increasingly cares less about his relative standing in society and becomes, at the limit, a selfish decision maker whose utility function u mimics standard economic theory.

To increase the precision of our predictions, we introduced additional assumptions about our ERC specification. The literature suggests that a utility function in the constant-relative-risk-aversion form is a suitable approximation for human behavior.[18] We used the simplest form of such a utility function: $u[x] = Sign[r]x^r$ for $r \neq 0$, $u[x] = LN[x]$ for $r = 0$. We further assumed that a constant-relative-inequality-aversion function is a good approximation for fairness preferences: $f[\sigma] = (\frac{1}{2} - \sigma)^2$. Substituting these functions into (1), the ERC formula becomes:

$$v[y] = Sign[r] \cdot x^r - k \cdot (\tfrac{1}{2} - \sigma)^2 \qquad (2)$$

We indicate relative risk aversion by r $(= -\frac{v''}{v'} + 1)$, where r <(>) 1 indicates risk-averse (risk-loving) preferences. A dictator who gives proportion p of an ex ante unknown realization of S (and thus keeps percentage 1-p) is assumed to have a motivation function given by Equation (2) where $(1-p)S$ is substituted for x and 1-p for σ:

$$v[p] = E\left[Sign[r] \cdot ((1-p) \cdot S)^r \right] - k \cdot (\tfrac{1}{2} - p)^2 \qquad (3)$$

As S is a random variable, the motivation function is in fact an *expected* motivation function. Taking the non-random terms outside of the expectation operator and differentiating (3) with respect to the proportion of giving, p, yields the first-order condition for optimal giving in the Dictator scenario:[19]

$$Sign[r] \cdot r(1-p)^{r-1} \cdot E[S^r] = 2k(\tfrac{1}{2} - p) \qquad (4)$$

In the Ultimatum game, a responder evaluates the offer p from the risky endowment S. The motivation function of a responder who evaluates the offer p of an ex ante unknown realization of S is found by taking the expectation over the right-hand side of (2) and substituting pS for x and p for σ:

$$v[p] = Sign[r] \cdot E[(pS)^r] - k \cdot (\tfrac{1}{2} - p)^2 \qquad (2')$$

[18] For standard stakes (such as the ones in our experiment) the constant-relative-risk-aversion form of the utility function can be rationalized experimentally by the results of Holt & Laury [5], p.1652, who suggest that it works as a "good approximation" of human behaviour. This approximation simplifies our theoretical arguments considerably.

[19] Note that always $p \leq \frac{1}{2}$. For any c, $0 < c \leq \frac{1}{2}$, any proportion of giving equal to $\frac{1}{2} + c$ is dominated by $\frac{1}{2} - c$, as the latter proportion of giving results in higher utility, $u[y]$, and equal inequality, $f[\sigma]$.

The responder rejects the offer when her motivation function is negative; that is, when her expected utility is lower than the disutility from having an unequal distribution. At the threshold, the lowest offer the responder accepts requires the motivation function to equal zero. Rearranging (2'), the condition is:

$$Sign[r] \cdot p^r \cdot E[S^r] = k(\tfrac{1}{2} - p)^2 \qquad (4')$$

Equations (4) and (4') imply that the level of risk may have an effect on choices in the Dictator and Ultimatum games. For subjects with a typical degree of risk aversion, $0 < r < 1$, increasing the risk of a lottery by a mean-preserving spread, implies $E[S_H^r] < E[S_L^r]$, where S_H is the high- and S_L the low-risk lottery. An increase in risk thus lowers the value of the expected motivation function. This makes the left-hand side of Equations (4) and (4') smaller. For the Dictator (Ultimatum) scenario, increasing p can restore the equality as the right-hand can be made arbitrarily small by letting p approach ½, while the left-hand side stays bounded below by a strictly positive number for any p ≤ ½. The model thus predicts an increase in giving (threshold) when risk is increased. For highly risk-averse ($r < 0$) and risk-loving ($r > 1$) preferences, the relationship is rather complicated and we thus exclude subjects with these preferences.[20]

Numerical simulations, however, show that the effect of even a considerable increase in risk should result in minor differences in Dictator and Ultimatum game choices.[21] Consistency thus requires subjects to respond similarly under different levels of risk. We test this for all subjects in our data set with a typical degree of risk aversion, $0 < r < 1$, and exclude subjects with a degree of risk aversion given by $r < 0$ or $r > 1$. We thus formulate:

- ***Hypothesis 2:*** *For people that have typical risk-averse preferences (*$0 < r < 1$*), the risk of the endowment does not have a significant effect:*
  - ***D.*** *On giving in the Dictator game*
  - ***U.*** *On the threshold in the Ultimatum game*

2.3. Consistency across Games

We also extend our analysis and exploit the within-subject feature of our design by deriving predictions for within-subject consistency across games (but keeping the endowment risk constant). Following Blanco *et al.* [20], the analysis is conducted on the aggregate and individual level. We here also use as point of departure the theory of ERC and our specification in Section 2.2, Equations (4) and (4') are used to derive the inequality aversion parameter *k*. Rearranging (4) gives:

---

[20] For risk-loving ($r > 1$) preferences, the relationship is inverted in the Ultimatum game, and depends on the values of *k* and *r* in the Dictator game. For highly risk-averse ($r < 0$) preferences, the relationship is inverted in the Dictator game and undefined in the Ultimatum game.

[21] See Appendix A.3.

Dictator scenario: $$k = \frac{Sign[r] \cdot r(1-p_D)^{r-1} \cdot E[S^r]}{1-2p_D} \tag{5}$$

Rearranging (4') gives:

Ultimatum game: $$k = \frac{Sign[r] \cdot p_U^r \cdot E[S^r]}{(p_U - \frac{1}{2})^2} \tag{5'}$$

Equation (5) derives *k* from optimal giving, p_D , in the Dictator scenario. Equation (5') derives *k* from the optimal threshold, p_U, in the Ultimatum game. Standard theories of fairness [10,11] assume that an individual has constant parameters for inequality aversion *k*, and risk aversion *r*, regardless of the game. We thus Equations (5) and (5') to derive Equation (6), which shows giving in the Dictator game p_D, as an implicit function of the threshold in the Ultimatum game, p_U.

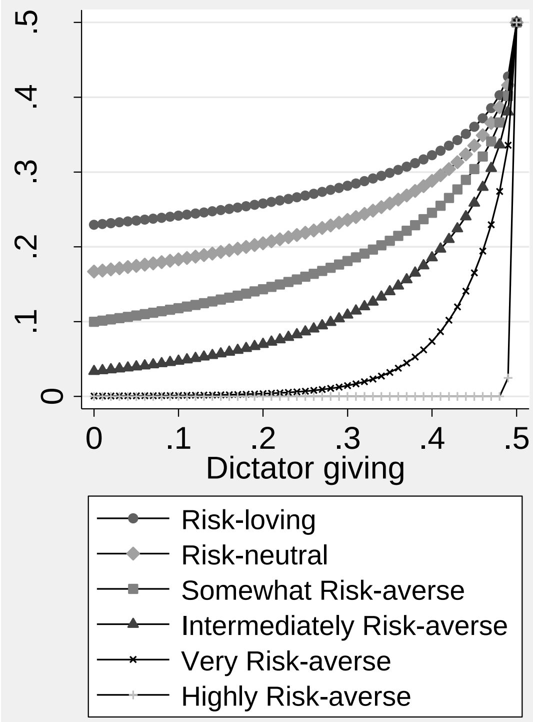
$$\frac{r(1-p_D)^{r-1}}{1-2p_D} = \frac{p_U{}^r}{(p_U - \frac{1}{2})^2} \tag{6}$$

We can then derive prediction intervals for giving choices in the Dictator game from the observed threshold choice in the Ultimatum game, and vice versa. As shown in Equation (6), the relationship depends on the degree of risk aversion, *r*. We cannot determine the precise value of the risk aversion parameter *r* using our version of Holt & Laury [5], but we can predict the upper and lower bounds. Our predictions thus consist of intervals. We can thus determine if an observed choice in the Dictator (Ultimatum) game falls in the prediction interval derived from choices in the Ultimatum (Dictator) game. Figure II.1 below illustrates our predictions.

Figure II.1 shows the relationship between giving in the Dictator game on the x-axis and thresholds in the Ultimatum game on the y-axis. The graph shows the predicted choices in the Ultimatum (Dictator) game, given the choices in the Dictator (Ultimatum) game. The prediction depends on the risk preferences; hence, several lines are drawn in the graph. For example, given some choice in the Dictator game, on the x-axis, the corresponding thresholds in the Ultimatum game can be read on the y-axis. We can see that highly and very risk-averse people have a low, mostly flat curve, predicting low Ultimatum (near 0) thresholds for people who give strictly positive amounts over the typical range in the Dictator game. Risk-loving people have a slowly increasing curve that starts at a relative high point (0.22). Likewise, the graph can be used to predict Dictator giving, given their threshold choices in the Ultimatum game.[22] Given the choice in the Ultimatum game, on the y-axis, the choice for giving in the Dictator game can be read on the x-axis.

[22] Note that for certain Ultimatum threshold levels no prediction of positive Dictator giving exists. For example, for Highly Risk-averse subjects, an Ultimatum threshold of 0.23 predicts zero Dictator giving. Ultimatum thresholds *lower* than 0.23 then predict *negative* Dictator giving, or "Dictator taking". As Dictator taking was not a choice option in the experiment, the 14 observations where the Ultimatum threshold predicts Dictator taking were excluded from the analysis.

Figure II.1 - Predictions across games based on the ERC formula



Relationship between the acceptance threshold in the Ultimatum game and the choice of giving in the Dictator game. If one goes from the x-axis to the y-axis, one predicts Ultimatum thresholds from Dictator giving choices, and *vice versa*.

The relation between Ultimatum thresholds and Dictator giving is non-decreasing, as increases in inequality aversion will increase both Dictator giving and threshold in the Ultimatum game. We thus formulate:

- ***Hypothesis 3:*** *The choices in Dictator game (scenarios) and Ultimatum game (scenarios) are consistent.*

3. Experimental Design and Implementation

To make hypotheses 1D and 1U experimentally testable, we specified endowments with increasing risk ("scenarios") for Dictator and Ultimatum games. In particular, we provided endowments in the form of lottery tickets S^i with $i \in \{N, L, H\}$ that realized a mean-preserving spread: $E[S^i] = 1000$. In the low-risk condition, the lottery S^L takes the value of 900 or 1100, with ½ probability each. In the high-risk condition, the lottery S^H takes the value of 300 or 1700, with ½ probability each. In the no-risk condition, the lottery S^N is fixed at 1000. S^H is thus a mean-preserving spread of S^L, which is in turn a mean-preserving spread of S^N. Table II.1 summarizes the different endowment risks.

Table II.1 - Operationalization of endowment risk

Games	Possible realizations of the pie sizes S^i (Dictator and Ultimatum)		
	No-risk condition S^N	Low-risk condition S^L	High-risk condition S^H
Dictator	-	900 or 1100	300 or 1700
Ultimatum	1000	900 or 1100	300 or 1700

We employed two different subject pools: students of the Czech Technical University (CTU) and employees of Prague City Hall (CH). The subject pool consisted of 44 CH employees and 116 CTU students. [23] Since the two subject pools are different and individual characteristics such as risk attitudes play an important role in our theoretical analysis, we controlled for these variables.

The experiment was conducted at CERGE-EI on a portable experimental laboratory. We conducted 14 experimental sessions. Table II.2 provides an overview of the experimental sessions:

Table II.2 - Overview of experimental sessions[24]

Session	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**Total**
Type of subjects	CTU	CTU	CTU	CTU	CTU	CTU	CTU	CTU	CTU^R	CTU^R	CH	CH	CH	CH	
Number of subjects	**14**	**12**	**16**	**14**	**12**	**16**	**16**	**16**	**16**	**16**	**10**	**8**	**16**	**10**	**192**

The experiment was programmed in z-Tree [26]. The experimental instructions were in Czech. The experimenter first read descriptions of the four games (see Appendix A.5 and A.8 for the scripted instructions and the instructions in the z-Tree program). All subjects then had to correctly answer two questions on payoff calculations (see Appendix A.7) before they could proceed to the 17 choices that constituted the experiment. All games, and their scenarios, were framed in abstract terms. Five choices were about risk aversion (see Holt & Laury [5]; explained in more detail below). The computer program randomly selected ex post

[23] We ran, at the tail end of the CTU sessions, two control sessions with an additional 32 CTU student subjects (sessions 9 and 10, indicated by CTU^R in Table II.2). The subjects in the control group read the same instructions but were given a different ordering of decisions: all low-risk treatments were switched to high-risk treatments and vice versa. This control group, which did not take part in any of the other sessions, was conducted to control for order effects. Two-sample Kolmogorov-Smirnov tests suggest that the control group differs statistically significantly from the CTU group in at least two respects (age and the measure of cognitive ability that we will discuss below). It is not clear to us why we find these differences in age and cognitive ability, although it might be due to the fact that we ran those sessions at the tail end of the CTU sessions. In addition, using again two-sample Kolmogorov-Smirnov tests, we found the following significant differences in the decisions of the control group: they gave less in the Dictator game, made a lower offer in the Trust game, and sent less back in the Trust game. We include the data in our analysis below to determine the extent of the estimated treatment effects attributable to decision-order effects with a dummy variable, CTU^R. This dummy is not significant in any of the regressions in the results section.

[24] Each session contained an even number of participants, and was constrained by the maximum lab capacity of 16 people. We attempted to have at least 12 participants in each session but scheduling the four CH sessions turned out to be difficult. There is no a-priori reason that we can think of that would suggest this was more than a logistic inconvenience.

one of these five choices to be paid out. The twelve remaining choices were paid in full. The fully paid choices were the offer and acceptance threshold in the Ultimatum game (under no-risk, low-risk and high-risk realizations of the pie size), dictator giving in the Dictator game, and investment and return decisions in the Trust game (under low-risk and high-risk realizations of the pie size).[25] Appendix A.6 contains the order in which these treatments were sequenced. No scenario realization was followed by the same scenario realization. We were aware that this kind of sequencing could lead to learning effects. Learning effects, however, are diminished when subjects are not informed about the outcomes of their decisions until the end of the session (e.g., Weber [27]), which is what occurred in our experiment. We thus ignore learning effects.

Each round random re-matching was used. As mentioned, subjects were informed of the outcomes of their decisions in the Dictator and Ultimatum games at the end of a session.[26] The exchange rate was 1 CZK per 20 experimental currency units (ECU); payoff per subject ranged from 190 CZK to 620 CZK (payoffs were rounded up to the nearest multiple of ten), and the average payoff was slightly below 400 CZK.[27] City Hall employees were paid an additional participation bonus of 150 or 200 CZK, as they had to commute to and from CERGE-EI and because sessions with CH employees lasted longer than those with student subjects (recall that the experiment proper started only once all participants in the session answered the questionnaire correctly, which on average took longer in the CH sessions).[28] At the end of each session, we asked all subjects to identify their age and gender, and to report their disposable income.

Since the model predicts that risk aversion influences individual decision making, we categorized - as mentioned - all subjects according to their risk aversion through an additional scenario (see Scenario Four in the Instructions, in Appendix A.5). Similar to the procedure in Holt & Laury [5], subjects had to choose between a series of safer and riskier options. In the first choice, the safer option had a higher expected value (EV) than the corresponding riskier one. In the following choice, the riskier option had a higher expected value than the corresponding safer one. With every choice, the EV of the riskier choice grew faster than the EV of the corresponding safer one (see Table II.3 below).[29] Standard theory predicts that an agent will make at most one switch, if any, across the five choices. Since we wanted the decisions to be independent, we did not provide the choices in a back-to-back manner, but

[25] Subjects were thus paid twice as recipients in the Dictator game (once for the low-risk and once for the high-risk condition).

[26] In the Trust game, subjects were informed of the amount they received once they were asked at the end of a session to make a decision as responder.

[27] When the experimental sessions were conducted, the exchange rate was about 23 CZK/1 USD, implying that our subjects – not counting appearance fees – earned on average 17–18 USD. The local purchasing power of these payoffs was about twice as much. Thus, it seems fair to say that the stakes were considerable for both students and City Hall employees. Since CH employees (and students) were told ex ante what average earning they could expect, we believe that only subjects that thought the money was worth their troubles signed up for the experiment.

[28] Sessions lasted from 60 to 100 minutes, with student sessions typically being in the lower half and the CH sessions in the upper half of the interval.

[29] As in Holt & Laury [5], subjects were not told the expected value of the options they were given.

interspersed them as questions 2, 6, 10, 12, and 16 respectively with the other Scenario questions. Table II.4 shows the risk aversion interval implied by the number of safe options the subject chooses.

Table II.3 - Elicitation of risk aversion

		Safer option			EV			Riskier option			EV
Choice 1:	1000	if n> 40	, 1250	otherwise	1100		60	if n> 40	, 2400	Otherwise	996
Choice 2:	1000	if n> 50	, 1250	otherwise	1125		60	if n> 50	, 2400	Otherwise	1230
Choice 3:	1000	if n> 60	, 1250	otherwise	1150		60	if n> 60	, 2400	Otherwise	1464
Choice 4:	1000	if n> 70	, 1250	otherwise	1175		60	if n> 70	, 2400	Otherwise	1698
Choice 5:	1000	if n> 80	, 1250	otherwise	1200		60	if n> 80	, 2400	Otherwise	1932

n is a random number between 1 and 100 and determines the payoff of the chosen option for each of the choices

Table II.4 - Implied ranges of risk aversion r

Number of safe choices	**Range of Relative Risk Aversion *r* for $U(x) = Sign(r)x^r$**		**Risk Preference Classification**
0	1.15	∞	Risk-loving
1	0.86	1.15	Risk-neutral
2	0.60	0.86	Somewhat Risk-averse
3	0.33	0.60	Intermediately Risk-averse
4	0.04	0.33	Very Risk-averse
5	$-\infty$	0.04	Highly Risk-averse

4. Results

As in Holt & Laury [5], we use the number of safe choices to characterize and measure risk aversion in our sample. Only slightly more than half our subjects (52.7% of the CTU and CTU^R groups, and 54.5% of the CH group, *i.e.*, 78 student subjects and 24 City Hall employees) made consistent choices.[30] The risk preferences of subjects that made inconsistent choices is either undetermined, or within a very wide range. Since we compare subjects with low and high risk aversion in the range of $0 < r < 1$, we need to determine risk preferences precisely to test our hypotheses. We thus excluded inconsistent subjects from our sample in the regressions. Table A.9 in Appendix A.1 shows the implied interval for risk aversion for all patterns of choice.

[30] We were aware ex ante (based, for example, on evidence reported in Hey & Orme [28]) that our procedure was likely to induce at least 25% inconsistent choices. Since we did not give our subjects the five risk-aversion measurement choices back to back, and did not give our subjects the EV of their options, with the benefit of hindsight, the fairly high percentage of inconsistent choices (not observed in the pilots in which we employed CERGE-EI students) is arguably not that surprising. It might simply reflect (as does also recent evidence on the stability of risk attitude assessment measures; see Dickhaut & Wilcox [29]) that other forms of elicitation may be confounded by subjects' attempts to be consistent. Harrison *et al.* [30] interpret inconsistent choices as an indication that subjects are indifferent between the different gambles and that their risk preference is thus part of a "fatter" interval.

To test whether inconsistent subjects are different from consistent subjects, we ran two-sample Kolmogorov-Smirnov tests on socio-demographic variables. The distribution of inconsistent subjects does not differ for the sex and income variables, but differ for the variables *Age*, *Time_to_answer* (the control questions), and *Safe* (the number of safe choices) at a statistically (but not necessarily otherwise) significant level. On average, inconsistent subjects were 1 day older, took 55 seconds longer to answer the control questions, and made 0.9 fewer safe choices than consistent subjects. With the possible exception of the number of safe choices, these differences seem inconsequential.

We note that, on average, Dictator giving is 19%, Ultimatum offers are 43%, and Ultimatum thresholds are 26%, which is in line with previous findings in the literature (see Camerer [8] and Güth & Ortmann [9]).

Before analyzing our game data, we characterize the determinants of risk aversion for "consistent" subjects through regression analyses. Table II.5 shows that socio-demographic characteristics influence risk aversion.

Table II.5 - Linear regression of risk aversion (defined as the safe choices) on socio-demographic variables

Variables	Effects
Income	**–0.11** (0.06)**
Age	0.05 (0.03)
Female	**0.67 * (0.37)**
Time_to_answer	0.06 (0.04)
City_Hall_employee	–0.63 (0.82)
CTU^R	0.07 (0.50)
Constant	**2.23 *** (0.75)**

Both students and City Hall employees are, on average, soundly risk-averse.[31] City Hall employees are in general less risk-averse than students, but the difference is not significant ($p > 0.42$).[32] Among our subjects, a person is predicted to be significantly more risk-averse when s/he earns a lower income ($p = 0.043$) or is female ($p = 0.065$), which is consistent with the literature (see Harrison & Rutström [31]). The variable *Time_to_answer* (the comprehension questions), which could be interpreted as a measure of cognitive ability, is not significant.

We use linear regression models clustered at the individual level to analyze the game data. We also use the robust Huber/White/sandwich estimator for the variance (Froot [32]).

[31] The average number of safe choices is 3.5 for students and 3.2 for City Hall employees. This result is not out of line with other evidence (e.g., [5]), which suggests the vast majority of subjects are rather risk-averse. Given the considerable stakes in our experiment, our risk-attitude results seem sensible.

[32] We ran, as a robustness test, ordered logistic regressions with the number of safe choices as the independent variable: Signs are unaffected and the significance levels are roughly the same. The effect of being female, however, is no longer significant ($p = 0.12$). The difference between City Hall employees and students is not significant ($p > 0.42$).

4.1. Risk Attitude and Inequality Aversion

(Hypothesis 1: Those with Higher Risk Aversion Show Higher Inequality Aversion in Their Responses.)

Table II.6 contains the results of the linear regressions with clustered standard errors models. It contains the determinants of the amounts given in the Dictator game and the threshold level in the Ultimatum game. We test hypotheses 1D and 1U by regressing the percentage of the amount given in the Dictator game and the threshold level in the Ultimatum game on various variables of interest, as shown in Table II.6. To measure the effect of risk aversion, we include a dummy variable, *Somewhat_versus_highly_risk_averse*, which equals zero for risk-loving and somewhat risk-averse subjects (0, 1 or 2 safe choices), and equals one for intermediately and highly risk-averse subjects (3, 4 and 5 safe choices).[33]

Table II.6 - Linear regression of giving on risk aversion and risk in the Dictator and Ultimatum games (with clustered errors)

	Dictator				**Ultimatum**			
	Model 1D1	**Model 1D2**	**Model 1D3**	**Model 1D4**	**Model 1U1**	**Model 1U2**	**Model 1U3**	**Model 1U4**
Dummy: Somewhat_versus_ highly_risk_averse	**9.59*** (3.60)**	**6.01* (3.46)**	**10.49*** (3.92)**	**6.91* (3.92)**	5.81 (3.96)	4.68 (4.08)	6.73 (4.10)	5.59 (4.24)
High_endowment_risk	−0.97 (1.81)	−0.97 (1.83)	0.30 (3.58)	0.30 (3.62)	**3.71*** (1.24)**	**3.71*** (1.25)**	**5.40* (2.93)**	**5.40* (2.95)**
Low_endowment_risk					1.39 (0.86)	1.39 (0.86)	1.63 (1.41)	1.63 (1.42)
Interaction Dummy: Somewhat_versus_ highly_risk_averse x High_endowment_risk			−1.80 (4.15)	−1.80 (4.19)			−2.40 (3.19)	−2.40 (3.21)
Interaction Dummy: Somewhat_versus_ highly_risk_averse x High_endowment_risk							−0.34 (1.77)	−0.34 (1.78)
Income		−0.00 (0.49)		−0.00 (0.50)		−0.64 (0.51)		−0.64 (0.51)
Age		−0.32 (0.32)		−0.32 (0.32)		−0.21 (0.27)		−0.21 (0.27)
Female		3.81 (3.65)		3.81 (3.66)		−1.95 (3.52)		−1.95 (3.53)
Time_to_answer		**1.80*** (0.37)**		**1.80*** (0.38)**		**0.61* (0.32)**		**0.61* (0.32)**

[33] As robustness tests, we rerun the regressions using, to capture the difference in risk preferences, the variable *safe*, the number of safe choices a subject made, instead of the dummy *Somewhat_versus_highly_risk_averse*. With the variable *safe*, the significance is overall lower: the relationship for Dictator giving in Model 1 stays significant ($p < 0.042$), but not in Model 2 ($p < 0.311$).

City_Hall_employees	**8.73*** **(4.62)**	3.98 (6.36)	**8.73*** **(4.63)**	3.98 (6.38)	−2.93 (4.08)	2.67 (7.08)	−2.93 (4.09)	2.67 (7.10)
CTU^R	−3.87 (5.16)	−6.46 (4.57)	−3.87 (5.18)	−6.46 (4.58)	−4.38 (4.94)	−5.26 (4.95)	−4.38 (4.96)	−5.26 (4.97)
Constant	**10.55***** **(3.27)**	**12.70*** **(7.45)**	**9.91***** **(3.51)**	**12.06*** **(7.20)**	**20.18***** **(3.71)**	**25.67***** **(7.28)**	**19.53***** **(3.80)**	**25.03***** **(7.28)**
N **(Independent clusters)**	204 (102)	204 (102)	204	204	306 (102)	306 (102)	306	306
R-squared	0.08	0.21	0.08	0.21	0.04	0.07	0.04	0.08

Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The models 1D1 and 1U1 test the hypotheses 1D and 1U with the simplest specifications. In addition, we run regressions including the socio-demographic variables (models 1D2 and 1U2), and we also run - as robustness tests - regressions with interaction dummies to test if there is a significant interaction between endowment risk and risk aversion (models 1D3, 1D4, 1U3 and 1U4).

Model 1D1 provides support for hypothesis 1D: subjects with relatively high risk aversion give significantly ($p < 0.01$) *more* in the Dictator game.[34] Model 1D2 shows that the effect is robust to the inclusion of the socio-demographic variables: the effect stays statistically significant ($p < 0.09$), albeit at a lower level of significance. Models 1D3 and 1D4 show that the interaction dummies, which have coefficients that are not significantly different from zero, do not affect the results.

