RESEARCH ON ECONOMIC INEQUALITY VOLUME 11

INEQUALITY, WELFARE AND INCOME DISTRIBUTION: EXPERIMENTAL APPROACHES

EDITED BY

FRANK COWELL

2004

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

INEQUALITY, WELFARE AND INCOME DISTRIBUTION: EXPERIMENTAL APPROACHES

RESEARCH ON ECONOMIC INEQUALITY

Series Editors: Yoram Amiel and John A. Bishop

ELSEVIER B.V.
Sara Burgerhartstraat 25
P.O. Box 211
1000 AE Amsterdam
The Netherlands

ELSEVIER Inc.
525 B Street, Suite 1900
San Diego
CA 92101-4495
USA

ELSEVIER Ltd
The Boulevard, Langford
Lane, Kidlington
Oxford OX5 1GB
UK

ELSEVIER Ltd
84 Theobalds Road
London
WC1X 8RR
UK

First edition 2004

Library of Congress Cataloging in Publication Data
A catalog record is available from the Library of Congress.

British Library Cataloguing in Publication Data
A catalogue record is available from the British Library.

ISBN: 0-7623-1113-4
ISSN: 1049-2585 (Series)

♾ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper). Printed in The Netherlands.

CONTENTS

LIST OF CONTRIBUTORS

Yoram Amiel	Department of Economics and Management, Ruppin Institute, Emek-Hefer, Israel
Steven R. Beckman	University of Colorado at Denver, Denver, CO, USA
Yoella Bereby-Meyer	Department of Education, Ben Gurion University of the Negev, Beer Sheva, Israel
Michele Bernasconi	Dipartimento di Economia, Università dell'Insubria, Varese, Italy
Friedel Bolle	Department of Economics, Europa-Universität Viadrina, Frankfurt (Oder), Germany
Kristof Bosmans	Katholieke Universiteit Leuven, Leuven, Belgium
Daniele Checchi	University of Milan, Milan, Italy
Frank Cowell	STICERD and Economics Department, London School of Economics and Political Science, London, U.K.
Valentino Dardanoni	Dipartimento di Scienze Economiche, Aziendali e Finanziarie, Università di Palermo, Palermo, Italy
Liema Davidovitz	Department of Economics and Management, Ruppin Institute, Emek-Hefer, Israel
Antonio Filippin	EUI, Florence, Italy
John P. Formby	Department of Economics, Finance and Legal Studies, University of Alabama, Tuscaloosa, AL, USA
Brit Grosskopf	Department of Economics, Texas A&M University, College Station, TX, USA

Alexander Kritikos	Department of Economics, Europa-Universität Viadrina, Frankfurt (Oder), Germany
Yoram Kroll	School of Business, Hebrew University, Jerusalem, Israel
Luigi Mittone	Computable and Experimental Economics Laboratory, Department of Economics, University of Trento, Trento, Italy
Erik Schokkaert	Katholieke Universiteit Leuven, Leuven, Belgium
W. James Smith	University of Colorado at Denver, Denver, CO, USA
Dan Slottje	Economics Department, Southern Methodist University, Dallas, TX, USA
Daniel Zizzo	Christ Church College, University of Oxford, Oxford, U.K.

PREFACE

The emerging literature on experimental methods in connection with economic inequality has shed fresh light on how to think about inequality, how important issues of equality are in comparison with other economic objectives and how individuals incorporate criteria of equality and fairness into their own decisions.

This special issue covers a wide range of topics in the field of income distribution and extends the approach to related issues such as mobility and attitudes to risk. The selection of papers includes both participatory experiments where individuals respond to economic incentives and questionnaire experiments designed to investigate the extent to which individuals' values and attitudes conform to the way that economists and others conventionally model inequality and welfare. The contributions appear by alphabetical order of author. All papers were refereed in the usual way. I have been greatly assisted by advice from:

Steven Beckman
Jordi Brandts
Guillermo Cruces
Kurt Devooght
Urs Fischbacher
John Formby
Uri Gneezy
Hans Grüner
Heike Hennig Schmidt
James Konow
Shlomo Maital
Sarah Maxwell
Astri Muren
Adam Oliver
Matthew Rabin
Xavi Ramos
Javier Ruiz-Castillo
Erik Schokkaert
Martin Sefton
Michael Tagler
Richard Wahlund
Paul Webley
George Wu
Rami Zwick

Frank Cowell
Volume Editor

WHY DO PEOPLE VIOLATE THE TRANSFER PRINCIPLE? EVIDENCE FROM EDUCATIONAL SAMPLE SURVEYS

Yoram Amiel, Frank Cowell and Dan Slottje

ABSTRACT

We run income inequality questionnaire in 17 universities in the USA. In the questionnaire we examine how students of economics compare inequality of income distributions, when transfers are made between income recipients. The results are analysed in terms of several personal characteristics of the respondents: family income, ethnicity, sex, geographic origin, number of siblings, age, and by ranking of the universities.

1. INTRODUCTION

The transfer principle has sometimes been seen as the cornerstone of the study of inequality. First enunciated by Pigou (1912) and subsequently developed by Dalton (1920) it is used as part of the characterisation of standard families of inequality measures and of ranking principles for income distributions. However, in recent years, the appropriateness of this axiom has been called into question. A series of contributions using questionnaire-experiments has demonstrated that individuals' rankings of income distributions violate the Dalton principle although

Inequality, Welfare and Income Distribution: Experimental Approaches
Research on Economic Inequality, Volume 11, 1–16

ISSN: 1049-2585/doi:10.1016/S1049-2585(04)11001-6

they usually accord with the original Pigou insight (Amiel & Cowell, 1992, 1999, 2002). In this paper, we focus on the proximate causes for this apparently aberrant behaviour and also provide further confirmation of the phenomenon.

The structure of the paper is as follows. Section 2 discusses the transfer principle in more detail and presents the research questions that we address. Our experiment and the data are discussed in Section 3. Section 4 presents the results and Section 5 provides some interpretation.

2. THE ISSUES

2.1. The Principle

The transfer principle (principle of transfers) has been extensively discussed in the inequality literature (see, for example, Amiel & Cowell, 1998, 1999) and so we only provide a brief sketch of the essentials here.

Pigou (1912) argued that, in a *two-person community*, an income transfer from the richer to the poorer person that did not reverse their positions would unambiguously increase social welfare; but he left open the questions raised by a community of more than two persons. However, Dalton (1920) built on this idea and argued that an appropriate principle for inequality measurement would be that, in an n-person community, an income transfer from any person to someone poorer (that did not switch these person's positions in the distribution) must reduce inequality. So if $(x_1, x_2, \ldots, x_i, \ldots, x_j, \ldots, x_n)$ represents a vector of incomes arranged in ascending order then the vector $\big(x_1, x_2, \ldots, x_i + \delta, \ldots, x_j - \delta, \ldots, x_n\big)$ must represent a distribution with lower inequality, as long as $\delta > 0$ and $x_i + \delta < x_j - \delta$.

This is clearly a much stronger position than that originally taken by Pigou and is one that has been challenged by some of the recent empirical literature. It forms the focus of the present study.

2.2. Research Questions

The heterodox views reported in the studies cited in the introduction are not the result of haphazard responses to questionnaire experiments. Consistency checks have been carried out by the authors in a number of studies and the phenomenon has been replicated by the works of others (Ballano & Ruiz-Castillo, 1993; Beckman, Cheng, Formby & Smith, 1994). Some of the main routes for further investigation for an explanation of the phenomenon are:

(1) Whether the economic context has an influence on perceptions of distributional comparisons. (Cf. for example, the investigation of the present authors in the context of social welfare, poverty and risk Amiel & Cowell, 1994, 1997, 2002.)
(2) Whether the position in the distribution plays a part in determining the perception of a specific pure richer-to-poorer as inequality-reducing. (See for example, the discussion in Amiel & Cowell, 1998.)
(3) Whether specific personal characteristics predispose an individual to respond in an orthodox or heterodox fashion. For example, it might be argued that if people had been properly educated in mainstream economics then they "ought to" respect the principles such as Dalton's transfer principle.

