

Becker, Nicole; Häger, Kirsten; Heufer, Jan

Working Paper

Revealed Notions of Distributive Justice II – Experimental Evidence

Ruhr Economic Papers, No. 444

Provided in Cooperation with:

RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Becker, Nicole; Häger, Kirsten; Heufer, Jan (2013) : Revealed Notions of Distributive Justice II – Experimental Evidence, Ruhr Economic Papers, No. 444, ISBN 978-3-86788-501-0, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Essen, <https://doi.org/10.4419/86788501>

This Version is available at:

<https://hdl.handle.net/10419/85068>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



RUHR

ECONOMIC PAPERS

Nicole Becker
Kirsten Häger
Jan Heufer

Revealed Notions of Distributive Justice II – Experimental Evidence

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, email: W.Leininger@wiso.uni-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Christoph M. Schmidt
RWI, Phone: +49 (0) 201/81 49-227, e-mail: christoph.schmidt@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #444

Responsible Editor: Wolfgang Leininger

All rights reserved. Bochum, Dortmund, Duisburg, Essen, Germany, 2013

ISSN 1864-4872 (online) – ISBN 978-3-86788-501-0

The working papers published in the Series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

Ruhr Economic Papers #444

Nicole Becker, Kirsten Häger, and Jan Heufer

**Revealed Notions of Distributive
Justice II – Experimental Evidence**



Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über:
<http://dnb.d-nb.de> abrufbar.

<http://dx.doi.org/10.4419/86788501>

ISSN 1864-4872 (online)

ISBN 978-3-86788-501-0

Nicole Becker, Kirsten Häger, and Jan Heufer¹

Revealed Notions of Distributive Justice II – Experimental Evidence

Abstract

We report the results of a combination of a dictator experiment with either a “social planner” or a “veil of ignorance” experiment. The experimental design and the analysis of the data are based on the theoretical framework proposed in the companion paper by Becker, Häger, and Heufer (BHH, 2013), in which we introduce a “notion of distributive justice” by which individuals trade off equality and efficiency. The purpose of the theoretical framework is to explain preferences in dictator experiments by a combination of selfishness and concerns for distributive justice. Most participants conform very well with the Agreement and Symmetry axioms proposed in BHH; we find that for 80% of participants the evidence is very strong. The experiment therefore demonstrates that most participants’ behaviour in dictator experiments can be explained by a combination of selfishness and concerns for distributive justice. We also provide a rough classification of preferences and notions of distributive justice and show that participants’ strength of the sense for justice (Karni and Safra 2002b) can be compared nonparametrically.

JEL Classification: C14, C91, D11, D12, D63, D64

Keywords: Altruism; dictator games; distributive justice; experimental economics; nonparametric analysis; preference decomposition; revealed preference; social preferences

October 2013

¹ Nicole Becker, RGS Econ and TU Dortmund University, Kirsten Häger, Friedrich Schiller University Jena, Jan Heufer, TU Dortmund University – Views expressed represent exclusively the authors’ own opinions. Thanks to Wolfgang Leininger for his support. Thanks to Christian Bredemeier, Katharina Eckartz, Dirk Engelmann, Yiquan Gu, Burkhard Hehenkamp, Linda Hirt, Jörg Franke, Oliver Kirchkamp, Anthony la Grange, Wolfgang Leininger, and Karl Schlag for helpful comments and discussions, and to the participants of the European Meeting of the Econometric Society 2011 in Oslo, the Dortmund Brown Bag Seminar 2011, the Brown Bag Seminar 2012 at the Max Planck Institute of Economics, the seminar at the Research Institute of Industrial Economics (IFN) in Stockholm 2013, and the World Meeting of the Economics Science Association 2013 in Zurich. Thanks to Ulrike Vollstädt for helpful advice on programming the experimental software, and Markus Prasser, Katharina Eckartz, and Severin Weingarten for the experimental assistance. Funding from ProChance (Friedrich Schiller University Jena), TU Dortmund University, the Ruhr Graduate School in Economics, and Friedrich Schiller University Jena is gratefully acknowledged. – All correspondence to: Jan Heufer, TU Dortmund University, Department of Economics and Social Science, 44221 Dortmund, Germany. E-Mail: jan.heufer@tu-dortmund.de

1 INTRODUCTION

1.1 *Motivation and Aim*

Economic experiments have shown that people do not only have selfish preferences, but also other social preferences. They have shown that altruistic preferences and a concern for justice do exist. On the one hand, most participants of economic experiments do give a fraction of their endowment in the dictator game (cf. Forsythe et al. 1994, Engel 2011). As trustees they usually return a positive amount in the trust game although they have no future benefit from this (cf. Berg et al. 1995, Burks et al. 2003, Sutter and Kocher 2007, Willinger et al. 2003). Most of them are not pure free riders in the public goods game. They even contribute in the last round of a game (cf. Andreoni 1995, Fischbacher and Gächter 2010, Isaac and Walker 1988) and also in a stranger matching (cf. Andreoni 1995, Fischbacher and Gächter 2010). They sacrifice own payoff to punish non-cooperators in a stranger matching (cf. Fehr and Gächter 2000, 2002), and they do this even as a third party who has no personal stake in the game (cf. Fehr and Fischbacher 2004, Henrich et al. 2006). On the other hand, experimental results also show some selfish tendencies. For example, the amount given to recipients in the dictator game is usually smaller than the fraction kept for oneself (cf. Forsythe et al. 1994, Engel 2011). Thus, people have shown both a concern for their own selfish needs as well as a motivation to give, share, or contribute, and the two types of preferences are reflected in their choices.

In this paper we decompose preferences shown by our experimental participants into a *notion of distributive justice* (or simply *notion of justice*) and a *selfishness* part. Further, we recover individual notions of distributive justice and classify our participants into prototypical types. Our experimental design and the analysis of the data are based on the theoretical framework proposed in a companion paper (Becker et al. 2013, henceforth BHH), in which we introduce a "notion of justice" by which individuals trade off equality and efficiency.

1.2 *Background*

BHH introduce two axioms on a person's preference and notion of distributive justice: *Symmetry* (SY) and *Agreement* (AG). Symmetry means that no individual is per se more deserving than any other individual, and that all permutations of allocations among different persons should be considered equally just when nothing specific is known about them. Agreement means that if an allocation x is at least as just as an allocation y and x gives a decision maker a higher payoff than y then the decision maker should not strictly prefer y over x . Thus, if the comparison of the "justness" of two allocations agrees with the comparison of the own payoff, the final preference over x and y will agree with both as well. An alternative (equivalent) interpretation is that if an allocation y yields a higher payoff for the decision maker than an allocation x and the decision maker prefers x over y although the payoff is lower for him, then this must be because he feels that x is more just than y .

In this paper, we will analyse data of an experiment which we designed to test the two aforementioned axioms and which consists of two different tasks for each participant (see section 2): A dictator experiment (henceforth D-experiment) will be used to recover a participant's preferences and either a social planner task (henceforth P-experiment) or an allocation task under the veil of ignorance (henceforth V-experiment) will be used to recover a participant's individual notion of distributive justice.

In the D -experiment participants choose distributions of money between themselves and another participant from budgets with varying transfer rates as in Andreoni and Miller (2002) and in the two-person experiment by Fisman et al. (2007). These transfer rates are generally called price ratios, as the rate at which one unit of payoff can be transferred to another participant can be interpreted as the price of giving. In the P -experiment participants distribute money between two other unknown persons. In the V -experiment participants distribute money between two persons labelled A and B. After their decision has been made, they will be assigned one of the two roles at random while another participant will be assigned the other role.

We collected choices from two of these tasks (from the D -experiment and either the P - or V -experiment) for each participant. We first test whether participants satisfy the most basic axiom, the *Generalised Axiom of Revealed Preference* (GARP), in both the D - and the P - or V -experiment. GARP is a necessary and sufficient condition for utility maximisation. We then go on to test whether participants choose according to the two axioms proposed in BHH. To do so, we test the analogue revealed preference formulations of the Symmetry and the Agreement axiom. These conditions are called the SYGARP (for Symmetry, tested only for the P - and V -experiment), the AGGARP (for Agreement, using data from both tasks), and the AGSYGARP (for Symmetry and Agreement together).

In addition to the two-person experiment, Fisman et al. (2007) also analyse the results from a three-person experiment. They provided participants with three dimensional budgets and asked them to distribute money between themselves and two others. This procedure allows them to distinguish preferences for giving from what they call “social preferences” (i.e., preferences about the trade-off between the payoffs of others). Our axioms allow us to provide additional interpretations of their results.

Furthermore, our approach has some advantages compared to the paper by Fisman et al. (2007): First, with our study we are able to capture notions of justice (or “social preferences”) also for those who give little or nothing at all in the D -experiment. Even if we look at the P - and V -experiment separately we find that most participants make sensible decisions. Second, our test for the Agreement axiom is more direct and robust. This is in particular valid as many of our participants do not give much to others when the price ratio is particularly low. In a multi-dimensional setting one would have less chances to learn anything about the notion of justice in situations with these budgets. Third, we can examine in how far the P - or V -experiment are suitable to find out about notions of justice. If the V -experiment is suitable to do this it has an important practical advantage compared to the three-person game: It is incentive-compatible and possibly easier to understand for participants than a three-dimensional task. It is incentive-compatible in the sense that the decision of the allocation between the *two* others in Fisman et al. (2007) does not have any effect on one’s own payoff, whereas all parts of the allocation in our V -experiment can have consequences for oneself. It is possibly easier to understand because money has to be distributed between two persons instead of three, price ratios are straightforward, and a two-dimensional graphical representation of a budget might be easier to understand than a three-dimensional one. Finally, we apply non-parametric techniques without relying on specific functional forms and without the need to impose separability. Our experimental design facilitates the revealed preference analysis in BHH and allows to test the axiomatic foundations of the theory developed therein.

1.3 Outline

The remainder of the article is organised as follows: Section 2 briefly summarises the aims and scope of our study (Section 2.1), reviews the theory proposed in BHH (Section 2.2), and provides

information on the experimental protocol (Section 2.3). Section 3 explains the efficiency tests for our experimental analysis. Section 4 presents our experimental results, divided into a test for consistency with our axioms (Section 4.1), the distribution of prototypical preferences and notions of justice (Section 4.2), a graphical inspection of participants' choices (Section 4.3), and non-parametric interpersonal comparisons (Section 4.4). We also test if there are any treatment specific differences in the data (Section 4.5). Section 5 concludes and discusses strength and limits of our study as well as future research. Appendix A provides additional material. Instructions and a screen shot of the experimental software can be found in Appendix B.

2 THEORY AND EXPERIMENTAL DESIGN

2.1 Aims and Scope

In our experiment each individual participated in two types of tasks: The D-experiment and one of the other two tasks (P-experiment or V-experiment). With the combination of these tasks we want to decompose our participants' behaviour into a notion of justice and a selfishness component. Their choices will also allow us to sort them into different justice types.

We keep the experimental design as simple as possible. In the D-experiment the decision maker allocates money to himself and one other person. In the V-experiment he allocates money to himself and one other person without knowing which of the two allocations he will receive. And in the P-experiment the decision maker allocates money to two other persons.

Generally it would be desirable to get allocations from as many choice sets as possible from each participant. However, due to participant fatigue after a certain time and financial constraints, the number of choice sets had to be kept reasonably small. We therefore limited the number of rounds to 15 per task.

2.2 Brief Summary of the Theory

To keep this part of the paper self-contained, we briefly summarise the theory developed in BHH here. We only consider the two-dimensional case here, as we conducted only two-dimensional experiments; this also simplifies the analysis. For a more detailed analysis and discussion the reader is referred to the companion paper (BHH).

Let $\mathbb{A} = \mathbb{R}_+^2$. We consider the *preference* and the *notions of justice* of a participant who is asked to allocate money (i) between himself and another participant, or (ii) between two other participants. An allocation is an element $a \in \mathbb{A}$. In case (i), a_1 is the participant's own payoff. He chooses according to his preference \succsim in case (i) and according to his notion of justice \succsim_J in case (ii). The *self payoff* relation \succsim_S is trivial and the same for all participants; it is defined as $a \succsim_S a'$ if $a_1 \geq a'_1$.¹

The first axiom is called *Symmetry* (Sy) and concerns the justice relation. As the participant does not know the identity of the other individuals, he should not feel that one is more deserving than another. Thus,

$$(a_1, a_2) \succsim_J (a_2, a_1) \text{ for all } a \in \mathbb{A}. \quad (\text{Sy})$$

¹The strict and indifference relations are defined as usual, that is, as the asymmetric and symmetric parts of the weak relations.

The second axiom, called *Agreement* (AG), postulates that if both \succsim_J and \succ_S agree, then \succsim agrees with them as well. A participant who feels that a is at least as just as a' , and a provides a higher payoff to the participant than a' , then the participant prefers a over a' . Thus,

$$[\succsim_J \cap \succ_S] \subseteq \succsim \quad (\text{AG})$$

which is equivalent to $[\succ \cap \prec_S] \subseteq \succsim_J$ (see BHH, Fact 1). Thus, if an allocation is preferred to another allocation although it yields a lower payoff for oneself, it must be considered to be more just. Axiom AG formalises the idea that the preference is *between* pure selfishness and pure justice.

Note that SY and AG have immediate implications for a parametric approach. Suppose \succsim is represented by a CES utility function $u(a) = (\alpha a_1^r + [1 - \alpha] a_2^r)^{1/r}$ and \succsim_J is represented by a justice function $v(a) = (\beta a_1^s + [1 - \beta] a_2^s)^{1/s}$. Then SY and AG are equivalent to $\beta = 1/2$, $r = s$, and $\alpha \in [1/2, 1]$ (see BHH, Fact 4). Thus, our axioms provide an immediate interpretation for the elasticity of substitution $r/(1 - r)$ as a representation of the notion of justice.

To formulate a testable version of SY and AG, we turn to revealed preference. We must first specify what kind of data we observe. In our experiments, a participant is asked to make decisions on several *budgets*. Budgets are of the form $B^i = B(\rho^i) = \{a \in \mathbb{A} : \rho^i \cdot a = 1\}$, where ρ is a *price vector*. In the experiment, we *observe* one choice per participant on each of the presented budgets. For each experimental task, we observe choices on $N = 15$ different budgets. The set of observations from the D-experiment is denoted $\Omega_x = \{(x^i, p^i)\}_{i=1}^{15}$, and the set of observations from a P- or V-experiment is denoted $\Omega_y = \{(y^j, q^j)\}_{j=1}^{15}$.

The main idea behind these experiments is that we can use Ω_x to elicit a part of a participant's preference and Ω_y to elicit part of a participant's notion of justice. Let $(Q)^+$ be the transitive closure of a binary relation Q . Given Ω_x , we say that x^i is *directly revealed preferred* to an allocation a , written $x^i R^0 a$, if $p^i x^i \geq p^i a$; it is *revealed preferred* if $x^i R a$, where $R = (R^0)^+$; it is *strictly directly revealed preferred* to a , written $x^i P^0 a$, if $p^i x^i > p^i a$; it is *strictly revealed preferred*, written as $x^i P a$, if $x^i R x^j$, $x^j P^0 x^k$, and $x^k R a$ for some observations x^j, x^k . Given Ω_y , the relations R_J^0, R_J, P_J^0 , and P_J are defined in the same way.

Let

$$\begin{aligned} \{z^k\}_{k=1}^{30} &= \{x^i\}_{i=1}^{15} \cup \{y^j\}_{j=1}^{15} \\ \{\xi^m\}_{m=1}^{60} &= \{(x_1^i, x_2^i), (x_2^j, x_1^j)\}_{i=1}^{15} \cup \{(y_1^j, y_2^j), (y_2^j, y_1^j)\}_{j=1}^{15}. \end{aligned}$$

To test SY, and to use it for further analysis, we impose it on the data. The *SY-closure* $C_{SY}(Q)$ of any binary relation Q on \mathbb{A} is simply defined as

$$a C_{SY}(Q) a' \text{ if } [a Q a' \text{ or } (a_2, a_1) Q a' \text{ or } a Q (a'_2, a'_1) \text{ or } (a_2, a_1) Q (a'_2, a'_1)].$$

For the AG-extensions, we define $\hat{R}_J^0 = R_J^0 \cup [R^0 \cap \prec_S]$ and $\hat{R}^0 = R^0 \cup [R_J^0 \cap \succ_S]$, and similarly for the strict relations and the transitive closure. So if a participant revealed that $a R^0 a'$ even though $a \prec_S a'$, then if we accept the axiom AG, we must conclude that $a R_J^0 a'$, etc. Similarly for the AGSY-extension, we define $\hat{R}_J^0 = C_{SY}(\hat{R}_J^0)$ and $\hat{R}^0 = \hat{R}^0 \cup [\hat{R}_J^0 \cap \succ_S]$.

We will test the data for consistency with various axioms which characterise a certain behavioural hypothesis. In particular, we are interested in the existence of utility functions that represent the revealed preference relations. We say that a utility function u *rationalises* a set of observations Ω_x if

$u(a) \geq u(a')$ whenever $a \mathring{R} a'$, and similarly, a justice function v *rationalises* a set of observations Ω_y if $v(a) \geq v(a')$ whenever $a \mathring{R}_j a'$. A justice function is *symmetric* if for every $a \in \mathbb{A}$, $v(a) \geq v((a_2, a_1))$.

A set of observations satisfies the *Generalised Axiom of Revealed Preference* (GARP) if whenever $x^i \mathring{R} x^j$ then $[\text{not } x^j \mathring{P}^0 x^i]$. There exists a continuous, monotonic, and concave utility function that rationalises a set of observations if and only if the set satisfies GARP (Afriat 1967, Varian 1982).

A set of observations satisfies the SYGARP if whenever $y^i [C_{\text{SY}}(\mathring{R}_j)]^+ y^j$ then $y^j C_{\text{SY}}(\mathring{P}^0) y^i$. There exists a symmetric, continuous, concave, and monotonic justice function which rationalises Ω_y if and only if Ω_y satisfies SYGARP (Corollary 1). Thus, the SYGARP provides a direct test for the hypothesis that participants' choices are consistent with SY.

Given Ω_x and Ω_y , and $\Omega = \Omega_x \cup \Omega_y$, we say that a utility function u *AG-rationalises* Ω if $u(a) \geq u(a')$ whenever $a \mathring{R} a'$, and it *AG-SY-rationalises* Ω if $u(a) \geq u(a')$ whenever $a \mathring{R} a'$, and similarly for justice functions with \mathring{R}_j and \mathring{R}_j .

The following definitions introduce conditions which we show are necessary and sufficient for extended rationalisability to test for AG and the combination of AG and SY: A set of observations Ω satisfies AGGARP if Ω_x and Ω_y satisfy GARP and if whenever $z^i \mathring{R} z^j$ then $[\text{not } z^j \mathring{P}^0 z^i]$, and it satisfies AGSYGARP if Ω_x satisfies GARP and $\hat{\Omega}_y$ satisfies SYGARP and if whenever $\xi^i \mathring{R} \xi^j$ then $[\text{not } \xi^j \mathring{P}^0 \xi^i]$.

In BHH (Theorem 2) we show that in two dimensions (only two persons), the following conditions are equivalent:

1. Ω satisfies AGGARP [Ω satisfies AGSYGARP].
2. There exists a non-satiated, continuous, concave, and monotonic utility function u which AG-rationalises [Ω AG-SY-rationalises] Ω , and a non-satiated, continuous, concave, and monotonic [and symmetric] justice function v which AG-rationalises Ω , such that for all $a, a' \in \mathbb{A}$ with $a >_S a'$, $u(a) \leq u(a')$ implies $v(a) \leq v(a')$ and $v(a) \geq v(a')$ implies $u(a) \geq u(a')$.

In our companion paper, the precise implications of the number of dimensions are discussed in more detail. There are conditions based on the Weak GARP (Banerjee and Murphy 2006) which are equivalent in the two-dimensional case, but for brevity, we only present the GARP-extensions here.

The companion paper also shows how to recover everything that can be said about a participant's preference and notion of justice. For that, we construct revealed preferred and worse, and revealed more just and less just sets based on the observations and the axioms. These sets are referred to as, for example, $\overline{\mathring{R}\mathring{P}}(x^0)$, which is the set of all allocations which a participant must prefer over the allocation x^0 if he satisfies AGSYGARP, based on his relations \mathring{R} and \mathring{R}_j recovered from his choices in the D- and P- or V-experiment. See BHH for more details.

2.3 Experimental Design and Protocol

We ran our experiment from June 2012 to November 2012 at the laboratory of the department of business administration and economics of the University of Jena, Germany. The majority of participants were students from the University of Jena, some were students from the University of Applied Sciences in Jena. Participants came from a broad range of fields. Recruitment was done via ORSEE (Greiner 2004). Our participants were aged 18-32; the average age was 22.7. Slightly more than half of our participants (53.2%) were female. Our experiment consists of four treatments. We conducted three sessions per treatment, that is, twelve sessions with a total of 188 participants.² The

²There were 16 participants in eleven of the twelve sessions. In one of the sessions we only had 12 participants because some of the people who registered did not show up. We ran two prior pilot sessions to test the software, the

experiment was programmed and conducted with z-tree (Fischbacher 2007). Sessions lasted between 90 and 120 minutes (104 minutes on average). The average payment was 16.50 Euros.

The experimental sessions were conducted in the following way: Participants came to the lab and were randomly assigned to seats. To ensure that all participants were capable of reading the instructions all participants passed a German language test; if they had not passed it during a previous experiment they had to do it prior to our experiment. Then all participants read the instructions for the first part (see Appendix B.1), in which it was announced that a second part would follow, on their own and filled out a comprehension test for this part. If participants had questions, they were answered by the experimenter in private. Once the experimenter had controlled the comprehension test, she explained to the respective participants again in private the issues they did not understand. Great care was taken to provide clarifications in the same way each time. After the first part of the experiment had been completed participants received the instructions and comprehension test for the second part knowing that this would be the last part. Other procedures were the same as in the first part. After both experimental parts had been completed participants filled out a questionnaire on socio-demographic issues and questions on the strategies they employed.

We had three kinds of tasks in four different treatments. In each treatment participants played the dictator experiment (D) and either the social planner experiment (P) or the veil of ignorance experiment (V). To account for possible ordering effects, we first conducted the dictator experiment and then the other experiment in treatments 1 and 2 and vice versa in treatments 3 and 4. See Table 1 for an overview of the treatments.

TREATMENT	GAME 1	GAME 2
1	dictator experiment D	social planner experiment P
2	dictator experiment D	veil of ignorance experiment V
3	social planner experiment P	dictator experiment D
4	veil of ignorance experiment V	dictator experiment D

Table 1: The treatments.

In the D-experiment each participant was asked to split up induced budgets with varying price vectors between himself and a randomly matched participant. Giving something to the matched participant meant forgoing own payoff. The amount one had to forego in order to give one unit to the matched participant, however, varied from budget to budget and therefore also from round to round. Since all other participants had to do the same task, each participant also received an amount that a randomly matched participant (not the one that he was matched to for his allocation) allocated to him.

In the P-experiment each participant was asked to allocate given budgets with varying price vectors to two other participants. This means that his choice did not affect his own payoff. However, since all participants faced the same task, each participant also received two allocations from two randomly matched participants (not the ones he allocated money to).

In the V-experiment each participant divided budgets with varying price vectors between himself and another participant, called person A and person B. He allocated the money to person A and person B and it was later randomly determined which person he would be and therefore which of

instructions, and the payoff calibration. The data of the pilot sessions will not be included in the data analysis. We had to repeat one session with a new set of participants due to technical problems.

the payoffs he allocated to person A and person B he would receive. Participants knew that they would receive these roles with equal probability. With this mechanism of randomly determining a role each person also received an amount of money that another participant allocated to the other role. That is, if role A was randomly chosen, each participant received the payoff he allocated to person A and the payoff that another participant allocated to person B.

In each of the three tasks participants received 15 budgets with different price relations for the payoffs of the two matched participants, that is, certain allocations were more efficient than other allocations. All of the participants in a given task received the same 15 budgets, but the order of the budgets was randomised for each participant to account for ordering effects. The 15 budgets were always the same 15 budgets within a task, but differed between the D-experiment and the P- and V-experiment. To avoid reputation effects, the possibility of revenge, and other confounds, players were rematched randomly after every allocation (random stranger matching). No feedback was given between rounds and participants never learned about the identity or choices of the other participants. At the beginning of each of the two parts of the experiment participants had two trial rounds to become familiar with the software which were not payoff relevant before the 15 payoff relevant rounds started.