Model 1U1 does not provide support for hypothesis 1U: while subjects with relatively high risk aversion set, in line with hypothesis 1U, higher thresholds in the Ultimatum game compared to those with relatively low risk aversion, the effect is not statistically significant. Surprisingly, endowment risk has a highly significant effect ($p < 0.01$) on the threshold level set in the Ultimatum game: an increase in endowment risk increases the threshold level. Model 1U3 shows that this effect occurs mostly independent of risk preferences: the effect of endowment risk on threshold remains significant ($p = 0.07$), albeit at a lower level of significance, in the presence of the interaction dummies that capture the interaction effects between endowment risk and risk aversion.

At first glance, there appears to be a subject-pool effect in Model 1D1, as City Hall employees tend to give significantly ($p = 0.06$) more in the Dictator game. Model 1D2 shows, however, that the subject-pool effect becomes insignificant when social-economic variables are included in the regression. In Model 1D2, the socio-economic variable *Time_to_answer* (the comprehension questions), which is possibly an indication of cognitive ability, is positive and significant ($p < 0.01$). We note that the coefficient of determination is quite low for all regressions in Table II.6.

[34] Running a Tobit regression, to account for the left-censoring of Dictator giving, gives the same significance levels for Models 1D1 and 1D3, and slightly higher ones for Models 1D2 and 1D4.

Even though, as mentioned, we cannot derive theoretical predictions for giving in the Ultimatum game, we ran a regression of Ultimatum giving on the same variables as in Table II.6 and we found that Ultimatum giving does not depend on risk aversion.

4.2. Consistency within Games

(Hypothesis 2: Increasing Endowment Risk does not have a Significant Effect)

Since we can only test hypotheses 2D and 2U by including somewhat risk-averse (but not risk-loving) subjects and very risk-averse (but not highly risk-averse) subjects, the number of observations is relatively small. In particular, we had to exclude not only inconsistent subjects, whose risk preference cannot be measured with sufficient precision to test hypothesis 1 and 2, but also highly risk-averse (5 safe choices) and risk-loving (0 safe choices) subjects, as they are not within the domain of hypothesis 2.[35]

To measure the effect of endowment risk, we include a dummy variable *High_endowment_risk* for testing hypothesis 2D and dummy variables *High_endowment_risk* and *Low_endowment_risk* for testing hypothesis 2U. Note also that the baseline is "low risk" in the Dictator game and "no risk" in the Ultimatum game.

Model 2D1 and 2U1 test the hypotheses with the simplest specifications. We also run regressions including the socio-demographic variables (models 2D2 and 2U2), and for hypothesis 2D we also run – as robustness tests – Tobit regressions (models 2D3 and 2D4) since many subjects made a choice of zero in the Dictator game.

Supporting hypothesis 2D, model 2D1 shows that the dummy for high-endowment risk is not statistically significant. The dummy for high-endowment risk is also insignificant in Model 2D2 (including the socio-demographic variables) and in models 2D3 and 2D4 (the robustness tests using Tobit regressions). Hypothesis 2D is thus supported by our data.

Contradicting hypothesis 2U, model 2U1 shows that both dummies for high-endowment risk ($p < 0.01$) and low-endowment risk ($p = 0.07$) are statistically significant.[36] This suggests that participants with typical risk-aversion preferences set significantly higher thresholds under risk than under certainty. Model 2U2 shows that including socio-demographic variables does not change the results. Hypothesis 2U is thus not supported by our data.

[35] Note that subjects with one safe choice include subjects who lean towards being somewhat risk-averse ($0.86 < r < 1$), subjects who are risk-neutral ($r = 1$), and subjects who lean towards being risk-loving ($1 < r < 1.15$); see Table II.4. Since theory predicts that an increase in risk aversion has an effect for risk-loving subjects (decrease giving) that is opposite to that for risk-averse subjects (increase giving), the inclusion of these subjects should lead us to underestimate the increase in giving. However, effects on giving are very small round the point where risk-loving changes to risk aversion ($r = 1$), and we can thus expect that the resulting underestimation will be small. Indeed, running a regression excluding also the subjects with 1 safe choice does not change the results qualitatively.

[36] An F-test shows that the dummies for high and low endowment risk are not significantly different ($p = 0.42$).

Table II.7 - "Weak" Patterns in Conditional Schedules, Absolute Linear regression of giving on risk aversion and endowment risk in the Dictator and Ultimatum games (with clustered errors)

	Dictator				Ultimatum	
	Model 2D1	**Model 2D2**	**Model 2D3 (Tobit)**	**Model 2D4 (Tobit)**	**Model 2U1**	**Model 2U2**
High_endowment_risk	−0.83 (1.96)	−0.83 (2.02)	−0.92 (2.52)	−0.93 (2.50)	**4.78***** **(1.65)**	**4.78***** **(1.68)**
Low_endowment_risk					**3.18*** **(1.73)**	**3.18*** **(1.76)**
Income		1.16 (1.17)		2.09 (1.42)		−2.05 (1.55)
Age		−0.49 (0.43)		−0.86 (0.59)		−0.08 (0.54)
Female		5.32 (5.88)		9.48 (7.71)		−1.20 (6.69)
Time_to_answer		**2.06**** **(0.89)**		**2.95**** **(1.25)**		0.55 (1.36)
City_Hall_employees	**14.90*** **(7.99)**	2.19 (12.07)	**18.80*** **(9.62)**	−3.74 (15.63)	−4.21 (8.58)	16.33 (17.12)
CTU^R	−3.53 (7.14)	−10.71 (7.88)	−8.78 (12.71)	−20.01 (13.97)	−1.88 (6.80)	−1.23 (7.80)
Constant	**15.59***** **(3.22)**	15.67 (10.53)	**10.04*** **(5.18)**	12.45 (13.50)	**24.47***** **(3.86)**	**30.28*** **(15.00)**

	Dictator				Ultimatum	
	Model 2D1	**Model 2D2**	**Model 2D3 (Tobit)**	**Model 2D4 (Tobit)**	**Model 2U1**	**Model 2U2**
N (Independent clusters)	80 (40)	80 (40)	80 (40)	80 (40)	120 60	120 60
R-squared (pseudo R-squared)	0.11	0.22	(0.02)	(0.04)	0.02	0.09

Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

As in the regressions for testing hypothesis 1, there appears to be a subject-pool effect in Model 2D1 and 2D3, as City Hall employees tend to set significantly higher thresholds (p = 0.07) in the Ultimatum game. As in testing hypothesis 1, when social-economic variables are included in the regression (see Models 2D2 and 2D4), the subject-pool effect becomes insignificant and the variable *Time_to_answer* (the comprehension questions) becomes positive and significant (p = 0.03). We note that the coefficient of determination is quite low for all regressions in Table II.7.

4.3. Consistency across Games

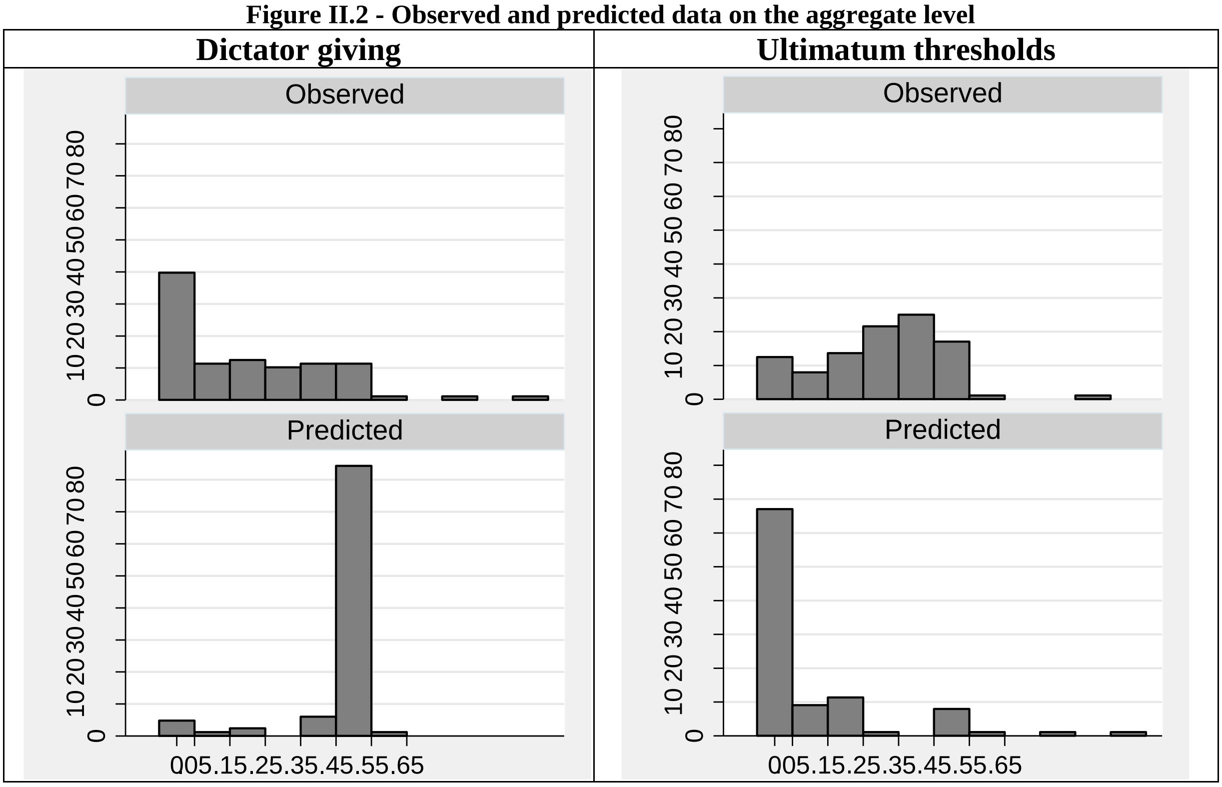
(Hypothesis 3: Subjects Are Consistent between Games)

We used Equation (6) and subjects' threshold choices in the Ultimatum game to derive prediction intervals from the observed threshold choices in the Ultimatum game for their giving choices in the Dictator game, and vice versa. This is a test of consistency between the two games on the individual level. Table II.8 shows that subjects do not make consistent choices in the two games: the percentages of successful prediction are very low. Only 10% of the responses observed are within the predicted interval.[37] We conclude that subjects are not consistent across games when measured on the individual level.

Table II.8 - Observed and predicted data on the individual level

	Successful	Unsuccessful	Total
Predict D from U and	8	75	83
visa versa	(10%)	(90%)	(100%)

Figure II.2 - Observed and predicted data on the aggregate level



[37] The success percentages are symmetric by design. The analysis in Table II.8 excluded the 14 responses where the choice for the Ultimatum threshold predicted Dictator *taking* (giving a negative amount, which was not a possible choice in the experiment). Including these 14 responses, and accepting the closest possible Dictator giving choice, zero, as a correct prediction, increases the correct percentage of predictions somewhat, from the 10% reported in Table II.8 to 20%.

We also examined how well the ERC formulation predicts aggregate behavior. Figure II.2 compares the predicted and observed choices for Dictator giving and Ultimatum thresholds, aggregated over all consistent subjects for both levels of endowment risk (low and high). Using the two-sample Kolmogorov-Smirnov tests, we reject the hypothesis that the observed and predicted distributions are the same for the Dictator game and the Ultimatum game ($p < 0.001$). The predicted distributions are thus poor approximations for the aggregate observed choices. Predictions at the aggregate level by and far fail. On the whole, this implies that subjects' behavior is not consistent across games, on both the individual level or on the aggregate level. These findings are only partially in line with Blanco *et al.* [20], who found that behavior measured at the individual level was not consistent, but behavior measured at the aggregate level was fairly consistent. Our predictions were, however, more fine-grained than those in Blanco *et al.* [20]. They were thus subjected to a stricter, more demanding test for consistency. We also note that our findings are based on specification of the extended ERC formula. It is therefore possible that other specifications fare better (or worse). Our specification is, however, one of the simplest ways to implement constant risk-averse preferences. Better specifications would have more complicated forms.

5. Concluding Discussion

Summarizing the results, Hypothesis 1D and 2D are supported. Specifically, those who are more risk-averse are also more inequality-averse in the Dictator game (1D). Though the sign is correct for the Ultimatum game responder (IU), it is not significant, possibly because the number of observations was too low. This tentatively supports our first hypothesis, that those with higher risk aversion are more inequality-averse (and thus possess a stronger preference for fairness) than those with lower risk aversion. Our finding that more risk-averse people tend to be more inequality-averse is roughly in line with the results in Ferrer-i-Carbonell & Ramos [3] and Carlsson *et al.* [4] who used survey data and non-incentivized choices between imagined societies and lotteries. Our somewhat weaker results may reflect the fact that findings of inequality aversion are likely to be susceptible to "price" sensitivity, *i.e.*, it is not difficult to be inequality-averse when there are no monetary consequences to making decisions.

Using the predictions derived from ERC theory, as implemented by a simple constant-relative-risk-aversion utility function (2), we find that endowment risk has no effect on giving in the Dictator game, confirming hypothesis 2D. Endowment risk does have a significant positive effect on the acceptance threshold in the Ultimatum game, contradicting hypothesis 2U. The effect, an increase in endowment risk leading to an increase in threshold, is in line with the prediction, but the size of the effect is much larger than predicted. This indicates that risk, which is not an ingredient of ERC theory, may affect acceptance thresholds in Ultimatum games.

We can thus conclude that the data corroborate the predictions from specification of the extended ERC formula for the Dictator game (1D and 2D), but not for the Ultimatum game (1U and 2U). We find that subjects are not consistent across games at the individual level, which contradicts hypothesis 3 but is consistent with the results of Blanco *et al.* [20]. At the aggregate level, we found that responses were not consistent across games, which is not in line with the results of Blanco *et al.* [20].

It is important to recall that people make inconsistent choices as viewed from the perspective of the theory that we use. One might conjecture that the functional form we used to incorporate risk preferences in the ERC formula may not be appropriate for the Dictator or Ultimatum game. The fact that our test of the effect of endowment risk on subject behavior, Hypothesis 2, is confirmed for the Dictator game, but contradicted for the Ultimatum game, suggests that the functional form we used may be appropriate for the Dictator game, but not for the Ultimatum game. We note that our test predicting the responses in one game from those in the other game is rather demanding test. These predictions are conditional on the risk preferences, which have been derived from our variant of the Holt & Laury [5] instrument. Basically, consistent choices in Hypothesis 3 thus require that the choices in *three* decision situations are consistent. That said, we know from other studies (like the one by Blanco *et al.* [20]) that others also have found that subjects make inconsistent choices.

Our data thus give tentative support to the claim that there is a positive relationship between risk aversion and fairness considerations. Our data also suggests that ERC theory, as formulated, does not seem particularly well suited to account for the effects of risk in the environment and risk preferences across the Dictator and Ultimatum games. The ERC formulation seems to fare better for the Dictator game.

We tested our theoretical predictions experimentally on two different subject pools: students of Czech Technical University – a subject pool we have drawn on previously that produced behavior in line with the behavior of student subjects elsewhere [33] – and employees of Prague City Hall. We generally did not find significant differences between the two groups, except for the regression of Dictator giving which indicated that employees of Prague City Hall give more in the Dictator game. This effect, however, is insignificant once socio-economic variables are incorporated in the regression. We included sessions where students were presented with the decision problems in a different order to control for order effects. These students show somewhat different responses on some of the variables of interest, indicating that order effects may play a role in the results of this study. The indicator variable for sessions with these students is, however, not significant in any of our tests, suggesting that order effects play a minor role in the results.

To summarize, we find that: (1) Those who are more risk-averse are also more inequality-averse in the Dictator game in that they give more. We believe this finding is novel. We find a similar result for the Ultimatum game but that result is statistically not significant; (2) Using the within-subject feature of our design, and in line with our theoretical prediction, varying risk does not influence behavior in the Dictator game, but does so in the Ultimatum game; (3) Using the within-subject feature of our design, subjects tend to make inconsistent decisions across games; this is true on the level of individuals (confirming the findings in Blanco *et al.* [20]) as well as in the aggregate (contradicting the findings in Blanco *et al.* [20]); (4) There are no subject-pool differences once we control for the elicited risk attitude and demographic variables that we collect.

Acknowledgments

Funds for the experiment were provided by GAUK (V. Babicky) and GACR (A. Ortmann). S. van Koten is grateful for the Jean-Monnet Fellowship and the financial support of the Robert Schuman Center for Advanced Studies at the EUI in Florence. We are grateful to Prague City Hall for helping us contact municipal employees and we thank our colleague Petra Brhlikova for her help in the organization of the experimental sessions. We benefited from comments on early versions of this article by Martin Dufwenberg, Werner Güth, Glenn Harrison, David K. Levine, Ondřej Rydval, Jakub Steiner, and participants of GfeW, FUR, and ESA meetings. We are also grateful for the very constructive contributions of the two (anonymous) referees and the guest editor, Dorothea Herreiner, of GAMES. The usual caveat applies.

Conflict of Interest

The authors declare no conflict of interest.

References

1. **Rawls, J., 1971.** *A Theory of Justice*; Belknap Press of Harvard University Press: Cambridge, MA, USA
2. **Chambers, C.P., 2012.** Inequality aversion and risk aversion. *J. Econ. Theory 147*, 1642–1651.
3. **Ferrer-i-Carbonell, A.; Ramos, X., 2010.** Inequality aversion and risk attitudes. SOEP papers 271, DIW Berlin: Berlin, Germany.
4. **Carlsson, F.; Daruvala, D.; Johansson-Stenman, O., 2005.** Are people inequality-averse, or just risk-averse? *Economica 72*, 375–396.
5. **Holt, C.; Laury, S., 2002.** Risk aversion and incentive effects. *Am. Econ. Rev. 92*, 1644–1655.
6. **Harrison, G.W.; Johnson, E.; McInnes, M.M.; Rutström, E.E., 2005.** Risk aversion and incentive effects: Comment. *Am. Econ. Rev. 95*, 897–901.
7. **Cherry, T.; Frykblom, P.; Shogren, J., 2002.** Hardnose the dictator. *Am. Econ. Rev. 92*, 1218–1221.
8. **Camerer, C.F., 2003.** *Behavioral Game Theory*; Princeton University Press: Princeton, NJ, USA.

9. **Güth, W.; Ortmann, A., 2006.** Behavioral approach to distribution and bargaining. In *Foundations and Extensions of Behavioral Economics: A Handbook*; Altman, M., Ed.; Sharpe Publishers: New York, NY, USA; pp. 405–422.
10. **Bolton, G.; Ockenfels, A., 2000.** ERC: A theory of equity, reciprocity and competition. *Am. Econ. Rev. 90*, 166–193.
11. **Fehr, E.; Schmidt, K., 1999.** Theory of fairness, competition and cooperation. *Quart. J. Econ. 114*, 817–868.
12. **Hoffman, E.; McCabe, K.; Smith, V.L., 1996.** Social distance and other-regarding behavior in Dictator games. *Am. Econ. Rev. 86*, 653–660.
13. **Bekkers, R., 2007.** Measuring altruistic behavior in surveys: The all-or-nothing Dictator game. *Surv. Res. Methods 1*, 139–144.
14. **Smith, V., 2010.** Theory and experiment: What are the questions? *J. Econ. Behav. Organ. 73*, 3–15.
15. **Engelmann, D.; Strobel, M., 2004.** Inequality aversion, efficiency and maximin preferences in simple distribution experiments. *Am. Econ. Rev. 94*, 857–869.
16. **Engelmann, D.; Strobel, M., 2007.** Preference over income distributions - Experimental evidence. *Public Finance Rev. 35*, 285–310.
17. **Engelmann, D., 2012.** How not to extend models of inequality aversion. *J. Econ. Behav. Organ. 81*, 599–605.
18. **Charness, G.; Rabin, M., 2002.** Understanding social preferences with simple tests. *Quart. J. Econ. 117*, 817–869.
19. **Cox, J.; Sadiraj, V., 2010.** Direct tests of individual preferences for efficiency and equity. *Econ. Inquiry 50*, 920–931.
20. **Blanco, M.; Engelmann, D.; Normann, H.T., 2011.** A within-subject analysis of other-regarding preferences. *Games Econ. Behav. 72*, 321–338.
21. **Güth, W.; Levati, M.V.; Ploner, M., 2008.** On the social dimension of time and risk preferences. *Econ. Inquiry 64*, 261–272.
22. **Cappelen, A.W.; Konow, J.; Sorensen, E.; Tungodden, B., 2013.** Just luck: An experimental study of risk taking and fairness. *Am. Econ. Rev. 103*, 1398-1413.
23. **Krawczyk, M.; Le Lec, F., 2010.** "Give me a chance!" And experiment in social decision under risk. *Exper. Econ. 13*, 500–511.
24. **Brennan, G.; González, L.G.; Güth, W.; Levati, M.V., 2008.** Attitudes toward private and collective risk in individual and strategic choice situations. *J. Econ. Behav. Organ. 67*, 253–262.
25. **Babicky, V., 2003.** Fairness under risk: Insights from dictator games. CERGE-EI Working Paper No. 217; The Center for Economic Research and Graduate Education, Economic Institute: Prague, Czech Republic.
26. **Fischbacher, U., 2007.** Z-Tree, Zurich toolbox for ready-made economic experiments. *Exp. Econ. 10*, 171–178.
27. **Weber, R.A., 2003**. "Learning" with no feedback in a competitive guessing game. *Games Econ. Behav. 44*, 134–144.

28. **Hey, J.; Orme, C., 1994.** Investigating Generalizations of expected utility theory using experimental data. *Econometrica 62*, 1291–1326.
29. **Dickhaut, J.; Wilcox, N., 2010.** Elicitation of risk attitudes, a new assessment battery. Mimeo.
30. **Harrison, G.W.; Lau, M.I.; Rutström, E.E.; Sullivan, M.B., 2005.** Eliciting risk and time preferences using field experiments: Some methodological issues. In *Field Experiments in Economics: Research in Experimental Economics*; Harrison, G.W., Carpenter, J., List, J.A., Eds.; Emerald Group Publishing Limited: Bingley, UK; Volume 12, pp. 125–218.
31. **Harrison, G.W.; Rutström, E.E., 2008.** Risk aversion in the laboratory. In *Research in Experimental Economics*; Cox, J.C., Harrison, G.W., Eds.; Emerald: Bingley, UK; Volume 12, pp. 41–196.
32. **Froot, K.A., 1989.** Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data. *J. Finan. Quant. Anal. 24*, 301–312.
33. **Rydval, O.; Ortmann, A., 2004.** How financial incentives and cognitive abilities affect task performance in laboratory settings: An illustration. *Econ. Lett. 85*, 315–320.

A. Appendix

A.1. Overview of Responses on the Holt-Laury Test

Table A.9 - The patterns of answers on the Holt-Laury test and the risk-aversion interval

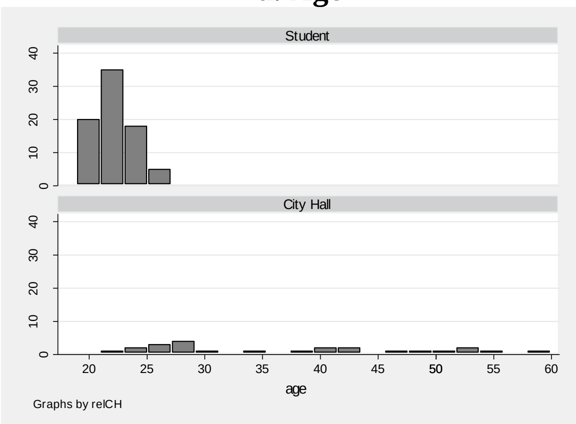
1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	Interval	Risk aversion	Occurrence
Subjects that were consistent in their choices							
RISKY	RISKY	RISKY	RISKY	RISKY	[1.15; +∞]	Risk-loving	9%
SAFE	RISKY	RISKY	RISKY	RISKY	[0.86; 1.15]	Risk-neutral	4%
SAFE	SAFE	RISKY	RISKY	RISKY	[0.60; 0.86]	Somewhat Risk-averse	4%
SAFE	SAFE	SAFE	RISKY	RISKY	[0.33; 0.60]	Intermediately Risk-averse	4%
SAFE	SAFE	SAFE	SAFE	RISKY	[0.04; 0.33]	Very Risk-averse	9%
SAFE	SAFE	SAFE	SAFE	SAFE	[−∞; 0.04]	Highly Risk-averse	24%
Subjects that had fat risk aversion intervals							
SAFE	RISKY	RISKY	RISKY	SAFE	[−∞; 1.15]	Undeterminable	2%
SAFE	RISKY	RISKY	SAFE	RISKY	[0.04; 1.15]	Undeterminable	2%
SAFE	RISKY	RISKY	SAFE	SAFE	[−∞; 1.15]	Undeterminable	3%
SAFE	RISKY	SAFE	RISKY	RISKY	[0.33; 1.15]	Undeterminable	2%
SAFE	RISKY	SAFE	RISKY	SAFE	[−∞; 1.15]	Undeterminable	1%
SAFE	RISKY	SAFE	SAFE	RISKY	[0.04; 1.15]	Undeterminable	2%
SAFE	RISKY	SAFE	SAFE	SAFE	[−∞; 1.15]	Undeterminable	3%
SAFE	SAFE	RISKY	SAFE	RISKY	[0.04; 0.86]	Undeterminable	3%
SAFE	SAFE	RISKY	SAFE	SAFE	[−∞; 0.86]	Undeterminable	1%
SAFE	SAFE	SAFE	RISKY	SAFE	[−∞; 0.60]	Undeterminable	2%
RISKY	RISKY	RISKY	SAFE	RISKY	[0.04; +∞]	Undeterminable	4%
RISKY	RISKY	SAFE	RISKY	RISKY	[0.33; +∞]	Undeterminable	3%
RISKY	SAFE	RISKY	SAFE	RISKY	[0.04; +∞]	Undeterminable	4%
RISKY	SAFE	SAFE	RISKY	RISKY	[0.33 ; +∞]	Undeterminable	1%
Subjects that made contradictory choices							
RISKY	RISKY	RISKY	RISKY	SAFE	[−∞; +∞]	Undeterminable	3%
RISKY	RISKY	RISKY	SAFE	SAFE	[−∞; +∞]	Undeterminable	4%
RISKY	RISKY	SAFE	SAFE	SAFE	[−∞; +∞]	Undeterminable	1%
RISKY	SAFE	RISKY	RISKY	SAFE	[−∞; +∞]	Undeterminable	1%
RISKY	SAFE	RISKY	SAFE	SAFE	[−∞; +∞]	Undeterminable	1%
RISKY	SAFE	SAFE	RISKY	SAFE	[−∞; +∞]	Undeterminable	1%
RISKY	SAFE	SAFE	SAFE	SAFE	[−∞; +∞]	Undeterminable	7%

A.2. Overview of the Socio-demographic Characteristics of the Subjects

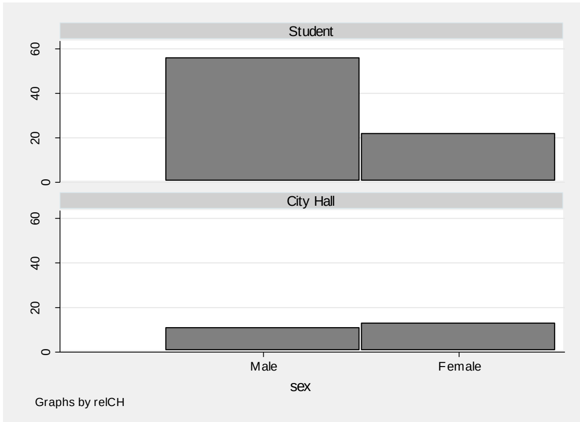
Figure A.3 shows the distribution of our two subject pools conditional on the socio-demographic variables.