The present paper focuses on a combination of points 2 and 3 above. In addressing the issue "why do people violate the transfer principle?" we raise a number of supplementary questions: Could this be the outcome of different economics education? What is the role of other personal attributes such as age and gender? The motivation for these further questions is not hard to find. Previous researchers have investigated whether economics as a discipline makes people selfish (Eckel & Grossman, 1998) or whether females are more or less selfish than males (Frey & Meier, 2001). Our own work has drawn attention to the importance of an economics background in shaping responses to questions about distributional comparisons (Amiel & Cowell, 1994). So it is reasonable to raise the question of whether these factors might play a significant part in determining whether or not inequality rankings conform to the transfer principle.

3. THE EXPERIMENT

3.1. The Questionnaire Study

The general approach used here is fully described in Amiel and Cowell (1999) – an anonymous questionnaire experiment carried out in supervised sessions with unpaid students. The specific questionnaire used in this is study of the "A4" type described in Amiel and Cowell (1999) but supplemented with a number of personal questions. The questionnaire is reprinted in full as the appendix to this paper. However, a brief summary of the format is as follows:

- *Scene setting*: A brief narrative is provided in order to describe to respondents the distributional comparisons on which they are asked to give their views.
- *A series of five linked numerical questions*: In each question a pair of income vectors is presented to the respondent who is invited to say which of the two

appears to be more unequal. Respondents can also indicate that they find the two distributions equally unequal.

- *A multi-option verbal question*: The issue presented is closely to that underlying the set of numerical questions. Respondents are allowed to choose more than one of the options.
- *Review*. Respondents are given the option of reviewing their responses to the numerical questions in the light of the reasoning of the verbal question.
- *Background and circumstances*: A number of questions about personal attributes and circumstances including family income, age, gender, area of domicile.

3.2. The Data

The set of questionnaire experiments was run in 17 universities in the United States during 1994. The student respondents were all undergraduates who had not yet been taught the conventional welfare-analytic approach to income distributions and the data were collected during a normal class or lecture session. The breakdown of the combined sample is as in Table 1.

The right-hand column of Table 1 is a simple ranking indicator derived from The Lombardi Program on Measuring University Performance (2001) which presents

Table 1. Composition of the Sample.

	Sample Size	Rank
American University, Washington, DC	36	10
Arizona State University	14	11
Boston College, Chestnut, MA	13	5
Brigham Young University	13	6
California State University, Hayward	22	9
Columbia University	15	1
East Carolina University, Greenville, NC	72	16
Georgia State University, Atlanta	47	15
New York University, New York	16	2
SUNY Binghamton, Binghamton NY	24	8
Southern Illinois University, Carbondale	18	14
Syracuse University	42	7
Texas AM University, College Station	47	17
University of Arkansas, Fayetteville	45	12
University of Maryland, College Park	14	4
University of Wisconsin, Madison	24	3
Western Michigan University	49	13
Total	510	

categorisations of universities into groups by nine measures based on research performance, teaching quality, and funding in various departments of research, amongst many others. The method used to categorise the universities reveals that there is no single rank-ordered list, and the ranking used for this study has used a number of such orderings to compile this consensus ranking. It would be interesting to know whether being a student enrolled at a higher-ranked institution is a factor predisposing a respondent to make more conventional responses when comparing income distributions.

3.3. The Model

Using the standard theory we can specify a particular response pattern or patterns that should be expected to emerge from individual responses if people (a) were to have views consistent with the theory and (b) were to respond according to their beliefs. We can imagine the probability of this pattern emerging as being determined, in part, by individuals' attributes, the type of question and so on. So it would be appropriate to specify a model of the following form:

$$\Pr(\text{Response Pattern}) = \Phi(b_1 y_1 + b_2 y_2 + \cdots + b_k y_k) \tag{1}$$

where the y variables are personal characteristics; the b-coefficients can then be estimated using a standard technique such as probit. From the data set we have the following data on personal circumstances:

- `faminc` – family annual dollar income in one of five income intervals. In contrast to previous studies (such as Amiel & Cowell, 2002) all the respondents came from the the same country so that we were able to include a variable in actual currency units, rather than just a categorical variable.
- `eth` – ethnicity categorisation. The original data provided six categories, but some of these were sparsely populated. These were simplified to three: White = 1, Asian-Pacific = 2, Black, Hispanic and the rest = 3.
- `sex` – a gender indicator: male = 1 and female = 0.
- `geog` – an indicator of the geographical area from which the respondent came: Northeast = 1, Southeast = 2, Midwest = 3, Southwest = 4, Northwest = 5, West = 6.
- `sib` – number of siblings.
- `age` – age in years.
- `res` – residence picks up a possible "home university" effect. It takes the value 1 if the person attends university in the same state as he/she originates from.

- `unirank` – university rankings. These are specified in the right-hand column of Table 1.

The "Response Pattern" on the left-hand side of (1) will depend on the way in which the distinction between orthodox and heterodox views of distributional comparisons is specified. This in turn depends on the way conformity with the transfer principle is to be interpreted. The possible interpretations include:

(1) Conformity on the various questions, taken separately.
(2) Conformity on the collection of numerical questions jointly, and on the verbal question separately.
(3) Conformity with all of the questions taken jointly.

Clearly interpretation 3 is very demanding; but the other interpretations may also have merit. We will discuss these further in Section 4.

4. RESULTS

4.1. Numerical Questions: Summary

We focus first on the five questions that are in the form of a simple comparison of two vectors, A and B. The overall pattern of responses is given in Table 2 for the whole sample and males and females separately. In each case the row labelled A or B respectively gives the proportions of the sample who responded that A or B

Table 2. Numerical Questions: Summary Responses.

	Q1 (%)	Q2 (%)	Q3 (%)	Q4 (%)	Q5 (%)
All data ($N = 511$)					
A	33.9	54.0	47.0	45.4	54.4
B	19.8	18.2	20.5	38.7	23.3
$A\&B$	44.2	25.8	30.3	14.3	19.6
Males ($N = 299$)					
A	36.5	56.2	49.2	46.8	51.5
B	21.1	20.7	23.4	38.5	27.4
$A\&B$	41.1	21.7	25.8	14.0	18.7
Females ($N = 210$)					
A	29.5	50.5	43.3	42.9	58.1
B	18.1	14.8	16.7	39.5	17.6
$A\&B$	49.0	31.9	37.1	14.8	21.0

Note: Two respondents did not identify themselves by gender.

respectively is the distribution with the higher inequality; the row labelled *A*&*B* gives the proportions of the sample who responded that the two distributions are equally unequal.

Two things immediately stand out:

(1) The pattern of conventional responses differs markedly as between questions: the proportion of *A*-responses is higher for questions 2 and 5 where the implied transfer involves the richest person or the poorest person or both.
(2) With the exception of question 5 males appear to be more disposed than females to giving a response that accords with the conventional view on distributional comparisons.

Furthermore only 9.8% of the sample responded *A* to all five questions (10.4% of males, 8.1% of females).