The budgets are given in Table 2. Income is normalised to 1, so that each budget is defined by the equation $p_1x_1 + p_2x_2 = 1$ for the D-experiment and similarly for the P- and V-experiment. Here, p_1 is the price for one Euro for the decision making participant (x_1), while p_2 is the price for allocating one Euro to the receiving other participant (x_2). The table also shows the price ratio and maximal values of x_1 and x_2 . Figure 1 represents the budgets graphically.

BUDGETS										
BUDGET	DICTATOR					SOCIAL PLANNER AND VEIL OF IGNORANCE				
	p_1	p_2	p_1/p_2	$\max x_1$	$\max x_2$	q_1	q_2	q_1/q_2	$\max y_1$	$\max y_2$
1	0.272	0.061	4.496	3.67	16.50	0.290	0.084	3.452	3.45	11.91
2	0.200	0.067	3.000	5.00	15.00	0.260	0.104	2.499	3.85	9.62
3	0.222	0.111	2.000	4.50	9.00	0.222	0.111	2.000	4.50	9.00
4	0.190	0.095	2.000	5.25	10.50	0.190	0.095	2.000	5.25	10.50
5	0.167	0.083	2.000	6.00	12.00	0.220	0.126	1.745	4.55	7.94
6	0.182	0.152	1.200	5.50	6.60	0.182	0.152	1.200	5.50	6.60
7	0.143	0.143	1.000	7.00	7.00	0.143	0.143	1.000	7.00	7.00
8	0.125	0.125	1.000	8.00	8.00	0.125	0.125	1.000	8.00	8.00
9	0.152	0.182	0.833	6.60	5.50	0.152	0.182	0.833	6.60	5.50
10	0.111	0.222	0.500	9.00	4.50	0.111	0.222	0.500	9.00	4.50
11	0.100	0.200	0.500	10.00	5.00	0.095	0.190	0.500	10.50	5.25
12	0.095	0.190	0.500	10.50	5.25	0.083	0.167	0.500	12.00	6.00
13	0.083	0.167	0.500	12.00	6.00	0.084	0.200	0.420	11.91	5.00
14	0.090	0.279	0.322	11.12	3.58	0.068	0.240	0.283	14.71	4.17
15	0.061	0.272	0.222	16.50	3.67	0.061	0.272	0.222	16.50	3.67

Table 2: The budgets used in the experiment.

To facilitate the task, participants were not only shown numbers indicating the budget and price vector, but the information about the budgets and price vectors were also visualised graphically. The layout of the screens and the procedure was the following (see Figure 25 in Appendix B.2): On the right hand side of the screen, participants could see in words the maximum amount they could give

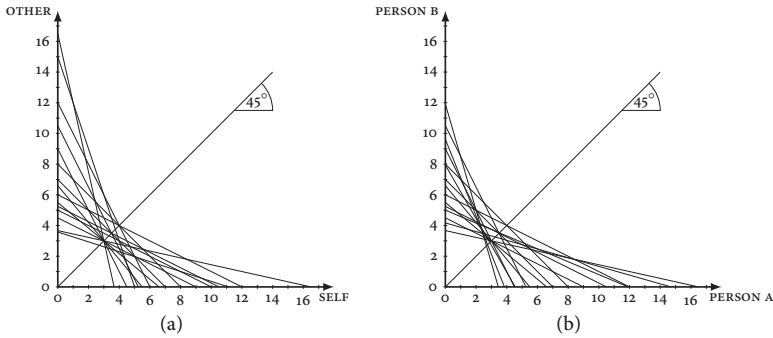


Figure 1: The budgets used in the D-experiment (a) and the P- and V-experiment (b).

to themselves and the other participant in the D-experiment or the maximum amount that they could give to person A and person B in the P-experiment and V-experiment, respectively. On the left hand side of the screen a figure showed the same information: The horizontal axis showed payoffs for oneself in the D-experiment or payoffs for “person A” in the P-experiment and the V-experiment, respectively. The vertical axis showed payoffs for the other person in the D-experiment or payoffs for “person B” in the P-experiment and the V-experiment, respectively. A blue line showed all possible allocations of the budget. To make a decision, participants could enter on the right hand side of the screen how much they wished to allocate to themselves (person A), respectively. When they pressed the “Show”-button, it was shown on the right hand side of the screen how much they (person A) and the other person (person B) would get if this choice was confirmed. On the left hand side of the screen, red lines in the graph indicated the currently chosen allocation of the budget. Participants could try out as many different allocations as they wished until they confirmed their choice with the “Confirm”-button. No feedback was given between rounds. Once everyone completed a round, the next round followed.

For the payment one round of the first task and one round of the second task were randomly chosen. Only these two rounds were paid out for each of the participants. This allows for larger stakes per choice assuming a fixed budget for conducting the experiment and at the same time avoids hedging of risky behaviour. Paying one round of the first task and one of the second task means that each participant received payoff for the D-experiment and either for the P-experiment (treatments 1 and 3) or the V-experiment (treatments 2 and 4). Thus, the payment consisted of four parts: For the D-experiment each participant received the amount he allocated to himself and the amount one of the other participants allocated to him (i.e. to his matched counterpart); for the P-experiment each participant received the amount that one matched participant distributed to player A and the amount that a second matched participant distributed to player B; for the V-experiment each participant received the amount he distributed to the person who’s role was randomly determined for him and the amount another participant distributed to the other role, respectively. As participants did not receive feedback between rounds, they were only informed about their earnings at the very end of the experiment. The amount earned in the experiment was paid out in Euros privately directly after the experiment had been completed.

3 EFFICIENCY TESTS

3.1 Efficiency and Test Power

3.1.1 Afriat Efficiency and Homothetic Efficiency

The data either satisfy GARP or they do not. If participants make minor errors when making decisions it is possible that they produce data that do not satisfy GARP. To account for this, we will use the *Afriat efficiency index* (AEI, Afriat 1972), also called *critical cost efficiency index*, and apply the same concept to the other axioms. The AEI is a goodness-of-fit measure which is one of the most popular for induced budget experiments. Reporting the value is very common for experimental studies.³

The main idea of the AEI is to shift budgets towards the origin and base a new revealed preference relation on the level of expenditure expressed by this new budget (where prices are kept constant). This is done until the data is consistent with GARP. Let $e \in [0, 1]$; for a set of observations $\Omega_x = \{(x^i, p^i)\}_{i=1}^{N_x}$ from a D-experiment, we define $x^i R^0(e) a$ if $e p^i x^i \geq p^i a$ and $x^i P^0(e) a$ if $e p^i x^i > p^i a$, and $R(e) = (R^0(e))^+$. We then say that Ω_x satisfies GARP(e) if $x^i R(e) x^j$ implies [not $x^j P^0(e) x^i$], and the AEI is the greatest number e such that GARP(e) is satisfied. Note that the AEI is a measure for wasted income, as a consumer with an AEI of e could have obtained at least the same level of utility which he actually did obtain by spending only a fraction of e of his income.

We can define, in straightforward analogy, the revealed notion of justice based on efficiency level e , and also the extended relations. The definition for R_j^0 should be obvious; the extended relations are obtained by appropriate replacements with the efficiency level relations. We do not, however, adjust the relation \succeq_S , as it is not revealed. So for example, we have $\tilde{R}_j^0(e) = R_j^0(e) \cup [R^0(e) \cap \prec_S]$. This then leads to straightforward efficiency measures for SYGARP, AGGARP, and AGSYGARP.

The *homothetic efficiency index* HEI is defined for two-dimensional data in analogy to the AEI. It was introduced by Heufer (2013b) and measures the extend of deviation from homothetic utility maximisation. The interpretation is similar to the AEI. We do not focus on homotheticity here, but will report the results nonetheless to compare them with results for data by two other dictator experiments (Andreoni and Miller 2002 and Fisman et al. 2007). See Heufer (2013b) for details.

3.1.2 Test Power

Even a set of random choices has a chance to satisfy GARP or any of the other axioms. To interpret the test results, it is therefore helpful to know the probability that some process other than utility maximisation will also lead to a set of observations that is consistent with an axiom, that is, to know the power of the test. Bronars (1987) suggested a Monte Carlo approach to determine the power against random behaviour: Generate many sets of random choices on the same set of budgets with which the participants were provided, and compute the percentage of random choices which violate GARP. This percentage is the approximate power.

However, if we allow participants to make minor mistakes and also accept those with an AEI of less than 1 as “close enough”, we obviously lose test power. It is therefore necessary to also compute the AEI of the random choices and find the test power by computing the percentage of random

³See, for example, Andreoni and Miller (2002), Choi et al. (2007b), Dickinson (2009), Février and Visser (2004), Fisman et al. (2007), Harbaugh et al. (2001), Mattei (2000), and Sippel (1997). See also Gross (1995) for a survey of other measures.

choice sets which exceed the same AEI-threshold as the one used for the experimental participants.⁴ In general, we will compare the distribution of the AEI in the random data with the distribution of the AEI in the actual experimental data to understand how significant our results are.

3.2 Testing

We will use the prefix x for results for Ω_x (i.e., data from a D-experiment) and y for Ω_y (i.e., data from a P- or V-experiment); for example, $x\text{AEI}$ is the AEI for a set of data Ω_x . The SY-efficiency for the Ω_y data will be called the $y\text{SYAEI}$, the AG-efficiency index the AGAEI , and the AGSY-efficiency index will be called the AGSYAEI .

We will first compute and report the distribution of the AEI for all participants for their choices in the D-experiment ($x\text{AEI}$) and the P- or V-experiment ($y\text{AEI}$). We will also report the $y\text{SYAEI}$ for the P- or V-experiment. We will then combine the two data sets for each participant and compute and report the AGAEI and the AGSYAEI . We will also repeat the first step with random data generated using Bronars' method and compare the efficiency of our participants with the efficiency of the random data.

We will repeat the entire analysis using only a subset of the "best" 14 choices of each task for each participant from the D- and the P- or V-experiment. We find the best 14 choices using a brute force method, that is, we compute the AEI for every participant of 14 choices in each experimental task and use the one which yields the highest AEI. We will refer to these values as the $x_{14}\text{AEI}$, $y_{14}\text{AEI}$, etc. This is done because it is possible that a single mistake made by a participant reduces the AEI substantially, even if GARP is satisfied without that mistake.

Furthermore, we will report the fraction of participants with efficiency indices of 1, and those greater than the thresholds .99, .975, .95, .925, and .9 for all choices, and the best 14 and (for comparison) best 13 out of each experimental task. These numbers will be presented in a way that illustrates set concordance: It will show the fraction of participants which exceed the threshold for any combination of $x\text{AEI}$, $y\text{AEI}$, $y\text{SYAEI}$, and AGSYAEI .

In Appendix A, we provide further analysis. In particular, in Appendix A.1.2, we describe Varian's (1993) improved violation index, and in Appendix A.1.3 we describe how it can be applied to our data. We use it to adjust individual choices to account for violations of GARP and SYGARP, and then use the adjusted data to compute the AGAEI and AGSYAEI . The results are reported in Appendix A.1.4.

We also provide an analysis based on Heufer's (2013a) method to generate random choices which satisfy GARP. This GARP-consistent random data can be tested for consistency with AGGARP and AGSYGARP. A description of this procedure and the results can be found in Appendix A.1.5.

4 RESULTS

We will report several results. First, we give a brief summary of the raw data. Then, in Section 4.1, we analyse the consistency of the individual choice sets with the proposed axioms. In Section 4.2 we provide a rough classification of participants' preferences and notions of justice. Section 4.3 shows the relationship between the log price ratio and the allocation share, which allows for a graphical analysis of choices. In Section 4.4, we apply the non-parametric interpersonal comparison proposed in BHH to compare the strength of participants' sense of justice. Finally, in Section 4.5, we will test if

⁴See also Sippel (1996) and Heufer (2012) for a discussion of this point.

there are any differences in the data between the treatments based on the raw data, efficiency, and interpersonal comparison results.

Table 3 shows the mean and median amount allocated to the other participant by the dictator (x_2) for all 15 budgets. As measures of dispersion, it also shows the coefficient of variation and interquartile range. Furthermore, the table shows the different means by treatment: The mean for treatments 1 and 2 is the mean amount of x_2 in the two treatments in which the D-experiment was the first task. Differences here would indicate an ordering effect. The mean for treatments 1 and 3 is the mean amount of x_2 in the two treatments with the P-experiment. We analyse these differences in more detail in Section 4.5.

DICTATOR EXPERIMENT								
BUDGET	ALL TREATMENTS				T1&2 VS. T3&4		T1&3 VS. T2&4	
	MEAN	CV	MEDIAN	IQR	MEAN	MEAN	MEAN	MEAN
1	5.07	1.09	3.01	8.63	4.95	5.20	4.85	5.31
2	3.23	1.28	1.50	6.00	3.41	3.05	3.17	3.30
3	1.75	1.23	1.00	3.00	1.83	1.67	1.78	1.73
4	2.16	1.20	0.50	3.50	2.27	2.05	2.15	2.18
5	2.38	1.22	2.00	4.00	2.74	2.01	2.33	2.43
6	1.14	1.19	0.60	2.64	1.26	1.02	1.12	1.17
7	1.24	1.12	1.00	2.50	1.38	1.10	1.32	1.15
8	1.37	1.21	0.50	3.00	1.59	1.14	1.32	1.42
9	1.04	1.11	0.50	2.17	1.14	0.94	0.93	1.17
10	0.72	1.31	0.16	1.25	0.81	0.62	0.70	0.74
11	0.70	1.42	0.00	1.12	0.83	0.57	0.65	0.75
12	0.73	1.44	0.25	1.25	0.81	0.63	0.72	0.73
13	0.94	1.36	0.19	1.50	1.09	0.78	0.87	1.00
14	0.46	1.49	0.04	0.69	0.50	0.42	0.46	0.46
15	0.44	1.67	0.00	0.68	0.47	0.42	0.43	0.46

Table 3: Mean and median amounts given in the D-experiment (x_2), the coefficient of variation (standard deviation divided by mean), and the interquartile range. The last four columns show the mean amount given in treatments 1 and 2 (D first) vs. 3 and 4 (P or V first), and treatments 1 and 3 (P) vs. 2 and 4 (V).

Table 4 contains the same information as Table 3, except that it shows the amount allocated to person B in the P- or V-experiment. Tables 9 and 10 in Appendix A.4 show the same information for the restricted set and for all treatments separately.

4.1 General Consistency

4.1.1 Consistency with Axioms

We start by testing the axioms. First, we test whether participants satisfy the most basic axiom, GARP, in both the D- and the P- or V-experiment, as this is a necessary and sufficient condition for the existence of a non-satiated utility function which rationalises the data. We will refer to these two conditions as the xGARP (D-experiment) and the yGARP (P- or V-experiment). Next, we are interested in whether or not participants conform to the new axioms introduced in BHH 2013: Symmetry (SY) and Agreement (AG). We test these two axioms independently and together: The data from the P-

SOCIAL PLANNER AND VEIL OF IGNORANCE EXPERIMENT								
BUDGET	ALL TREATMENTS				T1&2 VS. T3&4		T1&3 VS. T2&4	
	MEAN	CV	MEDIAN	IQR	MEAN	MEAN	MEAN	MEAN
1	7.11	0.51	7.34	6.90	6.93	7.29	6.77	7.46
2	5.17	0.50	4.62	4.36	5.00	5.35	4.74	5.62
3	4.78	0.48	4.31	3.10	4.76	4.81	4.79	4.78
4	5.79	0.45	5.50	4.00	5.88	5.69	5.57	6.02
5	4.07	0.42	3.58	1.57	3.87	4.28	3.98	4.16
6	3.33	0.36	3.00	0.24	3.37	3.28	3.23	3.43
7	3.36	0.31	3.50	0.00	3.40	3.31	3.54	3.17
8	3.95	0.27	4.00	0.00	4.09	3.81	4.02	3.88
9	2.61	0.36	3.00	0.35	2.58	2.65	2.66	2.56
10	2.00	0.56	2.23	2.00	1.95	2.05	2.20	1.79
11	2.30	0.55	2.62	2.25	2.14	2.46	2.32	2.28
12	2.48	0.60	3.00	3.00	2.29	2.68	2.69	2.26
13	2.18	0.62	2.48	2.50	2.07	2.30	2.37	1.99
14	1.57	0.77	1.34	2.55	1.44	1.70	1.69	1.44
15	1.37	0.84	1.22	2.42	1.28	1.46	1.54	1.19

Table 4: Mean and median amounts allocated to person B in the p- and v-experiment, the coefficient of variation (standard deviation divided by mean), and the interquartile range. The last four columns show the mean amount allocated to person B in treatments 1 and 2 (D first) vs. 3 and 4 (P or V first), and treatments 1 and 3 (P) vs. 2 and 4 (V).

and v-experiment are tested for consistency with SYGARP, and the combined data are tested for consistency with AGGARP and AGSYGARP.

Table 5 shows the fraction of participants who fully satisfy the axioms we are interested in. It also shows the fraction of participants who have a very high efficiency index greater than .999 – here, a very small change in the choices made would lead to full consistency. The second part of the table shows the same results restricted to subsets of the best 14 choices from each experimental task. The table also shows the percentage of random choice sets who satisfy the conditions (i.e., the approximate power of the test).

Note that our test has an extremely high power. Even for the best 14 choices, none of the random choice sets satisfies 14AGSYGARP and only .01% (that is, 1 out of 10,000) has a 14AGSYAEI of at least .999.

Given that participants were asked to make 30 choices overall (15 per task), we believe that allowing one erroneous choice per task can be acceptable. The test power here is still very high. We note that the vast majority of participants are in line with utility maximisation when based on the $x_{14}AEI$ and $y_{14}AEI$ (94.68% for the D-experiment and 89.89% for the P- or v-experiment), while the test power is still very high (94.98% and 96.09%, respectively).

Consistency with SY in the p- or v-experiment is not as common as we would have expected. However, we should note that the choices of many participants imply strong Rawlsian (max-min) preferences. Those participants will choose allocations around the 45° line, and very small errors will lead to inconsistency with SY. Given that consistency with SYGARP is a necessary condition for consistency with AGSYGARP, the results for AGSY are unsurprising.

We report the results for AGGARP as well for the sake of completeness. The best 14 result for AGGARP is rather high, but without consistency with SY we find this difficult to interpret.

ALL OBSERVATIONS				
	FULL CONSISTENCY		EFFICIENCY > 0.999	
	EXPERIMENTAL	RANDOM	EXPERIMENTAL	RANDOM
xGARP	132 (70.21%)	0.58%	149 (79.26%)	0.61%
yGARP	120 (63.83%)	0.34%	134 (71.28%)	0.47%
ySYGARP	67 (35.64%)	0.00%	74 (39.36%)	0.01%
AGGARP	91 (48.40%)	0.00%	111 (59.04%)	0.00%
AGSYGARP	52 (27.66%)	0.00%	66 (35.11%)	0.00%

BEST 14 OBSERVATIONS				
	FULL CONSISTENCY		EFFICIENCY > 0.999	
	EXPERIMENTAL	RANDOM	EXPERIMENTAL	RANDOM
x14GARP	134 (71.28%)	4.86%	178 (94.68%)	5.02%
y14GARP	122 (64.89%)	3.76%	169 (89.89%)	3.91%
y14SYGARP	71 (37.77%)	0.34%	109 (57.98%)	0.41%
14AGGARP	94 (50.00%)	0.00%	159 (84.57%)	0.02%
14AGSYGARP	56 (29.79%)	0.00%	103 (54.79%)	0.01%

Table 5: Number of participants consistent with the axioms.

As with other induced budget experiments which test for revealed preference axioms and collect many choices (such as Choi et al. 2007a and Fisman et al. 2007), we find that it is necessary to analyse the distribution of the efficiency index in more detail in Section 4.1.2.

4.1.2 Efficiency Indices and Power

We find that even for γ SYAEI the vast majority of participants have higher efficiency values than almost all random choice sets. We find that 90% of all random choice sets have an xAEI of less than .905 and a γ AEI of less than 0.903; 95% of all random choice sets have an xAEI of less than 0.934 and a γ AEI of less than 0.934. On the other hand, 94.15% of participants have an xAEI of more than .903 and 88.30% of participants have a γ AEI of more than 0.903; for the 95% threshold, these numbers are 92.02% and 86.70% of participants, respectively. For γ SYAEI, we find that 90% of all random choice sets have an efficiency index of less than .789, and 95% of less than .845, whereas 88.83% and 85.64% of all participants exceed these thresholds, respectively.

For the best 14 choices, the test power is still very high. For several participants, one major mistake was responsible for a low efficiency score. For example, we now find that 88.30% of participants have a γ SYAEI of more than .9, while this is only the case for 8.17% of the random choice sets. The histograms for our participants' efficiency indices and the random choices can be found in the appendix in Section A.1.1.

Figure 2 shows the results for AGAEI and SYAGAEI as well as 14AGAEI and 14SYAGAEI for the best 14 choices. We find that 90% (95%) of random choice sets have an AGAEI of less than .796 (.828). Using .796 (.828) as a critical value, we find that 93.62% (91.49%) of participants exceed this threshold. While 85.11% (79.79%) of participants have an AGAEI of more than .9 (.95), only .03% (.58%) of random choice sets exceed those thresholds.

We find that 90% (95%) of random choice sets have an AGSYAEI of less than .682 (.734), while 91.49% (89.36%) of participants exceed those thresholds. While 74.47% (66.49%) of participants have an AGSYAEI of more than .9 (.95), .00% (.08%) of random choice sets exceed those thresholds.

For the best 14, the test power remains very high while participants' efficiency increases substantially. For example, 86.17% (79.26%) of participants now have a 14AGSYAEI of more than .9 (.95).

Finally, Figure 3 shows the number and fraction of participants who satisfy various efficiency thresholds between 1 and .9, for all choices, the best 14, and the best 13. The lowest bar represents xAEI, the second one yAEI, the third one ySYAEI, and the upper bar represents AGSYAEI. The intersection of any two bars shows the number of participants who satisfy both requirements simultaneously. For example, the rectangle at the upper left corner shows that 177 (94.15%) of participants have an xAEI greater than .9. For yAEI, there are 166 (88.30%) participants who satisfy that requirement. It also shows that six participants satisfy the requirement for yAEI but not for xAEI, and that 17 participants satisfy the requirement for xAEI but not for yAEI, while five participants satisfy neither.

The results presented in the figures provide information for readers to decide for themselves which level of efficiency is close enough to the proposed axioms. As a comparison, Fisman et al. (2007) and Choi et al. (2007a) use all participants with an AEI of at least .8 for their parametric estimation and argue that those participants were close enough to utility maximisation. However, for both of these papers, the authors collected 50 choices from each participant, leading to more possibilities for violations. In the choice under uncertainty experiment of Choi et al. (2007a), 9.7% of participants are below the .8 threshold, while in the dictator experiment of Fisman et al. (2007), 26.3% of participants were below the threshold.

Based on the latter reference, we find that our result, with 74.5% of our participants above the AGSYAEI-threshold of .9 is good evidence of the theory proposed in BHH. Given the strong effect a single mistake can have on the efficiency score, we would also like to highlight the results for the best 14 choices. We believe that an efficiency threshold of .95 is, in both economic and statistical terms (compared to the Monte Carlo results), a strong requirement, and we find that 79.26% of participants satisfy it based on the best 14 choices. For a less demanding threshold of .9, we have 86.17% of participants who satisfy it.

Our participants also generally come close to homothetic utility maximisation. In the D-experiment, 59.57% (79.26%) of participants have an HEI of greater than .95 (.9). For the best 14 choices, 81.91% (92.55%) of participants exceed this threshold. In the P- or V-experiment, 64.89% (78.72%) of participants have an HEI of greater than .95 (.9). For the best 14 choices, 81.38% (92.02%) of participants exceed this threshold. The results are similar to those for data from Andreoni and Vesterlund (2001) and Fisman et al. (2007) as reported in Heufer (2013b).

4.1.3 Conclusions from the Efficiency Analysis

We have provided a detailed analysis of the efficiency levels in the data, and readers are invited to judge for themselves which thresholds are meaningful. Based on the finding that 79.26% of participants have a 14AGSYAEI of more than .95, and 86.17% of more than .9, we believe that we have found strong evidence for the axioms introduced in BHH.