Figure A.3 - Socio-demographic characteristics of City Hall employees and students

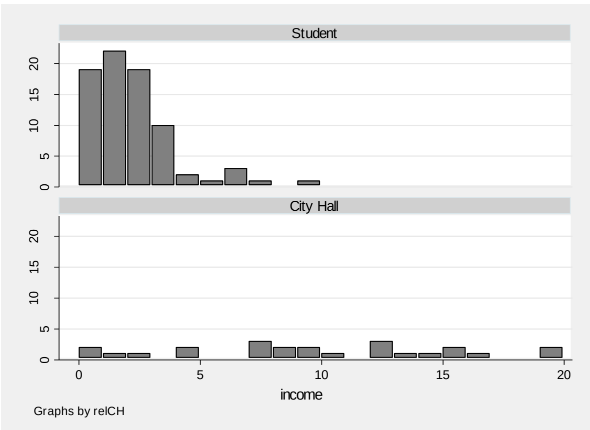
a. Age



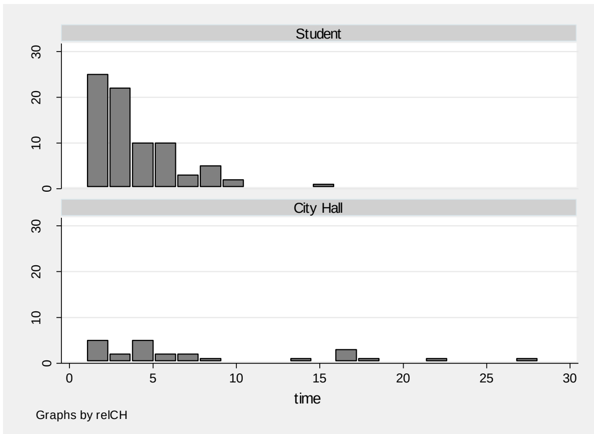
b. Female



c. Income in 1000 CKZ

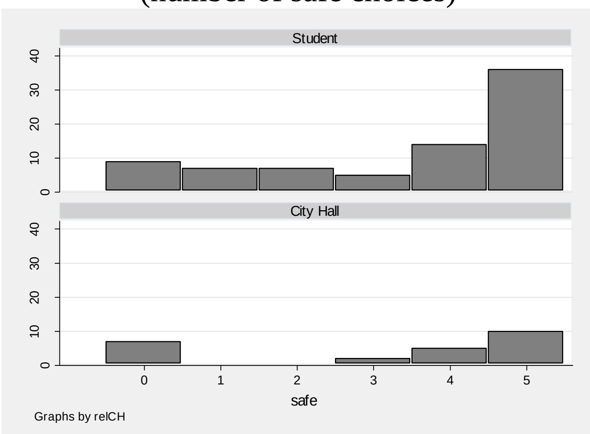


d. Time to answer the control questions in minutes



e. Risk Preferences

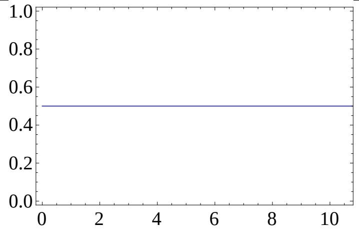
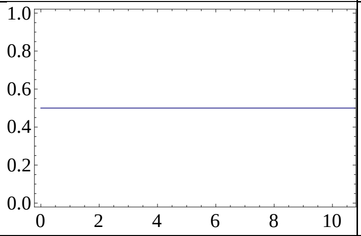
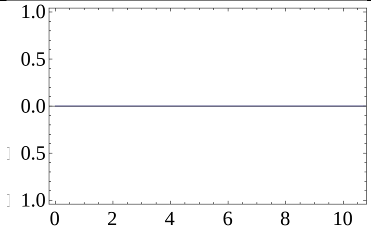
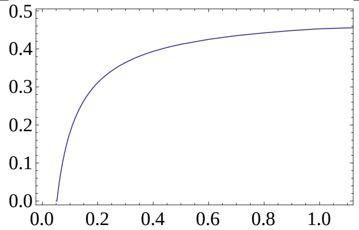
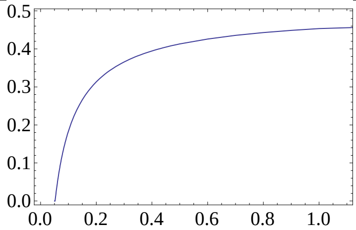
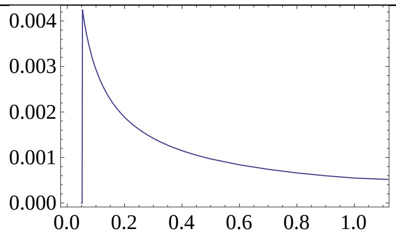
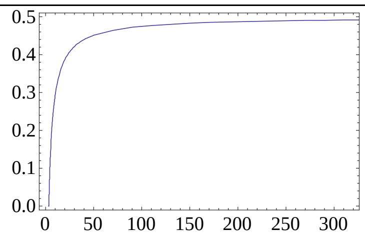
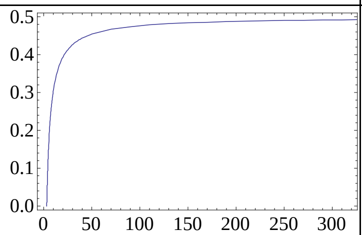
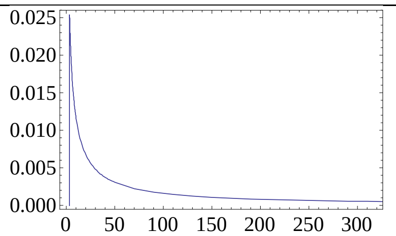
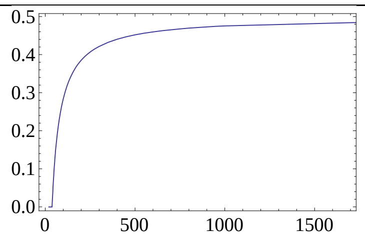
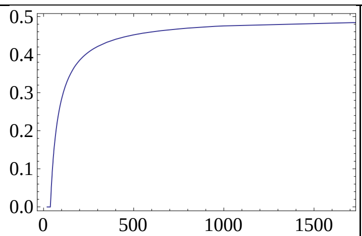
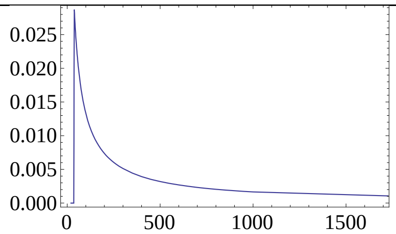
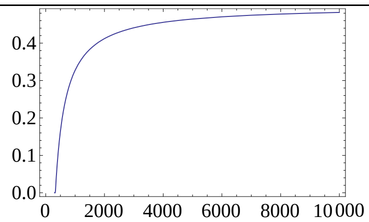
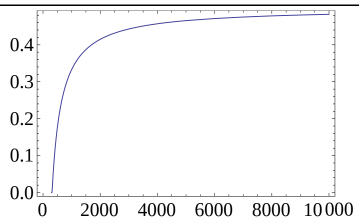
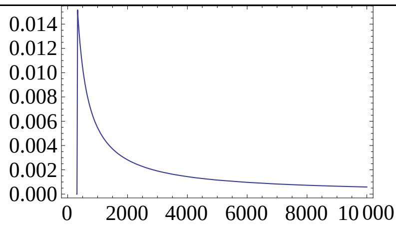
(number of safe choices)

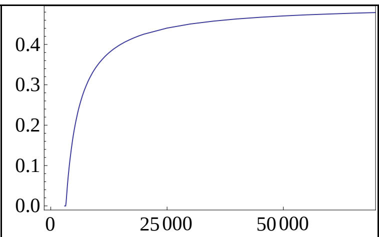
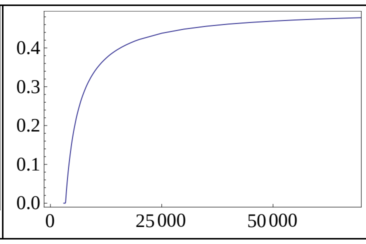
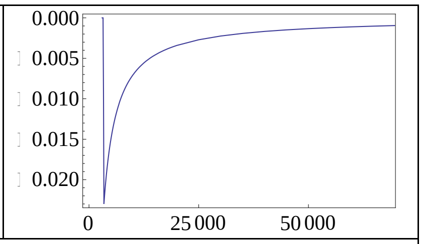


As shown in the histograms, the socio-demographic characteristics differ markedly across the two subject pools, though there is also considerable overlap. In Figure A.3.e, we show the number of safe choices in the risk-attitude assessment task. The number of safe choices in tandem with Table II.4 allows us to determine a subject's degree of risk aversion.

A.3. Simulations on the Effect of Risk in the Dictator and Ultimatum Games on Decisions

Figure A.4 - Effect of risk on giving in the Dictator game

Risk aversion	**Column 1: Low risk** (S = 900 or S = 1100 with equal likelihood)	**Column 2: High risk** (S = 300 or S = 1700 with equal likelihood)	**Column 3: Difference in optimal giving *p* between High and Low risk**
r = −0.14			
r = 0.04			
r = 0.33			
r = 0.60			
r = 0.86			

r = 1.15			

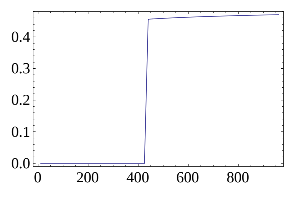
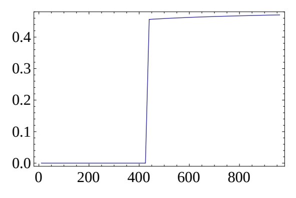
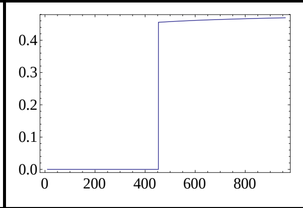
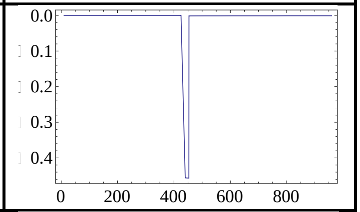
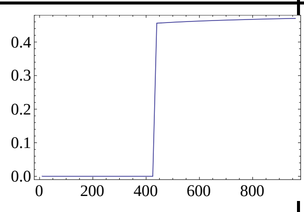
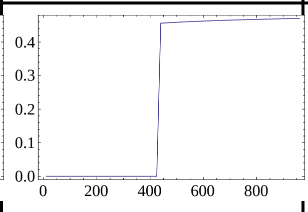
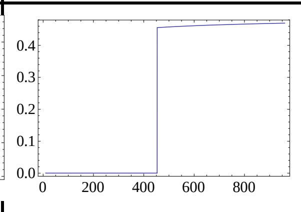
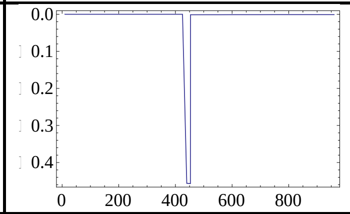
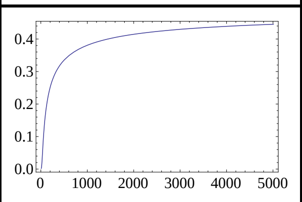
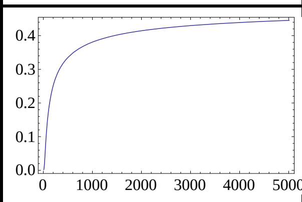
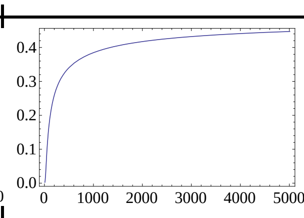
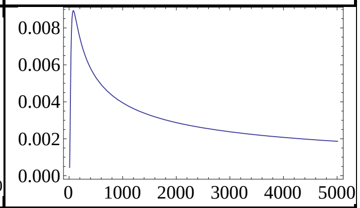
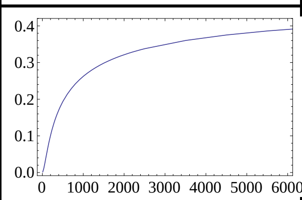
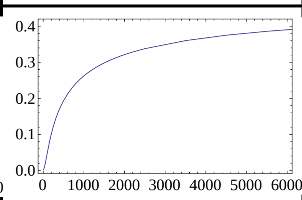
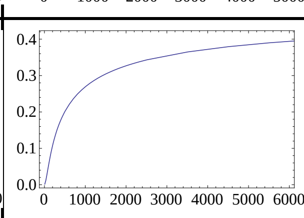
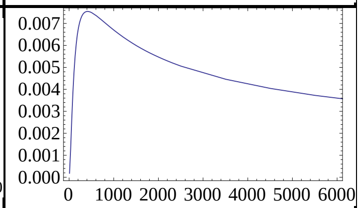
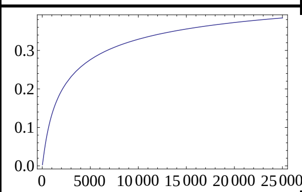
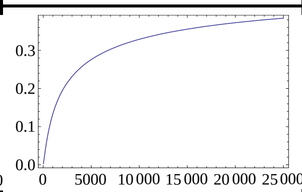
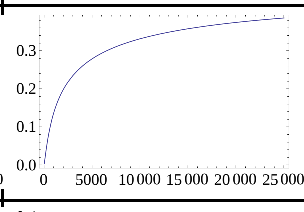
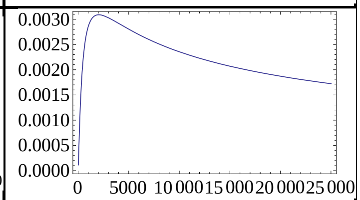
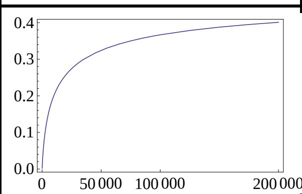
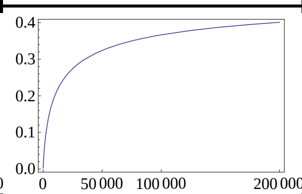
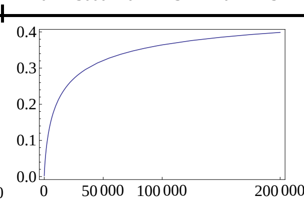
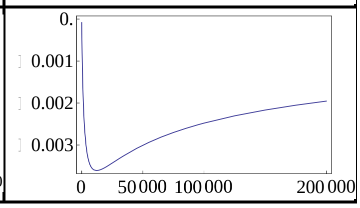
On the vertical axes is the proportion of giving *p*; On the horizontal axes is the inequality aversion parameter *k*.

Figure A.4 shows giving in the Dictator game as a function of the inequality aversion parameter *k* and the risk aversion parameter *r* for our formulation of the ERC model.[38] The relationship is shown for the low-risk condition in the first column and for the high-risk condition in the second one. The third column shows the difference between the two graphs, which shows that the absolute difference is never larger than 0.03 for risk-averse subjects, and is never larger than 0.025 for risk-loving subjects. These differences are very small, and moreover, they are maxima over all possible inequality aversion parameters for the five degrees of risk aversion we consider. For example, differences are predicted to be no larger than 0.0043 for very risk-averse subjects (r = 0.04) and no larger than zero for highly risk-averse subjects (r = −0.14). We thus expect that risk will not affect dictator giving in the experiment.

Figure A.5 shows acceptance thresholds in the Ultimatum game as a function of the inequality aversion parameter *k* and the risk aversion parameter *r* for our formulation of the ERC model. The relationship is shown for the no-risk condition in the first column, for the low-risk condition in the second one, and for the high-risk condition in the third one. The fourth column shows the difference between the high-risk and the no-risk graphs (as they are the most different), and shows that for risk-loving subjects, the absolute difference is never larger than 0.004. The difference is never larger than 0.009 for somewhat risk-averse preferences. For very and highly risk-averse subjects, the difference can be substantial – up to almost 0.5. This is the result of a sudden switchover in the threshold from zero to close to 0.5 when the inequality aversion parameter *k* passes a certain value. With larger risk, the switchover occurs at a slightly higher *k*, thus resulting in a large difference. This difference is, however, predicted to exist only for a very narrow range of the inequality aversion parameter *k*. Over all values of *k* that are outside of this very narrow range, the difference is virtually zero. We thus expect that risk will not affect the acceptance thresholds set in the experiment.

[38] For each of the levels of risk aversion we studied, we created a grid of values for the inequality aversion parameter *k* and, using formula (4), calculated the optimal giving for each of the values of *k*, given the level of risk aversion and the level of risk. We use these coordinates to draw the figures in Figure A.4. We used a likewise procedure, using formula (4’) for drawing the figures in Figure A.5 for the Ultimatum games. We programmed the algorithms in Mathematica.

Figure A.5 - Effect of risk on acceptance threshold in the Ultimatum game

Risk aversion	**Column 1: No risk** (*S* = 1000 with certainty)	**Column 2: Low risk** (*S* = 900 or *S* = 1100 with equal likelihood)	**Column 3: High risk** (*S* = 300 or *S* = 1700 with equal likelihood)	**Column 4: Difference in optimal giving *p* between High and No risk**
r = -0.14				
r = 0.04				
r = 0.33				
r = 0.60				
r = 0.86				
r = 1.15				

On the vertical axes is the proportion of giving *p*; On the horizontal axes is the inequality aversion parameter *k*.

A.4. Prediction Errors on the Individual Level

Taking into account that subjects may have made errors in their choices, we look at the size of the error between the observed choices and the prediction interval.

Figure A.6 - Prediction error

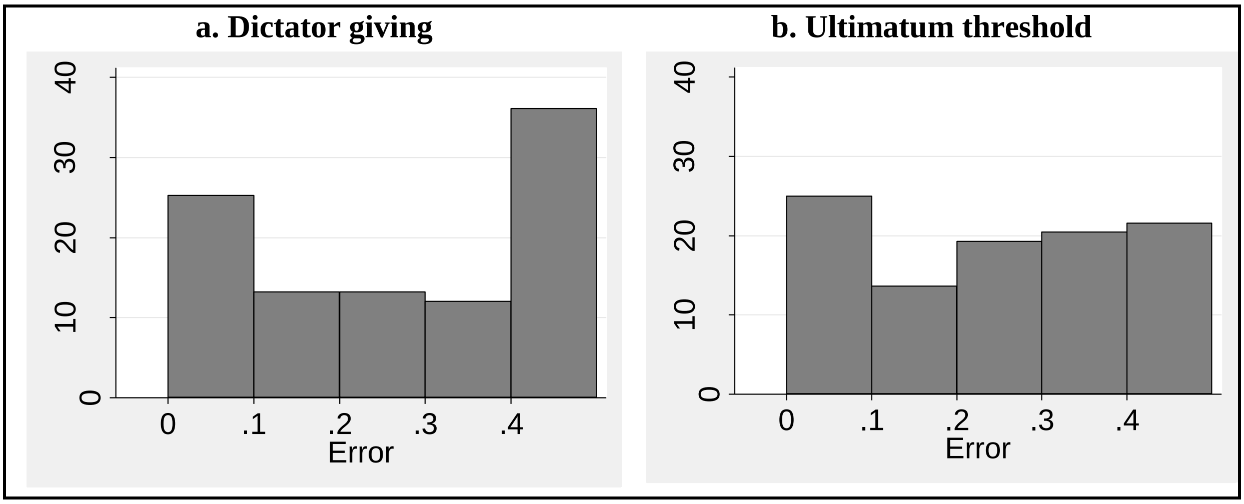


Figure A.6 shows that the overall prediction error is high both for Dictator giving and Ultimatum thresholds. The first bar in the histogram in Figure A.6.a accounts for all predictions of Dictator giving that have an error equal to 10 percentage points or less. It can be read from the vertical axis that these responses account for 25% of the total responses in the Dictator game. The first and second bar together account for all predictions that have an error equal to 20 percentage points or less: together they account for less than 40% of all responses in the Dictator game. Likewise, Figure A.6.b shows that the predictions for Ultimatum thresholds that have an error equal to 10 percentage points or less account for 25% of all responses and those that have an error equal to 20 percentage points or less account for 40% of all responses. The prediction error is thus considerable. We conclude that subjects are far from consistent across games when measured on the individual level.

A.5. Scripted Instructions

Welcome! You are about to participate in an economics experiment. You will be asked to make a series of decisions. Your decisions will have payoff consequences that will also depend on other participants' decisions. You will be paid privately in cash immediately after the experiment is over. You will get 1 CZK for each 20 ECU (experimental currency units) that you earn during the experiment.

We ask that from now on you refrain from any communication, whether verbal or nonverbal, with other participants. If you have a question, raise your hand and an experimenter will come to assist you.

Throughout the experiment you will, for every single decision (where applicable), be matched randomly with one other participant. The probability that you will be matched with the same participant for more decisions is therefore rather low.

All in all you will be asked to make 17 decisions. You will be informed about the payoff consequences of any of these decisions only after you have made your last decision.

[Any questions?]

During the experiment we will use the following <u>three basic scenarios</u> (labeled One, Two, and Three). The computer is instructed to match you randomly with some other participant of the experiment in each of these three scenarios (for every decision you will get a new match). You will also be asked about your preferences (<u>Scenario Four</u>). In this Scenario your payoff cannot be affected by the decision of another participant.

<u>Scenario One</u> involves a pie of size S that is being divided between two participants that we call Participant A and Participant B. Task of Participant A is to split the pie of size S in any way he or she sees fit. Participant B is the recipient of what Participant A allocates; he or she will not make any decision in this scenario. <u>Participant A will be asked to state her or his decision as a number between 0 and 100, *i.e.*, as a percentage of pie size S that he or she allocates to Participant B.</u>

[Any questions?]

<u>Scenario Two</u> involves a pie of size S that is being divided between two participants which we call Participant C and Participant D. Task of Participant C is to split the pie of size S in any way he or she sees fit. But now Participant D may either accept the offer or reject it. Participant C will be asked to state her or his decision of how to split the pie as a number between 0 and 100 (<u>the "offer"</u>), representing <u>a percentage of pie size S that he or she offers to Participant D</u>. Participant D will also be asked to state her or his decision whether he or she accepts the offer in a similar way as a number between 0 and <u>100 (the "acceptance threshold"</u>) representing <u>the minimal offer for which Participant D will not reject the offer</u>. If the acceptance threshold of Participant D is higher than the offer that Participant C made, then the offer is not accepted, and both participants will be paid nothing for this scenario. Otherwise, they will be paid in accordance with the split that Participant C proposed.

[Any questions?]

<u>Scenario Three</u> involves Participants E and F. Participant E is endowed with 500 ECU out of which he or she can send any amount of his or her choice (from 0 to 500 ECU) to Participant F (the rest of the 500 ECU endowment stays on the account of Participant E). The amount sent to Participant F will be multiplied by a factor X before it reaches Participant F. It is then task of Participant F to split the amount received (*i.e.*, X times the amount sent) in any way he or she sees fit. Participant E is the recipient of what Participant F allocates.

[Any questions?]

<u>Scenario Four.</u> The computer assigned to you at the beginning of the experiment a natural number N from 1 to 100 (any number 1, 2, 3, …, 100 is equally likely). This number will be revealed to you only at the end of the experiment. You will have to choose one of the two options "+" or "*". On the screen, you have to fill a blank box with your choice of "+" or "*"

and then press the “OK” button. Once you have pressed the “OK” button, you will not be able to go back. The computer is programmed to randomly select one of five such decisions you made during the whole experiment at the end of the experiment. For this purpose, the program uses a generator of random numbers. Choosing any of the five decisions in Scenario Four is equally likely. You will be paid at the end according to your choice in the selected decision and your personal N. (Note that in Scenario Four you do not interact with any other player.)

Example: choice +: 1000 ECU if N>40, 1250 ECU otherwise

or *: 60 ECU if N>40, 2400 ECU otherwise

(note that numbers will vary across decisions)

[Any questions?]

[Please turn your attention now to the computer screen but keep these hard copy instructions readily accessible.]

Thank you for participating in the experiment.

A.6. Sequencing of decisions

The sequencing of the decisions was the same for all participants, except those in sessions 9 and 10; see footnote 6 for an explanation:

Decision 1: Ultimatum proposal with no risk (pie size 1000)
Decision 2: Risk attitude measurement (n>40, *i.e.*, Choice 1)
Decision 3: Dictator with low risk (pie size 900 or 1100)
Decision 4: Trust game sending with high risk (factor 1.2 or 2.8)
Decision 5: Ultimatum proposal with high risk (pie size 300 or 1700)
Decision 6: Risk attitude measurement (n>50, *i.e.*, Choice 2)
Decision 7: Ultimatum threshold with high risk (pie size 300 or 1700)
Decision 8: Trust game sending with low risk (factor 1.8 or 2.2)
Decision 9: Ultimatum proposal with low risk (pie size 900 or 1100)
Decision 10: Risk attitude measurement (n>60, *i.e.*, Choice 3)
Decision 11: Ultimatum threshold with low risk (pie size 900 or 1100)
Decision 12: Risk attitude measurement (n>70, *i.e.*, Choice 4)
Decision 13: Dictator with high risk (pie size 300 or 1700)
Decision 14: Ultimatum threshold with no risk (pie size 1000)
Decision 15: Trust game return with high risk (factor 1.2 or 2.8)
Decision 16: Risk attitude measurement (n>80, *i.e.*, Choice 5)
Decision 17: Trust game return with low risk (factor 1.8 or 2.2)

A.7. Control Questions

Consider scenario One with the total amount of S=200 ECU. You have the role of participant A and you made a choice of 84.