4.2. Numerical Questions: Regression Analysis

Scrutinising the tabular break-down of responses by characteristics is suggestive but not decisive. In order to disentangle the impact of the various personal factors available in the data we use the econometric model (1) outlined above. Applying this to the numerical problems alone it is clear that – as we noted in Section 3.3 – there are two principal ways in which we can interpret the idea of an orthodox response pattern. These correspond to the two leading specifications for the equations for numerical responses alone:

(1) We can treat the observations as five separate unrelated responses (each of which could be *A* or *B* or *A*&*B*). This gives us in principle a single equation to be fitted to 5*N* observations where *N* is the sample size. An "orthodox response" is then just a simple *A* in any one of these 5*N* observations.
(2) We can take the observations as a single response pattern in which only *AA* . . . *A* is consistent with economic orthodoxy. Clearly, there are just *N* observations here.

We deal with each of these two cases in turn. Running the standard model of Prob(Response *A*) for the numerical questions on the 5*N* observations (case 1) gives the pattern reported in Table 3.[1] Column 5 (labelled $P > |z|$) gives the probability that the true value of the coefficient is greater than the critical value *z* given in column 4. Clearly `sex`, `geog` and `unirank` are all significant at the 5% level. Agreement with the transfer principle, solely on the basis of the numerical questions (case 2 described above) is summarised in Table 4. Again it is clear that `sex` and `unirank` are significant at the 5% level.

Table 3. Numerical Questions: Probit Regression of Response *A*.

	Coef.[a]	Std. Err.	*z*	$P > \|z\|$	[95% Conf. Int]	
faminc	0.0311	0.0202	1.54	0.124	−0.0086	0.0707
eth	−0.0590	0.0355	−1.66	0.097	−0.1286	0.0107
sex	0.1338	0.0395	3.38	0.001	0.0562	0.2113
geog	0.0503	0.0142	3.53	0.000	0.0224	0.0782
sib	−0.0110	0.0136	−0.81	0.417	−0.0376	0.0156
age	0.0055	0.0056	0.99	0.324	−0.0054	0.0164
res	−0.0389	0.0550	−0.71	0.480	−0.1467	0.0690
unirank	−0.0113	0.0057	−1.98	0.048	−0.0224	−0.0001
const	−0.2947	0.1874	−1.57	0.116	−0.6620	0.0725

Note: Number of obs = 2550.
[a]Marginal effects.

On either interpretation we have a clear message. The university ranking coefficient implies that the higher-ranked the university from which the student respondent comes, the more likely is he/she to respond in an orthodox fashion; furthermore the probability of responding in this fashion is higher if the person is a he rather than a she.

This message is broadly confirmed when we examine what is going on in each of the questions 1–5. A summary of the results of the regressions for individual numerical questions is presented in Table 5. Only `sex`, `geog` and `unirank` are ever significant; none of these is significant for the responses to question 5.

Table 4. Numerical Questions: Probit Regression of Agreement with Transfer Principle.

	Coef.	Std. Err.	*z*	$P > z$	[95% Conf. Int.]	
faminc	0.0610	0.0476	1.28	0.201	−0.0324	0.1543
eth	−0.1148	0.0845	−1.36	0.174	−0.2805	0.0509
sex	0.2219	0.0967	2.30	0.022	0.0324	0.4114
geog	0.0594	0.0326	1.82	0.068	−0.0045	0.1233
sib	−0.0384	0.0317	−1.21	0.225	−0.1005	0.0237
age	0.0041	0.0130	0.32	0.753	−0.0214	0.0296
res	−0.1516	0.1276	−1.19	0.235	−0.4017	0.0984
unirank	−0.0261	0.0132	−1.99	0.047	−0.0519	−0.0004
const	−0.4652	0.4377	−1.06	0.288	−1.3230	0.3926

Note: Number of obs = 510.

Table 5. Probit Regression of Response "A" for Individual Questions.

	Q1	Q2	Q3	Q4	Q5
faminc	0.0610	0.0310	0.0377	0.0450	−0.0122
	0.0476	0.0453	0.0455	0.0453	0.0450
eth	−0.1148	−0.1009	−0.0046	−0.1288	0.0406
	0.0845	0.0799	0.0795	0.0806	0.0792
sex	0.2219**	0.1812*	0.1881*	0.1261	−0.0019
	0.0967	0.0966	0.0970	0.0865	0.0789
geog	0.0594*	0.0758**	0.0641**	0.0318	0.0290
	0.0326	0.0329	0.0320	0.0320	0.0319
sib	−0.0384	−0.0153	−0.0144	0.0401	−0.0311
	0.0317	0.0306	0.0305	0.0306	0.0303
age	0.0041	0.0158	−0.0077	0.0099	0.0059
	0.0130	0.0128	0.0125	0.0125	0.0124
res	−0.1516	−0.1083	−0.0113	0.1223	−0.0562
	0.1276	0.1235	0.1234	0.1242	0.1231
unirank	−0.0261**	−0.0028	0.0015	−0.0280**	−0.0036
	0.0132	0.0129	0.0127	0.0128	0.0127
const	−0.4652	−0.3909	−0.3210	−0.3895	0.0427
	0.4377	0.4272	0.4219	0.4219	0.4170

Note: Number of observations for each column = 510. Standard errors in small type below coefficient estimate.

*Significant at 10% level.

**Significant at 5% level.

4.3. Verbal Question

Question 6 provides an interesting additional perspective on the transfer principle. By asking respondents to reason verbally on issues to do with income distribution and then inviting them to change their numerical responses in the light of this reasoning we have tried to ensure that unorthodox responses are not the result of a trick from the particular pattern of income distributions presented in the numerical questions.

The summary of responses for the verbal question (question 6) by gender breakdown is given in Table 6. Rows *a*–*e* give the percentage who responded by checking just one of the five options *a*–*e* respectively. Note that the option that corresponds to the pure transfer principle is *d*. Given that the respondents were allowed to make multiple selections it is of interest also to see the proportion who indicated a *d*-response combined with something else (the row *d*-comb) and the proportion who indicated other combinations.

Table 6. Verbal Question: Summary Responses.

	All Data (%)	Males (%)	Females (%)
a	12.3	13.0	11.4
b	28.0	29.1	26.2
c	15.1	10.7	21.4
d	25.4	28.1	21.9
e	13.1	12.4	13.8
d-comb	1.6	1.7	1.4
other comb	2.9	2.7	3.3
N	511	299	210

In interpreting this table the following should be noted:

- Option *b* is the "favourite": this is the case where the person views a rich*est* to poor*est* transfer as inequality reducing but is non-committal about richer-to-poorer transfers elsewhere in the distribution. This is, broadly, the Pigou position.
- The pure transfer principle (row *d*) is the next most favoured response.
- Again the proportion of males responding *d* is higher than the proportion of females.
- There is a small number of *d*-combination responses. Taken together with the "pure *d*" case they amount to about the same proportion as the *b* responses.

However, in this case, the apparent gender split may be misleading. The regression results for the verbal question are presented in Table 7 (for the "pure" response *d*) and Table 8 (for the *d* and *d*-combination responses together). The story is now different from the numerical responses and, arguably, even simpler: only `eth` is significant.

Table 7. Verbal Question: Probit Regression of Response "*d*."

	Coef.[a]	Std. Err.	*z*	$P > \|z\|$	[95% Conf. Int.]	
faminc	0.0389	0.0492	0.79	0.430	−0.0576	0.1353
eth	−0.1881	0.0917	−2.05	0.040	−0.3679	−0.0083
sex	0.0335	0.0870	0.38	0.700	−0.1371	0.2041
geog	0.0321	0.0347	0.92	0.356	−0.0360	0.1002
sib	0.0238	0.0328	0.73	0.468	−0.0405	0.0881
age	−0.0173	0.0139	−1.24	0.214	−0.0446	0.0100
res	0.0224	0.1349	0.17	0.868	−0.2420	0.2868
unirank	−0.0122	0.0137	−0.89	0.372	−0.0390	0.0146
const	−0.2570	0.4620	−0.56	0.578	−1.1625	0.6485

Note: Number of obs = 510.
[a]Marginal effects.