For further analysis in Sections 4.2 and 4.4, we need to decide on efficiency thresholds to distinguish between results for all participants and only those which come reasonably close to optimisation while maintaining a high test power. For the results in these two sections, we decide to base the thresholds on the best 14 choices, allowing participants to make one mistake in both the D-experiment

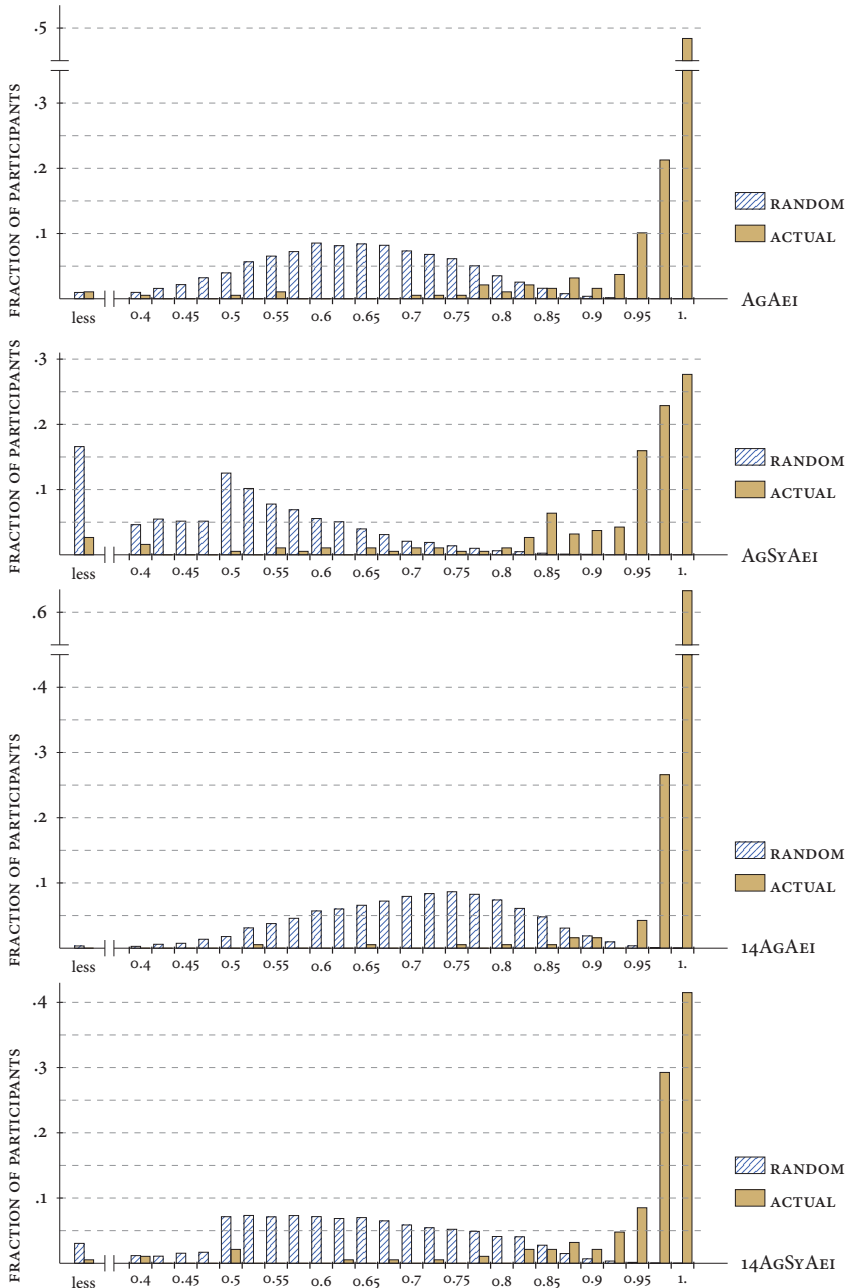


Figure 2: The histograms for AGAEI, AGSYAEI, 14AGAEI, and 14AGSYAEI (experimental data and random choice sets).

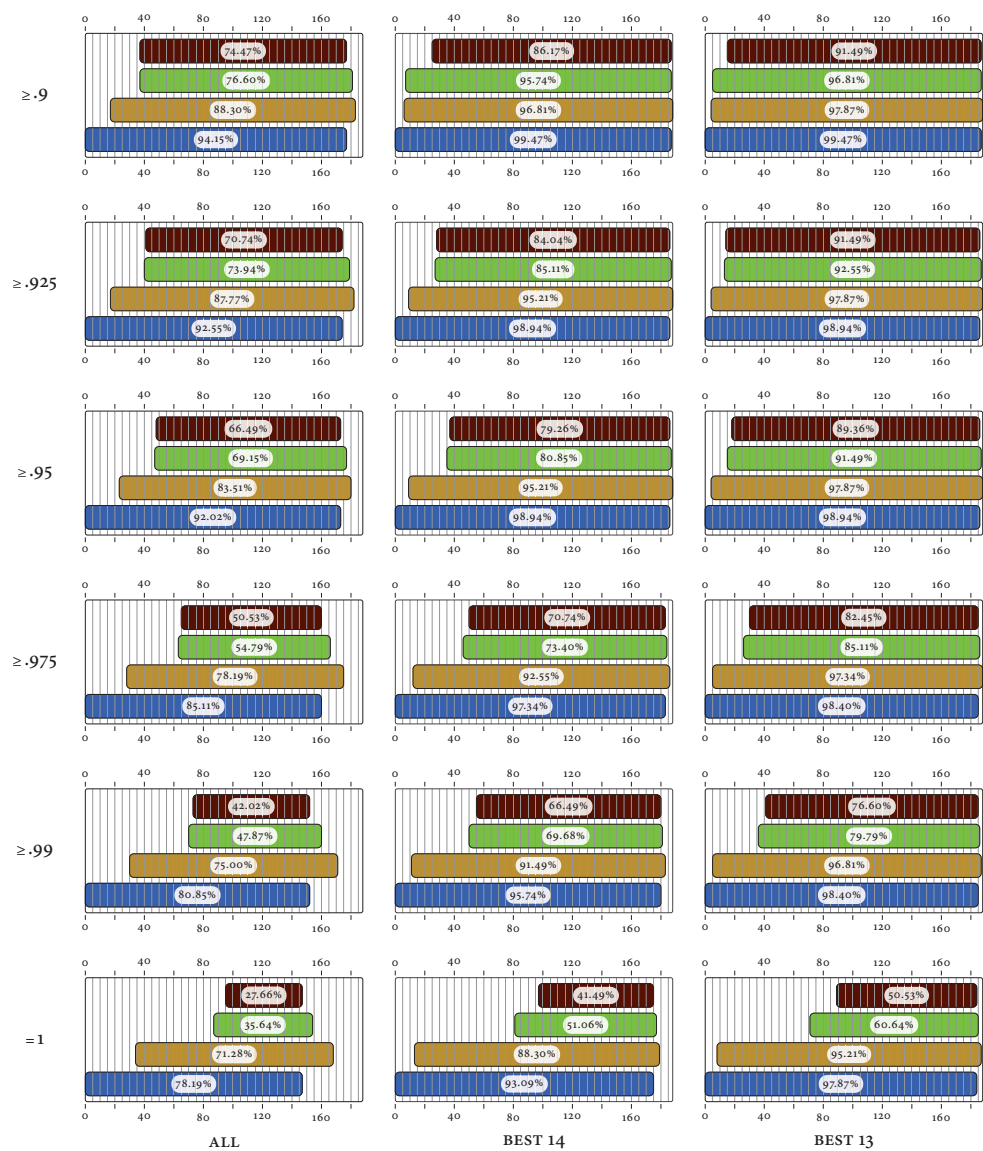


Figure 3: Number and percentage of participants who satisfy certain thresholds and set concordance. The lower bar represents the participants whose $xAEI$ satisfies the threshold on the left, the second bar is for $yAEI$, the third for $ySyAEI$, and the upper bar represents the participants whose $AgSyAEI$ satisfies the threshold. The intersections of any two bars show how many participants satisfy the threshold for both efficiency indices.

and the p- or v-experiment. In particular, we require an $x_{14}AEI$ and a $y_{14}SYAEI$ greater than .9; below this threshold, the deviation from optimising behaviour should be considered economically significant. We furthermore require that both $x_{14}AEI$ and $y_{14}SYAEI$ exceed that of at least 90% of the random choice sets to maintain a high test power. This leads to a threshold for $x_{14}AEI$ of .975419, which is satisfied by 183 (97.34%) of the participants, and a threshold for $y_{14}SYAEI$ of .9, which is satisfied by 166 (88.30%) of the participants.⁵ There are 164 (87.23%) participants who satisfy both requirements simultaneously.

4.2 Distribution of Prototypical Preferences and Notions of Justice

We now want to try to classify participants' preferences, similar to Andreoni and Miller (2002) and Andreoni and Vesterlund (2001). A detailed parametric analysis is beyond the scope of this paper, but we will provide a rough classification based on Euclidean distance to choices by prototypical utility and justice functions.

We consider four types of utility functions for the D-experiment: $U_R(x_1, x_2) = \min\{x_1, x_2\}$ (Rawlsian or max-min), $U_{PS}(x_1, x_2) = x_1 + x_2$ (perfect substitutes), $U_S(x_1, x_2) = x_1$ (selfish), and $U_N(x_1, x_2) = x_1x_2$ (Nash). We call the latter function Nash because it is an affine transformation of a utility function which maximises the product of utilities of the dictator and the beneficiary if both have the same concave Bernoulli utility function for money. Maximising U_N is equivalent to implement the Nash bargaining solution if both players have the same utility function and status quo (Nash 1950). For the p- or v-experiment, we only consider Rawlsian, perfect substitutes, and Nash.

Note that the four utility functions are all special cases of the CES utility function. In particular, for $u(a) = (\alpha a_1^r + [1 - \alpha] a_2^r)^{1/r}$, we obtain U_S for $\alpha = 1$. When $\alpha = 1/2$, we obtain U_R as $r \rightarrow -\infty$, U_{PS} for $r = 1$, and U_N as $r \rightarrow 0$.

Andreoni and Miller (2002) and Andreoni and Vesterlund (2001) do not consider U_N . For comparison, we therefore also provide classification results without Nash utility. They call a participant's fit with one of the functions "strong" if there is no deviation at all from the choices dictated by these utility functions. As our budgets are designed to maximise test power and extract as much information on participants' preferences as possible, there were some budgets for which it was not an easy task to choose an allocation exactly on the 45° line. We therefore categorise a participant as "strong" if the maximum deviation from any of the prototypical choices is .1 or less. The "weak" classification is based on the minimum of the sum of distances to the prototypical choices. To illustrate the different prototypical preferences, Figure 4 shows some examples from the D-experiment.

Table 6 reports the number and percentage of participants for the different categories. The table shows the result for all participants, and the weak classification for all participants who satisfy the efficiency requirement given in Section 4.1.3. In particular, for the D-experiment, the table reports the results for a restriction by $x_{14}AEI$; it also reports the results for the restriction that participants satisfy both the requirement for $x_{14}AEI$ and $y_{14}SYAEI$. This is the same for the p- and v-experiment, except that the first restriction is for $y_{14}SYAEI$.

We note that we have a relatively large percentage of participants who are classified as selfish compared to Andreoni and Miller (2002) who report 47.2% of participants as selfish. However, we have a lower percentage of participants who are classified as strongly selfish (22.7% vs. 16.5%). Part of the difference may be due to differences in budgets and the number of choices. Furthermore,

⁵Of the random choice sets, 90% have a $y_{14}SYAEI$ of less than .892376, so the threshold for economic significance is relevant. However, this does not change the total number of participants who satisfy the threshold.

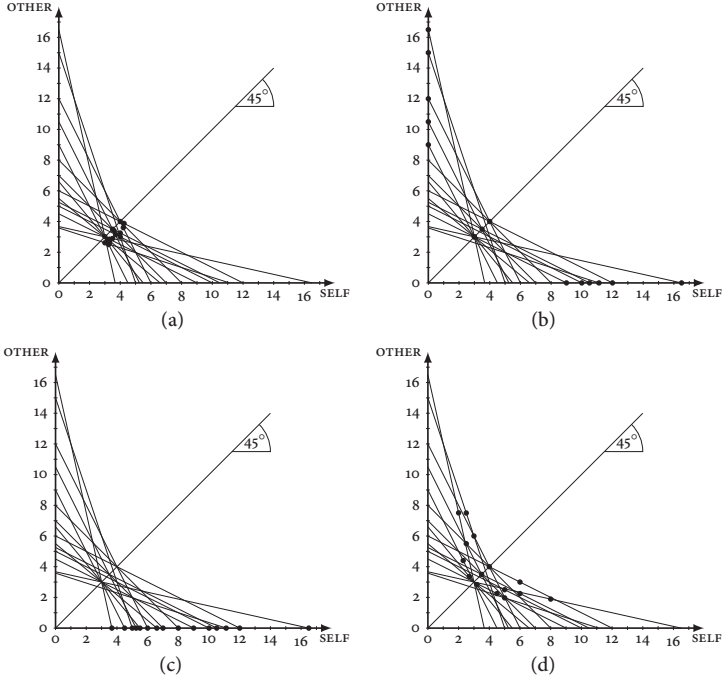


Figure 4: Choices of participants in the D-experiment: Classified as (a) weak Rawls (participant 60), (b) weak perfect substitutes (participant 137), (c) strong selfish (e.g., participant 3), (d) weak Nash (participant 139).

we conducted our experiment in Eastern Germany, and Ockenfels and Weimann (1999) and more recently Brosig-Koch et al. (2011) have shown that Eastern Germans behave somewhat more selfishly and show less solidarity than Western Germans in experiments.

In the P- and V-experiment, we find that without Nash notions of justice, slightly more than 1/3 of our participants exhibit Rawlsian preferences. As Nash preferences are a middle ground between perfect substitutes and Rawls, it is perhaps not surprising that we find that if Nash preferences are included, about half of our participants exhibit these preferences.

4.3 Log-Price Ratio vs. Allocated Share and Recovered Preferences

To plot individuals' choices, we can utilise the same tool as in the experimental paper on choice under uncertainty by Choi et al. (2007a): We plot the relationship between the log-price ratio $\log(p_2/p_1)$ or $\log(q_2/q_1)$ and the share allocated to the other participant ($x_2/(x_1 + x_2)$ for the D-experiment) or to person B ($y_2/(y_1 + y_2)$ for the P- and V-experiment). Some predictions for this relationship based on SY and AG are derived and illustrated in the appendix in Section A.3. We also present the results for all participants by treatment in the appendix.

Figure 5 shows the revealed preferred and worse sets next to the revealed more just and less just sets of four participants, based on the methods described in BHH. For example, in 5.(a), the participant exhibits a strong Rawlsian notion of justice, and his preference is consistent with this

D-EXPERIMENT						
UTILITY FUNCTION	FIT		STRONG + WEAK		AEI RESTRICTED	
	STRONG	WEAK	WITH NASH	WITHOUT NASH	ONLY X	BOTH
Rawls	0 (0.00%)	8 (4.26%)	8 (4.26%)	36 (19.15%)	7 (3.83%)	5 (3.05%)
Perfect Substitutes	0 (0.00%)	15 (7.98%)	15 (7.98%)	20 (10.64%)	15 (8.20%)	14 (8.54%)
Selfish	31 (16.49%)	97 (51.60%)	128 (68.09%)	132 (70.21%)	126 (68.85%)	113 (68.90%)
Nash	0 (0.00%)	37 (19.68%)	37 (19.68%)		35 (19.13%)	32 (19.51%)

P- AND V-EXPERIMENT						
JUSTICE FUNCTION	FIT		STRONG + WEAK		AEI RESTRICTED	
	STRONG	WEAK	WITH NASH	WITHOUT NASH	ONLY Y	BOTH
Rawls	7 (3.72%)	51 (27.13%)	58 (30.85%)	69 (36.70%)	49 (29.52%)	48 (29.27%)
Perfect Substitutes	9 (4.79%)	28 (14.89%)	37 (19.68%)	119 (63.30%)	32 (19.28%)	32 (19.51%)
Nash	2 (1.06%)	91 (48.40%)	93 (49.47%)		85 (51.20%)	84 (51.22%)

Table 6: Participant classification by utility function for the D- and the P- and v-experiment.

high concern for equality. While the preference is biased in the participants' own favour, the choices are largely unaffected by the price ratio.

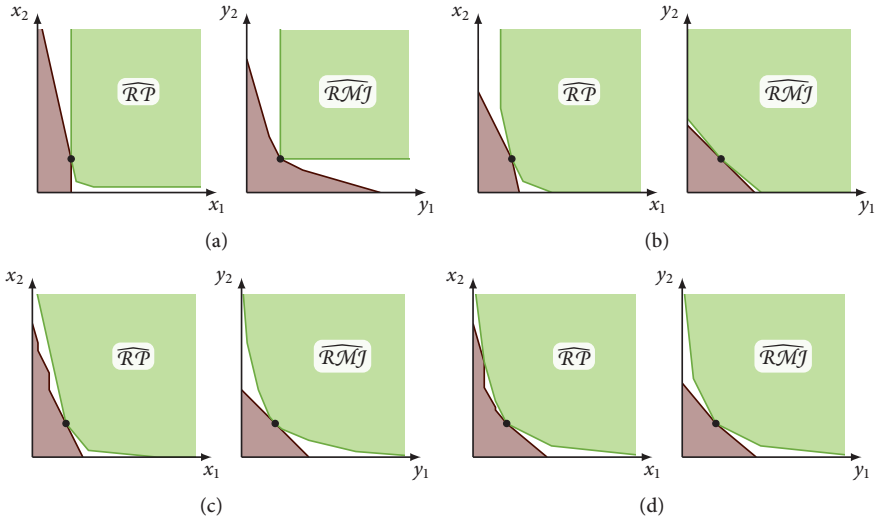


Figure 5: Four examples of revealed preferred, worse, more just, and less just sets of four different participants. (a) Participant 171. (b) Participant 157. (c) Participant 183. (d) Participant 1.

4.4 Non-Parametric Interpersonal Comparisons

As described in BHH, we can compare any two participants with each other in a non-parametric way to test if one has a stronger sense of justice than the other. This approach is similar to the theoretical

ideal proposed in Karni and Safra (2002b).⁶ To recall the concept introduced in BHH, let \succeq be the *stronger sense of justice than* relation. For two preference-notion of justice pairs $(\succsim^1, \succsim_j^1)$ and $(\succsim^2, \succsim_j^2)$ which satisfy AG, we define

$$(\succsim^1, \succsim_j^1) \succeq (\succsim^2, \succsim_j^2) \text{ if } (\succsim^1 \cap \succsim_j^1) \subseteq (\succsim^2 \cap \succsim_j^2).$$

That is, a participant with the preference-justice pair $(\succsim^1, \succsim_j^1)$ has a stronger sense of justice than a participant with the preference-justice pair $(\succsim^2, \succsim_j^2)$ if, for every allocation $a \in \mathbb{A}$, the set of allocations which the first participant prefers over a even though he feels that they are less just than a is a subset of the same construction for the other participant. In other words, the set of allocations for which the first participant violates his own notion of justice by preferring them is a subset of the set of allocations for which the second participant violates his own notion of justice. Even more succinctly, the “selfish set” of the first is a subset of the “selfish set” of the second. See BHH for more details, in particular about the operational aspect when dealing with experimental data.

As we have 188 participants with a total of 30 choices each, the non-parametric approach based on \hat{R} and \hat{R}_j is computationally infeasible. Instead, we will rely on the alternative simpler approach also suggested in BHH, and present results for the *partially revealed stronger sense of justice* relation \succeq_{Rev} . Note that two participants can both have a partially stronger sense than the other. We use choices adjusted by the relevant improved violation index (IVI) vectors. See Appendix A.1.2 for a definition and further details.

Comparisons between two participants fall into one of four categories: *Stronger (weaker)*, meaning that the first participant has a partially stronger sense of justice than the second participant but not vice versa (that the second participant has a partially stronger sense of justice than the first participant but not vice versa); *neither*, meaning that neither the first participant has a partially stronger sense of justice than the second participant nor vice versa; *both*, meaning that the first participant has a partially stronger sense of justice than the second participant, and the second participant also has a partially stronger sense of justice than the first participant.

The first two categories are unambiguous. “Neither” is ambiguous. It indicates that, whenever the participants agree about the justice comparison of two allocations, they also have the same preferences. However, two participants with very different notions of justice will rarely agree about which allocation is more just than the other and will generally fall into this category. “Both” is also ambiguous. It indicates that while the two participants do at least sometimes agree about the justice comparison, in at least two of these cases they disagree about which one is preferable. In any case, the categories “neither” and “both” indicate that it is not possible to establish that one of the two participants has a stronger sense of justice than the other.

Table 7 presents the results for the comparisons of all participants with each other. For example, in 61.41% of all of the possible pairs of two participants, we find that one participant clearly has a stronger sense of justice. However, this also means that for almost 40% of pairs of participants, we cannot establish one as having a stronger sense, and this is mostly due to the fact that they fall into the “both” category. This demonstrates that it is often impossible to compare the strength of the sense of justice of two individuals, as they often differ either in their notion of justice, or they are more just in some situations and less just in others, or both. The table also shows the same result restricted to the 164 participants which satisfy the best 14 efficiency requirements given in Section 4.1.3.

⁶See also Karni and Safra (2002a).

COMPARABILITY OF PARTICIPANTS			
	STRONGER/WEAKER	NEITHER	BOTH
All participants	61.41%	3.35%	35.24%
Restricted set	60.55%	3.27%	36.18%

Table 7: Comparability of participants, for the set of all participants and restricted for the 164 participants who satisfy efficiency requirements for $x_{14}AEI$ and $y_{14}SYAEI$.

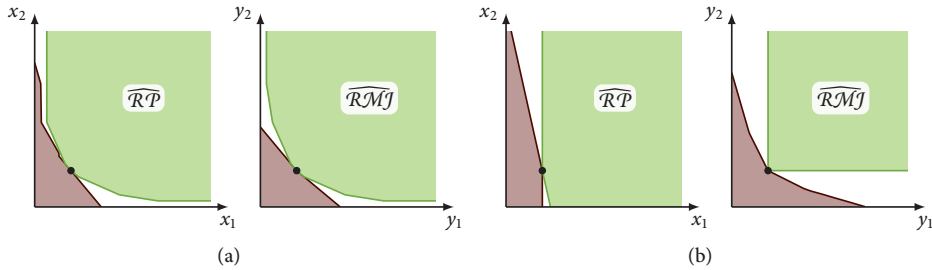


Figure 6: Revealed preferred, worse, more just, and less just sets. (a) A participant with a stronger sense of justice than most other (participant 93). (b) A participant with a weaker sense of justice than most other (participant 3).

Figure 6 shows the revealed preferred, worse, more just, and less just sets of two participants. The first one is the participant who most often falls into the stronger category, while the second one is one of the participants who most often fall into the weaker category.

We can also compare each participant to choices generated by maximising arbitrary utility and justice functions. We do so for a set of CES functions with different parameters. The results are shown in the appendix in Section A.2.

4.5 Treatment Effects

A detailed analysis of these results using socio-economic explanatory variables will be the subject of further studies. In this paper, we would like to focus on treatment-specific differences. For a better interpretation of the preceding analysis, an evaluation of the methodology, and potential improvements in the design for further studies, we will compare the effect of the different treatments (see Table 1) on various outcomes. There are two main questions about treatment effects: First, does the ordering of the tasks matter, that is, are there differences depending on whether the p-experiment was conducted first or second? Second, are there differences between the p-experiment and the v-experiment?

In particular, we will first look at the raw data and compare choices per budget between different treatments. We then analyse potential treatment effects on efficiency (see Section 4.1). Finally, we compare the results of the interpersonal comparison of the strength of justice.

4.5.1 Raw Data

We first compare the choices per budget in the different treatments. We will use the Kruskal-Wallis (KW) test (a non-parametric one-way analysis of variance; Kruskal and Wallis 1952) and the 2-sample

Anderson-Darling (AD) test (a non-parametric test of whether two samples of data are drawn coming from a single population; Anderson and Darling 1952, Scholz and Stephens 1987). We focus on the KW test as the appropriate non-parametric alternative to standard ANOVA to analyse the treatment effects. This will indicate whether or not there are significant differences in the median between two (or four) treatments.⁷ The AD test is concerned with differences in the distribution. We include the test mainly as a robustness check as we are more concerned about significant differences in the median. We only report results for the restricted set (i.e., for those subjects who satisfy the efficiency requirement; see Section 4.1.3).