- Question 1: What are your earnings from this scenario?
- Question 2: What are the earnings of participant B, who has been selected randomly and assigned to you for this scenario. Please fill out your answer in the space above and confirm.

A.8. Instructions in the z-Tree Program

In the z-Tree program, participants were given instructions and referred to the printed instructions to guide their understanding of the decision tasks. In Table A.10 below, we give the full instructions as they appeared in the z-Tree program. In the left column are the original Czech instructions and in the right column the corresponding English translations. Each part starts with a code in brackets [], that identifies the task (Dictator, Ultimatum or Trust game or Holt-Laury task), the role for the participant (Proposer or Respondent role) and the degree of risk in the task (none, low, or high). The code in brackets was not presented to the participants.

Abbreviations used

D = Dictator Scenario	P = Proposer role	none = no-risk condition
U = Ultimatum Game	R = Respondent role	low = low-risk condition
T = Trust Game		high = high-risk condition
HL= Holt-Laury task		

Table A.10 - Instructions in the z-Tree program

Original (Czech)	English translation
[Questionnaire] Nyni overime, ze vsichni porozumeli zakladnim scenarum, ktere popisuji instrukce. Odpovedi na nasledujici dve otazky nebudou mit dopad na Vasi vyplatu z experimentu, ale pokracovat budete moci jen po jejich spravnem zodpovezeni.	[Questionnaire] We will now ensure that everybody understood the basic scenarios from the instructions. Your answers on the two following questions will not have any effect on your earnings from the experiment, but you will be able to continue only after you have answered them correctly.

Mate-li otazku, zvednete ruku a experimentator Vam prijde odpovedet. Uvazujte scenar Jedna s celkovou castkou S=200 ECU. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Jste v roli ucastnika A a sve rozhodnuti jste vyjadril cislem 84. Otazka 1: Kolik je Vas zisk z pouziti tohoto scenare? Otazka 2: Kolik je zisk ucastnika B, ktery k Vam byl pro pouziti tohoto scenare nahodne prirazen? Vase odpovedi prosim vyplnte v jednotkach ECU do okenek vyse a potvrdte!	If you have an answer, please raise your hand and an experimenter will come. Consider scenario One with the total S=200 ECU (If you need to, you may return to the written instruction that describes this scenario) You have the role of participant A and you made a choice of 84. Question 1: What are your earnings from this scenario? Question 2: What are the earnings of participant B, who has been selected randomly and assigned to you for this scenario? Please fill out your answer in ECUs in the space above and confirm!
[U_P-none] Od nynejska budou mit vsechna Vase rozhodnuti dusledky na vysi Vasi vyplaty z tohoto experimentu. Proto davejte dobry pozor na ukoly, o ktere Vas budeme zadat. Uvazujte Scenar Dva s celkovou castkou S= 1000 ECU. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Jste v roli ucastnika C a pro ucely scenare jste byl pocitacem nahodne prirazen k jinemu ucastnikovi experimentu. Rozhodnuti 1: Jaka je Vase nabidka? Prosim vyplnte Vasi odpoved do okenka a potvrdte!	[U_P-none] From now on, all your decisions will affect your earnings from this experiment. Therefore, pay close attention to the tasks we will present to you. Consider Scenario Two with the total S=200 ECU (If you need to, you may return to the written instruction that describes this scenario) You have the role of participant C and for this scenario you have been randomly assigned to another participant in this experiment. Decision 1: What is your proposal? Please fill out your answer in the space above and confirm!

[HL-40] Uvazujte scenar Ctyri. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Volba "+": Ziskate 1000 ECU pokud N>40, jinak ziskate 1250 ECU Volba "*": Ziskate 60 ECU pokud N>40, jinak ziskate 2400 ECU Rozhodnuti 2: Kterou volbu preferujete?	[HL-40] Consider scenario Four (If you need to, you may return to the written instruction that describes this scenario) Choice “+”: You receive 1000 ECU when N>40, otherwise you receive 1250 ECU Choice “*”: You receive 60 ECU when N>40, otherwise you receive 2400 ECU Decision 2: What choice do you prefer?

[D] Nyni uvazujte scenar Jedna. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Celkova castka S je nahodna a se stejnou pravdepodobnosti je bud S=900 ECU nebo S=1100 ECU. Jste v roli ucastnika A a pro ucely scenare jste byl pocitacem nahodne prirazen k jinemu ucastnikovi experimentu. Rozhodnuti 3: Jake je Vase rozhodnuti, kolik procent prevest?	[D] Now consider Scenario One. (If you need to, you may return to the written instruction that describes this scenario) The total amount S is random and is, with equal likelihood, equal to either S=900 ECU or S=1100 ECU. You have the role of participant A and for this scenario you have been randomly assigned to another participant in this experiment. Decision 3: What is your decision, how much as a percentage will you transfer?
[T_P] Uvazujte nyni scenar Tri. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Faktor X je nahodne cislo a se stejnou pravdepodobnosti je bud X=1.2 nebo X=2.8. Jste v roli ucastnika E a pro ucely scenare jste byl pocitacem nahodne prirazen k jinemu ucastnikovi experimentu. Rozhodnuti 4: Kolik ECU posilate ucastnikovi F?	[T_P] Consider Scenario Three. (If you need to, you may return to the written instruction that describes this scenario) Factor X is a random number and is, with equal likelihood, equal to either X=1.2 or X=2.8. You have the role of participant E and for this scenario you have been randomly assigned to another participant in this experiment. Decision 4: How much ECU will you send to participant F?
[U_P-high] Uvazujte Scenar Dva. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Celkova castka S je nahodna a se stejnou pravdepodobnosti je bud S=300 ECU nebo S=1700 ECU. Jste v roli ucastnika C a pro ucely scenare jste byl pocitacem nahodne prirazen k jinemu ucastnikovi experimentu. Rozhodnuti 5: Jaka je Vase nabidka? Prosim vyplnte Vasi odpoved do okenka a potvrdte!	[U_P-high] Consider Scenario Two. (If you need to, you may return to the written instruction that describes this scenario) The total amount S is random and is, with equal likelihood, equal to either S=300 ECU or S=1700 ECU. You have the role of participant C and for this scenario you have been randomly assigned to another participant in this experiment. Decision 5: What is your proposal? Please fill out your answer in the space above and confirm!

[HL-50] Uvazujte scenar Ctyri. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Volba "+": Ziskate 1000 ECU pokud N>50, jinak ziskate 1250 ECU Volba "*": Ziskate 60 ECU pokud N>50, jinak ziskate 2400 ECU Rozhodnuti 6: Kterou volbu preferujete?	[HL-50] Consider scenario Four (If you need to, you may return to the written instruction that describes this scenario) Choice “+”: You receive 1000 ECU when N>50, otherwise you receive 1250 ECU Choice “*”: You receive 60 ECU when N>50, otherwise you receive 2400 ECU Decision 6: What choice do you prefer?
[U_R-high] Uvazujte Scenar Dva. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Celkova castka S je nahodna a se stejnou pravdepodobnosti je bud S=300 ECU nebo S=1700 ECU. Jste v roli ucastnika D a pro ucely scenare jste byl pocitacem nahodne prirazen k jinemu ucastnikovi experimentu. Rozhodnuti 7: Jaky je Vas prah akceptovatelnosti? Prosim vyplnte Vasi odpoved do okenka a potvrdte!	[U_R-high] Consider Scenario Two. (If you need to, you may return to the written instruction that describes this scenario) The total amount S is random and is, with equal likelihood, equal to either S=300 ECU or S=1700 ECU. You have the role of participant D and for this scenario you have been randomly assigned to another participant in this experiment. Decision 7: What is your proposal? Please fill out your answer in the space above and confirm!
[T_P-high] Uvazujte nyni scenar Tri. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Faktor X je nahodne cislo a se stejnou pravdepodobnosti je bud X=1.8 nebo X=2.2. Jste v roli ucastnika E a pro ucely scenare jste byl pocitacem nahodne prirazen k jinemu ucastnikovi experimentu. Rozhodnuti 8: Kolik ECU posilate ucastnikovi F?	[T_P-high] Consider Scenario Three. (If you need to, you may return to the written instruction that describes this scenario) Factor X is a random number and is, with equal likelihood, equal to either X=1.8 or X=2.2. You have the role of participant E and for this scenario you have been randomly assigned to another participant in this experiment. Decision 8: How much ECU will you send to participant F?
[U_P-low] Uvazujte Scenar Dva. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Celkova castka S je nahodna a se stejnou pravdepodobnosti je bud S=900 ECU nebo S=1100 ECU. Jste v roli ucastnika C a pro ucely scenare jste byl pocitacem nahodne prirazen k jinemu ucastnikovi experimentu.	[U_P-low] Consider Scenario Two. (If you need to, you may return to the written instruction that describes this scenario) The total amount S is random and is, with equal likelihood, equal to either S=900 ECU or S=1100 ECU. You have the role of participant C and for this scenario you have been randomly assigned to another participant in this experiment.

Rozhodnuti 9: Jaka je Vase nabidka? Prosim vyplnte Vasi odpoved do okenka a potvrdte!	Decision 9: What is your proposal? Please fill out your answer in the space above and confirm!

[HL-60] Uvazujte scenar Ctyri. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Volba "+": Ziskate 1000 ECU pokud N>60, jinak ziskate 1250 ECU Volba "*": Ziskate 60 ECU pokud N>60, jinak ziskate 2400 ECU Rozhodnuti 10: Kterou volbu preferujete?	[HL-60] Consider scenario Four (If you need to, you may return to the written instruction that describes this scenario) Choice "+": You receive 1000 ECU when N>60, otherwise you receive 1250 ECU Choice "*": You receive 60 ECU when N>60, otherwise you receive 2400 ECU Decision 10: What choice do you prefer?
[U_R-low] Uvazujte Scenar Dva. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Celkova castka S je nahodna a se stejnou pravdepodobnosti je bud S=900 ECU nebo S=1100 ECU. Jste v roli ucastnika D a pro ucely scenare jste byl pocitacem nahodne prirazen k jinemu ucastnikovi experimentu. Rozhodnuti 11: Jaky je Vas prah akceptovatelnosti? Prosim vyplnte Vasi odpoved do okenka a potvrdte!	[U_R-low] Consider Scenario Two. (If you need to, you may return to the written instruction that describes this scenario) The total amount S is random and is, with equal likelihood, equal to either S=900 ECU or S=1100 ECU. You have the role of participant D and for this scenario you have been randomly assigned to another participant in this experiment. Decision 11: What is your acceptance threshold? Please fill out your answer in the space above and confirm!
[HL-70] Uvazujte scenar Ctyri. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Volba "+": Ziskate 1000 ECU pokud N>70, jinak ziskate 1250 ECU Volba "*": Ziskate 60 ECU pokud N>70, jinak ziskate 2400 ECU Rozhodnuti 12: Kterou volbu preferujete?	[HL-70] Consider scenario Four (If you need to, you may return to the written instruction that describes this scenario) Choice "+": You receive 1000 ECU when N>70, otherwise you receive 1250 ECU Choice "*": You receive 60 ECU when N>70, otherwise you receive 2400 ECU Decision 12: What choice do you prefer?
[D-high] Nyni uvazujte scenar Jedna. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Celkova castka S je nahodna a se stejnou pravdepodobnosti je bud	[D-high] Now consider Scenario One. (If you need to, you may return to the written instruction that describes this scenario) The total amount S is random and is, with equal likelihood, equal to either

S=300 ECU nebo S=1700 ECU. Jste v roli ucastnika A a pro ucely scenare jste byl pocitacem nahodne prirazen k jinemu ucastnikovi experimentu. Rozhodnuti 13: Jake je Vase rozhodnuti, kolik procent prevest?	S=300 ECU or S=1700 ECU. You have the role of participant A and for this scenario you have been randomly assigned to another participant in this experiment. Decision 13: What is your decision, how much as a percentage will you transfer?

[U_R-none] Uvazujte Scenar Dva s celkovou castkou S= 1000 ECU. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Jste v roli ucastnika D a pro ucely scenare jste byl pocitacem nahodne prirazen k jinemu ucastnikovi experimentu. Rozhodnuti 14: Jaky je Vas prah akceptovatelnosti? Prosim vyplnte Vasi odpoved do okenka a potvrdte!	[U_R-none] Consider Scenario Two. With the amount S=1000 ECU. (If you need to, you may return to the written instruction that describes this scenario) You have the role of participant D and for this scenario you have been randomly assigned to another participant in this experiment. Decision 14: What is your acceptance threshold? Please fill out your answer in the space above and confirm!
[T_R-high] Uvazujte nyni scenar Tri. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Faktor X je nahodne cislo a se stejnou pravdepodobnosti je bud X=1.2 nebo X=2.8. Jste v roli ucastnika F a pro ucely scenare jste byl pocitacem nahodne prirazen k jinemu ucastnikovi experimentu. Obdrzel jste castku ECU (to je X krat mnozstvi poslane ucastnikem E, ktery k Vam byl pro toto rozhodnuti nahodne prirazen.) Rozhodnuti 15: Kolik ECU prevadite zpet na ucastnika E?	[T_R-high] Consider Scenario Three. (If you need to, you may return to the written instruction that describes this scenario) Factor X is a random number and is, with equal likelihood, equal to either X=1.2 or X=2.8. You have the role of participant F and for this scenario you have been randomly assigned to another participant in this experiment. You received the amount ECU: … (this is X times the amount send by the participant E, who was for this decision randomly assigned to you.) Decision 15: How much ECU will you send back to participant E?
[HL-80] Uvazujte scenar Ctyri. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Volba "+": Ziskate 1000 ECU pokud N>80, jinak ziskate 1250 ECU	[HL-80] Consider scenario Four (If you need to, you may return to the written instruction that describes this scenario) Choice “+”: You receive 1000 ECU when N>80, otherwise you receive 1250 ECU

Volba "*": Ziskate 60 ECU pokud N>80, jinak ziskate 2400 ECU Rozhodnuti 16: Kterou volbu preferujete?	Choice “*”: You receive 60 ECU when N>80, otherwise you receive 2400 ECU Decision 16: What choice do you prefer?

[T_R-low] Uvazujte nyni scenar Tri. (Pokud potrebujete, vratte se k tistenym instrukcim, ktere tento scenar popisuji.) Faktor X je nahodne cislo a se stejnou pravdepodobnosti je bud X=1.8 nebo X=2.2. Jste v roli ucastnika F a pro ucely scenare jste byl pocitacem nahodne prirazen k jinemu ucastnikovi experimentu. Obdrzel jste castku ECU (to je X krat mnozstvi poslane ucastnikem E, ktery k Vam byl pro toto rozhodnuti nahodne prirazen.) Rozhodnuti 17: Kolik ECU prevadite zpet na ucastnika E?	[T_R-low] Consider Scenario Three. (If you need to, you may return to the written instruction that describes this scenario) Factor X is a random number and is, with equal likelihood, equal to either X=1.8 or X=2.2. You have the role of participant F and for this scenario you have been randomly assigned to another participant in this experiment. You received the amount ECU: … (this is X times the amount send by the participant E, who was for this decision randomly assigned to you.) Decision 17: How much ECU will you send back to participant E?
[Demographics] Zatimco my a pocitacovy program urcujeme celkove vydelky z dnesniho experimentu, prosime Vas odpovedet na nekolik otazek o Vas. Vsechna data budou povazovana za prisne duverna a budou pouzita pouze pro tuto studii. Po vyplneni a az Vas experimentator pozada, predstupujte jednotlive s Vasim obcanskym prukazem (pripadne jinou ID kartou) k vyplate. - Vase rodne cislo - Kolik je Vas mesicni disponibilni prijem (to je, kolik penez muzete utratit pote, co zaplatite za sve ubytovani)?	[Demographics] While we and the computer program are calculating the total earnings from today’s experiment, we would like to ask you to answer some questions. All data will be considered as strictly confidential a will be used only for this study. After you have filled out the questionnaire and once the experimenter ask you to, please come one by one to the pay desk with your identity card (or with another form of ID). - Your birth number - How much is your monthly disposable income (that is, how much can you spend after you have paid for lodging?)
[Results] Vase N bylo Z rozhodnuti 2,6,10,12,16 bylo nahodne vybrano rozhodnuti Z toho rozhodnuti Vas vydelek cinil ECU Z rozhodnuti 1 jste vydelal ECU Z rozhodnuti 3 jste vydelal ECU V roli ucastnika B jste z rozhodnuti 3 Vam	[Results] Your N was From decisions 2, 6, 10, 12, 16 has been randomly chosen decision: … From this decision you earning is in ECU: … From decision 1 you earned in ECU: … From decision 3 you earned in ECU: … In the role of participant B, you have for

nahodne prirazeneho ucastnika A vydelal ECU	decision 3 received from a randomly assigned participant A in ECU: …
Z rozhodnuti 4 jste vydelal ECU	From decision 4 you earned in ECU: …
Z rozhodnuti 5 jste vydelal ECU	From decision 5 you earned in ECU: …
Z rozhodnuti 7 jste vydelal ECU	From decision 7 you earned in ECU: …
Z rozhodnuti 8 jste vydelal ECU	From decision 8 you earned in ECU: …
Z rozhodnuti 9 jste vydelal ECU	From decision 9 you earned in ECU: …
Z rozhodnuti 11 jste vydelal ECU	From decision 11 you earned in ECU: …
Z rozhodnuti 13 jste vydelal ECU	From decision 13 you earned in ECU: …
V roli ucastnika B jste z rozhodnuti 13 Vam nahodne prirazeneho ucastnika A vydelal ECU	In the role of participant B, you have for decision 3 received from a randomly assigned participant A in ECU: …
Z rozhodnuti 14 jste vydelal ECU	From decision 14 you earned in ECU: …
Z rozhodnuti 15 jste vydelal ECU	From decision 15 you earned in ECU: …
Z rozhodnuti 17 jste vydelal ECU	From decision 17 you earned in ECU: …

Chapter III. Public Goods Financing, Tax Avoidance, Reciprocity, and Inequality Aversion with Earned Wealth: An Experimental Investigation

(joint work with V. Semerak)

1. Introduction

Willingness to contribute funds used for the provision of public goods and the efficiency of various mechanisms that states and other institutions use to collect such funds have been a perennial topic in many debates on reforms of public finance. Here we focus on a specific facet of this discussion and analyze how willingness to contribute to public goods depends on the individual's position in the income distribution.

Our analysis is based on an experimental approach. We designed an experiment during which the participants (1) had a chance to earn one of three levels of incomes, and (2) were then asked to contribute to a public good. Our experiment allows us to analyze the contribution behavior of "poor" subjects in "rich" groups and vice versa. Both relative (percentage of income) and absolute contribute schemes were analyzed. We also studied the contribution behavior of poor and rich agents under efficient and inefficient state spending. Our analysis, in addition to providing empirical evidence on tax avoidance, inequality aversion, and reciprocity, under various institutional regimes, provides an experimental test-bed for the study of a number of related issues.

The participants in our experiment show significant deviations from the pure individual (selfish) income maximization; a more detailed analysis of the data provides evidence in favor of the reciprocity motive in motivation to contribute to public good provision, whereas the results for inequality aversion are ambiguous. The data also show that the configuration of a tax collection system can influence the gravity of tax avoidance; the tax regime which was more compatible with reciprocity considerations (equal lump sum tax) led to lower tax avoidance than the alternative (flat rate tax).

The Chapter is structured as follows: section 2 describes the theoretical foundations of our hypothesis and previous research done in this field; section 3 provides detailed information on the design of our experiments and on the reasoning behind the design. Section 4 contains analysis of the collected data; section 5 analyzes differences in performance under the two different tax regimes, and section 6 concludes.

2. Theoretical Background and Previous Research

Empirical data often show that substantial deviations from perfectly rational and selfish behavior are very common. Although some of these deviations might be attributed to errors, less than perfect information, or bounded rationality, there is substantial evidence from experiments with ultimatum games (Güth at al., 1982 or Güth and Ortmann, 2006) that suggests systematic deviations that can be explained by inequality aversion and reciprocity.

These deviations are not confined to ultimatum games. Public goods experiments also often find higher levels of contributions than predicted by standard theory of rational and selfish agents (e.g. Fischbacher, Gächter and Fehr, 2001). The evidence suggests that people show some level of reciprocity and fairness – both in responding to good and bad behavior – (e.g. Falk & Fischbacher, 2002, and Fischbacher, Gächter and Fehr, 2001).

Deviations from purely selfish behavior have been empirically identified in the tax evasion literature as well. Traditional models – e.g. Allingham and Sandmo (1972)[39] – explain tax avoidance as a very rational behavior. Indeed, the traditional deterrence models predict far too little compliance and far too much tax evasion (Frey & Feld, 2002). The differences between models and empirical data seem to be related to culture, the social values of individuals and organization of society. Alm & Torgler (2006) analyzed differences in tax compliance in the USA and 15 European countries; they find significant differences related to the size of shadow economies. Kleven et al. (2011) analyzed data from a tax enforcement field experiment with over 42,000 participants in Denmark and found that tax avoidance is significant for self-reported income, but that marginal tax rates also have positive, albeit modest, effects on avoidance.

There is significant evidence that group (social) interaction matters substantially in the decision to avoid taxation (e.g. Elster (1989)). Social norms, habits, and also direct reciprocity related to the observed behavior of others increases or decreases willingness to underreport income or undercontribute to public goods provision. For instance, in Traxler's (2010) model, taxpayers' evasion depends on others' compliance.

The taxpayers' belief that they are receiving adequate services for taxes paid matters as well (e.g. Feld and Tyran, 2002). A better understanding of these deviations from the standard model of the rational selfish agent and finding out whether they may be caused by reciprocity or other fairness related motives such as inequality aversion is thus important for the design of efficient systems of tax collections.

Unfortunately, it is often very difficult if not impossible to discern the role of the three main motives (reciprocity, inequity preferences, and "quid pro quo") when non-experimental data are used and even in many experimental environments. Our experiment was inspired by the experimental setup used in Falk & Fischbacher (2002), and Fischbacher, Gächter and Fehr (2001), who experimentally analyzed the role of social interaction in "criminal activities" and in public good games respectively, using a version of Selten's (1967) strategy method for

[39] See also Sandmo (2005) for an overview of the development of the theory of tax evasion.

collecting information on the strategic behavior of the participants, and combining the strategy method with an initial phase used to generate legitimate (deserved) initial wealth of the participants.[40] The main objective of the experiment was to separate and compare the effects of reciprocity, inequity preferences, and "quid pro quo" motives in the decision to contribute to public goods provision when collective punishment can occur.

2.1. Analyzed Cases of Other-Regarding Preferences

We focus on situations in which economic agents are provided with a public good which they know is financed from their contributions ("taxes"), and they know that they can decide to pay less (or more) than they are asked for.

We study whether, when deciding on contributions to public goods, economic agents indeed have a sense of "fairness" (in the form of either inequality aversion or reciprocity), i.e. depending on the environment, whether they will decide to pay what they are asked.

Specifically, we hypothesize that:

- When contributing funds/paying taxes used for provision of public goods, people do show signs of "altruism", i.e. they may be willing to contribute sums that significantly differ from the optimal solution of the maximization problem based on the assumption of a traditional "selfish" utility function.
- When contributing funds/paying taxes, agents take into account their environment, and they are brought up with a certain sense of fairness. When they perceive themselves as relatively richer "in their neighborhood", they have a tendency to contribute more than is asked, and when they perceive themselves as poor, they have a tendency to contribute less.

We construct a model and conduct an experiment on taxation and public goods provision in groups with income inequality. We test how agents would decide in a laboratory environment where many of the other ("non-altruistic") causes of behavior can be filtered out, specifically the following disturbances:

- Opportunities to avoid paying taxes may differ depending on income.
- The effect of the quality of public goods on willingness to pay is removed by the use of uniformly distributed pecuniary public goods (participants receive equal amounts of money which represent the public goods financed from their contributions).
- The threat of punishment for avoiding full contribution can be viewed as less/more strict for agents with different levels of income.
- If earned income depends positively on intellectual abilities, then agents with higher income are also those that are most able to understand the importance of contributions to the public good.

[40] The importance of asset legitimacy is shown e.g. in Cherry et. al (2002).

- Behavior induced by repeated interaction in small groups, a “social contract”, would mean that participants contribute less when they are short of funds and more when their situation improves (assuming the same behavior of other agents in the group).
- When contributions are public, paying higher contributions could be theoretically understood as a type of “conspicuous consumption” that shows the status of the agent.

When expressed in terms of utility functions, the traditional *homo oeconomicus* should maximize his pay-off without regard to the “neighborhood” or “group”. That is, the utility function depends positively on the consumption of public good (*PG*) and strictly negatively on the initial contribution (*x*) as well as subsequent additional contribution (*f*) collected if the taxes were not sufficient to finance the public goods. In this case, the maximization problem of each member of the group can be roughly described as follows:

$$\max_{x_i}\left[U_i(PG, payoff_i - x_i - f_i)\right]$$

with $\frac{\partial U_i}{\partial x_i} < 0$, $\frac{\partial U_i}{\partial f_i} < 0$, $\frac{\partial U_i}{\partial PG} \geq 0$, $\frac{\partial U_i}{\partial payoff} \geq 0$

As usual, we also assume concavity, i.e.: $\frac{\partial^2 U_i}{\partial payoff^2} < 0$ and $\frac{\partial^2 U_i}{\partial PG^2} < 0$

The alternative specification that provides a rational explanation for the systematic deviations from selfish behavior would assume a utility function that at least partially depends on the utility/final income of the other members of the group. We assume that participants believe that their decision has no direct influence on the quality and quantity of provided public goods. In this case we can omit the public good as a variable from the utility functions.

While many such utility functions can be found, we will be interested in two basic forms of the functions:

- **Utility function with reciprocity considerations**. In this case the participants want to contribute as much as other participants. “As much as” has two possible interpretations:
    - (i) they prefer equal contributions in absolute terms:

$$U_i(income_i - x_i - f_i, \left|x_i - \bar{x}_j\right|) \text{ where } \frac{\partial U_i}{\partial \left|x_i - \bar{x}_j\right|} < 0$$

    - (ii) they prefer to contribute equal shares of income:

$$U_i(income_i - x_i - f_i, \left|r_i - r_j\right|) \text{ where } \frac{\partial U_i}{\partial \left|x_i - \bar{x}_j\right|} < 0 \text{ and } r_i = \frac{f_i}{income_i}$$

It is not clear which type of reciprocity should be more relevant. We assume that this may depend on the type of situation and environment, i.e. in our setup, we can expect that participants in the relative version of the experiment may be more likely to focus on their share of income.