Table 8. Verbal Question: Probit Regression of Response "d" and d-Combinations

	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf. Int.]	
faminc	0.0461	0.0493	0.9400	0.3500	−0.0505	0.1426
eth	−0.2057	0.0918	−2.2400	0.0250	−0.3857	−0.0257
sex	0.0446	0.0859	0.5200	0.6030	−0.1237	0.2129
geog	0.0444	0.0332	1.3400	0.1810	−0.0206	0.1094
sib	0.0261	0.0324	0.8000	0.4210	−0.0374	0.0896
age	−0.0063	0.0123	−0.5200	0.6060	−0.0304	0.0177
res	0.0436	0.1362	0.3200	0.7490	−0.2234	0.3107
unirank	−0.0072	0.0128	−0.5600	0.5740	−0.0323	0.0179
cons	−0.5909	0.4735	−1.2500	0.2120	−1.5191	0.3372

Note: Number of obs = 510.

4.4. The Overall Picture

Finally, we ought to examine the overall pattern of responses, both numerical and verbal. Note that according the strictest interpretation very few responded in conformity with the transfer principle – just 4.3% of the overall sample (5.0% of males 3.3% of females) gave the response *A* to each of the five numerical questions and *d* to the verbal question.

In the light of the discussion in Section 3.3 it would be useful to consider both the regression model where the orthodox/heterodox response to each question is taken as a separate observation and that where the entire pattern (*AAAAAd*) is considered against all other combinations. The results are in Tables 9 and 10 respectively: clearly `eth`, `sex`, `geog` and `unirank` are significant in the former interpretation, but nothing is significant (even at the 10% level) in the latter.

Table 9. Probit Regression of Orthodox Responses: All Questions Separately.

	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf. Int.]	
faminc	0.0315	0.0186	1.70	0.090	−0.0049	0.0679
eth	−0.0741	0.0327	−2.27	0.023	−0.1381	−0.0100
sex	0.1101	0.0337	3.26	0.001	0.0440	0.1762
geog	0.0459	0.0130	3.53	0.000	0.0204	0.0713
sib	−0.0057	0.0124	−0.46	0.644	−0.0301	0.0186
age	0.0025	0.0051	0.48	0.630	−0.0075	0.0125
res	−0.0300	0.0505	−0.59	0.552	−0.1289	0.0689
unirank	−0.0111	0.0052	−2.14	0.032	−0.0213	−0.0009
const	−0.2954	0.1716	−1.72	0.085	−0.6317	0.0408

Note: Number of obs = 3060.

Table 10. Probit Regression of Orthodox Responses: All Questions Jointly.

	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf. Int.]	
faminc	0.0350	0.2649	0.13	0.895	−0.4842	0.5542
geog	−0.1935	0.2871	−0.67	0.500	−0.7562	0.3692
sib	−0.3019	0.3269	−0.92	0.356	−0.9426	0.3388
age	0.0266	0.0765	0.35	0.728	−0.1233	0.1765
res	−0.4166	0.5760	−0.72	0.470	−1.5456	0.7124
unirank	−0.0144	0.0753	−0.19	0.848	−0.1619	0.1331
const	−1.9334	2.3789	−0.81	0.416	−6.5959	2.7291

Note: Number of obs = 221. Because of multicollinearity several observations had to be dropped.

5. DISCUSSION

First we note what may be oddity. In several of the models the coefficient on `geog` is positive and significant – so if you are from the Midwest or West you are more likely to respond in line with economic orthodoxy. But this has nothing to do with going to university in your home state – `res` is not significant in any of the models. We also note that the background variables `faminc`, `sib`, and `age` are never significant.

There may be an issue of the interpretation of and reasoning in the English language. This is a possible explanation of the significance of the variable `eth` in the case of the verbal question, and *only* with the verbal question.

However, the most striking thing about the results is their consistency with broad conclusions that have emerged from other research. Specifically:

- In most of the numerical questions we have the same phenomenon as was evident the inequality-risk experiment reported in Amiel and Cowell (2002) – being male means that you are more likely to answer in an orthodox fashion to numerical questions about distributional comparisons. On the other hand gender turns out not to be significant when the issue is put in verbal form.
- The teaching of economics appears to be a significant factor in people's interpretation of distributional comparisons. Here the significance of the negative coefficient on `unirank` may have an intuitive and appealing explanation. For this variable the lower numbers mean higher ranking and so it is not surprising that higher-ranking universities answer more in line with orthodoxy.

Whether this indicates brainwashing by the profession is of course a moot point.

NOTE

1. One respondent left all numerical answers blank and so was dropped from the regression analysis.

ACKNOWLEDGMENTS

We are grateful to Sanghamitra Bandyopadhyay for computational assistance and to a referee for helpful comments.

REFERENCES

Amiel, Y., & Cowell, F. A. (1992). Measurement of income inequality: Experimental test by questionnaire. *Journal of Public Economics*, *47*, 3–26.

Amiel, Y., & Cowell, F. A. (1994). Income inequality and social welfare. In: J. Creedy (Ed.), *Taxation, Poverty and Income Distribution* (pp. 193–219). Edward Elgar.

Amiel, Y., & Cowell, F. A. (1997). The measurement of poverty: An experimental questionnaire investigation. *Empirical Economics*, *22*, 571–588.

Amiel, Y., & Cowell, F. A. (1998). Distributional orderings and the transfer principle: A re-examination. *Research on Economic Inequality*, *8*, 195–215.

Amiel, Y., & Cowell, F. A. (1999). *Thinking about inequality*. Cambridge: Cambridge University Press.

Amiel, Y., & Cowell, F. A. (2002). Attitudes towards risk and inequality: A questionnaire-experimental approach. In: F. Andersson & H. J. Holm (Eds), *Experimental Economics: Financial Markets, Auctions, and Decision Making* (Chap. 9, pp. 85–115). Dewenter: Kluwer.

Ballano, C., & Ruiz-Castillo, J. (1993). Searching by questionnaire for the meaning of income inequality. *Revista Española de Economía*, *10*(43), 233–259.

Beckman, S., Cheng, D., Formby, J. P., & Smith, W. J. (1994). Preferences for income distributions and redistributions: Evidence from experiments with real income at stake. Technical report, University of Colorado – Denver and University of Alabama.

Dalton, H. (1920). Measurement of the inequality of incomes. *Economic Journal*, *30*(9), 348–361.

Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men?: Evidence from dictator experiments. *Economic Journal*, *108*, 726–735.

Frey, B. S., & Meier, S. (2001). Political economists are neither selfish nor indoctrinated. Working Paper, Institute for Empirical Economic Research, University of Zürich, Blümlisalpstrasse 10, 8006 Zürich.

Pigou, A. C. (1912). *Wealth and welfare*. London: Macmillan.

The Lombardi Program on Measuring University Performance (2001). Top American universities. Annual report, The Center, University of Florida, http://thecenter.ufl.edu/research2001.pdf.

APPENDIX: QUESTIONNAIRE

Following is the questionnaire used in the experiment reported in the main text.

Income Inequality Questionnaire

This questionnaire concerns people's attitude to income inequality. We would be interested in *your* view, based on hypothetical situations. Because it is about attitudes there are no "right" answers. Some of the possible answers correspond to assumptions consciously made by economists: but these assumptions may not be good ones. Your responses will help to shed some light on this, and we would like to thank you for your participation. The questionnaire is anonymous.

Alfaland consists of five persons who are identical in every respect other than their income. Two economic policy proposals A and B are being considered for implementation in Alfaland next year. It is known that – apart from their impact on personal incomes – the two policies would have the same effect on the population. The impact upon incomes would depend upon the particular state of the Alfaland economy at the time the policy (A or B) is to be introduced.