Table 11 in the appendix (Section A.5) shows the results for differences in the amount allocated to the other participant in the D-experiment (x_2). For ordering effects, we have to consider comparisons between treatments 1 and 3 (P-experiment with D-experiment as first and second task, respectively), treatments 2 and 4 (V-experiment with D-experiment as first and second task, respectively), and between the pooled data of T1 and T2 vs. T3 and T4 (see also Table 1 for an overview of treatments). First, we find that there are no significant differences between T2 and T4, but there are some between T1 and T3. There is still a significant difference between the pooled data of T1 and T2 vs. T3 and T4. There are no significant differences for the KW test with all four treatments, however we cannot completely ignore the other result as the differences occur mainly for budgets 8 and 9, which are the budgets with relative prices of 1 and 0.833.

We have to be careful in our interpretation of this result because it is likely that out of 15 budgets there will be some with significant differences merely by chance. We therefore use a custom randomisation test (a random permutation test; see for example Westfall and Young 1993, Good 2005): When we compare two groups of participants, G_1 and G_2 , with n_1 and n_2 participants, respectively, we first pool the data to get $G = G_1 \cup G_2$. We then randomly sample without replacement n_1 of the participants in G and assign them to a group F_1 , and assign the remainder of G to a group F_2 . This eliminates treatment effects between F_1 and F_2 . We compare the choices made by participants in F_1 with those in F_2 for each of the budgets and compute the p-values for the KW test. We repeat this 10,000 times. We then compute the percentage of random permutations for which the number of budgets with significant differences is greater than or equal to the number of budgets with significant differences for the original groups G_1 and G_2 . This is an estimate of the probability with which we can expect to observe the differences between the choices of two groups G_1 and G_2 if there actually are no differences.

For example, G_1 and G_2 could be participants from T1 and T3. Then we find that in 17.91% of all random permutations, there is at least one budget with a significant difference at the 5% level and two additional budgets with differences at the 10% level. We therefore conclude that the differences between T1 and T3 are still within the limits of what one can expect to occur by chance. The differences in the comparison of T1 and T2 vs. T3 and T4 are less likely, as they only occur with a probability of 7.93%. The last line in table 11 shows all the probabilities with which differences at the same or stricter significant levels occur by chance.

It appears that if the ordering has any effect on the median amount of x_2 , this is more likely for the P-experiment than the V-experiment. The amount given, x_2 , is indeed higher in T1 than in T3 for all budgets. Participants knew that a second task would follow the first one, but were not informed about the nature of the second task. Thus, a possible interpretation is that participants who played the D-experiment first expected another similar task in which they could allocate money to themselves,

⁷Tables 9 and 10 in Appendix A.4 show, among other information, the median values of x_2 and y_2 for each treatment separately.

leading to more generous behaviour in the first task. Such an effect has been observed previously by Cox (2009), who finds that when participants are informed that there will be another unspecified task after the dictator task, participants show greater generosity in the dictator task. Participants who played the *p*-experiment first knew that there would be no more tasks after the *D*-experiment.⁸ This interpretation is also in line with the fact that the difference only occurs for treatments with the *p*-experiment. While participants did not know their label, the *v*-experiment might still be more likely to create the impression that it is the individual participant himself who decides about his own payoff.

Table 12 in the appendix (Section A.5) shows the results for differences in the amount allocated to person B in the *p*- or *v*-experiment (y_2). Again, the last line shows the probability that the differences could have occurred at chance based on the randomisation approach. There are some minor weak ordering effects which can be arguably ignored as they are very likely to occur by chance. More importantly, there are some significant differences between the *p*- and *v*-experiment. Interestingly, this does not show up for T1 and T2 (where the *D*-experiment was conducted first), but only for T3 and T4. In T4, participants choose higher y_2 when $q_1/q_2 > 1$ and lower y_2 when $q_1/q_2 < 1$ in comparison with participants in T3. The choices are therefore more concerned with maximising expected payoff in the *v*-experiment than in the *p*-experiment. As this does not occur when the *D*-experiment is conducted first, it indicates that participants who have already allocated money to themselves in the first task are less concerned about expected payoff. It is unlikely that this is a portfolio effect as that should lead to lower risk aversion in the second task. Therefore, a possible interpretation which is also in line with the ordering effects in the *D*-experiment is that participants are concerned with securing some payoff for themselves before considering others. If the *v*-experiment is considered, to a large extent, as choice under uncertainty without regard for the other participant, the *v*-experiment would already serve as a task that secures a certain payoff.⁹ If that is the case, then the *v*-experiment would appear to be less well suited to assess the notion of justice.

4.5.2 Efficiency

It is also helpful to compare the efficiency between treatments to see if, for example, participants had more problems with the *p*- or the *v*-experiment, or to find out if there are reasons to believe one of the treatments is less incentive compatible than the other. This would also confound the findings about treatment effects.

We analysed the efficiency levels for *p*- vs. *v*-experiments (T1 vs. T2 and T3 vs. T4), and between treatments in which the *D*-experiment preceded the *p*- or *v*-experiment and treatments in which it came second (T1 vs. T3 and T2 vs. T4). We compared the x_{AEI} , x_{14AEI} , y_{SYAEI} , $y_{14SYAEI}$, AG_{SYAEI} , and $14AG_{SYAEI}$. The KW test finds that for all comparisons the medians are not significantly different (all *p*-values are greater than 0.1534). We omit a detailed table as all results show that there are no differences.

⁸This interpretation is also supported by some of the answers given in a post-experimental questionnaire. For example, one participant in treatment 3 explained the choices in the *D*-experiment: "I don't know how the others decide and the risk would be too great to choose in accordance to the first part of the experiment."

⁹Anecdotal evidence from a post-experimental questionnaire indicates that there are some differences in the motivation by our participants in the *p*- and in the *v*-experiment: While many participants explain their choices using terms such as "justice" or "fairness" in the *p*-experiment, participants in the *v*-experiment also mention terms such as "uncertainty", "risk", or "luck", albeit "fairness" and "justice" are also mentioned.

4.5.3 Interpersonal Comparison

Finally, we analyse if there are differences in the strength of justice between the treatments. The results in Section 4.5.1 suggest that participants who first played the D -experiment and then the P -experiment (T_1) have a stronger sense of justice than participants who first played the P -experiment. As the variable of interest, we use the number of times a participant has a stronger sense of justice than another participant (see Section 4.4). We only report the result for the restricted set of participants who satisfy the efficiency threshold (see Section 4.1.3).

We conducted one-way and two-way ANOVA but found that the residuals are far from being normally distributed. We therefore turn to the non-parametric version of the one-way ANOVA, the KW test. Table 13 in the appendix (Section A.5) shows the results. We also include the 2-sample AD test as a robustness check. Evidently, there is a weakly significant difference between the pooled data from T_1 and T_2 vs. the pooled data from T_3 and T_4 . Thus, we find a weakly significant ordering effect: If the D -experiment is conducted first, participants are more likely to exhibit a stronger sense of justice than others.

5 CONCLUSION AND DISCUSSION

In BHH, we have proposed that individuals possess a notion of distributive justice by which they trade off efficiency and equality and that it is useful to examine such notions. With an appropriately designed experiment researchers would be able to assess a part of participants' notions of justice: the *revealed* notion of justice, which can then be analysed as any other observed revealed preference relation. We have suggested two experiments to collect the necessary data. The first experiment is a social planner experiment (P) in which a participant is asked to allocate money between two other participants. The second one is a veil of ignorance experiment (V) in which a participant is asked to allocate money between the labels "person A" and "person B", and the participant is one of the two persons but does not know which one. We have suggested to combine such an experiment with a standard dictator experiment (D) in which participants reveal their preference over allocations if they have an own stake. In all these experiments, we suggested to use several budgets with different price ratios (as in Andreoni and Miller 2002) in order to gather information about how participants trade off efficiency (sum of monetary payoffs) and equality.

We have postulated that the notion of justice should satisfy *Symmetry* (Sy): Given anonymity, no individual should per se be more deserving than others and therefore any permutation of an allocation should be considered as just as any other. Furthermore, the revealed preference relation from the D -experiment and the revealed notion of justice from the P - or V -experiment together should satisfy *Agreement* (AG): If an allocation x yields a higher payoff for the participant than an allocation y and the decision maker prefers y over x although the payoff is lower for him, then this must be because he feels that y is more just than x . Equivalently, if a participant considers an allocation x to be more just than an allocation y and x provides the participant with more monetary payoff than y , the participant prefers x over y . Therefore, a participant might make choices in the D -experiment which he himself considers to be unjust, but if he does so, it is to his own advantage.

The experimental analysis shows that participants in P - and V -experiments do indeed make choices consistent with a well-behaved notion of justice. Notions of justice are very diverse, ranging from perfect substitutes (about 20%) to the intermediate Nash welfare function (about 50%) to very egalitarian Rawlsian ideals (about 30%). We have found evidence that about 80% of participants are

consistent or almost consistent with the postulated Agreement axiom AG, which implies that their choices in dictator games can be explained by a combination of selfishness and concern for justice.

There are some differences between the P- and V-experiment, and the ordering of tasks could possibly affect choices as well, but differences are at most weakly significant. Participants in the V-experiment tend to be more efficiency than equality concerned, and giving in the D-experiment is greater when the D-experiment is conducted first. These results are far from conclusive, and we remain agnostic as to whether the P- or the V-experiment is better suited to assess the notion of justice.

Using the non-parametric interpersonal comparison proposed in BHH, we found that in about 60% of all cases, participants can be unambiguously compared in terms of their strength of justice. However, in about 40% of all cases, it is not possible to find such a clear ranking between two participants, which demonstrates that different ideas about what is just make it impossible to conclude that one person has a stronger sense of justice than another.

A APPENDIX: ADDITIONAL MATERIAL

A.1 Additional Efficiency Analysis

A.1.1 Histograms

Figure 7 shows the distribution of the $xAEI$, $yAEI$, and the $ySyAEI$, both for the actual experimental data and the random choice sets. Figure 8 shows the same result for the best 14 choices.

A.1.2 The Improved Violation Index

The AEI is a uniform lower bound for the efficiency of a set of observations. A more elaborate approach for a set of N observations would be to find a vector $f \in [0, 1]^N$ that provides information about the efficiency of each choice individually. We then say that $x^i R^0(f^i) a$ if $f^i p^i x^i \geq p^i a$ and similarly for the other relations. Then Ω_x satisfies $GARP(f)$ if $x^i R(f^i) x^j$ implies $[\text{not } x^j P^0(f^j) x^i]$. Varian (1993) provides an algorithm to compute such a vector, the *Improved Violation Index* (IVI), which is the algorithm we adopt for our data. It is a “minimum” vector in the sense that there does not exist an $f' \geq f$ for which $GARP(f')$ is satisfied. Note, however, that a minimum vector is not necessarily unique.

One way to use the IVI is to adjust choices in a minimal way such that they satisfy $GARP$ and use the adjusted data to test for consistency with the other axioms. This way, for each participant, we get sets of choices from a D- and a P- or V-experiment which have been made consistent with $GARP$. As mentioned before, the AEI from the D- and P- or V-experiment is an upper bound for the other efficiency indices. Using IVI-adjusted data instead will, ideally, provide information about the AG- and SY-efficiency if the data had been consistent with $GARP$. Generally, if the IVI-adjustment is only minor, this will complement the analysis. If a major adjustment is required, however, we should be careful in interpreting the adjusted data. We will discuss this point in more detail in Section A.1.3 below.

As a further adjustment, we introduce a very simple concept, the *Own Violation Index* (OVI). It is meant to be a rough adjustment for choices which even in isolation violate the SY axiom, that is, a choice y^i such that $y_j^i > y_k^i$ even though $q_j^i > q_k^i$. This clearly violates SY, as $y^i C_{SY}(P_j^0) y^i$. We define the OVI as a vector g such that $g^i = \min\{1, q_1^i y_2^i + q_2^i y_1^i\}$. For example, if $y_1^i > y_2^i$ and $q_1^i \leq q_2^i$, we have $g^i = 1$. But if instead $q_1^i > q_2^i$, we have $g^i = q_1^i y_2^i + q_2^i y_1^i < 1$. In any case, we will always have $y^i C_{SY}(P_j^0(g^i)) y^i$.

A.1.3 The Procedure

The procedure we will use for testing and screening the data, making use of the IVI and OVI, is summarised in Figure 9. The results of the first step, in which no adjustments are made, are reported in Section 4.1.

For the second step, we adjust the data from the individual experimental tasks (i.e., Ω_x and Ω_y) by the IVI and OVI to correct for mistakes already made in isolation and then use the adjusted data to test AG and SY-efficiency. As explained above, any set of data can be adjusted using the IVI such that $GARP$ is satisfied. However, for some sets, this adjustment would have to be unreasonably high. Therefore, before we apply the IVI, we will screen the data and only use a subset of our participants for the remaining analysis.

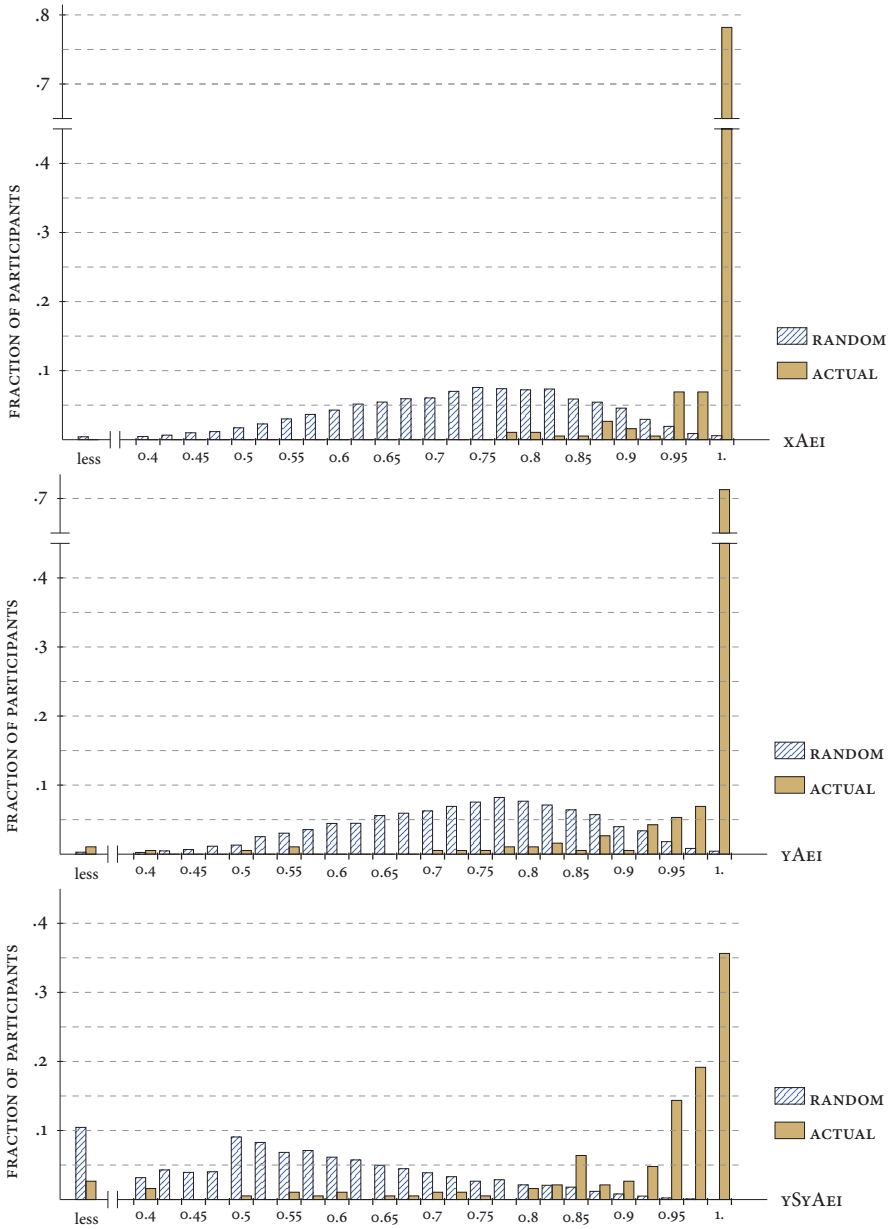


Figure 7: The histogram for the x_{AEI} , y_{AEI} , and y_{SyAEI} , both for the actual experimental data and for the random choice sets.

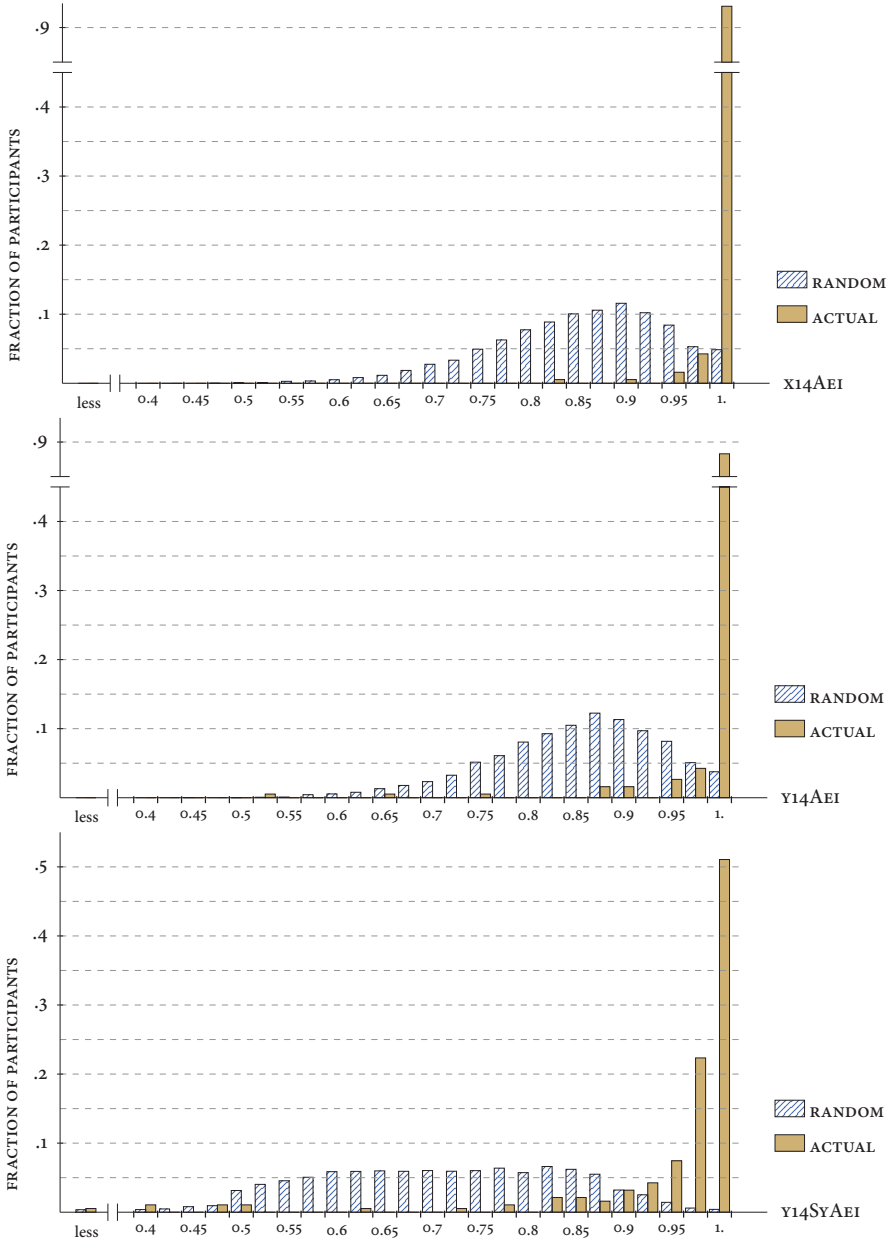


Figure 8: The histogram for the $x_{14}AEI$, $y_{14}AEI$, and $y_{14}SYAEI$, both for the actual experimental data and for the random choice sets.

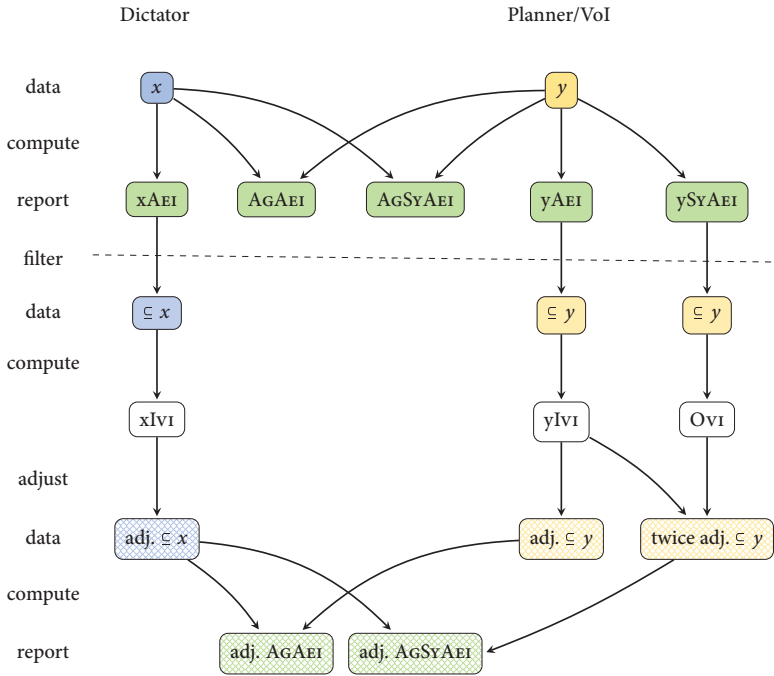


Figure 9: A procedure for testing and screening the data.

In particular, we will only use the data from participants whose AEI is greater than .9 and exceeds the AEI of at least 90% of the random choice sets generated with Bronars' method. Thus, to obtain the adjusted AGAEI, we first limit the analysis to those participants who have an xAEI greater than 90% of the random choice sets and a yAEI greater than 90% of the random choice sets. We then adjust the choices in the D-experiment using the xIVI and the choices in the P- or V-experiment using the yIVI and then compute the adjusted AGAEI based on the adjusted data. For the AGSYAEI, we will additionally require that participants have a ySYAEI greater than .9 and greater than 90% of the random choice sets, and we will adjust the data sets which pass this threshold using both the yIVI and the OVI.

A.1.4 Efficiency Indices and Power: Adjusted

Figure 10 presents the same results as Figure 2, but with adjustments by IVI and OVI as described in the second part of Section 3.2 and Figure 9. As mentioned before, for the actual experimental data and AGAEI, we only used those participants who have an xAEI greater than 90% of the random choice sets and a yAEI greater than 90% of the random choice sets. This leaves us with 160 (85.11%) participants for AGAEI. For AGSYAEI we additionally require participants to have a ySYAEI greater than those of 90% of the random choice sets, which leaves us with 140 (74.47%) participants.

Note that we used all generated random choice sets for the distribution of the efficiency indices in Figure 10. That means that most of these random choices were heavily IVI-adjusted, which leads to a substantial loss of power. If we were to apply the same stringent requirements for screening out random choice sets, we would be left with very few sets. The reason is that once most budgets are adjusted by a factor of, say, .8, there are few possibilities left to violate AG- or AGSYGARP.

Figure 11 shows the same results for the best 14 choices. Here, we use 173 (92.02%) participants for 14AGAEI and 160 (85.11%) of participants for 14AGSYAEI. Some of the changes compared to Figure 10 may seem counter-intuitive; for example, there is now a slightly lower percentage of participants who have a 14AGSYAEI of 1. There are two reasons: First, due to different thresholds, Figure 11 is based on the data of more participants than Figure 10. Second, and this also applies to the random choice data, the adjustments made by IVI and OVI are smaller because we consider only a subset of choices.

Figure 12 shows the same results as Figure 3, but with adjusted choices for y₁₄SYAEI and 14AGSYAEI.