In our definition of reciprocity, we do not differentiate between positive and negative reciprocity (unlike e.g. Falk, Fehr, and Fischbacher, 2008). Our reciprocity is more similar to "tit-for-tat" like behavior; i.e. the agent does not want to find herself contributing much more or much less than other agents in her neighborhood.

- **Utility function with inequality aversion.** In this case, the participants prefer a final outcome with lower dispersion of final payoffs. One of the simplest versions of such a utility function is this:

$$U_i(income_i - x_i - f_i, |y_i - \bar{y}_j|) \text{ where } \frac{\partial U_i}{\partial x_i} < 0 \text{ , } \frac{\partial U_i}{\partial PG} \geq 0 \text{, } \frac{\partial U_i}{\partial |y_i - \bar{y}_j|} < 0$$

with y being the final income of the participants: $y_j = income_j - x_j - f_j$

2.2. Role of Weights, Functional Forms and Budget Constraints

In principle it is possible to imagine any type of separability and relative weights of the roles of income maximization, reciprocity, and inequality aversion; this implies a continuum of shapes and slopes of response functions of the participants. In order to simplify the space of types of behaviors and to derive basic "ideal types" of response functions, we assume that:

- Utility function is separable in the three motives
- The utility function is concave in payoffs and convex in reciprocity and inequality, but for the range of values of contributions analyzed in our experiment it can be approximated by a linear relationship (possibly with a quadratic term).

Another important factor that influences the observed types of behavior is the assumed closure rule describing the relationship between individual contributions and provision of public goods, most importantly the behavior of the provider in cases of surplus/deficit. For example, if the participants are to receive a fixed volume public good regardless of their individual contributions, and excess contributions are not redistributed, participants with dominant inequality aversion would be likely to contribute the difference between their current income and the expected value of the income of the poorest member. The same participant would behave differently in situations in which, e.g., excess contributions are being redistributed back to participants.

In our setup, we assume that if the sum of contributions is not sufficient to finance the costs of the public goods, the state will collect additional "forced" contributions. The collection can either be efficient, in which case it will only be necessary to collect what is missing, or inefficient, in which case it will be necessary to collect more. When the excess contributions are collected, the participants are asked to contribute an equal lump sum in the absolute version and an equal percentage of their income in the relative version (for more details see section 3.4).

2.3. Expected Behavior for Basic Types of Preferences

The following table summarizes the differences in reported decisions for the three "pure" types of behavior in the absolute version.

Preferences	**Reported Guess**	**Unconditional Contribution**	**Shape of Conditional Schedule**	**Sensitivity to Differences in Income**
Purely selfish	0	0	Flat at 0	No
Strong reciprocity	Any value	Corresponds to reported guess	Increasing	No
Strong inequality aversion	Positive values	Reported guess + correction for differences in initial income	Increasing, intercepts "neighborhood dependent"	Yes

Table III.11 - Expected behaviour in the "absolute" version of the experiment

As the Table III.11 shows, the two basic hypotheses on preferences lead to different expected behavior. The crucial element useful for econometric testing is the difference in the sensitivity to the difference between one's own income and the average income in the group. The two intercept-related effects are of less use, because of their possible interaction with individual unobserved heterogeneity.

3. Design and Implementation of the Experiment

The experiment was designed with the intention to motivate the participants to reveal their preferences towards contributing to a public goods provision. Therefore, the design of the contribution scheme had properties of a prisoner's dilemma – while each participant had an individual motivation to contribute as little as possible (i.e., 0). The contribution scheme also included a mechanism for forced collection of missing funds for the public goods provision that could render collective avoidance to pay an inefficient strategy (see sections 3.3 and 3.4 for more details on the payoff calculations).

Three types of preferences were assumed to matter in this case.

1. **Selfish profit-maximizing preferences**. Participants would contribute nothing.
2. **Reciprocity – "tit for tat"**. Participants would contribute if they anticipated the other members would contribute. They would avoid contributing if they anticipated others would not to contribute.
3. **Inequality aversion**. Participants with high incomes were hypothesized to tolerate lower contributions from poorer participants and to contribute more than what they expect the others to contribute.

In order to differentiate between the three motives, we designed an experiment that could induce controllable income inequality (more details on this in section 3.2), and included a contribution scheme that allowed us to infer reciprocity (section 3.3). We have also tried to collect information about the expectations of the participants.

After an introductory phase during which the test participants were informed about the experiment (see the attached instructions) and could decline to continue, the experiment consisted of two phases:

1. **Introductory general knowledge quiz.** During the first stage participants took part in a quiz[41] in which they earned points that were converted into money at the end of the first stage. The role of this stage was to generate "legitimate" incomes that the participants would use responsibly in the second stage.[42]
2. **Contribution stage.** This stage had eight rounds in total. In each round, the 15 participants were randomly distributed into three groups. Knowing their own income and the average income in their group, they were asked to contribute to the public goods provision. In order to analyze the relationship between contributions and inequality, we designed a simple random sorting mechanism that guaranteed the desired sort of inequality in the groups.

[41] Samples of the questions are attached. Although the quiz was originally prepared in English, it was later translated and the questions were presented in Czech in the final version of the experiment.

[42] Legitimacy is crucial if the participants are to reveal their real preferences – e.g. Cherry et al. (2002).

In each round of the contribution stage, the participants started with the same initial income that they obtained in the first stage, as we wanted to avoid income effects that could arise if the participants simply continued with the same “account” during all eight rounds. Results of one randomly selected round were used to determine the final payoffs of the participants. The payoff was paid out in Czech currency at the end of the experimental sessions.

3.1. First Stage – Earning Income and Producing Inequality

The role of the first stage was purely auxiliary: to generate initial incomes for the participants and to establish the “legitimacy” of the incomes at the same time. A simple way of establishing legitimacy is to offer the participants a way to “earn” income; a general knowledge quiz was used. However, although a quiz may help to create “legitimate income” and some income inequality, the resulting income inequality is difficult to predict. Therefore, the income that the participants could earn during the first stage was actually composed of two parts:

1. Payments for each correct answer (1 ECU per answer) and deductions per incorrect answer (-0.6 ECU per answer). The participants could also decide not to answer in the given time and neither gain nor lose money. There were 20 questions in the quiz, which gives a maximum possible payoff of 20 ECU.
2. Final rank related premium. At the end of the first round, all participants were ranked by their accumulated income and received an additional premium. Those with rank 1-5, 6-10, and 11-15, received 450 ECU, 300 ECU and 150 ECU respectively.

Thus there were three income “classes”, each with five participants. The expected initial income for members of the classes was about 450 ECU, 300 ECU, and 150 ECU respectively, for an expected average income of 300 ECU at the end of the first stage.

3.2. Second Stage: Group (Neighborhood) Composition

In order to avoid possible signaling or multi-round cooperation (and adaptive behavior) in the second stage, the groups were drawn randomly at the beginning of each round and the composition of the groups was kept secret. By doing this, we basically turned the second stage into a sequence of one-shot games; the participants had no guarantee that they would be in the same neighborhood in the next round. We believe this is similar to real-world situations for large groups (states) and provides sufficient anonymity.

Random sorting prevented multistage games; but in order to achieve a controlled environment, it was also necessary to generate neighborhoods with predictable levels of income inequality. We did this by imposing a fixed income structure on the resulting groups (neighborhoods). Results of stage one generated a society with five “rich” agents with average expected income of about 450 ECU, five “middle-income” agents with average expected income of 300 ECU, and five “poor” agents with average expected income of 150 ECU. When assigning members to the groups, we imposed the following simple constraints:

- The first group, “**high income neighborhood**” is always inhabited by three “rich” agents, one “middle income”, and one “poor” agent.
- The second group, “**middle class neighborhood**” consists of three “middle income” agents, one “rich”, and one “poor” agent.
- The third group, “**poor neighborhood**” is inhabited by three poor agents, one “rich”, and one “middle income” agent.

The whole sorting procedure is summarized in Figure III.7. Theoretically, there should be 8,000 different results of the sorting procedure, which is more than sufficient to prevent any attempts at collusion by the participants.

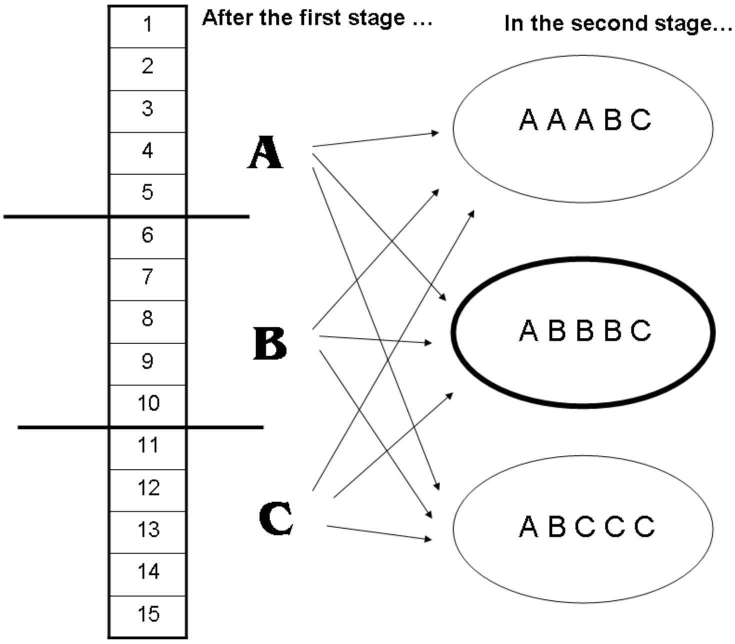


Figure III.7 - Sorting and generation of inequality

At the same time, each of these results guarantees the emergence of neighborhoods with the following properties:

- A “**Rich neighborhood**” with an average expected income per capita of about 360 ECU, i.e. the “rich” participants have above average income, “middle income” agents are slightly below average, and the “poor” agent is substantially below the average in the group.
- A “**Middle income neighborhood**” with an average expected income per capita of about 300. This means that the “rich” participant is substantially above average, the “middle income” has more or less the same expected income as the average income per capita in the group, and the “poor” agent is again substantially below average.

- A “**Poor neighborhood**” with an average expected income per capita equal to 240 ECU. This means that both “rich” and “middle income” agents have higher than average income, but the “poor” participants are still below average.

3.3. Second Stage: Contribution Schemes

After earning their initial income in the first stage and after being sorted into “neighborhoods” at the beginning of the each round of the second stage, the participants were asked to contribute 150 ECU (in the absolute version of the experiment) or 50% of their income (in the relative version of the experiment); thus in both cases the required contribution equaled 50% of the overall average income of all participants. To demonstrate the fact that the taxes were indeed used to finance a public good, the participants also immediately received additional revenue equal to 150 ECU (or 50% of their income respectively) symbolizing their benefits from the consumption of the public good.

In order to motivate the participants to contribute as suggested, the participants were told that if they (within their group) together contributed less than would be necessary to finance the public goods, the missing money would be collected in the form of “forced contributions” (see section 3.4 for details). Two different states of the collection of “forced contributions” could happen: in the efficient state, the “state” had to collect only the missing sum of money, in the inefficient state, the “state” had to collect twice as much as was missing. In both cases the costs were distributed evenly among members of the group.

The information that we asked the participants to provide was fairly complex. They knew their initial income; they were also informed of the average income of their group (“neighborhood”) for the particular round, and they were told whether the state of collection of the forced contributions would be efficient or inefficient. With this information set, they were asked to report three types of information:

- Their **unconditional contribution**, i.e. how much they were willing to contribute directly to finance the provision of public goods in their group.
- Their **guess** about the average contribution of the other members. In order to provide as earnest an estimate as possible, a motivation premium (10 ECU) was given to the member of the group whose estimate was closest to the actual number.
- Their **conditional contributions**, or rather **conditional contribution schedule**. This means that the participants were asked to report how much they would like to contribute from their own income in five different situations with given ranges of average contributions of the other members of their group. The five ranges of the contributions of the others were defined as [0 – 40], [41 – 80], [81 – 120], [121 – 160], [more than 161] in the absolute version; and [0 – 15%], [16% - 30 %], [31% - 45%], [46% - 60%], and [61% - 100%] in the relative version. The weighted averages (weighted by income) of the percentage rates were used in the relative version.

In order to improve readability, the unconditional and conditional contributions were reported in two separate screens [43] (the unconditional decision on the first one and the conditional schedule on the second).

In each round and each “neighborhood”, one randomly selected participant had to contribute according to the conditional schedule, with the range for her/his contribution being determined from the average unconditional contributions of her/his peers. This arrangement was used in order to motivate the participants to reveal their true preferences [44]; if the final payoffs depended on the unconditional contributions only, the reliability of the reported schedules could be doubted.

3.4. Second Stage: Calculation of Payoffs

After the participants keyed in the information on their unconditional and conditional contributions, and estimates of the average contributions of the others, the payoffs for the respective round were calculated. The payoff had five components:

Initial “endowment”, i.e. the income from the first stage. The participants started with the same initial income in every round, i.e. final payoffs of the previous rounds did not influence the initial endowment for the next round.

+ **Benefit from public goods consumption**. Each participant received 150 ECU (absolute version) or 50% of her income (relative version), which symbolized her consumption of the public good.

- **“Direct” contribution** to the public good provision. In each group, four participants paid according to the unconditional decisions that they had reported and one randomly selected participant paid according to the conditional decision reported by her (the choice of the rate was based on the average unconditional contribution of other members of the particular group).

- **“Forced” contribution** to the public good provision. If “direct” or “voluntary” contributions were not sufficient to finance the public goods provision, “forced” contributions were imposed, i.e. the participants had to pay what was lacking. The forced contributions were distributed evenly, which meant an equal amount from everyone in the absolute version, and an equal rate in the relative version, but they could not exceed the initial income. The forced contributions depended on the net difference between funds raised and the costs of the public goods provision, and on the “efficiency” with which the “forced” contributions were used.

+ In each group one participant also received a small **motivating reward** for the closest guess to the average of other members’ contributions. As this reward was small and applied to

[43] Examples of the four screens (two for the absolute version, two for the relative version) can be found in the appendix. Again, English was used in the preparatory and testing stages of the experiment; but all the screens and instructions were carefully translated into Czech for the experiment proper.

[44] A very similar procedure was used, e.g., in Falk & Fischbacher (2002) and Falk, Fehr, and Gächter (2001).

only one member of the group, it will be omitted from the following description of the details of the payoff calculation.

Forced Contributions and Payoffs in the Absolute Version

Let us denote the direct ("voluntary") contribution of each member of the group ("neighborhood") as C_i and her income as I_i (both in ECU). We also define an efficiency factor (ξ) that equals 1 or 2 (1 means efficient and 2 inefficient tax collection), and is known to participants before they report they decisions.

If the sum of direct contributions is not sufficient (i.e. 750 = 5 x 150 ECU), then the state collects indirect (forced) contributions. The total sum of forced contributions in ECU in each group equals $\max[\xi \cdot \left(750 - \sum_i C_i\right), 0]$ and every member of the group will be asked to pay $\min\left\{I_i - C_i + 150, \max[\left(1/5\right) \cdot \xi \cdot \left(750 - \sum_i C_i\right), 0]\right\}$. The "forced" contributions thus cannot decrease the participants' income below zero.

We can thus write the payoff function as follows:

Equation 1

$$\pi_i = \max\left\{I_i - C_i + 150 - \left(1/5\right) \cdot \xi \cdot \max[\left(750 - \sum_i C_i\right), 0]; \quad 0\right\}$$

Forced Contributions and Payoffs in the Relative Version

Let us again denote the "voluntary contributions" C_i with the difference that C_i is reported in percentage points this time. For the sake of brevity, we will be using $c_i = C_i/100$ in the subsequent formulas. I_i again stands for the participants initial income and we again define an efficiency factor (ξ) that equals 1 or 2 (1 means efficient and 2 inefficient tax collection), and is known to participants before they report they decisions.

In the relative version, each member of a group is asked to contribute 50% of her income, so "forced contributions" will have to be collected if the voluntary contributions are lower than 50% of the sum of the incomes of all members of the group.

The total sum of "forced" contributions in each group will thus equal $\max\left\{\xi \cdot \left[0.5 \cdot \sum_i I_i - \sum_i (c_i \cdot I_i)\right], 0\right\}$, and the rate for the "forced contributions will be calculated as follows[45]:

[45] Note that excessive contributions are not returned to the participants, resembling many real-world tax systems.

Equation 2

$$r_f = \frac{\max\left\{\xi \cdot \left[0.5 \cdot \sum_i I_i - \sum_i \left(c_i \cdot I_i\right)\right], 0\right\}}{\sum_i I_i}$$

We again imposed a non-negativity constraint on the final payoffs, the payoff function of each participant thus becoming:

Equation 3

$$\pi_i = \max\left\{I_i - c_i \cdot I_i + 0.5 \cdot I_i - r_f \cdot I_i \; ; 0\right\} = \max\left\{I_i \cdot \left(1 - c_i + 0.5 - r_f\right); 0\right\}$$

3.5. Final Payoffs to the Participants

As we tried to prevent possible income effects and problems with generating groups with comparable degrees of inequality in subsequent rounds, results from just one randomly selected round were used to calculate the final payoffs. In every round of the second phase, the participants started from scratch with the same initial income earned in the first stage, and results of all of the eight rounds were recorded. Only after the eighth round were the participants informed about their results in all rounds.[46] Then a number from the range [1-8] was drawn randomly,[47] participants were informed about the choice and they could check that their payoff was correct. The payoffs were converted from Experimental Currency Units into Czech crowns (CZK) at a 1:1 exchange rate. The final payoffs differed slightly from the results of the selected round; all participants received a participation bonus of 150 CZK. The average final payoff was about 450 CZK for about 90 minutes of activity, which we assumed was more than sufficient to motivate the participants to fully participate in the experiment.[48]

3.6. Participants Recruiting

The participants were recruited randomly from among students of Prague-based universities. Two approaches to reaching the students were combined:

1. Direct advertising of the experiments on the electronic bulletin boards of some of the universities.
2. Contacting a database of participants who had previously voiced interest in participating in experiments at CERGE-EI.

[46] They already knew the result of rounds 5-8, for which they received feedback when the round was finished; now they received information about the results of all eight rounds. The information came in a form of a simple table.
[47] This was done automatically in the z-Tree code.
[48] The average gross monthly salary in the Czech Republic, was according to the Czech Statistical Office, 19,020 CZK in 2005.

In both cases the basic advertisement included the following information:

- That the experiment was about economic decision-making.
- Estimated time costs.
- Estimated earnings.
- Time and location of the planned sessions.
- Description of the application procedure.

Prospective participants were informed that if they wanted to participate, they should respond as soon as possible, i.e. send an e-mail to the organizers and indicate which day/time would suit them. They were informed that the final selection would be done on a first-come-first-served basis. Applications were gathered in a spreadsheet and sorted according to the suggested day and time of experiment. More than the necessary number of participates were invited to allow for the possibility some might decline to continue once details on the organization of the experiment were explained. [49] Redundant participants were selected randomly.

3.7. Problems with the Design and Interpretation

In spite of running pilot experimental sessions after which we fine-tuned the experiment, several problems remained that may complicate the interpretation of results, most importantly[50]:

- **Complexity of the design and instructions for participants of the experiment.** The two stage experiment with its fairly complex contribution scheme (especially with respect to the conditional decisions) was not easy to explain to the participants, especially if their behavior was not to be too influenced by the wishes of experimenters. Some patterns in the data as well as questions and comments that we received after the experiment suggest that some participants did not understand all details of the experiment, even though the instructions included a fairly detailed description (including numerical examples) of the contribution schemes, and sufficient time for preparation was provided.
- **Possible problems with self-fulfilling wording of instructions.** We decided not to avoid using words such as "public goods" in the instructions[51], which might have motivated some participants to rely on stereotypes from their daily life instead of analyzing the problems thoroughly. However, neither the data, nor the subsequent comments of participants of the experiment brought any conclusive evidence that this actually happened.

[49] We intended to have some slack mainly as a precaution for the possible changes in the decision to participate (our setup could not be run without 15 participants without changes in the z-Tree code). However, none of the participants who attended the sessions and read the instructions refused to participate.

[50] This part builds upon comments that we received during the presentations of the research and from our colleagues at CERGE-EI (see the acknowledgment for details).

[51] See the instructions for both the relative and absolute version of the experiment in appendix.

- **Randomness of efficient/inefficient states.** In an attempt to find a simple way to avoid a specific order effect of switching efficient and inefficient states of the contributions collection, we resorted to a simple random draw of the state.[52] In retrospect this seems an unfortunate decision, which led to relatively few observations of situations with some specific combinations of parameters. The solution to this problem may be of two types: run even more sessions and rely on the law of large numbers, or prepare pseudo-random patterns of switching of the efficiency states beforehand.
- **Decision to include both rounds without feedback (first four rounds) and with feedback (information on final payoffs available after rounds 5-8)**. While this decision made it possible to analyze effects of feedback and of learning from comparing original guesses about the behavior of others with reality, together with the randomness of efficiency states, this could have caused troubles with insufficient representation of situations with specific combinations of parameters.
- **Possible order effects in reporting unconditional/conditional decisions.** The order of reporting unconditional and conditional decisions was identical in all rounds.

[52] It was of course done automatically in the z-Tree experimental software.

4. Analysis of Experimental Data

We obtained observations on 120 participants (8 sessions, 15 participants in each session), out of which 60 participated in the “absolute” and 60 in the “relative” version of the experiment. Each session yielded data for 8 rounds, which gives us 480 observations in each type of session. The experiment took place on the premises of CERGE-EI during four days of December 2005. Two sessions were run every day. The experiment was run on a computer network and it was programmed in the z-Tree experimental software. No fixed and enforced time constraints were set for decision-making in the second stage, although suggested time limits appeared both in the instructions and in the computer screens. One session took approximately 60-90 minutes.

The use of z-Tree experimental software made it possible to collect a fairly detailed set of data for each observation; besides information on the “unconditional” and “conditional” contributions, we also obtained the participants’ estimates of the average behavior of other members of the group and of the characteristics of the group itself. The richness of the data made it possible to calculate quantitative characteristics of the analyzed behavior, to carry out more detailed econometric tests of the subjects’ behavior, and to construct simple tests that reveal information about the consistency and rationality of the participants’ behavior.

4.1. Consistency/Rationality Analysis

In each round the participants reported their “unconditional” decisions, i.e. how much in Czech crowns (in the absolute version), or what share in percentage points (in the relative version) of their income they were willing to contribute. They also reported their estimate of the average contribution of the other members of their group.[53] Subsequently, they also reported their “conditional” decision, i.e. what their contribution would be if the other members decide to contribute the amount/share in pre-specified brackets. This enabled us to construct a simple test of consistency of the participants’ decision-making: the unconditional contribution should not be too different from the conditional contribution reported for the bracket that contains the participant’s prediction of the contribution of the other members of the group.

Even in the best case the fit cannot be perfect because the forms used in the experiment worked with five relatively broad brackets.[54] Because the participants did not know for sure whether they would pay what they suggested as a conditional or unconditional contribution, we presumed that they should be motivated to report their best decision in both situations. Any substantial differences in consistency of the unconditional and conditional decisions will

[53] See sample screens in the appendix A.1. Samples show the English version; Czech versions were used in the experiment proper.
[54] Our experience from previous test versions suggested that more complicated forms lead to loss of interest and to a mechanical attitude towards the forms.

therefore signal that the participants either (1) did not understand the instructions and the experiment, or (2) they did not paid sufficient attention to reporting their decisions, or (3) there was some other influence (e.g. the framing of the questions) that caused them to behave differently in the two situations. The first two causes would lead to the presence of high "noise" in the data, while the third problem would imply some systematic bias that could render the data unsuitable for our analysis. The consistency check was therefore of the utmost importance.

The test we used was simple. For each participant and each round of the second stage we took the reported guess of the average contributions of the others, and used it to obtain the estimate of "conditional contribution" from the reported conditional contribution schedule reported by the same participant. The range of values centered on the "conditional contribution" was then compared with the unconditional contribution. If the unconditional contribution fell into the range, the behavior was marked as "rational". It was necessary to use the range, because the fact that the original conditional schedules were reported for five pre-specified brackets made a direct comparison of the point estimate of "conditional contribution" with the reported unconditional contribution unreasonable. A participant's behavior was then described as consistent if more than 50% of the reported decisions (i.e. decisions from at least 5 rounds) were "rational". This criterion allows for possible initial misunderstandings and a "trembling hand".

Consistency of Data from Absolute Versions

For the absolute version of the experiment, we compared the reported unconditional decision (reported by the participants in ECU, "Experiment Currency Units")[55] with a range centered on the point estimate of "conditional behavior". The width of the range was 40 ECU (±20 ECU), and thus, it was the same as the width of the original bracket on the form used to report conditional decisions.[56]

Using the width of range and the condition of at least five rational decisions, we can conclude that 31 of 60 participants in the absolute version of the experiment were consistent in their reported conditional and unconditional decisions. This suggests that our results may be afflicted by substantial noise. Table III.12 shows how the number of "consistent" participants changes when we set the range at 30, 40, and 50 ECU, respectively, and the minimum number of consistent decisions of one participant at 4, 5, 6, and 7 respectively.

[55] For the calculation of payoffs we used the exchange rate of 1 Czech crown (CZK) for 1 Experimental Currency Unit (ECU).

[56] See the sample form in the appendix.