In each of the questions (1)–(5) two alternative lists of incomes A and B (in Alfaland local currency) are given. Each of these pairs represents the outcomes of the A-policy and the B-policy in each of five different situations in which Alfaland might find itself. In each case please state which policy you consider would result in higher *inequality* in Alfaland by circling A or B. If you consider that the two policies will result in the same inequality than circle both A and B.

(1) A = (2, 5, 9, 20, 30) B = (2, 6, 8, 20, 30)
(2) A = (2, 5, 9, 20, 30) B = (3, 5, 9, 20, 29)
(3) A = (2, 5, 9, 20, 30) B = (2, 6, 9, 20, 29)
(4) A = (2, 5, 9, 20, 30) B = (2, 10, 9, 15, 30)
(5) A = (10, 10, 10, 10, 30) B = (10, 10, 10, 20, 20)

In question 6 you are presented with a hypothetical income change and some possible views about that change. The views are labelled (a)–(e). Please circle the letter alongside the view that corresponds most closely to your own. You can check more than one answer, provided that you consider they do not contradict each other. Feel free to add any comment which explains the reason for your choice.

(6) Suppose we transfer income from a person who has more income to a person who has less, without changing anyone else's income. After the transfer the person who originally had more income still has more.

(a) Income inequality in this society has fallen if the ranking of the income of all the people remains the same. If there is any change in the rank of all the incomes then it is possible that income inequality increases or remains the same.
(b) If the transfer was from the richest to the poorest, and after the transfer the richest remains the richest and the poorest remains the poorest, than income inequality has fallen. In other cases we cannot say *a priori* how inequality has changed.
(c) The relative position of others has also been changed by the transfer. So we cannot say *a priori* how inequality has changed.
(d) Inequality in this society has fallen, even if there is a change in the ranking of the income of people as a result of this transfer, and even if the transfer is not from the richest in the society to the poorest.
(e) None of the above.

In the light of your answer to question 6, would you want to change your answer to question 1–5? If so, please state your new response here.

(1)
(2)
(3)
(4)
(5)

(7)	a. family income	1. below and up to 15,000	
		2. 15,000–25,000	
		3. 25,001–31,000	
		4. 31,001–59,999	
		5. 60,000–andup	———
	b. ethnicity	1. White,non-hispanic	
		2. Black, non-hispanic	
		3. Hispanic, Mexican	
		4. Hispanic, other	
		5. Asian or Pacific Island	
		6. Native American Indian	———
	c. sex	1. Female	
		2. Male	———
	d. geographic origin	1. Northeast	
		2. Southeast	
		3. Midwest	

4. Southwest
5. Northwest
6. West

e. number of siblings _____
f. age _____

EFFICIENCY, EQUITY AND DEMOCRACY: EXPERIMENTAL EVIDENCE ON OKUN'S LEAKY BUCKET

Steven R. Beckman, John P. Formby and W. James Smith

1. INTRODUCTION

Income redistribution and its consequences have been the subject of intense debate over the last three decades. This is nowhere better evidenced than in the motivations which are variously ascribed to such redistributions. The social welfare approach, for example, starts from the premise that redistribution of income from the rich to the poor enhances social welfare, a fact from which redistribution derives its motivation. In contrast Tullock (1983, p. 2) argues that the major impetus for income redistribution is simply that the beneficiaries of transfer programs want larger incomes and greater wealth and have the political power to realize their goals. Buchanan (1984, p. 187) finds a constitutional basis for redistribution arguing that, if the voting franchise is universal and the constitution allows collective decisions concerning income transfers, then the basic property right to income in a society inheres in the voting franchise.

One implication is clear; we should expect substantial income redistribution in democratic societies. And, in fact, such transfers have become so common that they now constitute a hallmark of modern democratic governments. However,

Inequality, Welfare and Income Distribution: Experimental Approaches
Research on Economic Inequality, Volume 11, 17–42

ISSN: 1049-2585/doi:10.1016/S1049-2585(04)11002-8

given that income and wealth distributions are positively skewed with the result that medians are substantially less than means, the extent of redistribution is substantially less than what might be expected if voters engaged in purely personal wealth maximizing political behavior.

Why do we not observe redistribution to an extent greater than we actually do? The answer is not at all obvious. One explanation may be that redistributions are not costless and that voters may be uncertain about their future position in the income distribution.[1] As a result, utility maximization may lead rational voters to refrain from some redistributions, not because of a general concern for efficiency, but because such reallocations hold the potential for adversely affecting positions in the income distribution which they may conceivably occupy in the future.

A second explanation suggests that equality and efficiency are both valued by individuals in general with neither taking precedence over the other. The question then becomes what is the rate at which voters will trade off efficiency for equity? Okun (1975) succinctly illustrates the issue with a simple conceptual experiment in which $4,000 on average is paid by the top 5% of American families to finance transfers to the bottom 20% of families.[2] If there are no inefficiencies, $1,000 could be transferred to each family. But Okun notes that such a redistribution

> has an unsolved technological problem: the money must be carried from the rich to the poor in a leaky bucket. Some of it will simply disappear in transit, so the poor will not receive all of the money that is taken from the rich. The average poor family gets less than $1,000, while the average rich family gives up $4,000. Suppose 10% leaks out; that leaves $900 for they average poor family instead of the potential $1,000. Should society still make the switch? If 50% leaks out? 75%? Even if 99% leaks out the poor get a little benefit; the $4,000 taken from the rich will yield $10 to each poor family. Where would you draw the line? (Okun, 1975, pp. 91–92).

Inspired by precisely this issue Amiel et al. (1999) ask respondents to a questionnaire what is the minimum amount that must be delivered to the poor if varying amounts are taken from someone that is rich. While there is some difference in subject pools, generally leakage rates of 20–30% are claimed by respondents as the maximum leakage rate tolerable. This work follows on closely related opinion surveys by Amiel and Cowell (1992, 1993, 1994) and Harrison and Seidl (1994) which shed some light on these issues by providing evidence concerning student attitudes toward income distributions. When asked to choose between alternative vectors of incomes substantial numbers of students report that they prefer distributions with less equality. This, of course, indicates disagreement with the Pigou-Dalton principle of transfers and suggests that voter preferences for income distribution may help explain why there is not as much redistribution as might be expected.[3] However, these surveys are subject to the well-known

shortcoming that what people say in response to hypothetical questions and what they actually do when income is at stake may be quite different. For example see Beckman et al. (2002).

In this paper we employ the methods of experimental economics to examine voter preferences regarding efficiency and redistribution. Subjects are asked to choose between two distributions of payoffs which taken together are the equivalent of income transfers from the top position to the bottom involving varying degrees of leakage from Okun's bucket. The nature of the experiment allows considerable control and permits individual choices to be studied both from behind the Rawlsian veil of ignorance and when the subjects know their positions in a payoff distribution before voting. Subjects were recruited from student volunteers in four locations, two in the U.S. (Denver and Tuscaloosa) and two in China (Guiyang and Tianjin).

The results turn out to be instructive. We find that voters are willing to support very large efficiency losses to transfer income if they do not have to pay and do not have any chance of obtaining the top positions. Support for redistribution, however, drops off markedly when voters are asked to contribute to the transfer or if they know their position before voting and they are at the top. Preferences for mandated redistribution thus are strongly conditioned by self interest and the opportunities which exist to occupy positions which receive the highest payoffs.