A.1.5 Comparison with Random Optimising Choices

In our case, the AEI in the random choice set for the D-experiment and P- or V-experiment is an upper bound for the other efficiency indices such as the AGSYAEI. Heufer (2013a) provides an efficient way to generate a set of N choices on N budget sets which are drawn from a uniform distribution on the budgets with the additional restriction that the resulting set of data happens to satisfy GARP. This is useful for our purposes as it allows us to generate GARP-consistent random data for the D- and the P- and V-experiments and then go on to compute the fraction of sets which also satisfy the SY- and AG-based axioms we are interested in. This provides a way to prevent the above mentioned problem and to find the conditional power of a test, given that the data already satisfies GARP. For our purposes, we generate 10,000 random choice sets for the the D-experiment as well as the P- and V-experiment.

Figure 13 shows the distribution of the ySYAEI, AGAEI, and AGSYAEI for random choice sets which happen to satisfy xGARP and yGARP, as described in Section 3.1.2. Still, 90% (95%) of these

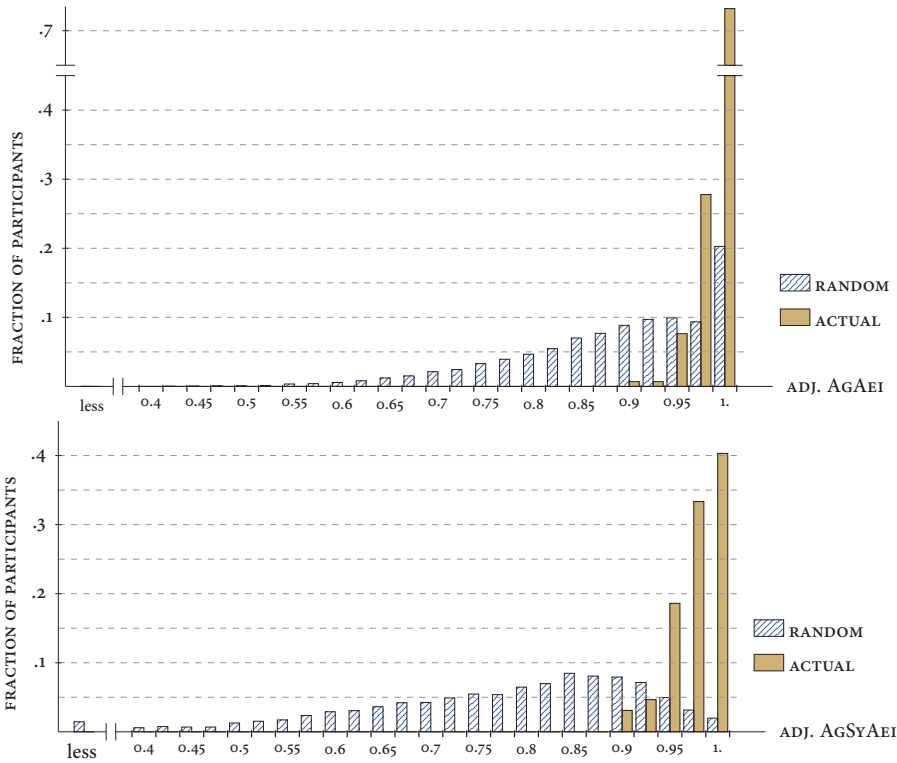


Figure 10: Adjusted AGÆEI and AGSYÆEI. Note that the random data includes heavily adjusted choices from the d- and p- or v-experiment, whereas the actual data does not.

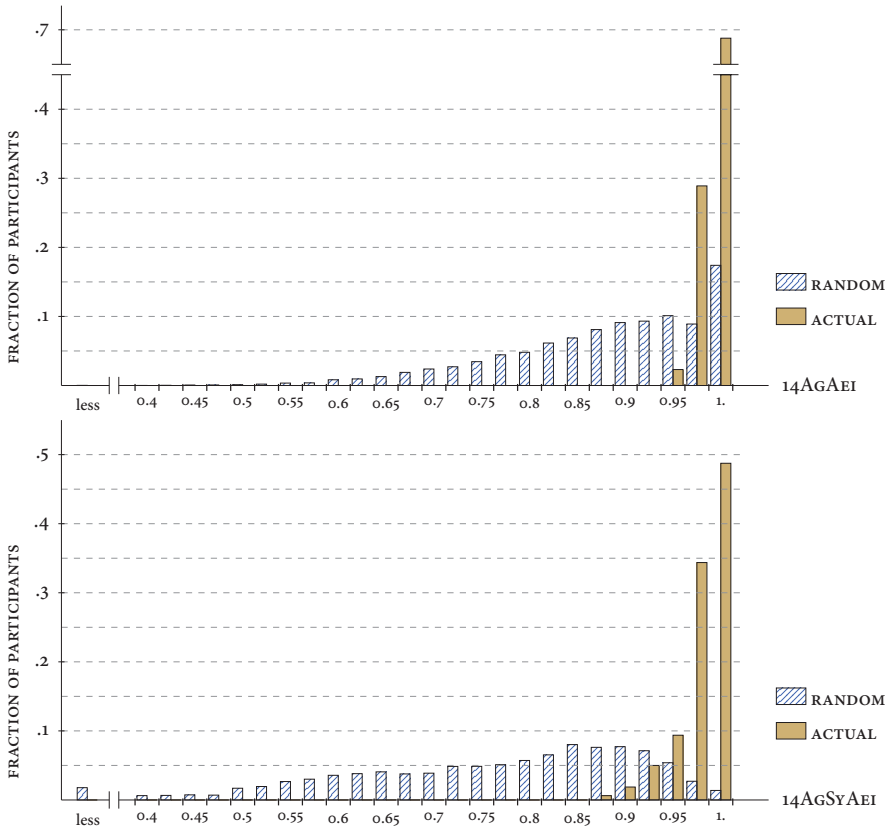


Figure 11: Adjusted 14AGAEI and 14AGSYAEI. Note that the random data includes heavily adjusted choices from the D- and P- or v-experiment, whereas the actual data does not.

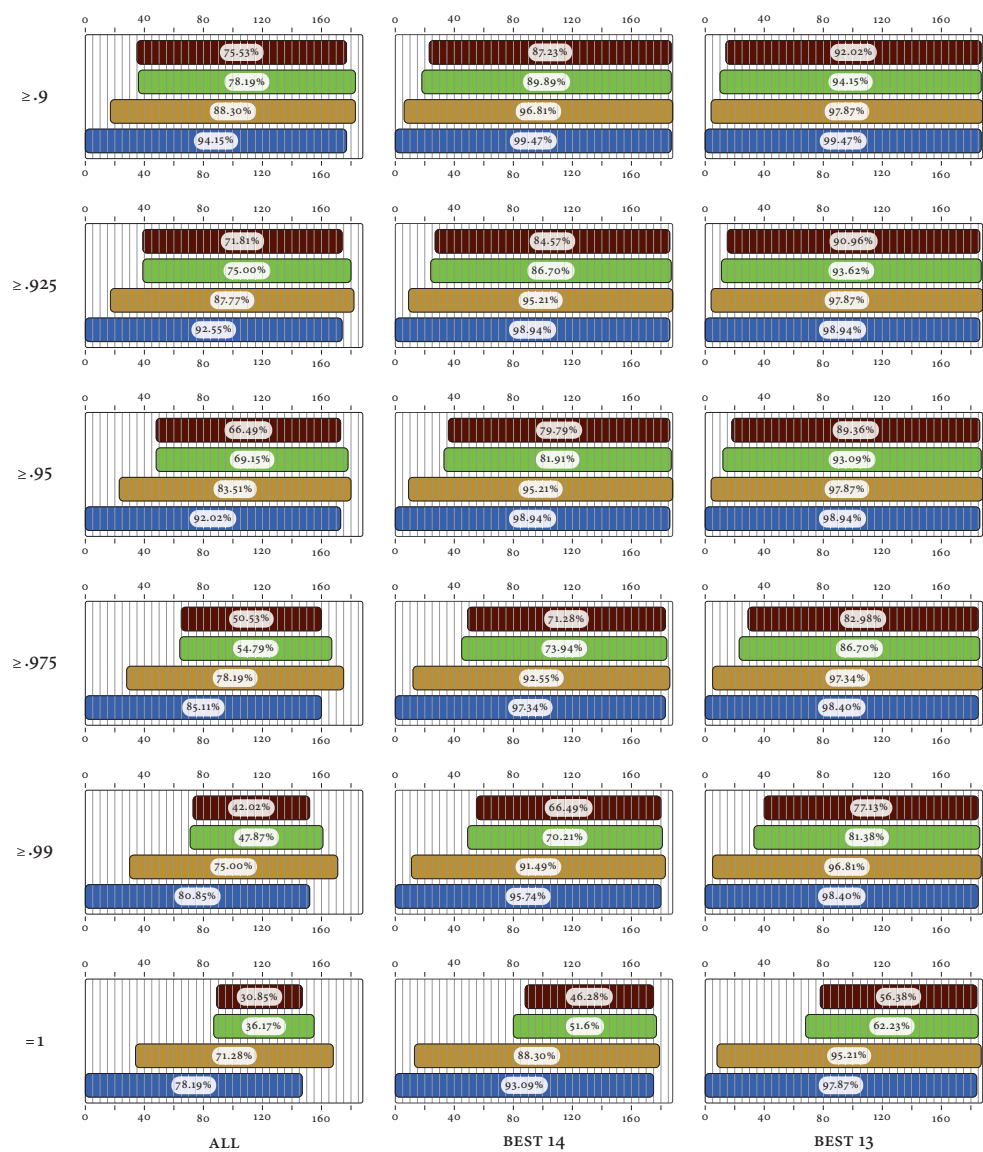


Figure 12: Number and percentage of participants who satisfy certain thresholds and set concordance, with adjustments.

random choice sets have AGSyAEI values of less than 0.885 (.917). For AGSyAEI, the histogram also shows the distribution for data that has been adjusted by the OVI as described in Section A.1.2.

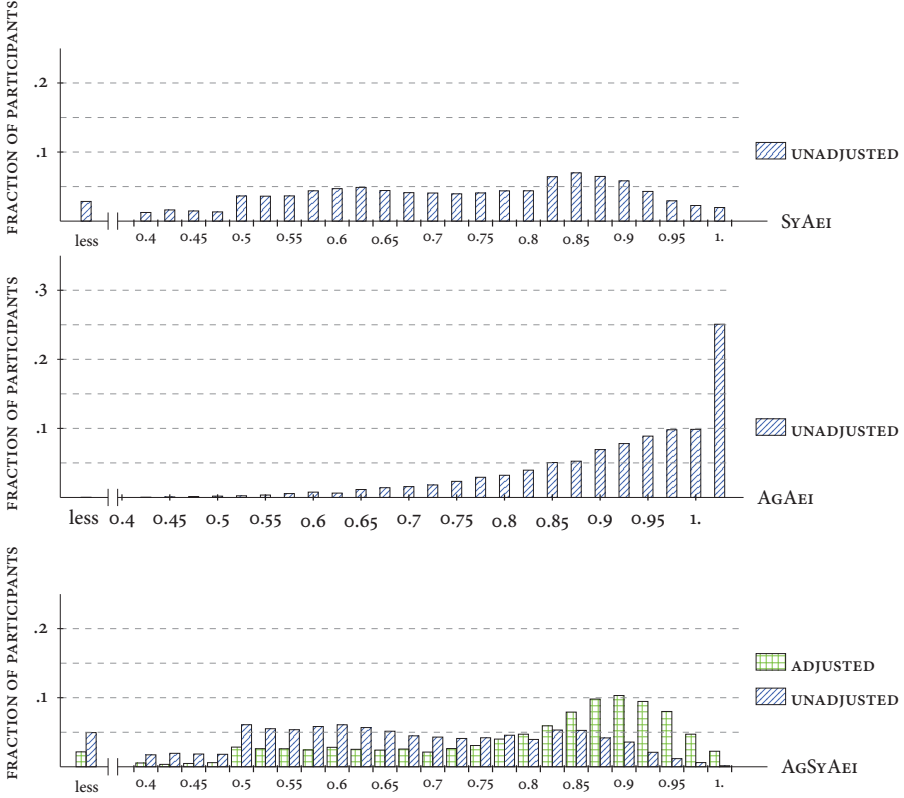


Figure 13: SYAEI, AGAEI, and AGSyAEI for random GARP consistent choice sets, unadjusted. For AGSyAEI, we also show the adjusted result (adjusted for Symmetry only).

A.2 Non-Parametric Interpersonal Comparisons

We compare each participant to choices generated by maximising a set of CES functions with different parameters. In particular, we generate choices on each p^i , $i = 1, \dots, 15$, by maximising $u(a) = (\alpha a_1^r + [1 - \alpha] a_2^r)^{1/r}$, and on each q^j , $j = 1, \dots, 15$, by maximising $v(a) = ([1/2] a_1^s + [1/2] a_2^s)^{1/s}$. We use six different values for α ($\alpha \in \{.5, .6, \dots, 1\}$) and 22 different values for r ($r \in \{-20, -1, -.9, \dots, .9, 1\}$). We then compare these choices with the choices of the participants.

The first part of Table 8 shows the number of participants for which there exists at least one combination of α and r such that, when compared to the choices generated with these parameters, the participant falls into one of the four categories. The second part of the table shows the number of

participants for which there exists at least one r such that, for a particular α , the participants fall into one of the four categories. Not surprisingly, the fraction of participants for which there exists a r for the stronger (weaker) category increases (decreases) in α . What is interesting is that for an intermediate value of α , say .8, only for 43.29% of participants there exist r such that the participant has a stronger sense of justice. That means that for more than half of the participants, we cannot establish that they have a stronger sense of justice than someone with a parameter of $\alpha = .8$ in his CES utility function. However, for 83.54% of participants, we can find an r such that the participant has at least a partially stronger sense of justice.

COMPARABILITY WITH CES: EXISTENCE OF PARAMETERS				
NUMBER OF PARTICIPANTS FOR WHICH (α, r) EXISTS SUCH THAT	STRONGER	NEITHER	BOTH	WEAKER
	All participants	155 (82.45%)	60 (31.91%)	182 (96.81%)
Restricted set	135 (82.32%)	54 (32.93%)	158 (96.34%)	163 (99.39%)
NUMBER OF PARTICIPANTS (RESTRICTED SET) FOR WHICH r EXISTS SUCH THAT				
	STRONGER	NEITHER	BOTH	WEAKER
$\alpha = 1.$	135 (82.32%)	32 (19.51%)	0 (0.00%)	0 (0.00%)
$\alpha = .9$	133 (81.10%)	31 (18.90%)	153 (93.29%)	63 (38.41%)
$\alpha = .8$	71 (43.29%)	5 (3.05%)	137 (83.54%)	94 (57.32%)
$\alpha = .7$	58 (35.37%)	1 (0.61%)	149 (90.85%)	94 (57.32%)
$\alpha = .6$	24 (14.63%)	1 (0.61%)	124 (75.61%)	122 (74.39%)
$\alpha = .5$	0 (0.00%)	17 (10.37%)	0 (0.00%)	163 (99.39%)

Table 8: Comparability of participants with choices generated by a CES utility function and a CES justice function with the same r . First part: The number of participants for which at least one combination of (α, r) exists such that the comparison falls into one of the four categories. Second part: The number of participants who satisfy the efficiency requirement (best 14 xAEI and ySYAEI greater than .9) such that for a certain α , there exists a r such that the comparison falls into one of the four categories.

Figure 14 shows the results for each combination of α and r .

A.3 Log-Price Ratio Plots

We utilise the same tool as in the experimental paper on choice under uncertainty by Choi et al. (2007a): We plot the relationship between the log-price ratio $\log(p_2/p_1)$ or $\log(q_2/q_1)$ and the share allocated to the other participant ($x_2/(x_1 + x_2)$ for the D-experiment) or to person B ($y_2/(y_1 + y_2)$ for the P- and V-experiment).

In our case, we can also derive predictions on the price and share pairs based on SY and AG. Assume that demand is homothetic, which in particular implies that the share is the same for a particular price ratio, irrespective of the wealth level. Consider Figure 15, and suppose we have a continuous demand function. The share should be a decreasing function of $\log(p_2/p_1)$. SY implies that (except in certain limit cases), $y_2/(y_1 + y_2) = 1/2$ in the P- or V-experiment when $\log(q_2/q_1) = 0$, so the share should be at least $1/2$ when $q_2 < q_1$ and at most $1/2$ when $q_2 > q_1$. Agreement implies that $x_2/(x_1 + x_2)$ is not greater than $y_2/(y_1 + y_2)$ for all price ratios. Note that this is neither a necessary nor a sufficient condition for AG (or AGARP), as it does require homotheticity (for necessity) and does not require even GARP to be satisfied (for sufficiency).

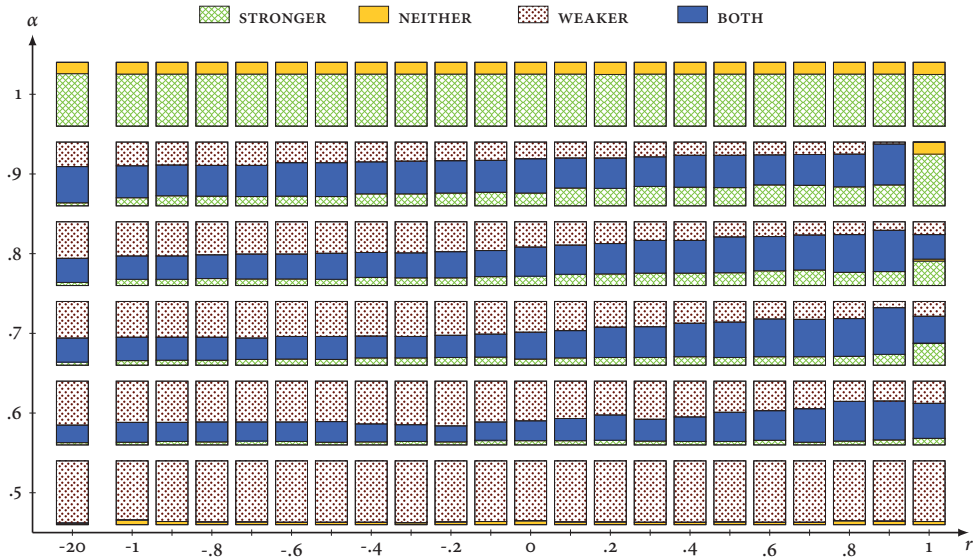


Figure 14: Comparability of participants with choices by CES functions with parameters α (for utility) and r (same for utility and justice). The figure shows the percentage of participants for which the comparison falls into each of the four possible categories.

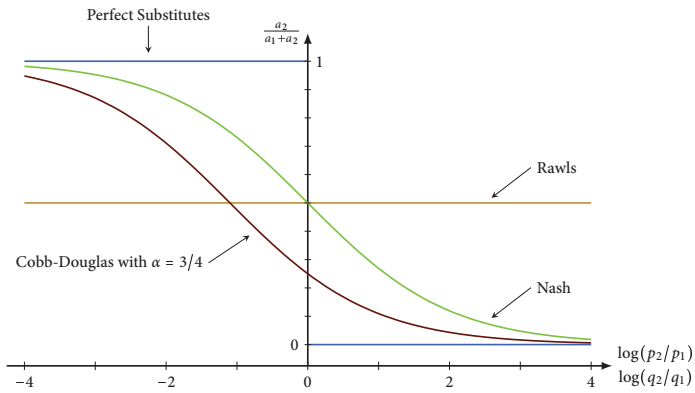


Figure 15: The relationship between the log-price ratio and the share allocated to the recipient or person B.

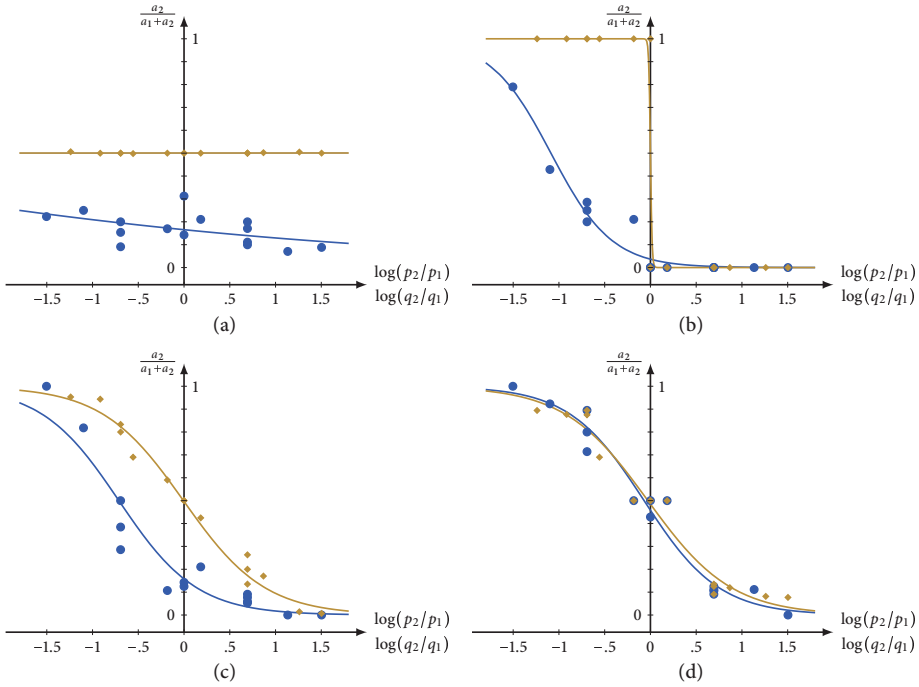


Figure 16: Relationship between log-price ratio and allocation share with logit estimator for the D-experiment (●) and the P- or V-experiment (◆). (a) Participant 171. (b) Participant 157. (c) Participant 183. (d) Participant 1.

Figure 15 shows three prototypical functions for the P- and V-experiment based on special cases of the CES function used in Section 4.2. It also shows one example of the share based on a Cobb-Douglas utility function $u(x_1, x_2) = x_1^{3/4} x_2^{1/4}$, which is an intermediate case between Nash with $\alpha = 1/2$ and purely selfish preferences with $\alpha = 1$.

Figure 16 shows the choices of four different participants from the D-experiment (●) and the P- or V-experiment (◆). The fitted curve is a simple estimated logit model and only intended as a visual aid. The participant in 16.(a) exhibits strong concern for equality (Rawlsian notion of justice), and his choices in the D-experiment reflect this as well. The participant in 16.(b) shows strong concern for efficiency (perfect substitutes type notion of justice), and again, this is reflected in his choices in the D-experiment. The participant in 16.(c) is an intermediate case. The participant in 16.(d) shows a very strong sense of justice in the sense that his choices in the D-experiment are almost perfectly in line with his choices in the P-experiment. The figures also illustrate the idea that the preference is “between” pure selfishness and pure justice.

Figure 5 shows the revealed preferred and worse sets next to the revealed more just and less just sets of the same four participants as in Figure 15, based on the methods described in BHH. Comparing Figure 5 with Figure 15 we find that the interpretation of the information contained in the plots of the allocation share is confirmed. For example, in 15.(a), the participant exhibits a strong Rawlsian notion of justice, and his preference is consistent with this high concern for equality. While

the preference is biased in the participant's own favour, the choices are largely unaffected by the price ratio.

Figures 17 to 24 show the relationship between the log-price ratio $\log(p_2/p_1)$ or $\log(q_2/q_1)$ and the share allocated to the second person $x_1/(x_1 + x_2)$ or $y_1/(y_1 + y_2)$ for the D-experiment (●) and the P- or V-experiment (◆).