Share of Consistent Participants (%)		Range		
		30	**40**	50
Number of consistent choices	4	53.3	68.3	73.3
	5	38.3	**51.7**	55.0
	6	21.7	35.0	43.3
	7	13.3	23.3	26.7

Table III.12 - Sensitivity Analysis of Consistency Test, Absolute Version

Consistency of Data from Relative Versions

For the relative version, we compared the reported unconditional decision (reported as percentage of income) with a range centered on the point estimate of "conditional behavior". The width of the range was 15 (±7.5 percentage points). The width of the range was again the same as the width of the original bracket on the form used to report conditional decisions.[57]

Using the width of range and the condition of at least five rational decisions, we can conclude that only 24 of 60 participants in the relative version of the experiment were consistent in their reported conditional and unconditional decisions. The result was thus even worse than in the case of the absolute version, which confirms our conjecture that the relative version may put higher requirements on the participants. Unfortunately, it also signals that the acquired data contain lots of noise, even more than in the case of the absolute version. We again carried out a sensitivity analysis of the consistency test by setting the range at 10, 13, 15, 17 and 20 respectively and the minimum required number of consistent decisions at 4, 5, 6, and 7. We included ranges that roughly correspond to the ranges used in the consistency test of the absolute version (10, 13, 17 percentage points), the originally suggested range that corresponds to the width of the "bracket" at the forms used during the experiment, and an additional 20 point range. The reason for inclusion of an even wider range was the fact that ranges in the relative version subjectively appear to be narrower and may therefore be subjectively stricter than ranges defined in absolute numbers and used in the test of the data from the absolute version.

Share of Consistent Participants (%)		Range (perc. points.)				
		10	13	**15**	17	20
Minimum number of consistent choices	4	60.0	61.7	65.0	68.3	88.3
	5	35.0	38.3	**40.0**	43.3	68.3
	6	18.3	21.7	26.7	26.7	46.7
	7	10.0	10.0	10.0	11.7	30.0

Table III.13 - Sensitivity Analysis of Consistency Test, Relative Version

[57] See the sample form in the appendix.

Considering these data together, one may argue that the consistency of participating subjects is considerably low, but this fact is due rather to the complex setting of the experiment as a whole and also to the fact that people mostly make their decisions intuitively and in a rationally inattentive way rather than with full application of the complete understanding of the whole setting. However, analyzing the same decision sets of a random generator with all decisions independently uniformly distributed, we can obtain a consistency level for such subjects considerably below the level of 0.1%, substantially different from the subjects participating in our experiments.

4.2. Characteristics of the Data

After analyzing the consistency of the reported decisions, we calculated the simple quantitative characteristics of the data and compared them with our expectations. The following tables summarize the behavior of participants with respect to their relative income and their placement in poor, middle-income or rich groups. The tables also include information on the incidence of "efficient" and "inefficient" states during the experiment.

We also provide a rough characterization of the reported conditional contribution schedules, although due to their complexity and variability they are more difficult to summarize.

Absolute Version

There were 60 participants in the absolute version of the experiment (4 sessions, each with 15 participants). The average income with which the participants entered the second stage was 307.7 ECU. Their average unconditional contributions amounted to 113.5 ECU (the participants knew that provision of the public good required 150 ECU from each participant and they were asked to contribute 150 ECU). The non-zero average was not caused by outliers (the mode interval was around 150). The histogram in Figure III.8 shows that a substantial proportion of participants deviated from zero in their contributions, although the presence of "selfish" participants is also significant.

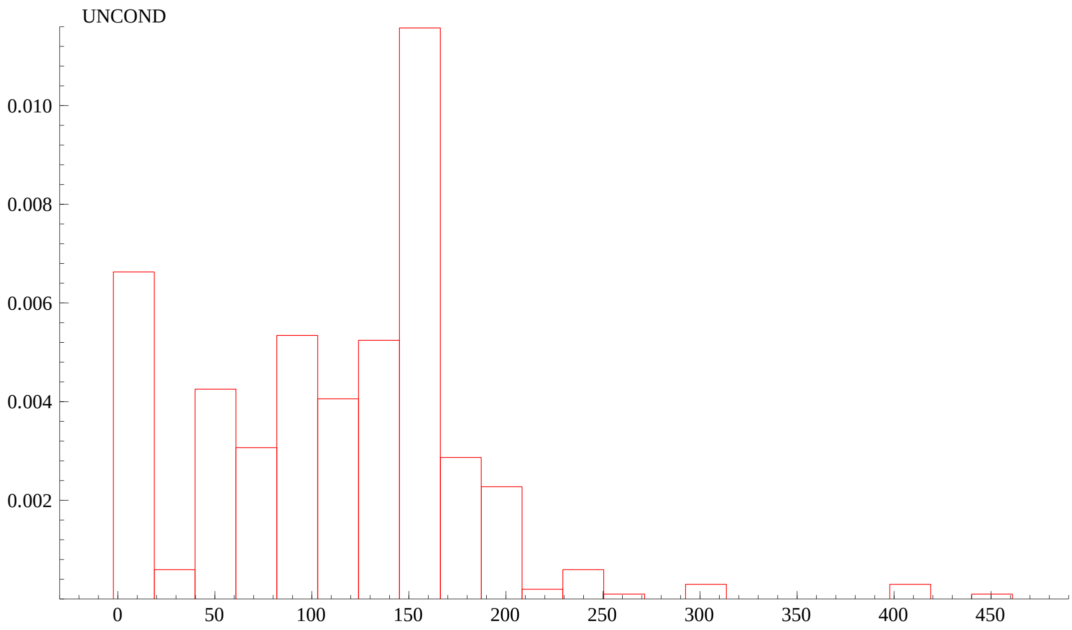


Figure III.8 - Histogram of Unconditional "Unforced" Contributions, Absolute Version

Table III.14 describes average income and average unconditional contributions of rich, middle-income and poor participants in the second stage. There is a clear difference between the average unconditional contributions of the three income groups.

Income Classification	Average Income	Average Unconditional Contribution	Number of Participants
Rich	461.2	143.8	20
Middle	306.3	104.1	20
Poor	152.5	92.7	20

Table III.14 - Summary Characteristics by Income Groups, Absolute Version

Table III.15 contains a more detailed break-down of the absolute unconditional contributions of the participants from particular income groups (rich, middle-income, poor) in particular neighborhoods (rich group, middle-income, poor group) for both the efficient and inefficient state. All observations were included; no difference was made between rounds with (rounds 6-8) and without (rounds 1-5) feedback.

Average Contributions	Total Average	Number of Observations	Average Efficient	Number of Efficient Observations	Average Inefficient	Number of Inefficient Observations
Rich in poor group	145.8	32	120.1	14	165.8	18
Rich in middle group	155.0	32	144.0	14	163.6	18
Rich in rich group	139.3	96	120.4	42	154.0	54
Middle in poor group	112.0	32	101.4	14	120.3	18
Middle in middle group	95.9	96	81.1	42	107.4	54
Middle in rich group	120.7	32	115.0	14	125.1	18
Poor in poor group	92.0	96	81.6	42	100.0	54
Poor in middle group	97.2	32	96.9	14	97.3	18
Poor in rich group	90.4	32	79.5	14	98.8	18

Table III.15 - Average Contributions in Absolute Version

Table III.15 suggests that there was a clear difference between the behavior of participants from the three income groups. The participants who entered the second stage with the higher income were in general willing to contribute more than participants with lower initial income (although there is a small overlap between the middle-income and poor participants). This seems to suggest that inequality aversion really can play an important role, though more formal tests are needed to confirm this conjecture. The data also suggest that agents on average contributed more in the “inefficient state”, which suggests that participants may have interpreted efficiency and inefficiency rather as severity of punishment for non-compliance with the suggested contribution.

Aggregated description of trends in the conditional schedules is more difficult. When we take the average of all participants and all rounds (Figure III.9), we find that the conditional schedule has an “inverted-v-pattern” on average. The contributions were increasing function of the other members’ contributions up to the range that roughly corresponds to the suggested contribution, and then they became decreasing. However, individual conditional schedules reported in individual rounds show more varied patterns. We have identified four basic patterns of conditional contribution schedules (see Figure III.10 for examples of the conditional decisions reported by participants):

1. **Flat schedules**; i.e. the same value, most often 0, reported for all five indicated brackets of suggested contributions of other members of the group.
2. **Strictly increasing or non-decreasing schedules**; i.e. schedules in which suggested contributions were either consistently increasing, or first constant and then increasing, or first increasing and later constant.
3. **Strictly decreasing or non-increasing schedules**; i.e. schedules in which suggested contributions were either strictly decreasing function of the estimated contributions of other participants, or first constant and then decreasing, or first decreasing and later constant.
4. **Inverted-V-patterns**; i.e. schedules with contributions first increasingly and then decreasingly dependent on the estimated contribution of others.
5. There were also observations that do not fit into any of the four categories. These observations are difficult to explain with respect to the possible effects of either

purely self-seeking behavior, or of some type of either reciprocity, or inequality aversion, and they may have been caused by errors or by a loss of focus of the participants caused by the length of the experiment.

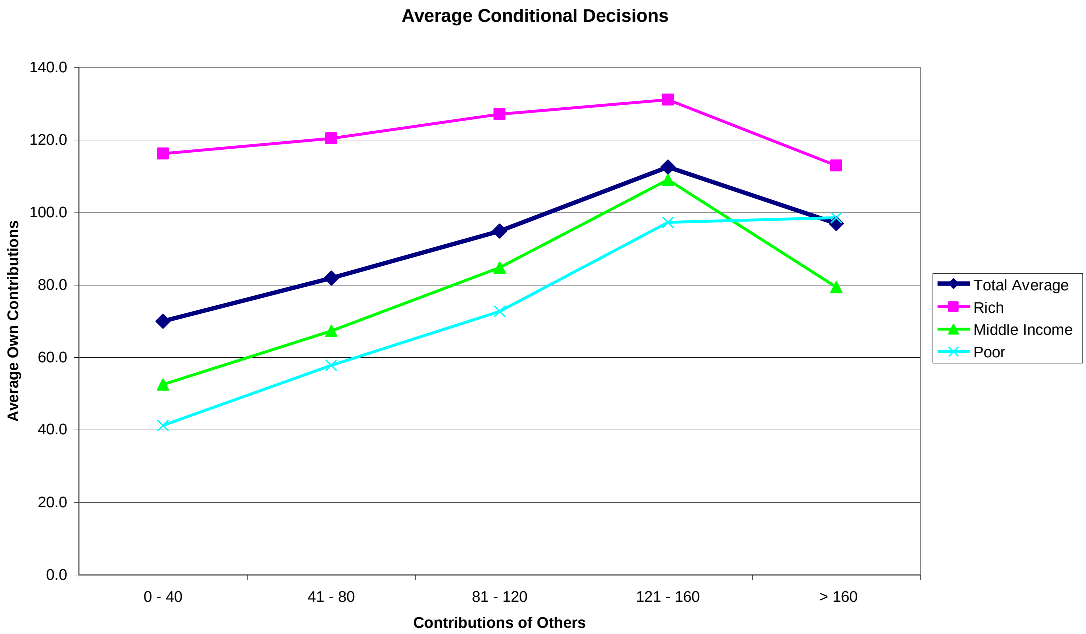


Figure III.9 - Average Conditional Contributions, Absolute Version

Exact classification of the behavior of individual participants was made complicated by the fact that we have eight observations for every participant; these eight conditional schedules from each participant in many cases do not have a stable pattern.

Table III.16 shows the incidence of pure types of agents defined as agents with purely increasing, decreasing or flat conditional schedules. In order to be labeled "increasing", the schedules reported by the participant had to be strictly increasing in at least six out of the eight rounds.[58] Using this criterion we are able to classify the behavior of only 35% of our agents; this suggests either fairly high "within-agent" variability of the schedules or a significant role of less strictly defined schedules (non-decreasing, non-increasing, v-patterns with flat sections, etc.).

Type	Incidence	Share [%]
Flat	2	3.3
Strictly increasing	7	11.7
Strictly decreasing	4	6.7
Strict inverted V-pattern	8	13.3

Table III.16 - Incidence of Pure Types of Conditional Schedules, Absolute Version

[58] This criterion allows for 25% of the reported schedules to be of a different type. Therefore we also report the results for a stricter criterion (threshold of seven consistent decisions) in the appendix A.3. The stricter version decreases the share of the strict types, and increases the share of unclassified agents to 18% of the sample.

When less strict definitions of types are used, i.e. when we analyze the incidence of non-decreasing (where at least six of the eight reported schedules are non-decreasing), non-increasing and "weak" inverted v-patterns (a part of the v-pattern is allowed to be non-increasing or non-decreasing)[59], the coverage of the sample increases, and only seven agents do not fall into any of the categories. The types of the schedule reported by these remaining participants were not sufficiently stable (i.e. less than six rounds with stable pattern).

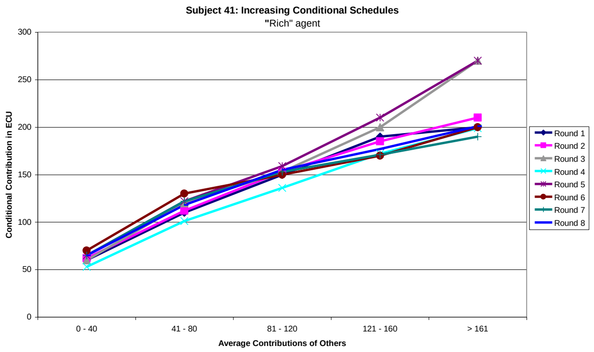
Type	Incidence	Share [%]
Flat	2	3.3
Nondecreasing (flat not included)	15	25.0
Nonincreasing (flat not included)	15	25.0
Weak inverted V-Pattern (flat, nonincreasing and nondecreasing not included)	21	35.0
Agents in no category	7	11.7

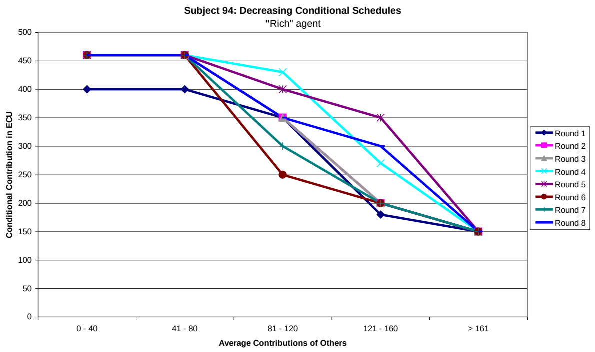
Table III.17 - "Weak" Patterns in Conditional Schedules, Absolute Version

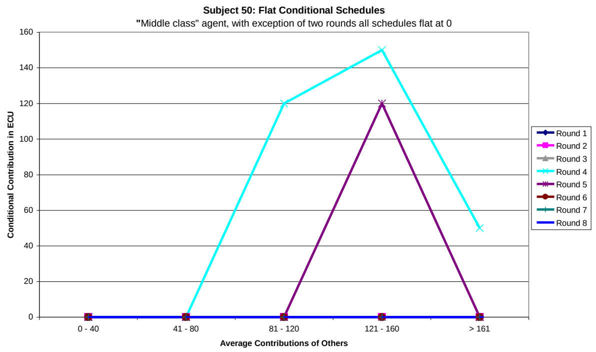
The low share of flat patterns and high share of weak patterns constitutes a piece of evidence in favor of the other-regarding preferences. All the non-flat categories can be explained as consistent with reciprocity or inequity aversion types of preferences; both increasing and decreasing schedules can be understood as parts of incomplete inverted v-patterns. The two types of preferences should differ in the role of intercept and in the position of the peak of the inverted v. While the peak of the average schedules (Figure III.9) corresponds well to the predictions of the reciprocity based model, the differences in intercepts of the group based averages would rather suggest the primacy of the inequity aversion motive. The relatively high share of participants who did not fall into any of the categories can either be understood as noise, or also a part of the indirect evidence in favor of the inequity aversion type of preferences.[60]

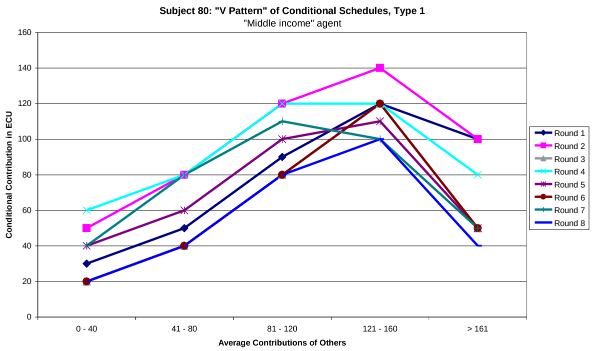
[59] The categories of non-increasing and non-decreasing overlap with the "weak" inverted V patterns. The classification can be a bit arbitrary. Here we give priority to the "non" types – i.e. weak V-patterns that fall into both categories are counted as non-increasing or non-decreasing.

[60] The interplay between variable intercept, budget constraint, and non-negativity constraint for contributions may cause switching between v-pattern and flat schedules even in the simple case with linear risk-neutral utility function.









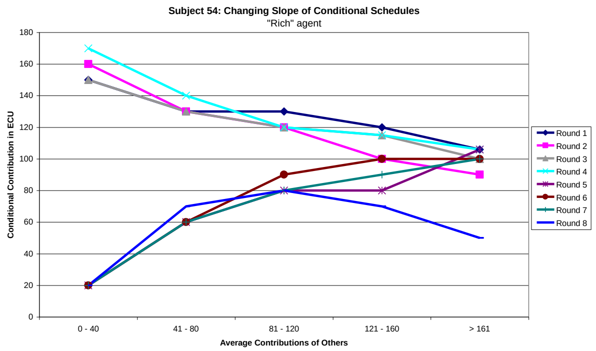


Figure III.10 - Conditional Schedules, Absolute Version

Relative Version

There were 60 participants in the relative version of the experiment (4 sessions, each with 15 participants). The average income with which the participants entered the second stage was 305.4 ECU. Their average unconditional contribution was 27.5% of their income (the participants knew that provision of the public good required 50% of income from each participant, and they were asked to contribute the same relative share of income, i.e. 50%).

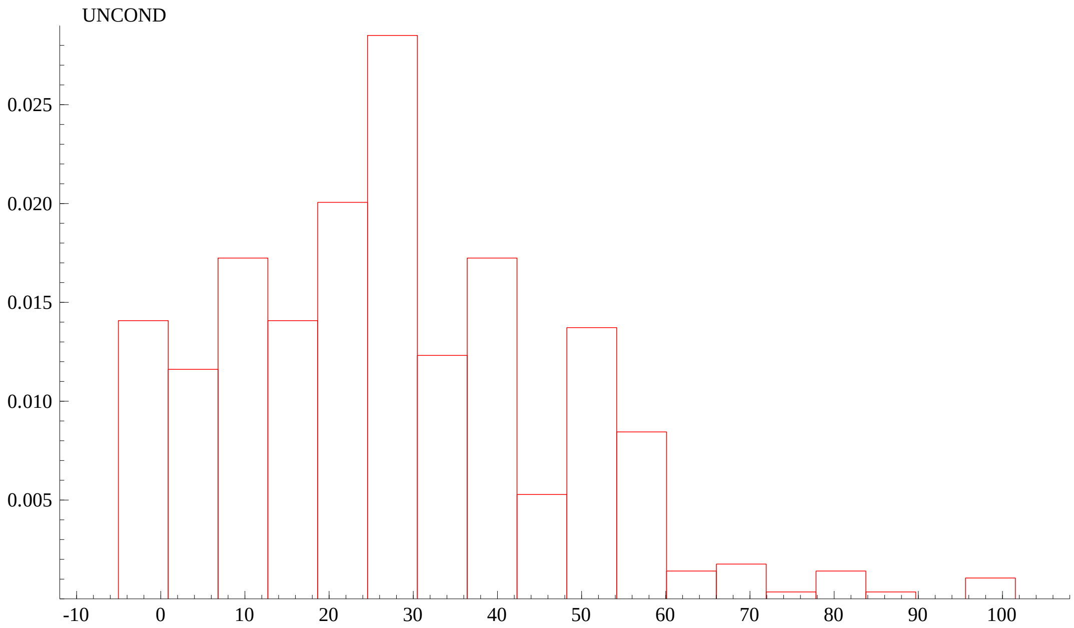


Figure III.11 - Histogram of Unconditional Contributions, Relative Version

A histogram of the unconditional contributions (Figure III.11) confirms that the positive averages were not caused by a limited number of outliers.

Income Classification	Average Income	Average Unconditional Contribution	Average Contribution in ECU	Number of Participants
Rich	459.0	26.6	120.0	20
Middle	305.1	29.8	91.0	20
Poor	152.0	26.2	40.4	20

Table III.18 - Average Contributions by Income Groups, Relative Version

Table III.18 shows average income and average unconditional contributions of rich, middle-income and poor participants in the second stage. The contributions are reported in percentage of income. The clear differences between the average contributions of the three income groups seems to be missing when we consider only the reported unconditional contributions as originally reported, i.e. as a percentage of income. In this case it would be the middle-income participants that are contributing the highest share of their income. However, when we used the original disaggregated data to calculate average unconditional contributions expressed in ECU, we get a very different picture – again, the participants who became richer in the first round contributed substantially more than the poor, although their average contribution remains lower than in the absolute version of the experiment. It seems that the "rich" participants were not willing to bear higher relative burdens.

Average Contributions (%)	Total Average	Number of Observations	Average Efficient	Number of Efficient Observations	Average Inefficient	Number of Inefficient Observations
Rich in poor group	29.6	32	28.1	14	30.8	18
Rich in middle group	28.9	32	30.3	14	27.8	18
Rich in rich group	24.1	96	21.3	42	26.4	54
Middle in poor group	30.5	32	30.2	14	30.7	18
Middle in middle group	30.0	96	28.0	42	31.5	54
Middle in rich group	28.8	32	26.9	14	30.3	18
Poor in poor group	26.4	96	26.8	42	26.1	54
Poor in middle group	28.6	32	28.0	14	29.0	18
Poor in rich group	25.1	32	20.4	14	28.7	18

Table III.19 - Average Contributions in Relative Version

Table III.19 reports average unconditional contributions (again expressed in the original form, i.e. percentage points) for 9 different combinations of the income group of the participant and "neighborhood" in which she contributes to public good provision. We again get a similar picture as in Table III.18; the distinction between group contributions is much less clear and it seems that the middle-income group bears the highest relative burden regardless of the "neighborhood". Also, the clear distinction between contributions in the "efficient" and "inefficient" state is much less clear in this case.

Average Contributions (ECU)	Total Average	Number of Observations	Average Efficient	Number of Efficient Observations	Average Inefficient	Number of Inefficient Observations
Rich in poor group	135.6	32	128.6	14	141.0	18
Rich in middle group	132.5	32	138.7	14	127.6	18
Rich in rich group	110.6	96	97.5	42	120.8	54
Middle in poor group	92.9	32	92.2	14	93.5	18
Middle in middle group	91.4	96	85.4	42	96.0	54
Middle in rich group	88.0	32	82.2	14	92.5	18
Poor in poor group	40.1	96	40.7	42	39.7	54
Poor in middle group	43.4	32	42.5	14	44.1	18
Poor in rich group	38.2	32	31.1	14	43.7	18

Table III.20 - Average Contributions in Relative Version - in ECU

As for the relative version, the average conditional contribution schedule (again with average taken over all participants and all rounds) has a shape that resembles an inverted V, although the "peak" is rather flat (Figure III.12). When looking at details, we again identify four basic patterns of conditional contribution schedules:

1. **Flat schedules**; i.e. the same value, most often 0, reported for all five indicated brackets of suggested contributions of other members of the group.

2. **Strictly increasing or non-decreasing schedules**; i.e. schedules in which suggested contributions were either consistently increasing, or first constant and then increasing, or first increasing and later constant.

3. **Strictly decreasing or non-increasing schedules**; i.e. schedules in which suggested contributions were either strictly decreasing function of the estimated contributions of other participants, or first constant and then decreasing, or first decreasing and later constant.

4. **Inverted-V-patterns**; i.e. schedules with contributions first increasingly and then decreasingly dependent on the estimated contribution of others.

5. As in the case of absolute version, there were also observations that do not fit simply into any of the four categories and observations for a single agent in many cases did not follow a stable pattern. See Figure III.13 for examples of the conditional decisions reported by participants.

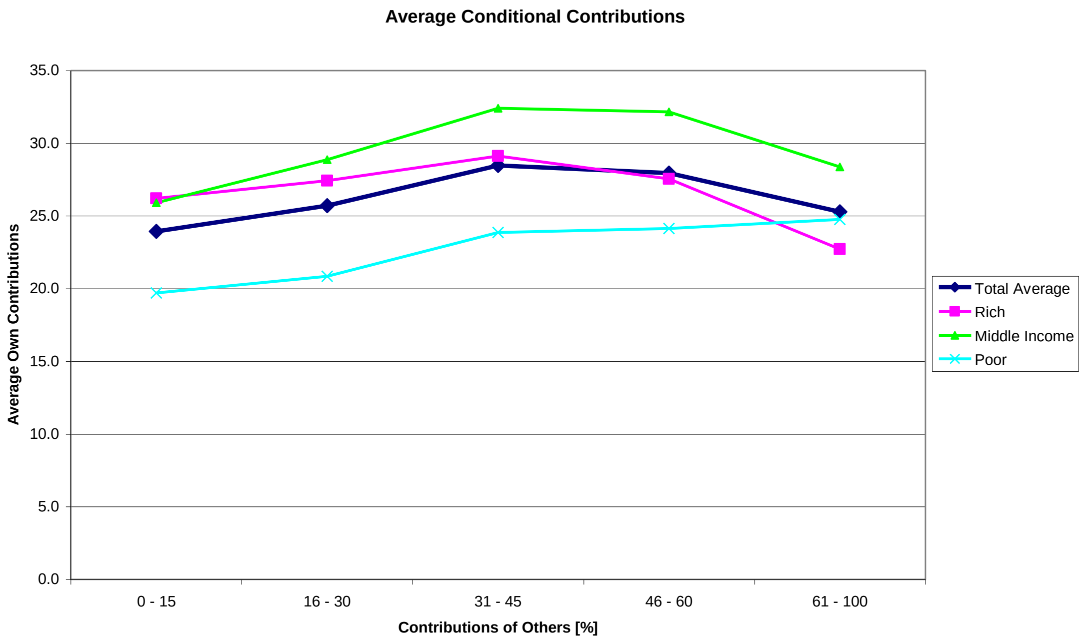


Figure III.12 - Average Conditional Contributions, Relative Version

Table III.21 shows the incidence of pure types of agents defined as agents with purely increasing, decreasing, inverted v-patterns or flat conditional schedules. In order to be labeled as such, the schedules reported by the participant had to be of the same type in at least six of the eight rounds.[61] The categories shown in the table

[61] We again report the results for a stricter criterion (threshold of 7 consistent decisions) in the appendix A.4. The stricter version decreases the share of the strict types, and increases the share of unclassified agents to 25% of the sample.

describe a higher share of the sample than in the absolute case (43% versus 35%), the share of pure inverted v-patterns is lower, and the share of the decreasing patterns higher than in the absolute version.