2. EXPERIMENTAL PROCEDURES

The objective of the experiment is to induce subjects to reveal their preferences on income redistributions involving varying degrees of efficiency loss. The experiment is conducted under two regimes: (1) when subjects make choices behind the Rawlsian veil of ignorance and do not know which positions they will eventually occupy in the distribution of payoffs; and (2) when subjects make choices knowing their positions with certainty beforehand. At the outset, we should point out that, unlike Rawls (1971), in using the veil of ignorance we make no normative judgements concerning what choices subjects ought to make. Instead, we use the veil as a positive device to introduce conditions of uncertainty about their future position in the income distribution. When making choices from behind the veil, the only knowledge subjects possess consists of the lists of payoffs, the mean value of the alternative payoff vectors and the probability of various outcomes. In the second setting, subjects know their position and, in four of five rounds, except for the richest and poorest, do not have a personal stake in the outcome.

We adopt a design in which subjects are members of a five person committee which uses a secret ballot and majority rule to choose between two ordered vectors of payoffs. In setting up the social choice aspects of the experiment it is important

for subjects to know that there are in fact other subjects in the experiment who will be paid. Thus members of each committee are placed in the same room during the experiment and are identified before the decisions are made and the votes announced. However, members of the committees are not permitted to communicate in any manner, precluding any opportunity for lobbying or logrolling. In an effort to obtain evidence across different cultures we repeated the experiments in four regional locations, two in the United States and two in China. One of the locations is in the Rocky Mountain West (Denver) and the second in the South (Tuscaloosa). In China one location is in the North (Tianjin) and the other in the South (Guiyang).

Multiple experiments were conducted at each location. Subjects were recruited from lower level college courses and possessed only limited exposure to economics. Moreover, the structure of the experiment minimized any influence such exposure may have had. The terms "equality," "income distribution," "leaky bucket" and "efficiency" were never used.

In the U.S. the subjects all attended large, state supported universities, the University of Colorado-Denver and the University of Alabama-Tuscaloosa, located approximately 1500 miles apart. In China, subjects were drawn from more diverse universities, Guizhou Minority University in the South and Nankai University in the North, which are also approximately 1500 miles apart.

In any given experiment 10 subjects were divided into two committees of five people. Instructions, included as an appendix, were read and a practice round conducted.[4] Each experiment contained 20 rounds of choices. In each round subjects were asked to vote for one of two alternative lists of payments. The list which garners majority support is used to make payments to subjects. For half of the rounds, subjects know their position and the payment which applies to them before voting. For the remaining rounds the assignment of positions is made after votes are taken and the majoritarian outcome announced. The same list of payment schedules are used under the two regimes to allow comparisons between voting behavior behind the Rawlsian veil of ignorance and when positions in the ordered payoff vector have been assigned.

The experimenter determines positions by shuffling 5 cards, ace through 5, fanning them out face down and walking before the subjects. The subjects point to one card, which the experimenter places face up in front of the subject. The highest payoff goes to the five and the lowest to the ace. Positions are assigned in this way during every round so that over the course of 20 rounds each subject is likely to earn relatively high and relatively low pay. For half the subjects in each location, 10 rounds of an experiment are conducted behind the veil followed by 10 rounds with positions known. For the other half of the subjects the order is reversed to test for order effects.

Four experiments involving 10 subjects each were conducted in Tuscaloosa and Denver. Thus, there are a total of 80 U.S. subjects. In China, two experimental sessions involving 20 subjects each were conducted in Guiyang and Tianjin. However, as explained below, two separate experiments involving the same subjects were carried out at each China session, so there are a total of eight China experiments, but only 40 subjects.

Subjects were not told in advance how much they would be paid for participating in the experiment, but were informed that it would be enough to keep their interest. The experiment lasted approximately one hour. In the U.S. payments averaged $18, just under four times the legal U.S. minimum wage. For one hour's work in the experiment U.S. subjects received on average about one half the amount that young, unskilled workers could earn in a day. In China the low per capita income and cost of living provided the opportunity to examine the effects of the size of monetary payments on revealed preferences concerning income distribution and redistribution. To investigate this issue we followed the general procedures of Katchelmeier and Shehata (1992) and used two payoff schemes, which are referred to below as "low pay" and "high pay." In the low pay experiment we converted U.S. dollars to Chinese yuan at the official exchange rate and paid one tenth of the amount we paid in the U.S. This resulted in payments that were equivalent in value to several days' wages of young, unskilled workers. At the end of the low pay experiment subjects were asked if they would like to repeat the experiment with payoffs 10 times larger. All subjects readily agreed. In the high pay experiment subjects received the equivalent of more than a month's normal earnings by a young, unskilled worker.

The first four rounds involve leaky bucket experiments in which income is transferred from the top income position to the bottom. In round 1 there is no leakage, and 100 points are transferred.[5] Round 2 transfers 200 points and 100 of those leak out with the result that the bottom position receives only 100 points. Thus, Okun's rate of leakage is 50%. Round 3 transfers 300 points from the top position and 200 points leak out, a rate of 67%. Round 4 transfers 300 points but the bottom position in the income distribution receives none of it, a leakage from Okun's bucket of 100%. Round 5 transfers 80 points to the bottom position with no leakage as in round 1 but with the difference that now all positions above the lowest must contribute in a progressive fashion. The last five rounds transfer points between adjacent positions with zero efficiency loss. They are included to construct an index of preferences for equality.

When positions are assigned before votes are taken and subjects are motivated entirely by self interest they need only examine the payoffs to their position in distributions A and B and vote for whichever is larger. But in each round there are either three or four positions that have the same earnings in the two payment

schedules. For subjects whose payoff is the same in both A and B, self interest predicts indifference concerning how the committee's income is distributed. To allow for this possibility subjects were permitted to abstain from voting by marking an X.[6] If subjects choose to vote, rather than abstain, their votes will reflect a true preference and concern over the rewards given to others and not just forced choices. Therefore, if positions are assigned before the vote is collected, subjects are in one of two possible situations, some of the subjects face a choice where their direct self interest is involved and others where only the interests of some other subjects who are members of their committee are affected.

3. EXPERIMENTAL RESULTS

Figure 1 reports results from rounds 1 through 4 conducted behind the veil of ignorance. The numbers in the charts for China reflect votes and abstentions in the two locations and are separated into high pay and low pay. Examining Fig. 1 and reading across each row and then comparing rows reveals a very clear and consistent pattern: regardless of country, location or high or low pay, support for the

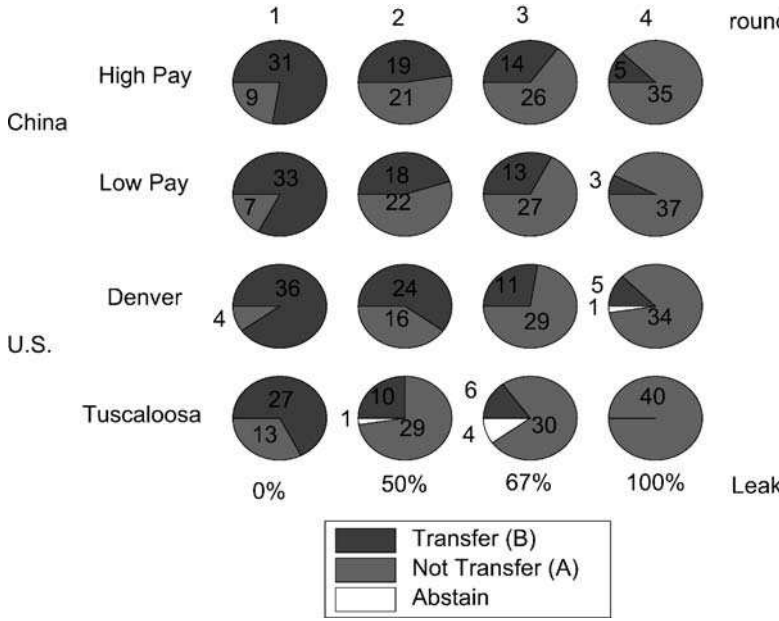


Fig. 1. Tests of the Leaky Bucket (Behind the Veil of Ignorance).

income transfer monotonically declines as efficiency losses rise. In the first round, in which 100 points are transferred from the top to the bottom with zero efficiency loss, all locations vote for the transfer by substantial majorities ranging from 90% in Denver to 67.5% in Tuscaloosa with the Chinese results lying intermediate between the two. In the next round (50% efficiency loss), Denver again approves the transfer by a majority of 60%, but the other three locations reject although in China the margins of defeat are small, 2.5% and 5%, for high and low pay groups respectively. Subjects in Tuscaloosa in contrast reject the transfer by a large majority, 74.5%. In the third round (67% efficiency loss) all locations reject the transfer with large majorities ranging from 83% in Tuscaloosa to 65% among the high pay group in China. Finally, in the fourth round (100% efficiency loss), all locations again reject but by larger majorities, ranging from unanimity in Tuscaloosa to 87% in Denver.