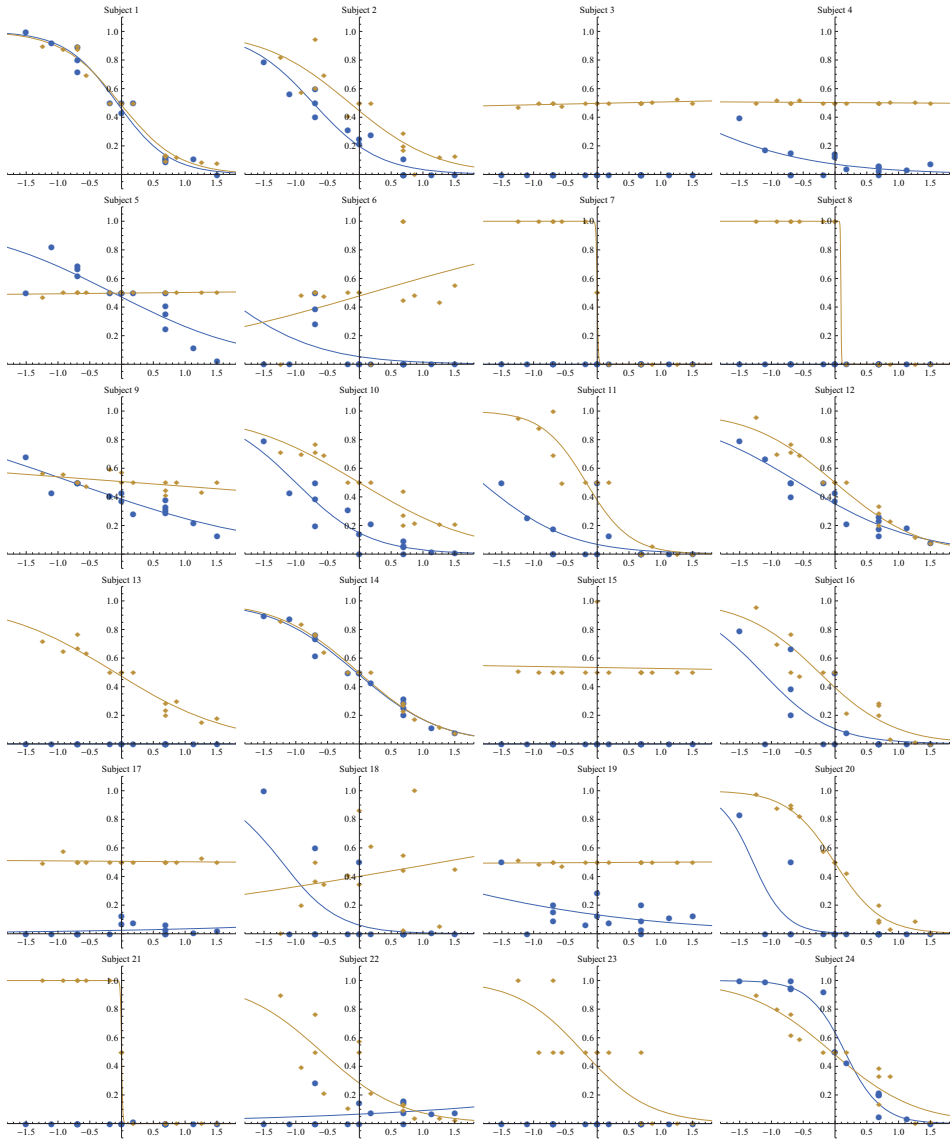


Figure 17: Treatment 1 (first D, then P), part 1: Relationship between log-price ratio and allocation share with logit estimation.

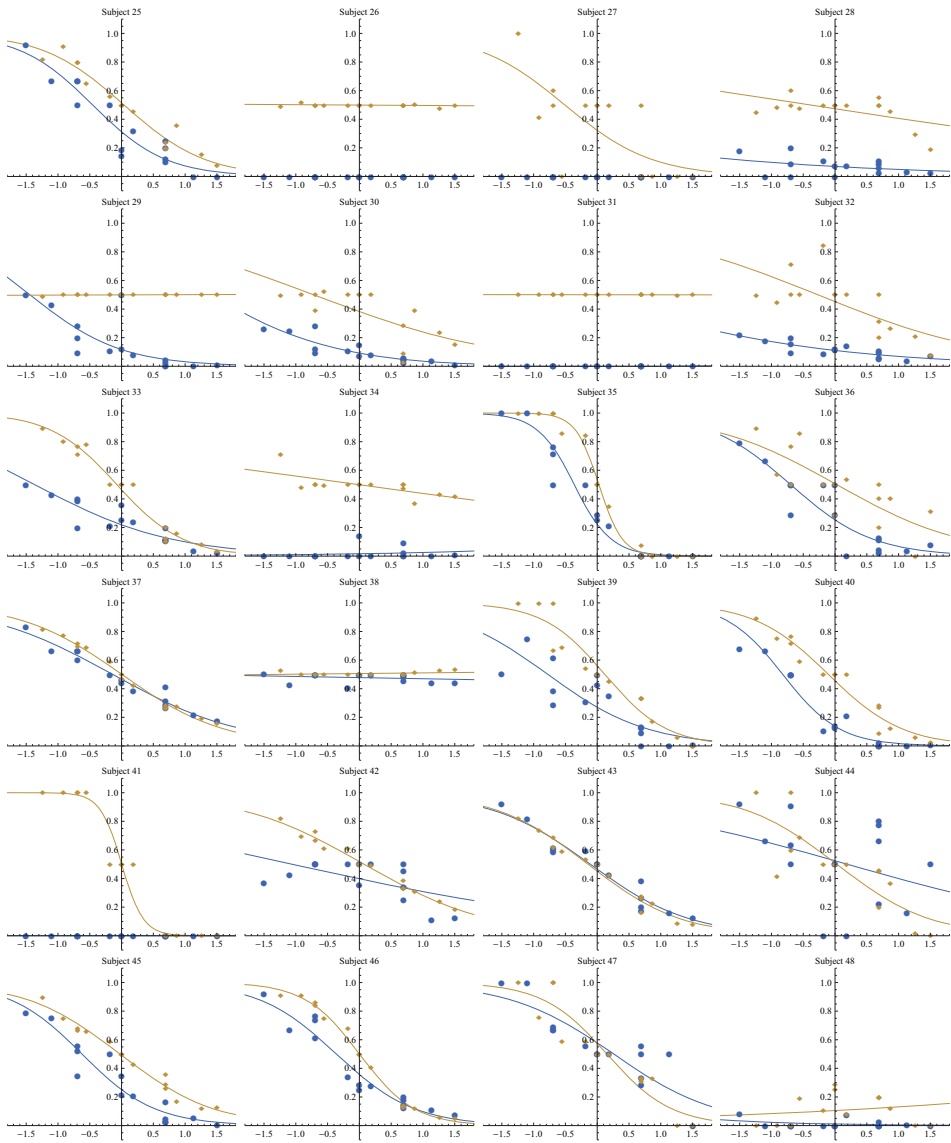


Figure 18: Treatment 1 (first D, then P), part 2: Relationship between log-price ratio and allocation share with logit estimation.

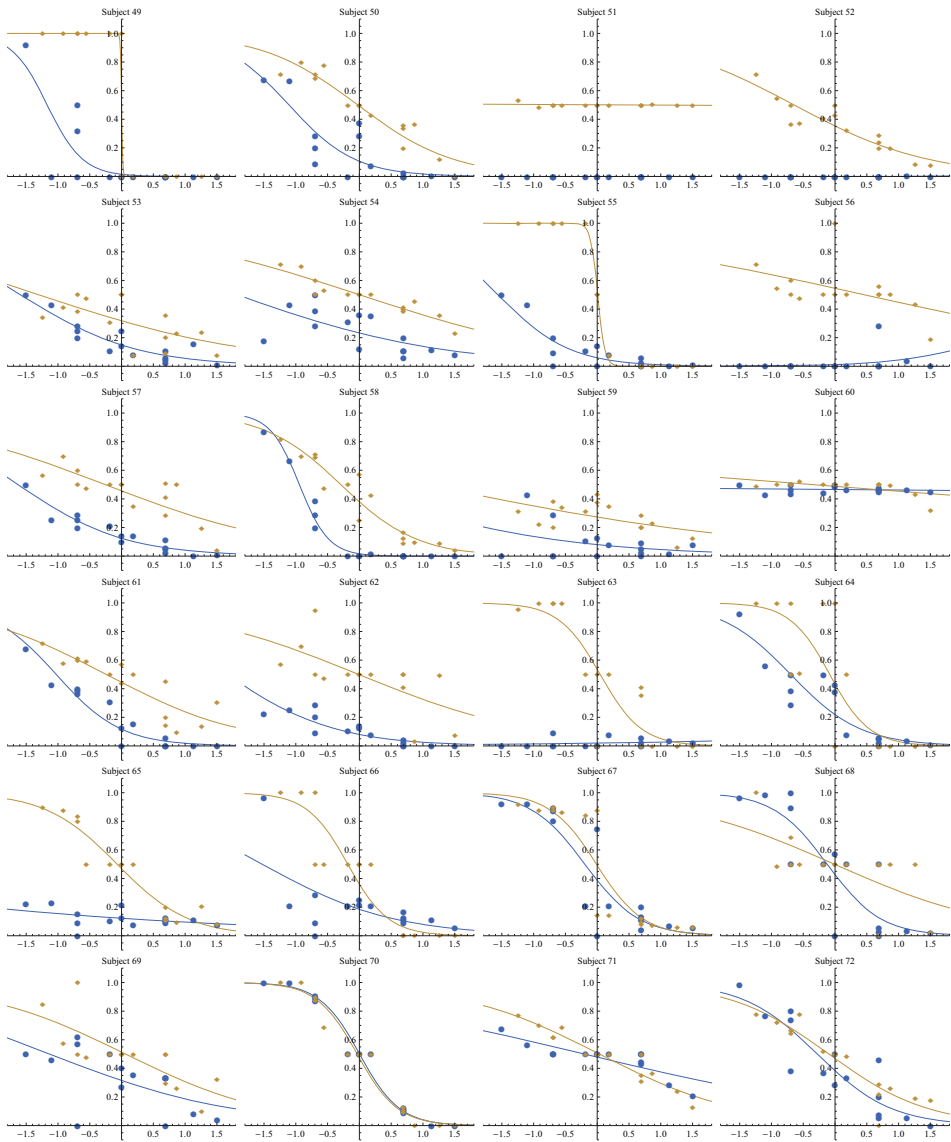


Figure 19: Treatment 2 (first D, then v), part 1: Relationship between log-price ratio and allocation share with logit estimation.

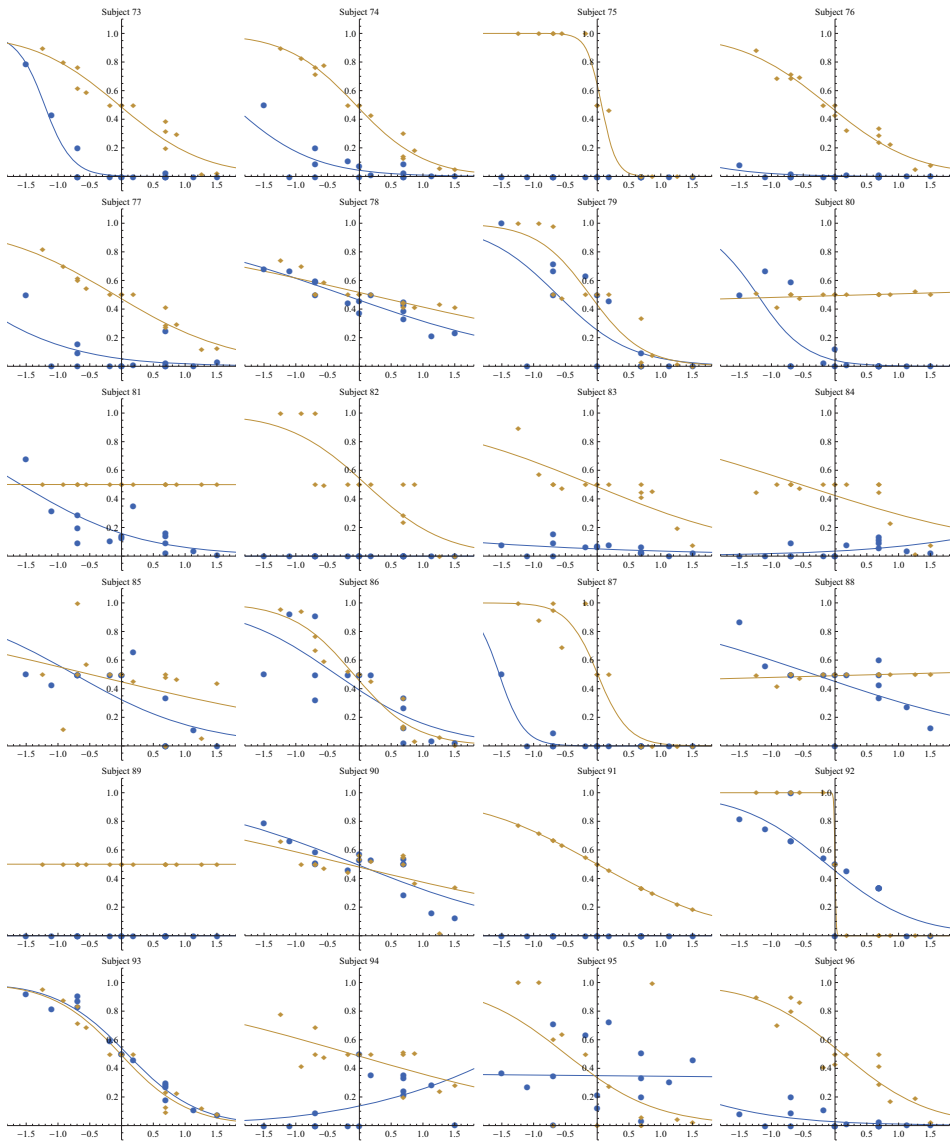


Figure 20: Treatment 2 (first D , then v), part 2: Relationship between log-price ratio and allocation share with logit estimation.

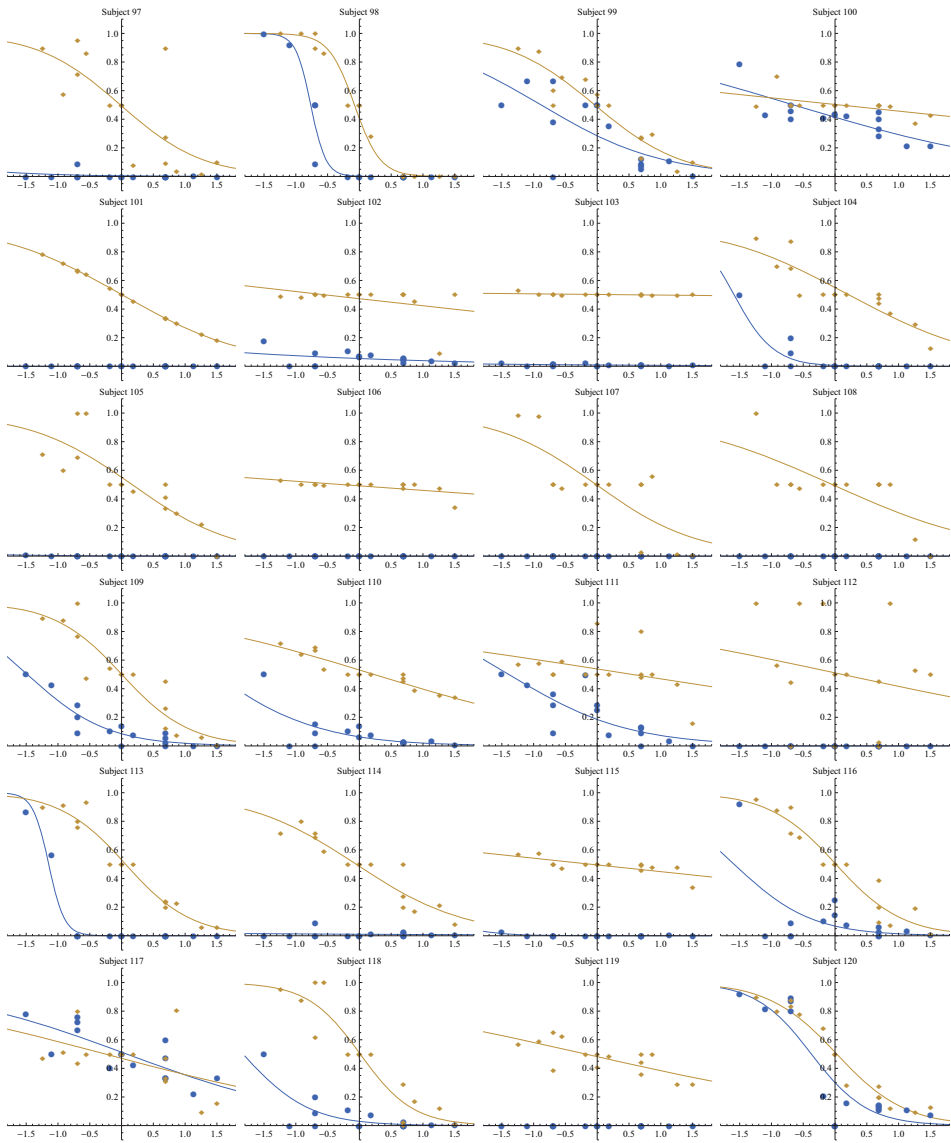


Figure 21: Treatment 3 (first P, then D), part 1: Relationship between log-price ratio and allocation share with logit estimation.

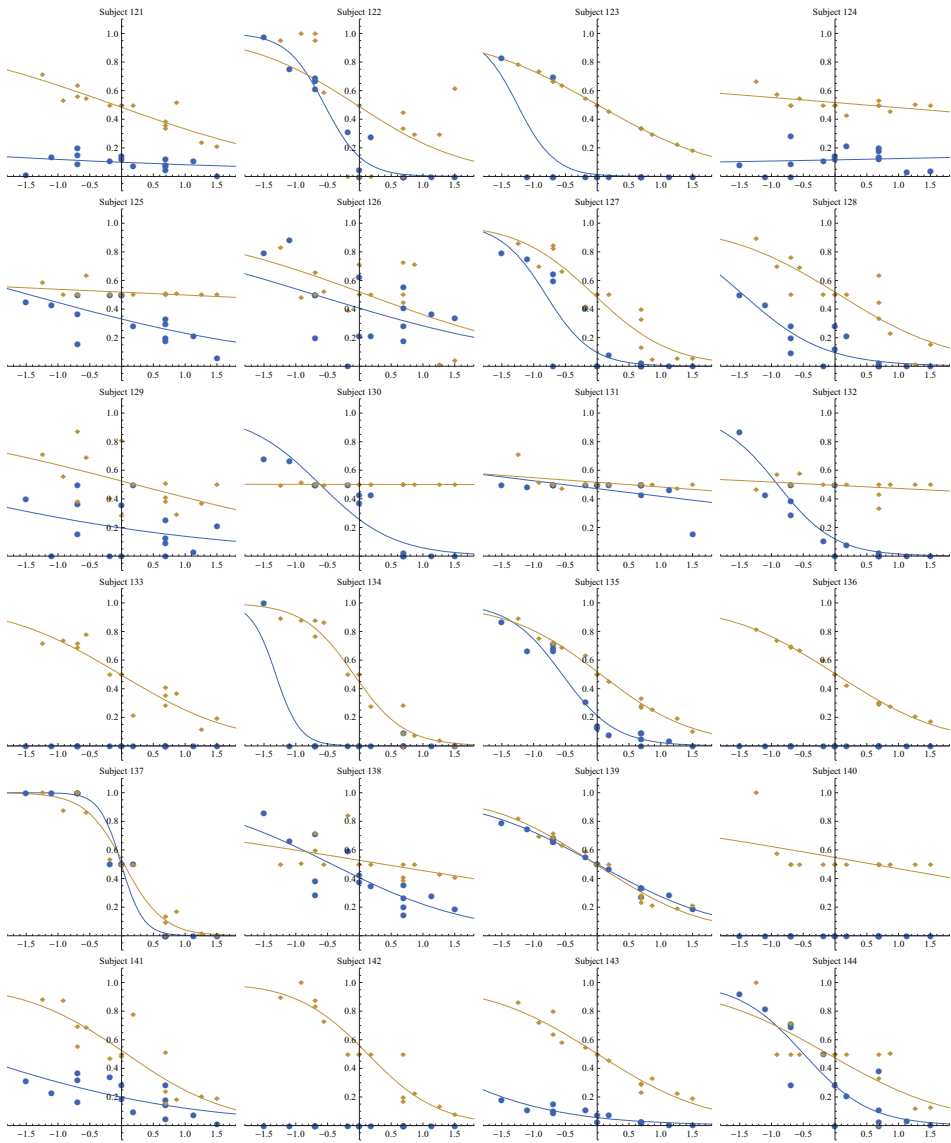


Figure 22: Treatment 3 (first P, then D), part 2: Relationship between log-price ratio and allocation share with logit estimation.

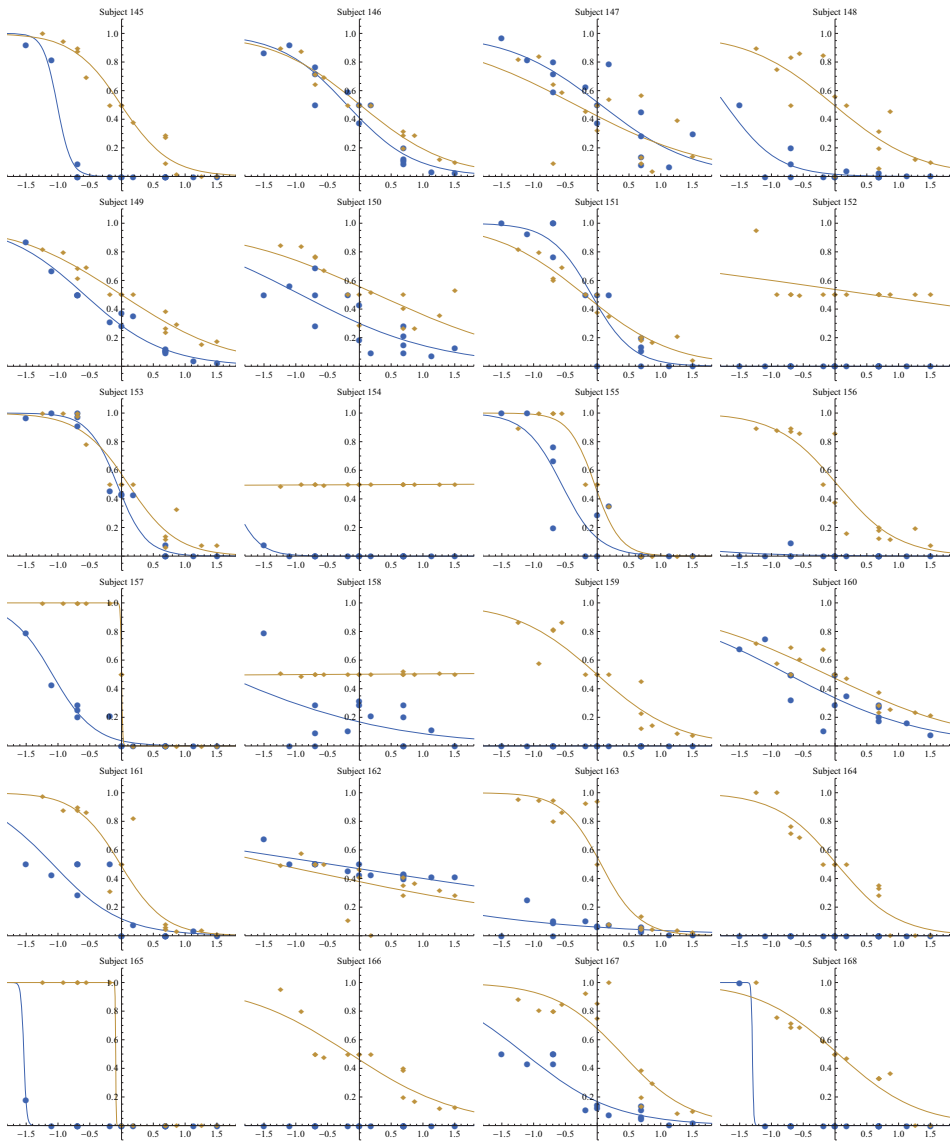


Figure 23: Treatment 4 (first v, then D), part 1: Relationship between log-price ratio and allocation share with logit estimation.