Type	Incidence	Share [%]
Flat	3	5.0
Strictly increasing	10	16.7
Strictly decreasing	10	16.7
Strict inverted V-pattern	3	5.0

Table III.21 - Incidence of Pure Types of Conditional Schedules, Relative Version

When less strict definitions of types are used, i.e. when we analyze the incidence of the non-decreasing, non-increasing, and weak v-patterns (again using the threshold of six consistent decisions), the coverage of the sample further increases (Table III.22). Only three agents did not fall into any of the categories.

Type	Incidence	Share [%]
Flat	3	5.0
Weakly increasing (flat not included)	16	26.7
Weakly decreasing (flat not included)	22	36.7
Weak inverted V-pattern (flat, weakly increasing and weakly decreasing not included)	16	26.7
Agents not included in any category	*3*	*5.0*

Table III.22 - Non-decreasing and Non-increasing Schedules, Relative Version

The low share of flat patterns and the high share of weak patterns can be again understood as evidence in favor of the presence of other-regarding preferences. As in the previous section, it is not possible to decide which of the motives prevails based on the shapes of the conditional schedules per se.

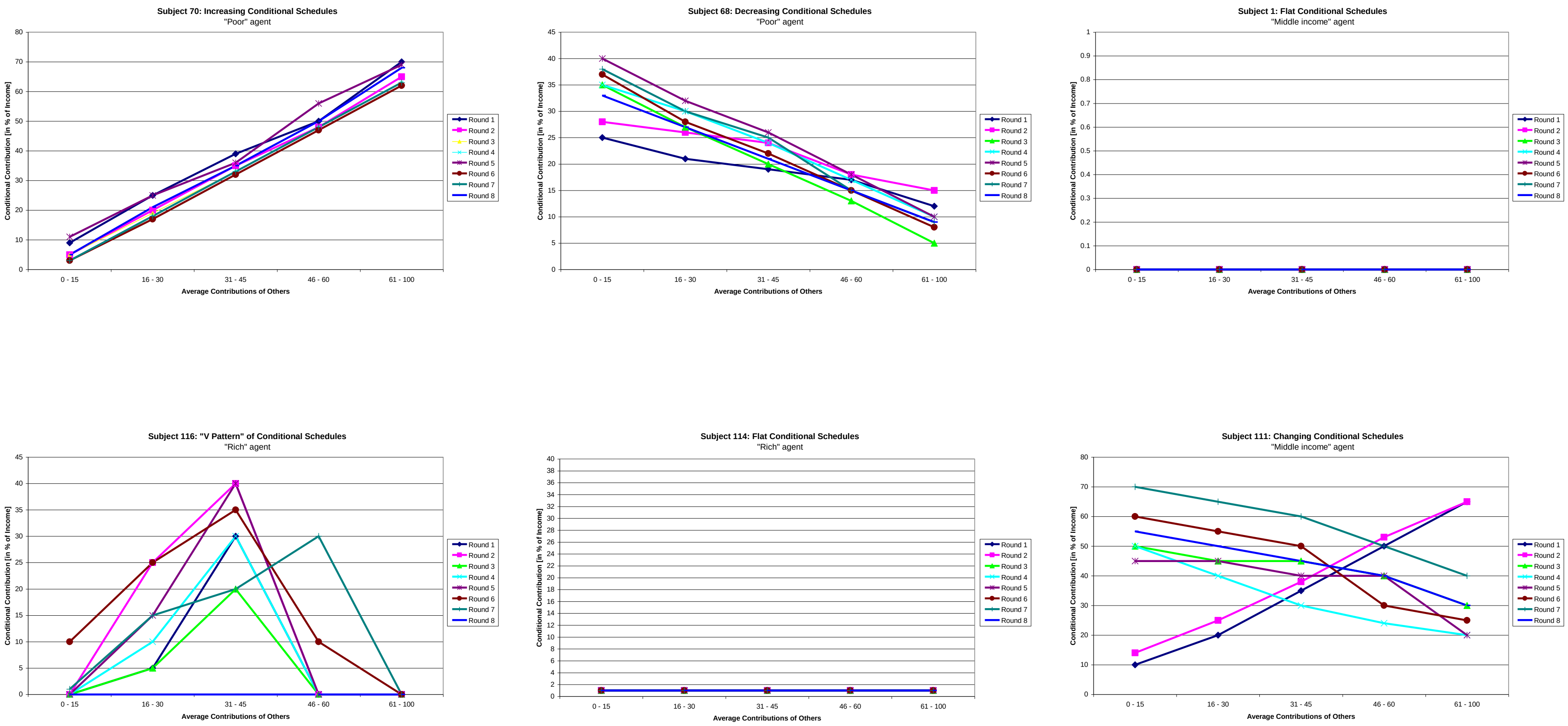


Figure III.13 - Conditional Decisions, Relative Version

4.3 Econometric Tests

The experiment generated a relatively rich data file with 960 observations (480 observations both for the absolute and the relative version). Given the fact that each of the rounds of the second stage started with the same initial balance and with fresh sorting into "neighborhoods", each round should be treated as a one-stage one-shot game. This means that the time dimension of the "panel" differs slightly from more traditional panels.

The most important variables used in the analysis are those that described the level of unconditional contribution, the expected contribution by other members of the group, the participant's own initial income, and average initial income in the topical reference group. We have also used dummies for the "efficient" and "inefficient" states; we also tested for the importance of feedback on learning and behavior changes of the participants by using the feedback dummy. Variables were used in the econometric models in the following forms:

- **Unconditional (UNCOND)**… reported unconditional contribution (in ECU in the absolute version; in percentage points or in equivalent ECU in the relative version)
- **Income difference (INC_DIFFERENCE)** … difference between one's own income at the beginning of the second stage and the average income of the group
- **Guess (GUESS)** … reported estimate of average contributions of other members of the group
- **Efficiency state dummy (STATE_DUMMY)** … dummy defined as 0 for "efficient" state and 1 for "inefficient" state
- **Feedback dummy (FEEDBACK)** … Participants were given no feedback for the first four rounds, they received information on the final payoff after the second state (and therefore also indirect information about the behavior of the other participants) only in periods 5-7. The feedback dummy was therefore set at 0 for rounds 1-5 (rounds, during which participants had no feedback information at the time of decision) and 1 for rounds 6-8.

A very general version of the suggested basic specification looks as follows:

Equation 4

$$UNCOND = f\left(INC\_DIFFERENCE, GUESS, STATE\_DUMMY, FEEDBACK, UNOBSERVED\right)$$

With the following expectations on the properties of the partial derivatives:

$$\frac{\partial UNCOND}{\partial INC\_DIFFERENCE} > 0 \text{ and } \frac{\partial^2 UNCOND}{\partial INC\_DIFFERENCE^2} \leq 0$$

$$\frac{\partial UNCOND}{\partial GUESS} > 0 \text{ and } \frac{\partial^2 UNCOND}{\partial GUESS^2} \leq 0$$

From the theoretical considerations we derived the following three claims that can be tested with the data:

1. If participants have **purely selfish profit maximizing preferences**, both their conditional and unconditional contributions should be as low as possible, i.e. 0. Any deviations from zero would be caused by misunderstanding or error and they cannot be systematic (apart from the fact that negative contributions were not allowed, so all errors would lead to positive contributions).
2. If the participants' behavior is dominated by **reciprocity considerations** or **inequality aversion**, it should be the variable **Guess**, i.e. their estimate of other people's contributions that would explain the variation in unconditional contributions. However, even if this relationship is found, an alternative explanation is possible: as their guess they can report what they assume is the optimal solution to the "game". In both cases we should identify strong significance of the coefficient of the variable **Guess**.
   We allow for non-linearity in the case of **reciprocity considerations** and assume:
   $$\frac{\partial UNCOND}{\partial GUESS} > 0 \text{ and } \frac{\partial^2 UNCOND}{\partial GUESS^2} \leq 0.$$
3. If it is the **inequality aversion** that matters most, the participants should consider the relationship between their income and the average income in their topical group. This relationship can be measured either in absolute numbers (variable **inc_difference**) or as a ratio between one's own income and the average group income. We would expect that one's own higher relative income would lead to higher contributions; lower relative income to lower contribution. This being true, we again allow for non-linearity of this relationship:
   $$\frac{\partial UNCOND}{\partial INC\_DIFFERENCE} > 0 \text{ and } \frac{\partial^2 UNCOND}{\partial INC\_DIFFERENCE^2} \leq 0.$$

As far as the efficiency state dummy is concerned, from the previous analysis (see section 4.2) it seems that rather than as an indication of efficiency, it was understood as an indicator of the severity of punishment for non-compliance with recommended contribution. If this is true, we should expect:

$$\frac{\partial UNCOND}{\partial STATE\_DUMMY} > 0$$

The provision of feedback information may matter in situations when we have mixed groups with adaptive behavior, e.g., groups consisting of purely selfish agents and of agents who value either reciprocity or inequality aversion, but who when they found that their estimates of other agents behavior were too optimistic decrease their subsequent contributions:

$$\frac{\partial UNCOND}{\partial FEEDBACK} < 0$$

Taking into account the assumed constraints on partial derivatives, we designed the following regression function, in which the squared terms are used to allow for concavity:

Equation 5

$$\begin{aligned} UNCOND_{is} &= \beta_{0,i} + \beta_1 \cdot INC\_DIFFERENCE_{is} + \beta_2 \cdot INC\_DIFFERENCE_{is}{}^2 + \\ &+ \beta_3 \cdot GUESS_{is} + \beta_4 \cdot GUESS_{is}{}^2 + \beta_5 \cdot STATE\_DUMMY_{is} + \beta_6 \cdot FEEDBACK_{is} + \\ &+ \gamma_s \cdot z_i + u_{is} \end{aligned}$$

where i identifies participants and s the round (note that s is not exactly a time index). z_i is the vector of unobserved individual characteristics that are assumed to remain constant during the course of the experiment.

Specifications similar to Equation 5 are fairly common in the literature, however they tacitly assume substantial homogeneity in the sample. If we assume that at least some of the participants behave differently (e.g. a sample contaminated with a student of economics behaving like a rational homo oeconomicus) or if we are actually dealing with a sample consisting of several different types of agents (purely selfish, agents caring about reciprocity, agents caring about inequality), the above specification will lead to biased results. Based on the diversity of observed behavioral patterns we should consider this to be a problem that needs to be dealt with. We can either use methods less sensitive to contamination of the sample (this would be an acceptable approach if we have reason to assume that only selected individuals could be different) or specifications/methods that explicitly incorporate the probable presence of several different types of agents (Equation 6).

Equation 6

$$\begin{aligned} UNCOND_{ir} &= \beta_{0i} + \beta_1 \cdot INC\_DIFFERENCE_{ir} + \sum_{j=1}^{k-1} \delta_j \left(D_j \cdot ID_{jr} \right) + \beta_3 \cdot GUESS_{ir} \\ &+ \sum_{j=1}^{k-1} \theta_j \left(D_j \cdot Guess_{jr} \right) + \beta_5 \cdot STATE\_DUMMY_{ir} + u_{ir} \end{aligned}$$

Using separate data files for the absolute and relative versions of the experiment, we have estimated a simple model of the relationship of the contribution of individual characteristics and differences between the individual and her “neighborhood”, and state and feedback dummies respectively. Fixed effect versions and random effect specifications, LSDV with additional dummies relevant for the differentiation according to the assumed types of agents and LAD estimators have been considered; the latter two as estimators useful for dealing with influential observations (LAD) and assumed possibility of heterogeneity of participants (LSDV with "slope" dummies).

Absolute Version without Feedback

We have focused on just the first five rounds (i.e. the rounds for which the participants did not have any feedback before they reported their decisions). There were two basic reasons for this approach:

1. The feedback information can bring additional noise to the data because of different ways the agents can interpret it. Moreover, if they decide to rely on the information, we should expect some correction of the role of the feedback in rounds 7 and 8 (the participants received the first feedback in round 6), which may necessitate the introduction of too many new dummy variables that would take care of the issue.
2. Later inspection also revealed a small numerical error in our z-Tree program. This error meant that final outcomes could have been slightly higher that they should have been for the given reported "voluntary" contributions. The error would not change the reported final outcomes in the purely selfish case, but it could emphasize the impact of non-zero contribution on income, and so give the participants a false signal about reciprocity in the group. Moreover, the size of the error was not dramatic,[62] and groups were drawn randomly for each round, so there was little motivation to rely on the feedback information in subsequent rounds.

We estimated a slightly modified version of the "contribution function" as the feedback dummy has to be omitted.

Table III.23 shows the results; the Hausman test did not reject the consistency of the random effect model, while we would prefer FE because of the suspected role of unobserved variables. We also report the simple LAD as an attempt to reduce the role of possible influential observations.

[62] The final outcome of a purely selfish participant would not change at all, the final outcome of a "middle-income" participant in a group where everyone contributed suggested 150 ECU would increase by 12.5% in the efficient state (and 25% in the inefficient state). The error influenced only the absolute version and only the reported final outcome used in the feedback after the end of rounds 6-8. No other variables were affected.

Equation 7

$$UNCOND_{is} = \beta_0 + \beta_{1i} \cdot INC\_DIFFERENCE + \beta_{2i} \cdot INC\_DIFFERENCE^2 + \beta_{3i} \cdot GUESS + \beta_{4i} \cdot GUESS^2 + \beta_{5i} \cdot STATE\_DUMMY + u_{is}$$

Variable	**FE**	**RE**	**Simple LAD**	**LSDV with additional dummies**
INC_DIFFERENCE	0.095**	0.136**	0.093**	1.376**
	(0.048)	(0.029)	(0.046)	(0.806)
$INC\_DIFFERENCE^2$	0.0004**	0.0004**	0.0003**	0.003**
	(1.171)	(0.000)	(0.000)	(0.002)
GUESS	1.171**	1.315**	1.326**	1.871**
	(0.192)	(0.165)	(0.309)	(0.562)
$GUESS^2$	-0.003**	-0.003**	-0.003**	-0.009**
	(0.001)	(0.001)	(0.001)	(0.003)
STATE_DUMMY	4.579	3.829	4.338	5.796
	(5.045)	(4.930)	(3.156)	(5.939)
INTERCEPT	12.243	-1.080	-697	123.704
	(14.562)	(12.968)	(36.182)	(101.831)
R^2	0.63			

Table III.23 - Absolute Version without Feedback

These results show the statistical significance of the guess variable and of the role of the difference between own income and of the difference between one's own income and average income in the group. The state dummy has the right sign, but we cannot reject that it is equal to zero. We also were not able to reject the hypothesis that the coefficient for GUESS was higher than 1. However, for high values of GUESS this was compensated by the concavity of the response to GUESS (see Figure III.14 which shows response to GUESS for the range of values present in our data).

Our attempt to account for the possible presence of agents of different types in the sample (see Equation 6) led to outcomes with a similar structure and estimated coefficients. The F-test for joint significance of the additional slope dummies would lead us to the conclusion that we cannot reject their significance at the 10% significance level.[63]

[63] The p-value was 0.051.

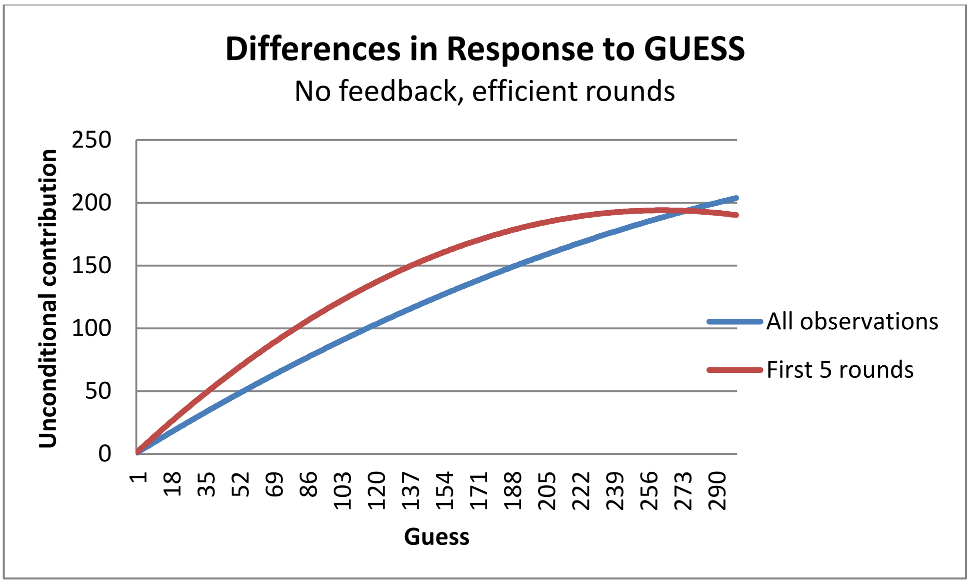


Figure III.14 - Response to GUESS, Absolute Version

Relative Version without Feedback

As in the absolute version, we have decided to focus on the first five rounds without the possible disturbing effects of feedback. We also converted the variables GUESS and UNCOND into ECU in order to facilitate direct comparisons with the absolute version.

Table III.24 summarizes the results for fixed effect specification[64]. When compared with the results for the whole dataset, we find a different sign for the state dummy and the square of GUESS (but we cannot reject the hypothesis that their influence is insignificant in both cases). There are again two variables with statistically significant influence (GUESS and INC_DIFFERENCE), and the slope coefficients of the model were also jointly statistically significant (F-test). The point estimates of the coefficients were similar as in the previous case, and we were not able to reject the hypothesis that the coefficient for INC_DIFFERENCE was equal to the previous results. However, we found an interesting difference in the result for GUESS – in this case we were not able to reject the hypothesis that it was equal to 1, which makes this result more similar to the absolute version.

Variable	Coefficient	Standard Error	P-Value
INC_DIFFERENCE	0.256*	0.047	0.000
INC_DIFFERENCE^2	0.0001	0.0002	0.383
GUESS	0.749*	0.203	0.000
GUESS^2	-0.0001	0.0009	0.904
STATE_DUMMY	-1.975	4.092	0.630

Table III.24 - Fixed Effects, Relative Version, No Feedback

[64] Random effect model was rejected by the Hausman test.

The explanatory power of the model was comparable to the previous cases, and the adjusted R^2 was equal to 0.798.

Summary of Econometric Analysis

The results for both versions of the experiment show that both inequality aversion (the influence of INC_DIFFERENCE) and reciprocity (GUESS) matter. In both cases the estimates had the expected sign and were statistically significant. The reciprocity motive seems to be stronger. In two of the cases we were not able to reject that the coefficient equals 1, which means that the participants preferred “quid pro quo” behavior, i.e. ceteris paribus contributing as much as they expect the others to contribute.[65] The estimates of the coefficients for INC_DIFFERENCE differed less. We cannot reject the hypothesis that they were identical (and positive) in all the treatments. The point estimates of the coefficient for INC_DIFFERENCE were 2.4-8.5 times smaller than for GUESS, which means that fairly high positive difference between one’s own income and the known average income of the others was necessary to motivate the participants to tolerate low expected contributions from others and to contribute more themselves. The difference between the values of the coefficients was lower in the relative versions of the experiment, and on the other hand, the average willingness to contribute was also lower in the relative version.

5. Comparative Analysis of the Two Tax Systems

The relative and the absolute version of our experiment can be interpreted as different tax systems applied to the same economy. Participants in the experiment had the same expected earnings in both versions and were asked to contribute either 150 ECU (in the absolute version) or 50% of the estimated average income, which was about 300 ECU (in the relative version). As the participants also received income (the “public goods”) which equaled the required contribution, any deviation from the required contribution also leads to a change in income distribution within the groups. We can therefore compare the impacts of the two systems on the efficiency of tax collection and on income distribution. The experimental design made it possible to abstract from the impact of differences in tax collection on investment/saving behavior, and motivation to participate in economic activities. Unlike studies based on

[65] There could be another possible explanation for this behaviour than reciprocity. If we assume that most of our participants reported as a guess their (erroneous) solution to the optimum strategy in the game and then they assumed that all the other participants behave rationally and contribute according to the same optimum strategy, we would find the same result. However, this explanation would rely on the assumption of very insufficient rationality (the game was not so difficult to solve).

real world data, we can therefore focus specifically on the impact of motives related to reciprocity and inequity aversion.

5.1. "Tax Avoidance" under the Two Regimes

First we analyze tax avoidance in all available observations. Table III.25 compares average tax arrears and average unconditional contributions under absolute and relative treatments[66]. The average was taken over all groups and periods. If we abstract from small differences in real average incomes from 300 ECU that were caused by motivation bonuses for correct guesses and by the results of the quiz[67], both systems would lead to the same result if the participants contributed as required. Instead we find that the average "voluntary" contributions were lower under the relative treatment (83.8 ECU) than under the absolute treatment (113.5 ECU)[68]. Consequently, average group tax arrears in the relative treatment were almost 1.9 times higher (344.5 ECU compared to 182.4 ECU in the absolute treatment).

Overall Data	Absolute Treatment	Relative Treatment
Average arrears of the groups	182.4	344.5
Average unconditional contribution (of an individual)	113.5	83.8

Table III.25 - Tax Avoidance Comparison, Absolute and Relative Treatment, All Rounds

This result is consistent with the previous conjecture that participants in the experiment cared about reciprocity and much less about inequality. The absolute version (lump sum tax) was much less "fair" in terms of inequality aversion than the relative treatment, but participants might have been more willing to respect the officially suggested contribution because it was more "fair" in terms of reciprocity.

Given the significant role of the efficiency of the collection of "forced contribution" and the fact that efficient and inefficient state were distributed randomly and independently over the absolute and relative treatments, we also analyzed the tax avoidance for the two efficiency states independently (Table III.26 and Table III.27).

[66] The contributions in relative treatments have been converted to ECU for the purpose of comparison.
[67] These two reasons caused the average income to be typically slightly higher than 300 ECU (on average 305 ECU in relative versions and 306 ECU in absolute versions). This means that the average required contribution in the relative version was actually 152.5 ECU instead of 150 ECU. However, these differences were too small to influence the result of the subsequent analysis.
[68] Although the tables contain calculated averages, more robust statistics such as medians led to very similar conclusions.

Only Efficient State	Absolute Treatment	Relative Treatment
Average arrears of the groups	247.9	368.4
Average unconditional contribution (of an individual)	100.4	79.1

Table III.26 - Tax Avoidance, Efficient State

However, the result remained the same. In both efficient and inefficient states, there were substantial and statistically significant[69] differences between average "voluntary" contributions in the relative and absolute versions respectively; the participants still tended to contribute more in the absolute treatment than in the relative treatments, so the "tax avoidance" was substantially higher in the relative treatment.

Only Inefficient State	Absolute Treatment	Relative Treatment
Average arrears of the groups	131.6	326.0
Average unconditional contribution (of an individual)	123.7	87.5

Table III.27 - Tax Avoidance, Inefficient State

5.2. Impact of Tax Avoidance on Inequality

Our two types of treatments were designed to be neutral with respect to inequality. However, systematic deviations from suggested contributions (150 ECU or 50%) should either increase or decrease inequality within the groups.

We measured inequality by variation coefficients (in %) of the incomes in each particular group. Figure III.15 shows the changes in inequality caused by the fact that participants decided to deviate from the suggested contribution (inequality in the case of observance of the suggested contribution is measured on the x axis[70], "observed inequality" on y axis) for both absolute and relative versions. The "observed" inequality describes inequality at the end of each round, i.e. after the participants contributed voluntarily, received their "public goods", and the state collected the "forced" contribution to cover the missing revenues of public goods provision.

There were three types of groups with respect to the hypothetical inequality (x axes in Figure III.15), two of which had similar levels of inequality (the "rich" and

[69] Using a simple t-test, we were able to reject the null of identical means of unconditional contributions for inefficient and efficient treatment at all standard levels of significance. Results of the test are in the appendix.
[70] It is also equal to inequality if no taxation/public goods provision took place.

"middle-income" neighborhoods (on average 31.8 and 33.5 respectively) and one with a higher level of inequality (the "poor" neighborhoods, 50.5 on average).

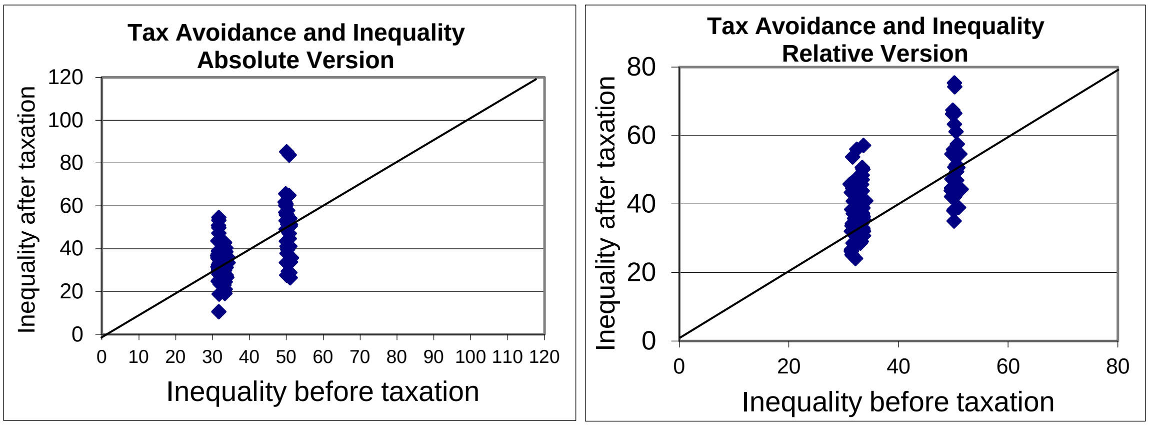


Figure III.15 - Changes in Inequality caused by Taxation

The "contributions" increased the dispersion of the observed inequalities in both absolute and relative versions, but the average impact on inequality was small and ambiguous under the absolute treatment (Table III.28 shows the averages from data in Figure III.15 for the absolute version), moreover, the standard deviations were so high compared to the average change that the effect was not statistically significant.

Neighborhood	Average inequality without taxation	Average observed inequality with taxation	Relative change in inequality (%)	Standard deviation of the relative changes
Poor	50.5	49.7	-1.5	29.1
Middle income	31.8	34.2	+7.7	31.0
Rich	33.5	32.1	-4.4	18.8

Table III.28 - Average Impacts of Taxation on Inequality, Absolute Version

Under the relative version, both the dispersion of results as well as average inequality increased (see Table III.29). The results seem to be less ambiguous, but the standard deviation of the changes in inequality was again rather high compared to the average change.