Figure 2 reports results for the same rounds when positions are known before votes are cast. The monotonic inverse relationship between inefficiency and voter support, readily apparent behind the veil, is no longer in evidence. In addition, there is a marked shift toward greater inefficiency. Rounds 2 and 3 (efficiency losses of 50% and 67% respectively), previously rejected by voters behind the veil

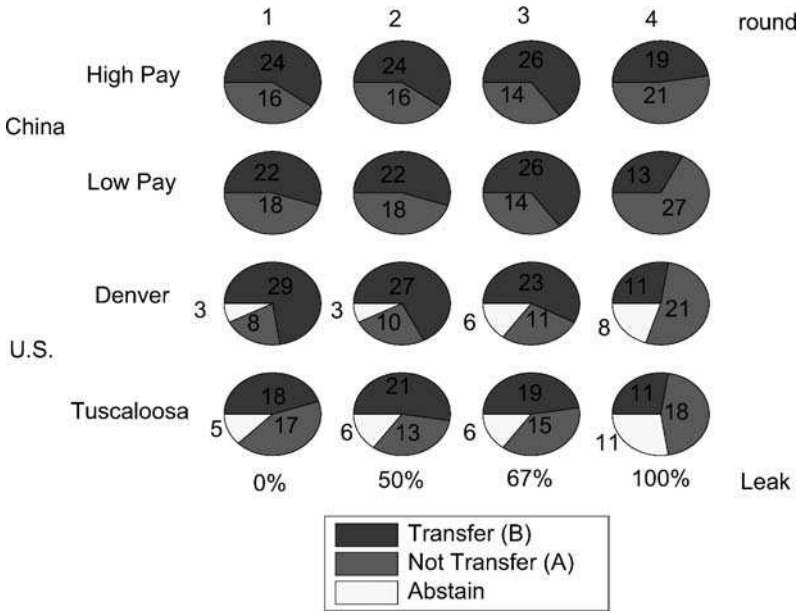


Fig. 2. Tests of the Leaky Bucket (Positions Known, All Votes).

except in Denver at the 50% level, now pass in all locations, U.S. and China with large majorities. Only the 100% efficiency loss draws strong enough opposition to bring about its defeat.[7] But even in this case, the shift toward support for greater inefficiency is striking. Behind the veil, the percentages of voters supporting the 100% efficiency loss range from 0 to 12.8%. With positions known, these percentages rise to 32.5–47.5%. Abstentions on a much larger scale than from behind the veil also surface in the U.S. In general, for the U.S., abstentions rise with efficiency losses. No abstentions are recorded in China.

The contrast between when positions are known before the vote is taken and when they are not is instructive when attempting to sort out motives underlying voting behavior. When their payoffs are affected and positions are known, individuals nearly always vote their self interest. In fact, we could find only 9 instances out of a possible 224 in which they did not (6 of the these are in one location, Denver, with none in Tuscaloosa and 3 in China).[8] Self interest is thus a very strong predictor of voting behavior. It however does not come into play when individuals' payoffs are unaffected by which vector is selected. Focusing on this group allows us to examine behavior when self interest is not at stake.

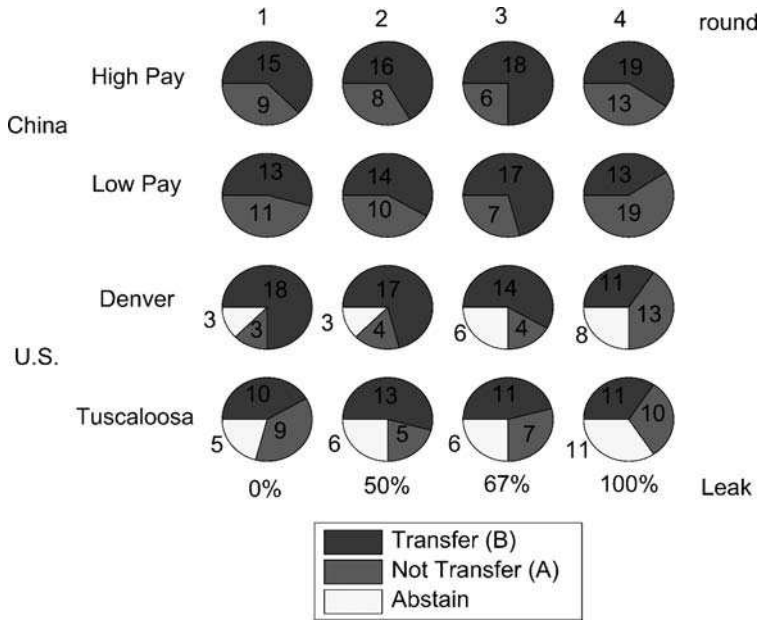


Fig. 3. Tests of the Leaky Bucket (Positions Known and Pay Unaffected).

Figure 3 reports the results.[9] Again reading across rows, we find that now only in Denver does support for the transfer decline monotonically with efficiency losses. In the other locations, within the first three rounds, support for transfers actually increases the greater the efficiency loss. The most surprising result, however, is that the 100% deadweight loss, soundly defeated in all locations behind the veil, is now approved by voters in Tuscaloosa and high pay China. Further, a large number of subjects in Denver and China low pay (40.6% and 45.8% respectively) also support the redistributions. These results suggest that when their self interest is not engaged significant numbers of subjects fail to support the Pareto principle, even though it would be difficult to construct a more straightforward opportunity for them to do so.

The differences in voting behavior appear to be quite significant; the question however is whether they are statistically significant. Table 1 reports results from

Table 1.

	0% (Round 1)	50% (Round 2)	67% (Round 3)	100% (Round 4)
(a) Behind the veil of ignorance				
Guiyang	33/40 (1)	26/40 (2)	18/40 (3)	4/40 (4)
Tianjin	31/40 (1)	11/40 (2)	9/40 (3)	4/40 (4)
Demver	36/40 (1)	24/40 (2)	11/40 (3)	5/39 (4)
Tuscaloosa	27/40 (1)	10/39 (2)	6/36 (3)	0/40 (4)
Sum of ranks	4	8	12	16
Page's $L = 4 + 2(8) + 3(12) + 4(16) = 120$, the 0.001 critical value is 117 and the probability of $L > 120$ given no treatment effect is 0.0005.				
(b) Positions known (total votes)				
Guiyang	27/40 (2)	25/40 (3)	29/40 (1)	17/40 (4)
Tianjin	19/40 (3)	21/40 (2)	23/40 (1)	15/40 (4)
Demver	29/37 (1)	27/37 (2)	23/40 (3)	11/39 (4)
Tuscaloosa	18/35 (3)	21/34 (1)	19/36 (2)	11/40 (4)
Sum of ranks	9	8	7	16
Page's $L = 9 + 2(8) + 3(7) + 4(16) = 110$, the 0.05 critical value is 111 and the probability of $L > 110$ given no treatment effect is 0.08.				
(c) Positions known (pay unaffected)				
Guiyang	17/24 (2.5)	17/24 (2.5)	20/24 (1)	17/32 (4)
Tianjin	11/24 (4)	13/24 (2)	15/24 (1)	15/32 (3)
Demver	18/21 (1)	17/21 (2)	15/18 (3)	11/24 (4)
Tuscaloosa	10/19 (3)	13/18 (1)	11/18 (2)	11/21 (4)
Sum of ranks	10.5	7.5	7	15
Page's $L = 10.5 + 2(7.5) + 3(7) + 4(15) = 106.5$, the 0.05 critical value is 111 and the probability of $L > 106.5$ given no treatment effect is 0.26.				