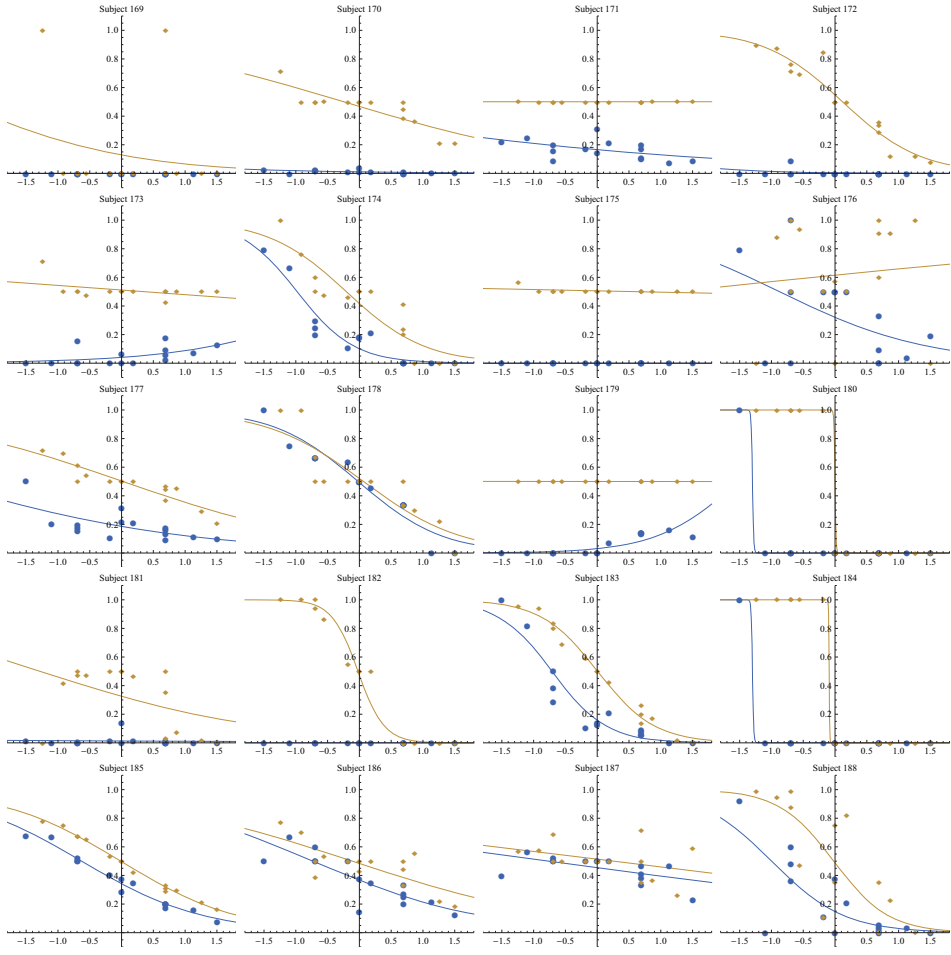


Figure 24: Treatment 4 (first v, then D), part 2: Relationship between log-price ratio and allocation share with logit estimation.

A.4 Averages and Measures of Dispersion by Treatment

Table 9 shows the mean and median amount allocated to the other participant by the dictator (x_2) for all 15 budgets, by treatment. As measures of dispersion, it also shows the coefficient of variation and interquartile range. Table 10 shows the same information as Table 9, except that it is the amount allocated to person B in the P- or V-experiment.

A.5 Treatment Effects

Tables 11 and 12 show the results for differences in the amount allocated to the other participant in the D-experiment (x_2) and in the amount allocated to person B in the P- or V-experiment (y_2), respectively.

Table 13 shows the results for the KW test and 2-sample AD test for the differences in the strength of justice between the treatments.

DICTATOR EXPERIMENT																
BUDGET	T1				T2				T3				T4			
	MEAN	CV	MEDIAN	IQR	MEAN	CV	MEDIAN	IQR	MEAN	CV	MEDIAN	IQR	MEAN	CV	MEDIAN	IQR
	1	5.09	1.05	3.01	7.51	4.95	1.06	3.01	7.38	4.54	1.15	3.01	8.34	6.02	1.07	3.01
2	3.90	1.17	3.00	6.00	3.36	1.26	1.50	6.00	2.64	1.41	0.00	4.31	3.62	1.28	1.35	6.00
3	1.95	1.03	1.00	3.50	1.86	1.30	1.00	3.00	1.71	1.33	0.50	3.00	1.66	1.42	0.75	3.00
4	2.51	1.06	2.50	4.50	2.05	1.24	0.50	3.00	1.76	1.44	0.50	2.95	2.04	1.30	0.50	3.50
5	2.85	1.00	2.26	4.00	2.82	1.21	2.00	5.00	1.81	1.52	0.10	2.00	2.06	1.39	1.00	4.00
6	1.42	1.07	0.60	3.00	1.15	1.14	0.60	2.64	0.99	1.25	0.60	2.30	0.96	1.37	0.33	1.80
7	1.58	0.87	1.25	3.00	1.18	1.23	0.50	2.12	1.08	1.28	0.01	2.00	0.99	1.27	0.12	2.00
8	1.74	0.95	1.10	4.00	1.48	1.22	0.80	3.40	0.98	1.54	0.00	1.38	1.06	1.33	0.00	2.50
9	1.14	0.95	0.88	1.96	1.18	1.02	0.50	2.23	0.78	1.28	0.50	1.33	0.95	1.24	0.48	2.17
10	0.79	1.20	0.50	1.50	0.82	1.19	0.50	1.56	0.57	1.53	0.00	1.00	0.56	1.38	0.00	1.10
11	0.75	1.27	0.12	1.50	0.84	1.30	0.25	1.31	0.52	1.77	0.00	0.69	0.53	1.35	0.00	1.00
12	0.88	1.29	0.25	1.50	0.84	1.35	0.25	1.25	0.56	1.73	0.25	0.44	0.59	1.32	0.05	1.25
13	1.03	1.21	0.50	2.00	1.08	1.26	0.50	2.01	0.69	1.64	0.00	1.00	0.74	1.40	0.00	1.00
14	0.50	1.41	0.12	1.00	0.44	1.48	0.04	0.70	0.42	1.65	0.04	0.36	0.31	1.58	0.02	0.68
15	0.44	1.51	0.11	1.00	0.40	1.65	0.11	0.51	0.32	2.03	0.00	0.11	0.36	1.71	0.00	0.33

Table 9: Mean and median amounts given in the D-experiment (x_2) for the restricted set (participants who satisfy the efficiency threshold), the coefficient of variation (standard deviation divided by mean), and the interquartile range, per treatment.

SOCIAL PLANNER AND VEIL OF IGNORANCE EXPERIMENT

BUDGET	T1												T2												T3												T4											
	CV			MEAN			IQR			MEDIAN			IQR			MEAN			CV			IQR			MEDIAN			IQR			MEAN			CV			IQR			MEAN			CV			IQR		
	CV	MEAN	IQR	CV	MEAN	IQR	CV	MEAN	IQR	CV	MEAN	IQR	CV	MEAN	IQR	CV	MEAN	IQR	CV	MEAN	IQR	CV	MEAN	IQR	CV	MEAN	IQR	CV	MEAN	IQR	CV	MEAN	IQR	CV	MEAN	IQR												
1	6.90	0.52	6.73	7.42	7.42	5.60	6.75	6.73	4.79	8.27	0.39	8.46	6.04																																			
2	5.08	0.48	4.62	4.37	4.12	4.88	4.62	3.70	6.21	0.41	5.95	5.00																																				
3	5.09	0.46	4.50	4.00	4.00	4.72	4.50	2.38	5.19	0.42	4.80	4.00																																				
4	6.08	0.41	5.80	3.50	5.00	5.45	5.50	3.00	6.16	0.46	5.50	5.00																																				
5	4.11	0.39	3.58	1.57	1.96	4.10	3.23	1.57	4.66	0.38	4.45	3.30																																				
6	3.49	0.30	3.00	0.60	0.00	3.18	3.00	0.18	3.69	0.36	3.00	0.60																																				
7	3.64	0.22	3.50	0.00	0.00	3.51	3.50	0.00	3.18	0.41	3.50	0.00																																				
8	4.09	0.15	4.00	0.00	0.00	3.98	4.00	0.00	3.79	0.35	4.00	0.00																																				
9	2.67	0.30	3.00	0.25	0.25	2.80	3.00	0.25	2.59	0.45	3.00	0.42																																				
10	2.04	0.52	2.12	1.50	2.00	2.37	2.00	1.00	1.78	0.62	2.00	2.25																																				
11	2.11	0.60	2.25	2.25	2.31	2.57	2.62	1.25	2.14	0.58	2.50	2.00																																				
12	2.33	0.61	2.38	3.00	3.00	2.89	3.00	1.75	2.26	0.62	2.50	2.25																																				
13	2.07	0.63	2.34	2.29	2.67	2.55	2.69	1.73	1.86	0.74	2.26	2.94																																				
14	1.59	0.79	1.34	2.83	2.13	1.84	1.34	1.96	1.48	0.82	1.34	2.46																																				
15	1.42	0.84	1.22	3.00	2.17	1.63	1.00	1.78	1.24	0.92	1.11	2.00																																				

Table 10: Mean and median amounts allocated to person B in the p- and v-experiment (y_2) for the restricted set (participants who satisfy the efficiency threshold), the coefficient of variation (standard deviation divided by mean), and the interquartile range, per treatment.

COMPARISON OF TREATMENTS: DICTATOR EXPERIMENT																
KRUSKAL-WALLIS																
BUDGET	1-2				1-3				1-2-3-4				ANDERSON-DARLING			
	1-2	2-4	3-4	1,2-3,4	1-3	2-4	3-4	1,2-3,4	1-2	1-3	2-4	3-4	1,2-3,4	2-4	3-4	1,2-3,4
1	0.9854 (0.000)	0.7239 (0.127)	0.4723 (0.525)	0.9247 (0.009)	0.5591 (0.344)	0.8997 (0.594)	0.6347 (0.230)	0.6438 (0.206)	0.3651 (0.902)	0.3902 (0.838)	0.5788 (0.366)	0.5659 (0.398)	0.6309 (0.639)	0.5287 (0.357)	0.6917 (0.065)	0.6917 (0.065)
2	0.6300 (0.236)	0.1854 (1.767)	0.4581 (0.559)	0.3269 (0.967)	0.8272 (0.048)	0.6232 (1.780)	0.6522 (0.184)	0.2301 (1.300)	0.6309 (0.241)	0.5287 (0.497)	0.3574 (0.922)	0.6917 (0.065)	0.6309 (0.639)	0.5287 (0.357)	0.6917 (0.065)	0.6917 (0.065)
3	0.5000 (0.462)	0.3716 (0.809)	0.7096 (0.001)	0.3290 (0.959)	0.6570 (0.199)	0.7175 (1.364)	0.4574 (0.670)	0.2941 (1.098)	0.6621 (0.158)	0.6747 (0.123)	0.3640 (0.904)	0.5500 (0.438)	0.6621 (0.158)	0.6747 (0.123)	0.3640 (0.904)	0.5500 (0.438)
4	0.7203 (0.130)	0.2896 (1.134)	0.7694 (0.088)	0.2366 (1.408)	0.9122 (0.012)	0.6710 (1.566)	0.2741 (1.158)	0.2839 (1.128)	0.3746 (0.878)	0.6321 (0.238)	0.3272 (1.004)	0.5934 (0.329)	0.3746 (0.878)	0.6321 (0.238)	0.3272 (1.004)	0.5934 (0.329)
5	0.6296 (0.237)	0.6621* (3.469)	0.3998 (0.167)	0.6852 (3.784)	0.0514* (0.005)	0.3383 (4.233)	0.5530 (0.437)	0.0557* (2.374)	0.5008 (0.565)	0.6549 (0.177)	0.4882* (2.500)	0.6791 (0.101)	0.5008 (0.565)	0.6549 (0.177)	0.4882* (2.500)	0.6791 (0.101)
6	0.5783 (0.314)	0.3358 (0.938)	0.4729 (0.523)	0.2423 (1.374)	0.6145 (0.256)	0.6421 (1.694)	0.6073 (0.301)	0.4317 (0.733)	0.5417 (0.465)	0.6311 (0.240)	0.3486 (0.946)	0.6246 (0.249)	0.5417 (0.465)	0.6311 (0.240)	0.3486 (0.946)	0.6246 (0.249)
7	0.1484 (2.100)	0.0758* (3.150)	0.6855 (0.167)	0.1118 (2.532)	0.2297 (1.450)	0.1755 (4.951)	0.0754 (2.146)	0.0457* (2.527)	0.6042 (0.310)	0.6675 (0.143)	0.1143 (1.844)	0.2444 (1.255)	0.6042 (0.310)	0.6675 (0.143)	0.1143 (1.844)	0.2444 (1.255)
8	0.3872 (0.758)	0.0220** (5.171)	0.3023 (1.077)	0.0182** (5.530)	0.7530 (0.168)	0.0902 (6.457)	0.3677 (0.896)	0.0086** (3.948)	0.2118 (1.366)	0.5496 (0.446)	0.0055** (4.397)	0.6577 (0.160)	0.2118 (1.366)	0.5496 (0.446)	0.0055** (4.397)	0.6577 (0.160)
9	0.8231 (0.051)	0.1148 (2.494)	0.1675 (0.025)	0.0375** (4.308)	0.8765 (0.122)	0.2224 (4.397)	0.5967 (0.328)	0.1424 (1.670)	0.1142 (1.835)	0.6184 (0.273)	0.0223** (3.120)	0.5948 (0.326)	0.1142 (1.835)	0.6184 (0.273)	0.0223** (3.120)	0.5948 (0.326)
10	0.9213 (0.010)	0.2451 (1.364)	0.1746 (1.856)	0.0764* (3.138)	0.9889 (0.006)	0.3677 (3.178)	0.7032 (0.042)	0.3175 (1.031)	0.1283 (1.748)	0.5655 (0.406)	0.0494* (2.482)	0.6746 (0.114)	0.1283 (1.748)	0.5655 (0.406)	0.0494* (2.482)	0.6746 (0.114)
11	0.8283 (0.048)	0.2685 (1.237)	0.3626 (0.841)	0.6931 (2.174)	0.6956 (0.154)	0.5136 (2.313)	0.6523 (0.184)	0.343 (0.961)	0.3589 (0.919)	0.6021 (0.315)	0.1534 (1.621)	0.6589 (0.157)	0.3589 (0.919)	0.6021 (0.315)	0.1534 (1.621)	0.6589 (0.157)
12	0.9151 (0.012)	0.1784 (1.823)	0.2272 (1.473)	0.0734* (3.202)	0.9803 (0.001)	0.3633 (3.202)	0.6474 (0.197)	0.2323 (1.293)	0.1085 (1.873)	0.4433 (0.705)	0.0599* (2.334)	0.6494 (0.183)	0.1085 (1.873)	0.4433 (0.705)	0.0599* (2.334)	0.6494 (0.183)
13	0.9773 (0.001)	0.1894 (1.736)	0.3118 (1.036)	0.0994* (2.717)	0.8981 (0.017)	0.4365 (2.741)	0.6512 (0.187)	0.248 (1.24)	0.3374 (0.976)	0.6583 (0.168)	0.098* (1.960)	0.6737 (0.116)	0.3374 (0.976)	0.6583 (0.168)	0.098* (1.960)	0.6737 (0.116)
14	0.9558 (0.003)	0.6308 (0.235)	0.1184 (2.447)	0.2945 (1.972)	0.5708 (0.324)	0.402 (2.952)	0.6002 (0.319)	0.5967 (0.328)	0.048 (2.485)	0.2297 (1.302)	0.173 (1.528)	0.6002 (0.312)	0.048 (2.485)	0.2297 (1.302)	0.173 (1.528)	0.6002 (0.312)
15	0.9244 (0.009)	0.2967 (1.101)	0.3198 (1.003)	0.1308 (2.288)	0.8111 (0.058)	0.9495 (2.004)	0.6589 (0.166)	0.3316 (0.992)	0.1819 (1.484)	0.5528 (0.438)	0.0898 (2.026)	0.6645 (0.142)	0.1819 (1.484)	0.5528 (0.438)	0.0898 (2.026)	0.6645 (0.142)
Prob.	-	17.91%	-	7.93%	-	-	-	25.08%	-	-	-	12.07%	-	-	-	-

Table 11: Comparison of amount allocated to the other participant in the D-experiment (x_2) between different treatments and per budget. 1,2 indicates the column with the results for treatment 1 vs. 2; 1,2-3,4 for treatment 1 and 2 (pooled) vs. 3 and 4 (pooled), etc. The table shows the p-values (test statistic in parentheses) for the Kruskal-Wallis and Anderson-Darling test. The last line is the probability of observing the number of significant differences by chance. Significance: ***, 1%, *, 5%, **, 10%.

COMPARISON OF TREATMENTS: SOCIAL PLANNER AND VEIL OF IGNORANCE EXPERIMENT															
KRUSKAL-WALLIS												ANDERSON-DARLING			
BUDGET	1-2	1-3	2-4	3-4	1,2-3,4	1,3-2,4	1-2-3-4	1-2	1-3	2-4	3-4	1,2-3,4	1,3-2,4		
1	0.4101 (0.688)	0.9267 (0.009)	0.3132 (1.030)	0.0391** (4.224)	0.6012 (0.275)	0.0489** (3.867)	0.1797 (4.895)	0.3876 (0.844)	0.4393 (0.714)	0.4709 (0.637)	0.0349** (2.733)	0.5344 (0.477)	0.0439** (2.574)		
2	0.4417 (0.601)	0.8749 (0.025)	0.2332 (1.436)	0.0272** (4.820)	0.3862 (0.756)	0.0437** (4.053)	0.1273 (5.685)	0.4925 (0.584)	0.2551 (1.217)	0.2994 (1.083)	0.0085** (3.954)	0.2642 (1.191)	0.0220** (3.130)		
3	0.9814 (0.001)	0.6897 (0.162)	0.7374 (0.115)	0.4079 (0.695)	0.9557 (0.003)	0.5778 (0.312)	0.8938 (0.620)	0.6915 (0.076)	0.5203 (0.517)	0.6101 (0.295)	0.5136 (0.533)	0.4996 (0.562)	0.6069 (0.295)		
4	0.9084 (0.083)	0.2964 (1.103)	0.9567 (0.003)	0.3270 (0.973)	0.5198 (0.418)	0.5293 (0.399)	0.7077 (1.406)	0.6391 (0.219)	0.3153 (1.357)	0.5854 (0.357)	0.2146 (1.355)	0.5051 (0.549)	0.4466 (0.693)		
5	0.4833 (0.499)	0.9756 (0.001)	0.0649* (3.399)	0.1363 (2.23)	0.1636 (1.947)	0.6212 (0.246)	0.2317 (4.301)	0.3252 (1.010)	0.6880 (0.085)	0.0446* (2.542)	0.1927 (1.439)	0.1762 (1.514)	0.3625 (0.908)		
6	0.2040 (1.627)	0.1969 (1.678)	0.4229 (0.652)	0.3485 (0.891)	0.8327 (0.045)	0.8205 (0.052)	0.4886 (2.447)	0.1291 (1.744)	0.1928 (1.439)	0.4798 (0.616)	0.2114 (1.367)	0.6568 (0.162)	0.3308 (0.994)		
7	0.3909 (0.747)	0.6154 (0.257)	0.3129 (1.031)	0.0426** (4.084)	0.2882 (1.135)	0.0513* (3.788)	0.1377 (5.507)	0.4250 (0.750)	0.6058 (0.305)	0.3563 (0.926)	0.0079** (4.021)	0.3025 (1.075)	0.0114* (3.712)		
8	0.7289 (0.122)	0.329 (0.965)	0.3693 (0.818)	0.8311 (0.046)	0.2057 (1.609)	0.9063 (0.014)	0.6312 (1.743)	0.6497 (0.191)	0.3442 (0.958)	0.3986 (0.816)	0.5657 (0.406)	0.1765 (1.513)	0.5854 (0.349)		
9	0.9701 (0.001)	0.9301 (0.008)	0.9384 (0.006)	0.9449 (0.005)	0.9184 (0.011)	0.9440 (0.005)	0.9995 (0.015)	0.7046 (0.038)	0.6324 (0.236)	0.5554 (0.432)	0.358 (0.921)	0.6659 (0.138)	0.5647 (0.402)		
10	0.4703 (0.529)	0.2045 (1.624)	0.757 (0.098)	0.017** (5.600)	0.4669 (0.533)	0.0283** (4.784)	0.0863* (6.553)	0.5669 (0.402)	0.2666 (1.181)	0.6429 (0.210)	0.0132** (3.550)	0.2223 (0.506)	0.0218** (3.141)		
11	0.4384 (0.609)	0.0989* (2.728)	0.521 (0.419)	0.1413 (2.175)	0.4938 (0.472)	0.6693 (0.184)	0.3590 (3.237)	0.5215 (0.514)	0.1129* (1.844)	0.5704 (0.395)	0.1643 (1.562)	0.4874 (0.592)	0.5409 (0.461)		
12	0.9196 (0.010)	0.0837* (2.991)	0.6688 (0.186)	0.0445** (4.010)	0.3557 (0.859)	0.2443 (1.363)	0.2185 (4.438)	0.5773 (0.377)	0.0636* (2.674)	0.3621 (0.910)	0.0474** (2.497)	0.1977 (1.424)	0.2735 (1.162)		
13	0.7201 (0.131)	0.0996* (2.716)	0.8640 (0.030)	0.0256** (4.920)	0.2736 (1.205)	0.0696* (3.290)	0.1061 (6.094)	0.6553 (0.176)	0.0788* (2.112)	0.6209 (0.267)	0.0137** (3.517)	0.2821 (1.135)	0.0638* (2.285)		
14	0.5820 (0.308)	0.3225 (0.991)	0.7992 (0.066)	0.1677 (1.916)	0.3099 (1.037)	0.1640 (1.943)	0.3621 (3.216)	0.5698 (0.395)	0.0607 (2.309)	0.6506 (0.189)	0.0542 (2.394)	0.2257 (1.319)	0.1942 (1.438)		
15	0.2997 (1.088)	0.3904 (0.748)	0.6903 (0.162)	0.1155 (2.486)	0.2850 (1.150)	0.0588* (3.565)	0.1922 (4.740)	0.2550 (1.218)	0.1897 (1.452)	0.5663 (0.405)	0.1001 (1.933)	0.3337 (0.986)	0.0738* (2.174)		
Prob.	-	18.21%	54.72%	2.57%	-	6.47%	57.37%	-	23.61%	65.40%	4.38%	-	8.75%		

Table 12: Comparison of amount allocated to person B in the p- or v- experiment (y_2) between different treatments and per budget. 1,2 indicates the column with the results for treatment 1 vs. 2; 1,2-3,4 for treatment 1 and 2 (pooled) vs. 3 and 4 (pooled), etc. The table shows the p-values (test statistic in parentheses) for the Kruskal-Wallis and Anderson-Darling test. The last line is the probability of observing the number of significant differences by chance. Significance: ***, 1%, ***, 5%, *, 10%.

STRONGER SENSE OF JUSTICE		
TREATMENTS	KRUSKAL-WALLIS	ANDERSON-DARLING
1-2	0.7861 (0.075)	0.4475 (0.694)
1-3	0.1604 (1.983)	0.1306 (1.735)
2-4	0.1799 (1.813)	0.1891 (1.454)
3-4	0.8915 (0.019)	0.567 (0.403)
1,2-3,4	0.0568* (3.621)	0.0573* (2.367)
1,3-2,4	0.9372 (0.006)	0.5218 (0.508)
1-2-3-4	0.3050 (3.640)	

Table 13: Comparison of the number of times a participant has a stronger sense of justice than another participant between different treatments and per budget. 1,2 indicates the column with the results for treatment 1 vs. 2; 1,2-3,4 for treatment 1 and 2 (pooled) vs. 3 and 4 (pooled), etc. The table shows the p-values (test statistic in parentheses) for the Kruskal-Wallis and Anderson-Darling test. Significance: ***1%, **5%, *10%.

B APPENDIX: INSTRUCTIONS AND EXPERIMENTAL SOFTWARE

B.1 Instructions

B.1.1 English Translation of Instructions

This is a translation of the original German instructions for the veil of ignorance experiment (v) played as the first part of a session (as in treatment 4). The instructions for the dictator experiment (D) and the social planner experiment (P) were as similar as possible and only changed where necessary (marked in italics). If a game was played as the second part of a session, the introduction and end of the instructions were changed accordingly.