Neighborhood	Average inequality without taxation	Average observed inequality with taxation	Relative change in inequality (%)	Standard deviation of the relative changes
Poor	50.3	51.8	+2.9	20.7
Middle income	31.8	36.5	+14.7	25.4
Rich	33.5	38.6	+15.3	19.7

Table III.29 - Average Impacts of Taxation on Inequality, Relative Version

Even though the differences in average contributions under relative and absolute versions were statistically significant (section 5.1), the directions of the changes in inequality were not. We can therefore conclude that while the average inequality seemed to be higher in the relative versions, this conjecture is not supported by the sufficiently robust evidence. This seems to be compatible with the results reported in Sections 4 and 5 – i.e. lower average willingness to contribute, but the possibility of higher relative importance of the difference between one's own income and the average income in the neighborhood in the relative version.

6. Conclusions

We experimentally investigated non-pecuniary motives to contribute to public goods provision in a situation when incomes are earned and stratified, the use of collected funds can be efficient or inefficient, and two different contribution schemes (absolute and relative) are used.

Subjects earned their initial income in the first stage of the experiment. Results of this stage implied the emergence of income inequality with three different income levels – rich subjects, middle income subjects, and poor subjects. In the second stage, in every round subjects were randomly sorted into groups such that one income level prevailed in every group, but every group consisted of subjects of all three income levels. In two different frameworks (absolute and relative), subjects were asked for a specific contribution to a public good. The absolute scheme worked on the lump sum principle, while the relative scheme was based on flat-rate taxation. We tested also for the differences between behavioral types according to the tax scheme chosen.

The design of the contribution scheme had properties of a prisoner's dilemma. If the amount of contributions collected within a group did not reach the defined level, punishment was imposed on the group according to the basic type of scheme (absolute or relative), and the collective avoidance to pay became an inefficient strategy. Two levels of punishment were exploited because we tested also for the total group income motivation (see also Charness & Rabin, 2002, and Engelmann & Strobel, 2004). While the efficient punishment scheme is zero-sum, the use of inefficient punishment scheme is costly for the group. We also tested hypotheses concerned to the behavioral differences across the two schemes.

Subjects made their choices based on the average income level of their group in their unconditional decision, and also conditionally on the average contribution level in their group. The strategy method was implemented because subjects must not be affected by the real contribution level of other participants. The experiment was performed according to the standard methodology (see Hertwig & Ortmann, 2001) on a group of 120 students recruited at technical and economics universities in Prague, using the portable experimental laboratory at CERGE-EI and z-Tree experimental software (see Fischbacher, 1999).

In the data analysis section, we first present a consistency test of the subjects' behavior where correct understanding of instructions and the complexity of the situation was measured. Given our definition, the decisions of 51.7% of participants of the absolute versions and 40% of participants of the relative version could be described as consistent. While this suggests possible problems that might have been caused by the complexity of the experiment, and even though the data were noisy, we could identify four patterns of conditional contributions: flat schedules, increasing schedules, decreasing schedules and inverted-V patterns. Because the optimum strategy in the absence of reciprocity and inequality considerations would lead to a flat schedule with all contributions equal to 0, the high incidence of alternative schedules suggests that the motivation of the participants was indeed more complex that simple individual income maximization. We were also not able to reject that unconditional contributions of the participants were significantly different from 0 (i.e. from the optimum "selfish" strategy). These differences persist if the possible influence of the differences between the two states (efficient and inefficient use of the collected funds) was considered.

Using the econometric analysis of the panel data, we provide results showing the relevance of both the reciprocity and inequality aversion motive. Participants derived their unconditional contributions primarily from the expected behavior of the others with coefficient equal to 1. The inequality considerations played a less important but nevertheless statistically significant role; a relatively high positive difference between one's own income and average income in the "neighborhood" was required to induce significant willingness to contribute more than the others. The results also do not exclude the possibility that different types of preferences were present in the sample.

The setup of the experiment also made it possible to analyze possible impacts of the form of collection of contributions for public goods (or taxation) on willingness to contribute. The results suggest that there are statistically significant differences that may be attributed to the parameters of the system of collection; the flat rate regime was found to be much less efficient in terms of raising revenue than the lump sum regime. The participant's deviation from the "benchmark" behavior also influenced inequality within the groups. Tax avoidance had marginally worse impacts on inequality under the flat rate than under the lump sum taxation; however, this result was not statistically significant because of the high variation of the effects of taxation on "within-group" inequality.

Looking at (the robustness of) our results, we consider our experiment to have been relatively successful. As far as possible extensions and modifications of the setup are concerned, it would be useful to replace random generation of efficient/inefficient states with a pre-determined environment and possibly to extend the number of rounds with feedback. Our experiment can be easily extended to analyze other tax regimes, including more traditional progressive taxation. While this is very easy to achieve from the technical point of view (simple modification of the z-

Tree code), our results suggest that the cognitive abilities (more complicated versions than the one presented here will require even more time for studying instructions) and limited attention span of the participants as the main challenges to extensions of the experiment.

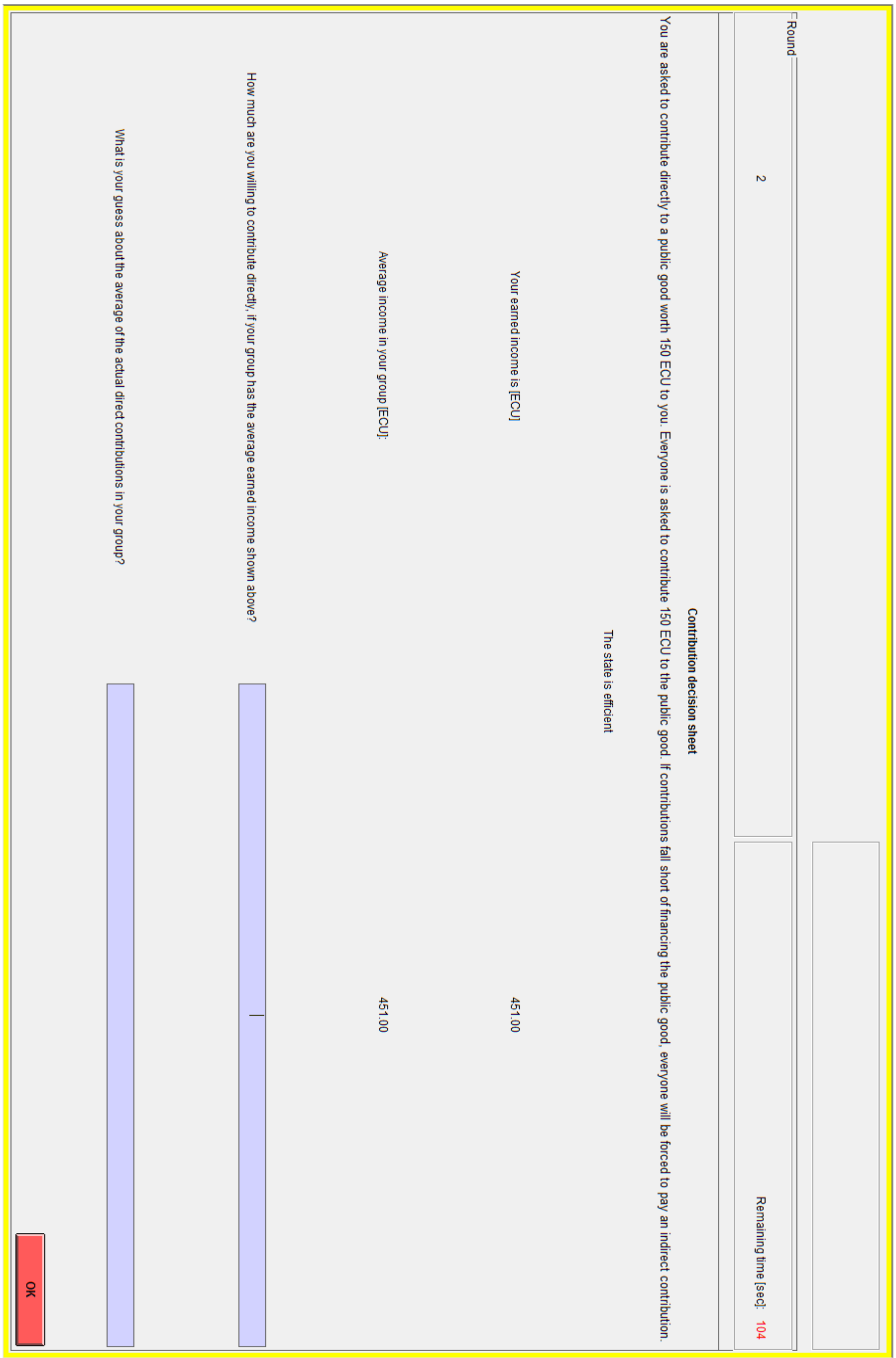
References

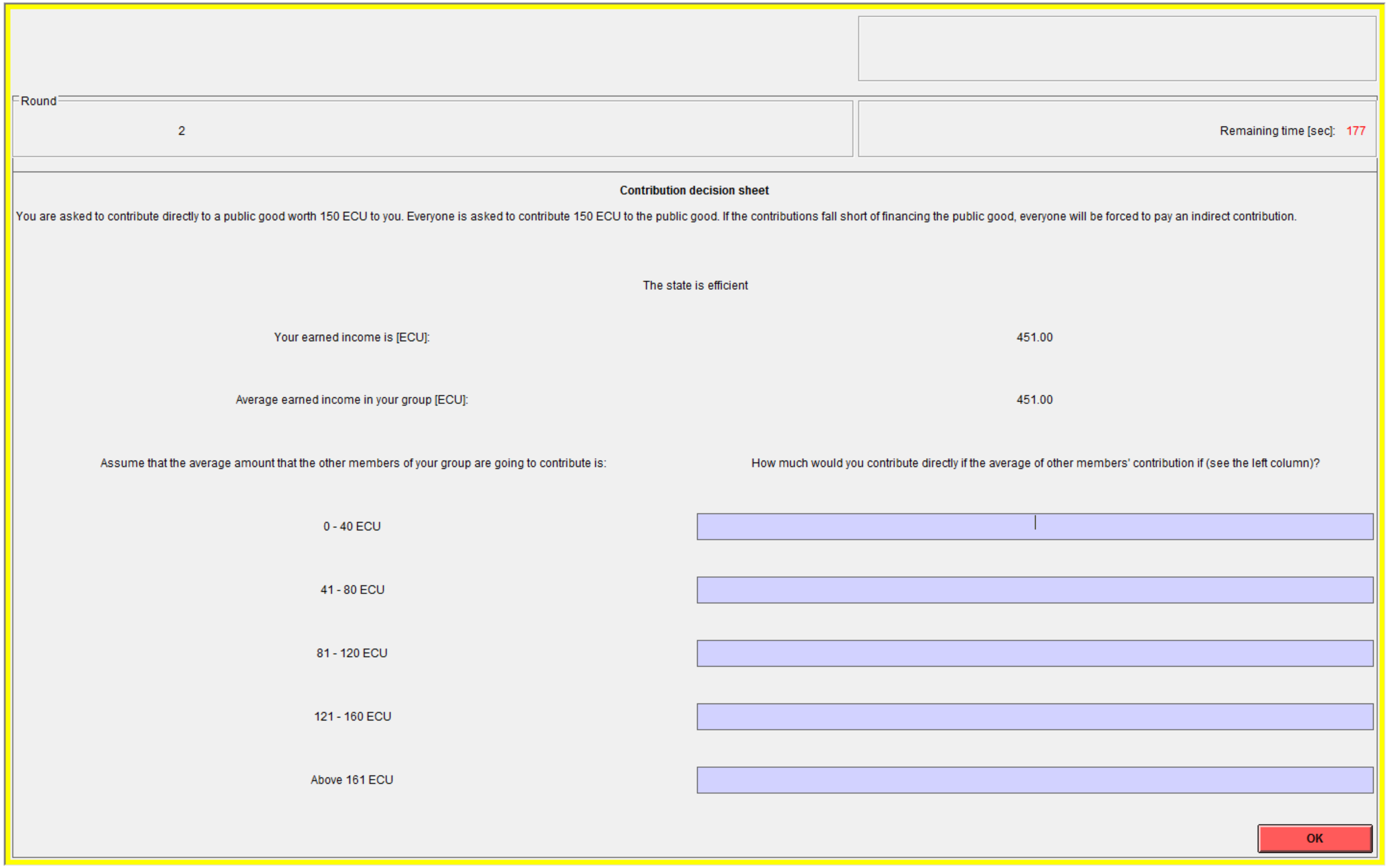
1. **Allignham M.G., Sandmo A., 1972.** Income tax evasion: a theoretical analysis. *Journal of Public Economics 1*, 323-338.
2. **Alm J., Torgler B., 2006.** Culture differences and tax morale in the United States and in Europe. *Journal of Economic Psychology 27*, 224-246.
3. **Alm J., 2012**. Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International Tax and Public Finance 19*, 54-77.
4. **Bolton G., Ockenfels A., 2000.** ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review 90*, 166-193.
5. **Charness G., Rabin M., 2002.** Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics 117*, 817-869.
6. **Cherry T. L., Frykblom P., Shogren J.F., 2002.** Hardnose the Dictator. The *American Economic Review 92*, 1218-1221.
7. **Cox J.C., Friedman D., Gjerstad S., 2007.** A Tractable Model of Reciprocity and Fairness. *Games and Economic Behavior 59*, 17-45.
8. **Dufwenberg M., Kirchsteiger G., 2004.** A Theory of Sequential Reciprocity. *Games and Economic Behavior 47*, 268-298.
9. **Elster, J., 1989**. The cement of society - a study of social order. Cambridge: Cambridge University Press.
10. **Engelmann D., Strobel M., 2004.** Inequality Aversion, Efficiency and Maximin Preferences in Simple Distribution Experiments. *American Economic Review 94*, 857-869.
11. **Falk A., Fehr E., Fischbacher U., 2008.** Testing Theories of Fairness – Intentions Matter. *Games and Economic Behavior 62*, 287-303.
12. **Falk A., Fischbacher U., 2002**. “Crime” in the Lab – Detecting Social Interaction, *European Economic Review 46*, 859-869.
13. **Falk A., Fischbacher U., 2006**. A Theory of Reciprocity. *Games and Economic Behavior 54*, 293-315.
14. **Fehr E., Schmidt K., 1999**. A Theory of Fairness, Competition and Cooperation. *The Quarterly Journal of Economics 114*, 817-868.
15. **Feld L.P., Frey B.S., 2002.** Trust breeds trust: how taxpayers are treated. *Economics of governance 3*, 87-99.
16. **Feld L.P., Tyran J.-R., 2002.** Tax Evasion and Voting: An Experimental Analysis. *Kyklos 55*, 197-221.
17. **Fischbacher U., 1999.** z-Tree, Toolbox for Readymade Economic Experiments: Experimenter’s Manual, IEW Working paper No.21, University of Zurich.
18. **Fischbacher U., Gächter S., Fehr E., 2001**. Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters 71*, 397-404.

19. **Güth W., Ortmann A., 2006.** A behavioral approach to distribution and bargaining in Altmann M. (ed.). Handbook of Contemporary Behavioral Economics. M.E. Sharpe, 405-422.
20. **Güth W., Schmittberger R., Schwarze B., 1982.** An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization 3*, 367-388.
21. **Hanousek J., Palda F., 2004.** Quality of government services and the civic duty to pay taxes in the Czech and Slovak Republics, and other transition countries. *Kyklos 57*, 237-252.
22. **Hertwig R., Ortmann A., 2001**. Experimental Practices in Economics: A Methodological Challenge for Psychologists? *Behavioral and Brain Sciences 24*, 383-451.
23. **Kleven H.J., Knudsen M.B., Kreiner C.T., Pedersen S., Saez E., 2011**. Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark. *Econometrica 79*, 651-692.
24. **Selten R., 1967.** Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments, Beiträge zur experimentellen Wirtschaftsforschung, Heinz Sauermann (ed.), Vol. I, Tübingen: J.C.B. Mohr (Siebeck), 136-168.
25. **Traxler, C., 2010.** Social norms and conditional cooperative taxpayers. European *Journal of Political Economy 26*, 89-103.

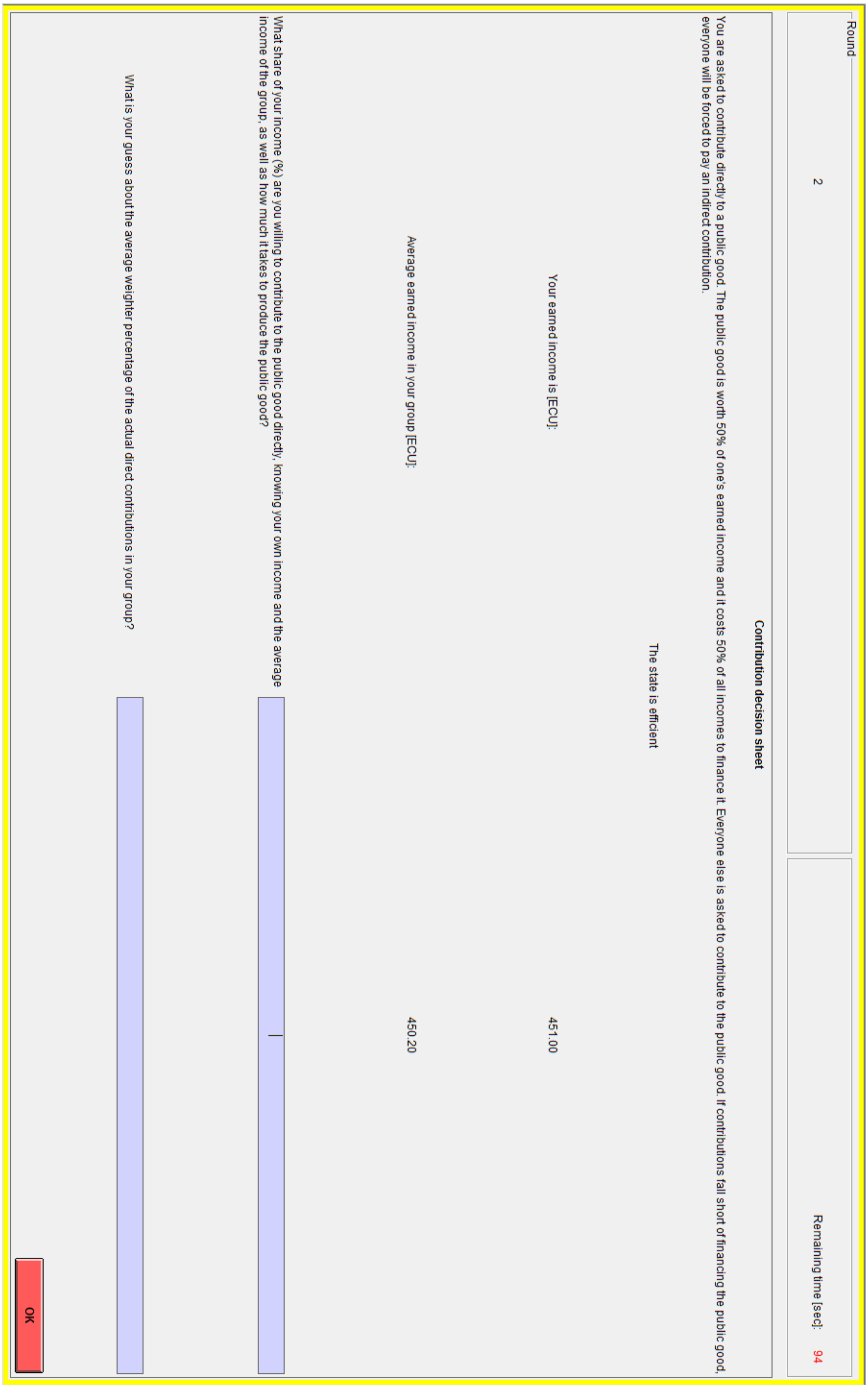
A. Appendix

A.1. Screens from z-Tree Program - Absolute Version




Round
2
Remaining time [sec]: 177
Contribution decision sheet
You are asked to contribute directly to a public good worth 150 ECU to you. Everyone is asked to contribute 150 ECU to the public good. If the contributions fall short of financing the public good, everyone will be forced to pay an indirect contribution.
The state is efficient
Your earned income is [ECU]:
451.00
Average earned income in your group [ECU]:
451.00
Assume that the average amount that the other members of your group are going to contribute is:
How much would you contribute directly if the average of other members' contribution if (see the left column)?
0 - 40 ECU
41 - 80 ECU
81 - 120 ECU
121 - 160 ECU
Above 161 ECU
OK

A.2. Screens from z-Tree Program - Relative Version



Round

2

Remaining time [sec]: 175

Contribution decision sheet

You are asked to contribute directly to a public good. The public good is worth 50% of one's earned income and it costs 50% of all incomes to finance it. Everyone else is also asked to contribute to the public good. If contributions fall short of financing the public good, everyone will be forced to pay an indirect contribution.

The state is efficient

Please, fill in the following table:

Your earned income is [ECU]:	451.00
Average earned income in your group [ECU]:	450.20

Assume that the average weighted percentage that other members of your group are going to contribute is:	How many percent of your earned income are you willing to contribute to the public good if the average contributions of the other members of the group lie in a particular range of values (see the left column)?
0% - 15%	
16% - 30%	
31% - 45%	
46% - 60%	
Above 61%	

OK

A.3. Types of Conditional Schedules in Absolute Version

Classification criterion: at least 7 consistent decisions required

Type	**Incidence**	**Share [%]**
Flat	*0*	*0.0*
Strictly increasing	3	5.0
Non-decreasing (flat not included)	*14*	*23.3*
Strictly decreasing	2	3.3
Non-increasing (flat not included)	*14*	*23.3*
Strict inverted V-pattern	4	6.7
Weak inverted V-pattern (flat not included)	25	41.7
Weak inverted V-Pattern (flat, non-increasing, and non-decreasing not included)	*21*	*35.0*
Agents in no category	*11*	*18.3*

A.4. Types of Conditional Schedules in Relative Version

Classification criterion: at least 7 consistent decisions required

Type	**Incidence**	**Share [%]**
Flat	*3*	*5.0*
Strictly increasing	8	13.3
Non-decreasing (flat not included)	*13*	*21.7*
Strictly Decreasing	7	11.7
Non-increasing (flat not included)	*18*	*30.0*
Strict inverted V-Pattern	2	3.3
Weak inverted V-Pattern (flat not included)	13	21.7
Weak inverted V-Pattern (flat, non-decreasing, non-increasing not included)	*11*	*18.3*
Agents not included in any category	*15*	*25.0*

A.5. Test of Differences in Mean Contributions in Absolute and Relative Version

Null hypothesis in both cases: mean contributions are identical.

Efficient State

	Absolute	***Relative***
Mean	100.4285714	79.07082857
Variance	1162.867456	1031.832417
Observations	42	42
Hypothesized Mean Difference	0	
df	82	
t Stat	2.954557384	
P(T<=t) one-tail	0.002042872	
t Critical one-tail	1.663649185	
P(T<=t) two-tail	0.004085744	
t Critical two-tail	1.989318521	

Inefficient State

	Absolute	***Relative***
Mean	123.6874	87.45016
Variance	807.6207	991.8474
Observations	54	54
Hypothesized Mean Difference	0	
df	105	
t Stat	6.277404	
P(T<=t) one-tail	3.94E-09	
t Critical one-tail	1.659495	
P(T<=t) two-tail	7.88E-09	
t Critical two-tail	1.982815	

Concluding remarks

The previous chapters were written (Chapter I) or at least conceptualized, designed, and experimentally implemented more than a decade ago. For a variety of reasons it has taken a long time to conclude them up. The theory chapter – already published in 2003 as working paper – has been authenticated through several citations, most noteworthy by Cappelen et al. in AER 2013.

The second chapter which builds on the first and for which the experimental sessions were conducted in 2004-5, was finally published in 2013 in Games, i.e. it has a fairly recent date of acceptance. Chapter III, for which the experimental sessions were conducted in 2005-6, is presented here for the first time. It was the most ambitious of all the projects, and we believe it to have been – certainly at the time of its conception and experimental implementation – state of the art. It was one of the very first real effort experiments (e.g., Dutcher, Salmon and Saral, 2015), and its results seem interesting, intuitive, and policy relevant.

Namely, and arguably most noteworthy, we found very robust evidence for the presence of systematic deviations from purely “selfish” preferences under earned wealth. This is an important result in light of the well-documented fact that earned wealth tends to shift people dramatically away from other-regarding behavior (see Cherry et al., 2002 for a prominent example). Through various types of robust econometric analysis, we provide evidence that reciprocity matters (both for unconditional and conditional decisions) and that inequality aversion matters, too. Based on our experimental results, it seems that inefficiency has an effect similar to punishment in that the avoidance of losses leads to the tendency to increase one’s own contributions, and type of the treatment also matters for behavioral patterns. Moreover, we found that flat tax regimes may have more undesirable impacts on inequality than lump sum regimes, mainly through lower total voluntary contributions under relative treatments.

We acknowledge important shortcomings of the experiment reported in Chapter III. The complexity of the environment had a considerable fraction of subjects (e.g., candidates are those who acted inconsistently by our measure) behaving often with bounded rationality or maybe rational inattentiveness. Because of that, and because various people behave differently under various patterns, our overall conclusions are not as clear as we would like. However, we believe these shortcomings are a reflection of the ambitiousness of our undertaking rather than a question of problematic design, implementation, or econometric evaluation.

References

1. **Babicky, V., 2003.** Fairness under risk: Insights from dictator games. CERGE-EI Working Paper No. 217; The Center for Economic Research and Graduate Education, Economic Institute: Prague, Czech Republic.
2. **Cappelen, A.W.; Konow, J.; Sorensen, E.; Tungodden, B., 2013.** Just luck: An experimental study of risk taking and fairness. *Am. Econ. Rev. 103*, 1398-1413.
3. **Cherry, T.; Frykblom, P.; Shogren, J., 2002.** Hardnose the dictator. *Am. Econ. Rev. 92*, 1218-1221.
4. **Dutcher E. G.; Salmon T.; Saral K., 2015.** Is “Real” Effort More Real? MPRA Paper 68394, University Library of Munich, Germany.
5. **Van Koten, S.; Ortmann A.; Babicky V., 2013.** Fairness in Risky Environments: Theory and Evidence. *Games 4*, 208-242.