Part 1

Welcome to the experiment and thank you for your participation! This experiment is jointly run by researchers of the Chair for Empirical and Experimental Economics (University of Jena), TU Dortmund University, and the Ruhr Graduate School in Economics (RGS).

You can earn an amount of money in this experiment which will depend on your decisions and on the decisions of other participants. **Therefore, it is very important that you thoroughly and accurately read these instructions.**

If you have a question please raise your hand. We will come to you and answer your question. Please do not pose your question(s) aloud. All participants of the experiments get the same instructions. However, the information provided on the screen during the experiment is only for the respective participant. Thus, you are **not allowed to look at the screen of other participants** and you are **not**

allowed to talk to other participants. If you violate these rules we are unfortunately forced to exclude you from the experiment. Please also switch off your mobile phones now.

The experiment comprises **two** parts. These are for the time being the instructions for the first part. Once the first part is finished you will receive the instructions for the second part. At the end of the experiment you will be paid for **both** parts. The payments for both parts are independent from each other, that is, your decisions in the first part will not influence your payments in the second part.

All information during the experiment will be in €.

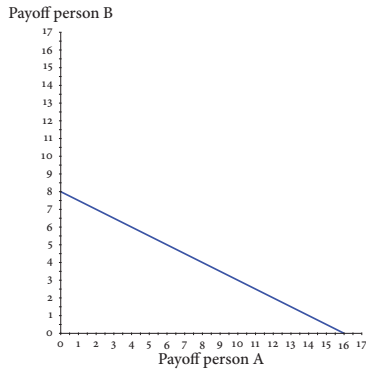
In the coming part of the experiment you will in each round divide one budget **between two participants** in this room for 15 rounds. *You will be one of them. In every round you will be randomly matched with a new person. You will never learn the identity of this person.* Also, your identity will never be disclosed.

In every round you will divide a given budget which varies from round to round *between two persons (henceforth person A and person B).* *You will be one of these persons, that is, you will be either randomly allocated to the role of person A or to the role of person B. You will be assigned each of the roles with equal probability. However, you do not know which role you will be assigned when you make your decision.*

The budget will be independent from your decisions as well as from the decisions of the other participants. The minimum amount for each of *these two persons* is 0 €. The maximum amount for each of *these two* will be shown on the right hand side of the screen. You find the maximum amount for *person A* at the top of the left column and the maximum amount for *person B* at the top of the right column. The maximum amount for *person A* and for *person B* can be unequal in a given round.

On the left hand side of the screen you find a graph. This graph displays the same information as shown on the right hand side in a different way: The payoffs of *person A* are shown on the horizontal axis and the payoffs of *person B* are shown on the vertical axis. The different payoff combinations for *person A* and *person B* are shown with a blue straight line. The maximum amount for *person A* is the intersection of the blue line with the horizontal axis and the maximum amount for *person B* is the intersection of the blue line with the vertical axis. Since the maximum amounts of *person A* and *B* vary from round to round, the trend and slope of the blue line which shows the possible payoff combinations changes from round to round, too.

The following graph shows an example: Here you can either allocate €16 to *person A* and *person B* then receives €0, you can allocate €0 to *person A* and *person B* then receives €8, or you can choose any allocation between these two values (up to two decimals) along the blue line. Here you have to keep in mind that one unit of payoff less for *person A* does not automatically mean that *person B* receives one unit of payoff more, but that this value can change from round to round. In the example €1 more for *person A* means €0.50 less for *person B*. The other way around, if you want to allocate €1 more to *person B*, you have to allocate €2 less to *person A*. This is due to the different maximum amounts for *person A* and *person B* and the resulting slope of the blue line.



Your task is now to divide the given budget between *person A* and *person B* without knowing which role will later be drawn for you. The payoff for *person A* or *person B* will be drawn with equal probability. To do this, enter into the respective field on the right hand side of the screen the amount you would like to allocate to *person A* (maximal 2 decimals possible). If you now press the button "Show" it will be shown on the right hand side of the screen how much *person A* and *person B* would each get if you confirm this distribution. At the same time you will see the distribution of the budget marked with red lines in the graph on the left hand side of the screen: On the horizontal axis the payoff for *person A* will be marked; on the vertical axis the payoff for *person B* will be marked. You can have as many distributions computed and shown in the graph as you wish until you decide for one distribution. If that is the case, press the "Confirm"-button which appears after a possible distribution has been shown for the first time. You will then get to the next round in which you will be matched with a *new person*.

The computation of the payoff will be done for all participants in the room in the same way: After the experiment has been completed **one** of these 15 rounds will be randomly drawn for you as well as your role (either *person A* or *person B*). For this round you will receive the amount that you allocated to the role that has been drawn for you, as well as the amount that another participant in this room (not the person that you assigned an amount to) allocated to the other role. If role *A* is randomly drawn for you, you will receive the amount that you allocated to *person A* in the respective round, as well as the amount that another participant allocated to *person B* in the respective round. If role *B* is randomly drawn for you, you will receive the amount that you allocated to *person B* in the respective round, as well as the amount that another participant assigned to *person A* in the respective round.

Before the real payoff-relevant experiment begins you can try the handling of the software and the experimental task for two rounds.

At the end of the experiment your earnings will be paid out in cash. This money will be paid privately. No other participant will learn from us how much you earn or which decisions you have taken. Neither will you learn about the decisions of the other participants.

Once you have read the instructions carefully, please start to answer the questions of the comprehension test. When you are done with this test, please raise your hand and one of the experimenters will come to you and correct the test. You can only participate in the experiment if you have understood the rules.

B.1.2 Original German Instructions

These are the original German instructions for the veil of ignorance experiment (v) played as the first part of a session (as in treatment 4). The instructions for the dictator experiment (D) and the social planner experiment (P) were as similar as possible and only changed where necessary (marked in italics). If a game was played as the second part of a session, the introduction and end of the instructions were changed accordingly.

Teil 1

Herzlich willkommen zum Experiment und vielen Dank für Ihre Teilnahme! Dieses Experiment wird gemeinschaftlich von Forschern des Lehrstuhls für Empirische und Experimentelle Wirtschaftsforschung (FSU Jena), der TU Dortmund und der Ruhr Graduate School in Economics (RGS) durchgeführt.

Sie können in diesem Experiment einen Geldbetrag verdienen, der abhängig ist von Ihren Entscheidungen und denen der anderen Teilnehmer. **Es ist daher sehr wichtig, dass Sie diese Instruktionen gründlich und genau durchlesen.**

Sollten Sie eine Frage haben, heben Sie bitte Ihre Hand. Wir werden dann zu Ihnen kommen und Ihre Frage beantworten. Bitte stellen Sie Ihre Frage(n) auf keinen Fall laut. Alle Teilnehmer des Experiments erhalten dieselben Instruktionen. Die Informationen im Verlauf des Experiments auf dem Bildschirm sind jedoch nur für die jeweiligen Teilnehmer bestimmt. Sie dürfen also **nicht auf den Bildschirm anderer Teilnehmer sehen und nicht mit anderen Teilnehmern reden**. Sollten Sie gegen diese Regeln verstoßen, sind wir leider dazu gezwungen, Sie vom Experiment auszuschließen. Bitte schalten Sie jetzt auch Ihre Mobiltelefone aus.

Das Experiment besteht aus zwei Teilen. Dies sind zunächst die Instruktionen für den ersten Teil. Wenn der erste Teil beendet ist, erhalten Sie die Instruktionen für den zweiten Teil. Am Ende des Experiments werden Sie für **beide** Teile entlohnt. Die Auszahlungen der beiden Teile sind unabhängig voneinander, d.h. Ihre Entscheidungen im ersten Teil haben keinen Einfluss auf Ihre Auszahlungen im zweiten Teil.

Alle Angaben während des Experiments sind in €.

Im kommenden Teil des Experiments teilen Sie 15 Runden lang jeweils ein Budget *zwischen zwei Teilnehmern* in diesem Raum auf. *Einer davon werden Sie sein. In jeder Runde wird Ihnen eine neue Person zufällig zugelost. Sie werden deren Identität nicht erfahren.* Auch Ihre Identität wird zu keinem Zeitpunkt bekannt gegeben.

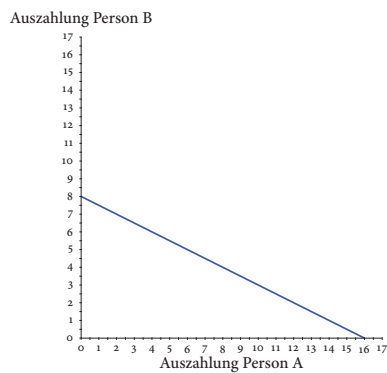
In jeder Runde teilen Sie ein gegebenes Budget, das von Runde zu Runde variiert, zwischen *zwei Personen (im Folgenden Person A und Person B) auf. Sie werden eine der Personen sein, also entweder die Rolle von Person A oder die Rolle von Person B zugelost bekommen. Jede der Rollen erhalten Sie mit gleicher Wahrscheinlichkeit. Allerdings wissen Sie zum Zeitpunkt Ihrer Entscheidung nicht, welche dies sein wird.*

Das Budget ist sowohl unabhängig von Ihren Entscheidungen als auch unabhängig von den Entscheidungen der anderen Teilnehmer. Der Mindestbetrag für *jede dieser beiden Personen* beträgt dabei € 0. Der Höchstbetrag für *jeden der beiden* wird Ihnen auf der rechten Seite des Bildschirms angezeigt. In der linken Spalte oben finden Sie den Höchstbetrag für *Person A*; in der rechten Spalte oben finden Sie den Höchstbetrag für *Person B*. Der Höchstbetrag in einer Runde kann für *Person A und für Person B* unterschiedlich hoch sein.

Auf der linken Seite des Bildschirms befindet sich eine Grafik. In dieser Grafik werden dieselben Informationen wie auf der rechten Seite in einer anderen Weise dargestellt: Die Auszahlungen

von *Person A* werden an der waagerechten Achse dargestellt, die Auszahlungen von *Person B* an der senkrechten Achse. Die verschiedenen Auszahlungskombinationen für *Person A und Person B* werden mit einer blauen Gerade dargestellt. Der Höchstbetrag für *Person A* ist der Schnittpunkt der blauen Geraden mit der waagerechten Achse und der Höchstbetrag für *Person B* ist der Schnittpunkt der blauen Geraden mit der senkrechten Achse. Da die Höchstbeträge von *Person A und Person B* von Runde zu Runde variieren, ändert sich auch der Verlauf und die Steigung der blauen Geraden, die die möglichen Auszahlungskombinationen abbildet, von Runde zu Runde.

Die folgende Grafik stellt ein Beispiel dar: Hier können Sie entweder *Person A* €16 zuteilen und *Person B* bekommt dann €0, Sie können *Person A* €0 zuteilen und *Person B* bekommt €8 oder Sie können jede beliebige Aufteilung zwischen diesen Werten (bis zu zwei Nachkommastellen genau) entlang der blauen Gerade wählen. Hierbei ist zu beachten, dass eine Einheit Auszahlung weniger für *Person A* nicht automatisch eine Einheit mehr Auszahlung für *Person B* bedeutet, sondern dass dieser Wert von Runde zu Runde variieren kann. Im Beispiel bedeutet €1 mehr für *Person A* €0,50 weniger für *Person B*. Umgekehrt heißt das, wenn Sie *Person B* €1 mehr zuteilen möchten, müssen Sie *Person A* €2 weniger zuteilen. Das liegt an den unterschiedlichen Höchstbeträgen für *Person A und Person B* und der daraus resultierenden Steigung der blauen Geraden.



Ihre Aufgabe ist es nun, das vorhandene Budget zwischen *Person A und Person B* aufzuteilen *ohne zu wissen, welche der Rollen Sie später zugelost bekommen*. Sie werden die Auszahlung von *Person A oder Person B* mit gleicher Wahrscheinlichkeit zugelost bekommen. Hierzu geben Sie nun in das entsprechende Eingabefeld auf der rechten Seite des Bildschirms ein, wie viel Sie *Person A* zuteilen (maximal 2 Nachkommastellen möglich). Wenn Sie nun auf den Knopf "Anzeigen" klicken, wird Ihnen auf der rechten Seite des Bildschirms angezeigt, wie viel *Person A und Person B* jeweils bekommen würden, falls Sie diese Aufteilung bestätigen. Gleichzeitig wird Ihnen die Aufteilung des Budgets in der Grafik auf der linken Seite des Bildschirms mit roten Linien angezeigt: An der waagerechten Achse wird die *Auszahlung für Person A* markiert; an der senkrechten Achse wird die Auszahlung für *Person B* markiert. Sie können solange verschiedene Aufteilungen berechnen und in der Grafik anzeigen lassen bis Sie sich für eine Aufteilung entschieden haben. Wenn dies der Fall ist, klicken Sie auf den "Bestätigen"-Knopf, der nach dem ersten Anzeigen einer möglichen Aufteilung erscheint. Danach gelangen Sie zur nächsten Runde in der Ihnen *eine neue Person* zugelost wird.

Die Berechnung der Auszahlung erfolgt für alle Teilnehmer im Raum auf die gleiche Weise: Nach dem Experiment wird **eine** dieser 15 Runden *sowie Ihre Rolle (entweder Person A oder Person B)* zufällig ausgelost. Für diese Runde bekommen Sie den Betrag ausgezahlt, den Sie *der Ihnen*

zugelosten Rolle zugeteilt haben sowie den Betrag, den ein anderer Teilnehmer in diesem Raum (nicht die Person, der Sie einen Betrag zugeteilt haben) der anderen Rolle zugeteilt hat. Falls Sie also Rolle A zugelost bekommen haben, erhalten Sie den Betrag, den Sie in der ausgelosten Runde Person A zugeteilt haben, sowie den Betrag, den ein anderer Teilnehmer in dieser Runde Rolle B zugeteilt hat. Falls Sie Rolle B zugelost bekommen haben, erhalten Sie den Betrag, den Sie in der ausgelosten Runde Person B zugeteilt haben, sowie den Betrag, den ein anderer Teilnehmer in dieser Runde Rolle A zugeteilt hat.

Bevor das eigentliche auszahlungsrelevante Experiment beginnt, können Sie zwei Runden lang die Handhabung der Software und die Experimentalaufgabe ausprobieren.

Am Ende des Experiments bekommen Sie Ihren Verdienst bar ausgezahlt. Dieses Geld wird Ihnen privat ausgezahlt. Kein anderer Teilnehmer wird von uns erfahren, wie viel Sie verdienen oder welche Entscheidungen Sie getroffen haben. Sie werden auch nichts über die Entscheidungen der anderen Teilnehmer erfahren.

Wenn Sie die Instruktionen gründlich gelesen haben, beginnen Sie bitte damit, die Fragen des Verständnistests zu beantworten. Wenn Sie mit dem Test fertig sind, melden Sie sich und einer der Experimentleiter wird zu Ihnen kommen und den Test kontrollieren. Sie dürfen nur dann am Experiment teilnehmen, wenn Sie die Regeln verstanden haben.

B.2 Screenshot

Figure 25 shows a screenshot from the experimental software.

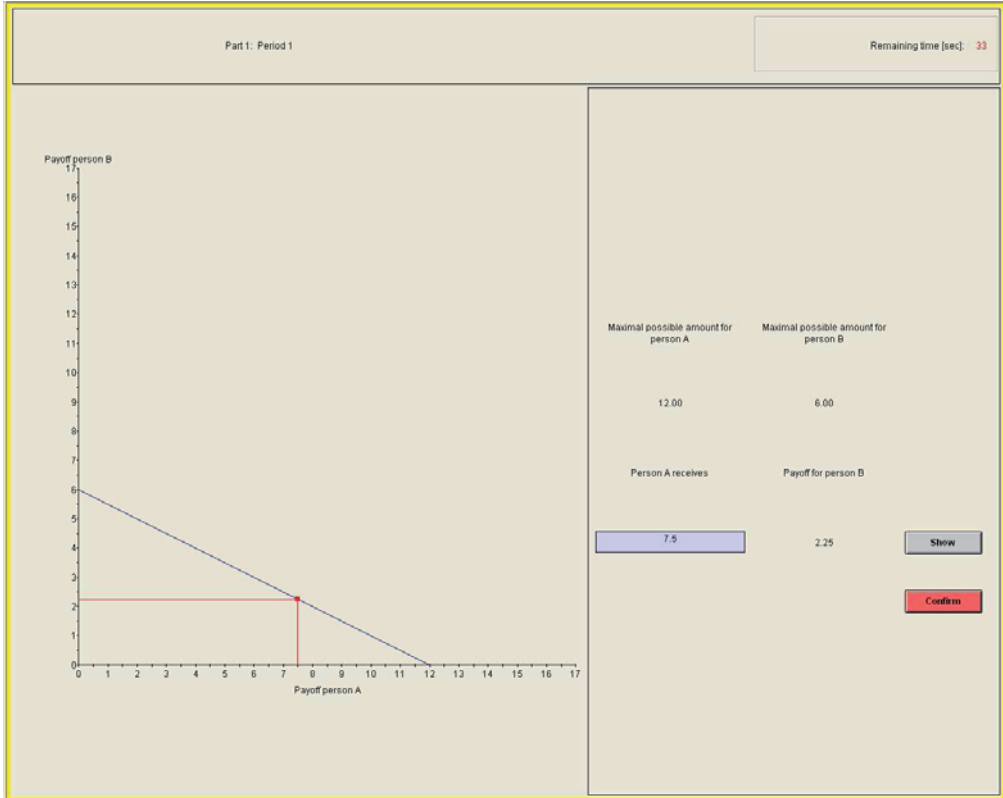


Figure 25: Screenshot of the experiment (translation of the original German screen), treatment 3/4, part 1 (social planner/veil of ignorance).

REFERENCES

- Afriat, S. N. (1967): "The Construction of Utility Functions From Expenditure Data," *International Economic Review* 8(1):67–77.
- (1972): "Efficiency Estimation of Production Functions," *International Economic Review* 13(3):568–598.
- Anderson, T. W. and Darling, D. A. (1952): "Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes," *Annals of Mathematical Statistics* 23(2):193–212.
- Andreoni, J. (1995): "Cooperation in Public-Goods Experiments: Kindness or Confusion?" *American Economic Review* 85(4):891–904.
- Andreoni, J. and Miller, J. (2002): "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica* 70(2):737–753.
- Andreoni, J. and Vesterlund, L. (2001): "Which is the Fair Sex? Gender Differences in Altruism," *Quarterly Journal of Economics* 116(1):293–312.
- Banerjee, S. and Murphy, J. H. (2006): "A Simplified Test for Preference Rationality of Two-Commodity Choice," *Experimental Economics* 9(9):67–75.
- Becker, N., Häger, K., and Heufer, J. (2013): "Revealed Notions of Distributive Justice I: Theory," Working paper.
- Berg, J., Dickhaut, J., and McCabe, K. (1995): "Trust, Reciprocity, and Social History," *Games and Economic Behavior* 10(1):122–142.
- Bronars, S. G. (1987): "The Power of Nonparametric Tests of Preference Maximization," *Econometrica* 55(3):693–698.
- Brosig-Koch, J., Helbach, C., Ockenfels, A., and Weimann, J. (2011): "Still different after all these years: Solidarity behavior in East and West Germany," *Journal of Public Economics* 95(11–12):1373–1376.
- Burks, S. V., Carpenter, J. P., and Verhoogen, E. (2003): "Playing both Roles in the Trust Game," *Journal of Economic Behavior and Organization* 51(2):195–216.
- Choi, S., Fisman, R., Gale, D., and Kariv, S. (2007a): "Consistency and Heterogeneity of Individual Behavior under Uncertainty," *American Economic Review* 97(5):1921–1938.
- (2007b): "Revealing Preferences Graphically: An Old Method Gets a New Tool Kit," *American Economic Review* 97(2):153–158.
- Cox, J. C. (2009): "Trust and reciprocity: implications of game triads and social contexts," *New Zealand Economic Papers* 43(2):89–104.
- Dickinson, D. L. (2009): "Experiment Timing and Preferences for Fairness," *The Journal of Socio-Economics* 38(1):89–95.

- Engel, C. (2011): "Dictator Games: A Meta Study," *Experimental Economics* 14(4):583–610.
- Fehr, E. and Fischbacher, U. (2004): "Third-party punishment and social norms," *Evolution and Human Behavior* 25:63–87.
- Fehr, E. and Gächter, S. (2000): "Cooperation and Punishment in Public Goods Experiments," *The American Economic Review* 90(4):980–994.
- (2002): "Altruistic Punishment in Humans," *Nature* 415:137–140.
- Février, P. and Visser, M. (2004): "A Study of Consumer Behavior Using Laboratory Data," *Experimental Economics* 7(1):93–114.
- Fischbacher, U. (2007): "z-tree: Zurich Toolbox for Ready-Made Economic Experiments," *Experimental Economics* 10(2):171–178.
- Fischbacher, U. and Gächter, S. (2010): "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments," *American Economic Review* 100(1):541–56.
- Fisman, R., Kariv, S., and Markovits, D. (2007): "Individual Preferences for Giving," *American Economic Review* 97(5):1858–1876.
- Forsythe, R., Horowitz, J. L., and Savin, N. (1994): "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior* 6(3):347–369.
- Good, P. (2005): *Permutation, parametric and bootstrap tests of hypotheses*, Springer Series in Statistics, New York: Springer-Verlag, third edition.
- Greiner, B. (2004): "The Online Recruitment System ORSEE 2.0 – A Guide for the Organization of Experiments in Economics," Technical report, University of Cologne, technical Report 10, Working Paper Series in Economics.
- Gross, J. (1995): "Testing Data for Consistency with Revealed Preference," *Review of Economics and Statistics* 77(4):701–710.
- Harbaugh, W. T., Krause, K., and Berry, T. R. (2001): "GARP for Kids: On the Development of Rational Choice Behavior," *American Economic Review* 91(5):1539–1545.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., and Ziker, J. (2006): "Costly Punishment Across Human Societies," *Science* 312(5781):1767–1770.
- Heufer, J. (2012): "Testing for Utility Maximization with Error and the Loss of Power," *German Economic Review* 13(2):161–173.
- (2013a): "Generating Random Optimising Choices," *Computational Economics* forthcoming.
- (2013b): "Testing Revealed Preferences for Homotheticity with Two-Good Experiments," *Experimental Economics* 16(1):114–124.
- Isaac, R. M. and Walker, J. M. (1988): "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism," *The Quarterly Journal of Economics* 103(1):179–199.

- Karni, E. and Safra, Z. (2002a): "Individual Sense of Justice: A Utility Representation," *Econometrica* 70(1):263–284.
- (2002b): "Intensity of the Sense of Fairness: Measurement and Behavioral Characterization," *Journal of Economic Theory* 105(2):318–337.
- Kruskal, W. H. and Wallis, W. A. (1952): "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association* 47(260):583–621.
- Mattei, A. (2000): "Full-scale real tests of consumer behavior using experimental data," *Journal of Economic Behavior & Organization* 43(4):487–497.
- Nash, J. F. (1950): "The Bargaining Problem," *Econometrica* 18(2):155–162.
- Ockenfels, A. and Weimann, J. (1999): "Types and Patterns: An Experimental East-West-German Comparison of Cooperation and Solidarity," *Journal of Public Economics* 71(2):275–287.
- Scholz, F. W. and Stephens, M. A. (1987): "K-Sample Anderson-Darling Tests," *Journal of the American Statistical Association* 82(399):918–924.
- Sippel, R. (1996): "A Note on the Power of Revealed Preference Tests with Afriat Inefficiency," SFB 303, Universität Bonn, Discussion Paper No. A-528.
- (1997): "An Experiment on the Pure Theory of Consumer's Behavior," *The Economic Journal* 107(444):1431–1444.
- Sutter, M. and Kocher, M. G. (2007): "Trust and Trustworthiness across Different Age Groups," *Games and Economic Behavior* 59(2):364–382.
- Varian, H. R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica* 50(4):945–972.
- (1993): "Goodness-of-Fit for Revealed Preference Tests," Working Paper, University of Michigan.
- Westfall, P. H. and Young, S. S. (1993): *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*, Wiley Series in Probability and Mathematical Statistics, New York: John Wiley & Sons.
- Willinger, M., Keser, C., Lohmann, C., and Usunier, J.-C. (2003): "A Comparison of Trust and Reciprocity between France and Germany: Experimental Investigation based on the Investment Game," *Journal of Economic Psychology* 24(4):447–466.