Page's Non-Parametric Test for Ordered Alternatives (1963) to determine whether the results indicate significant ordered preferences for different degrees of leakage from Okun's bucket. Page's test assumes rows are independent; therefore the appropriate division is the two locations in China, Denver and Tuscaloosa. The upper number in each cell is the number of votes for the transfer and the bottom the total number of votes. The ratio is the proportion of votes for the transfers which is required for Page's test. The test also requires specification of a particular rank order within a block of data, as we have done in parenthesis. The rank orders are then summed within a column, multiplied by the column position and summed across the row. The resulting statistic is known as Page's L and is compared to a table of critical values for designs with the same number of treatments and blocks.

The first panel reports results from behind the veil. The sample Page's L for the data from behind the veil is 120 with a critical value of 117 at the 0.001 significance level. We reject the null hypothesis of no round effect in favor of the alternative of an ordered effect, that is, behind the veil there are ordered preferences for the redistributive transfers with zero being the most preferred and 100% the least preferred.

The second panel in Table 1 reports results for total votes when positions are known. The sample Page's L is 110 and the critical value 111 for the 5% significance level. We therefore fail to reject the null hypothesis of no treatment effect at the 5% significance level but would reject at 10%. The third panel reports results for position known, pay unaffected. Again we fail to reject the null hypothesis at the 5% level but this time at 10% as well. Thus, when positions are known no systematic ordered preferences on efficiency losses is in evidence.[10]

What motivations which underlie these results? On the basis of Fig. 1, it could be reasonably argued that there exists a preference in general for equality and efficiency which is evident in every location as Okun surmised. As inefficiency rises, the willingness to transfer income becomes accordingly less.

This interpretation would seem to be, at least in part, belied by results when positions are known and pay is unaffected as shown in Fig. 3. Here, when no gain or loss is at stake, results markedly different from those behind the veil emerge. In the U.S., only Denver moves in a direction consistent throughout with the previous explanation, and in Tuscaloosa, the magnitude of the efficiency loss does not appear to impact the results in the expected manner at all. In both Denver and Tuscaloosa, however, there exist significantly larger numbers of individuals voting for the transfer in rounds 3 and 4 when positions are known than from behind the veil.

It is plausible that the explanation may lie in the fact that behind the veil there exists an equal chance of obtaining any position. This together with risk

aversion and a preference for a higher average payoff would provide a consistent explanation of these results. Self-interest, strong throughout, coupled with uncertainty and risk aversion certainly seem to be at play here.

Self interest, however, cannot explain the results in Fig. 3. When own pay is unaffected and positions are known, efficiency losses seem to have a positive value at least up to a point. One could argue that it is a new found concern for relative equality because round 4 is undeniably the most relatively equal. Unfortunately, this explanation is at odds with Amiel and Cowell's findings of widespread disagreement with the principle of transfers. It is also in disagreement with results of rounds, not reported here, which tested the same concepts as Amiel and Cowell.[11]

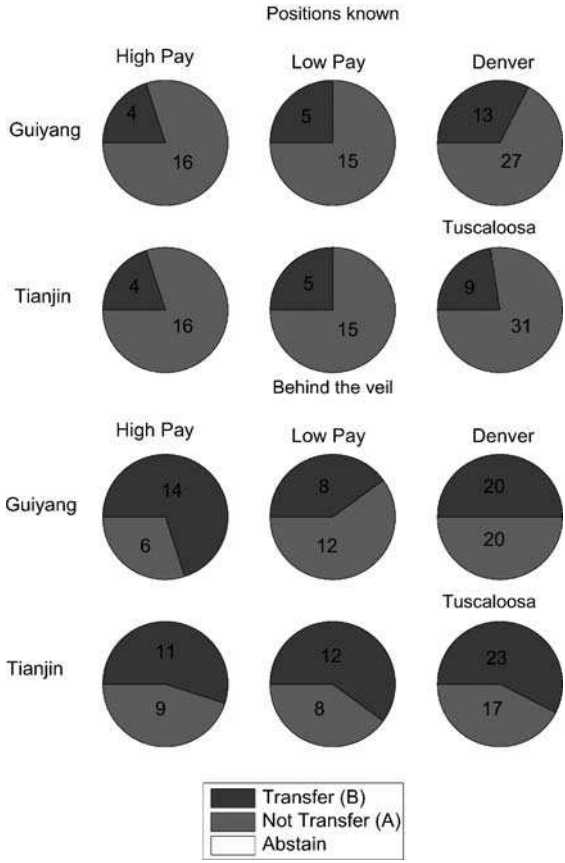


Fig. 4. Willingness to Pay for Transfers Round 5.

Envy would seem to be a plausible motivating factor and would, at least in part, account for the marked differences of results from behind the veil as opposed to known positions, pay unaffected.[12] It is important to inquire as to how strong these other motivations are relative to self interest. A final round sheds additional light on the issue.

In round 5, subjects are asked to choose between distributions A and B in which there is no efficiency loss. Round 5 differs from the previous four in that, if a transfer to the bottom takes place, all subjects in positions above the bottom must contribute with amounts increasing with income. Figure 4 summarizes the results. When positions are known, the transfer is resoundingly defeated in all locations by majorities ranging from 80% in both China high and low pay to 67.5% in Denver. This sharply contrasts with the results from round 1 in which the transfer was approved in all locations by margins ranging from 90 to 67.5%, even though both rounds have zero Okun leakages. It is clear then that redistribution garners much less support when subjects are asked to actually pay for it, even at "low" rates of contribution. Behind the veil, the margins of support for round 5 and transfers are very much smaller than in round 1 with low-pay China rejecting and Denver tied.

Table 2 reports results from chi-squared tests of the hypothesis that rounds 1 and 5 are equivalent. The hypothesis is easily rejected both from behind the veil and with positions known (in round 5 of course all subjects' payoffs are affected). Once positions are known, it is clear that support for the transfer is very little different than what pure self interest would predict. Under self interest, the bottom 20% would support the transfer while the transfer finds favor with only 25% of voters. This all suggests that "altruism" is much less prevalent among subjects when it comes at their expense than when it comes at another's. Alternatively, satisfaction from transferring income may not come so much from helping those in lower payoff positions as from soaking the rich.

Table 2. Will Subjects Support Taxes on the Middle Class to Help Others?

	Behind the Veil of Ignorance		Positions Known, All Votes	
	Transfer	Not to Transfer	Transfer	Not to Transfer
Round 1	127	33	93	60
Middle untaxed Round 5	88	72	40	120
Middle is taxed H_0: No change	$\chi^2 = 22$	$p = 3 \times 10^{-6}$	$\chi^2 = 41$	$p = 1.5 \times 10^{-10}$

Note: p is probability of a larger χ^2 given H_0.

4. LOGIT ESTIMATIONS

To further examine the motives underlying the observed voting behavior, we estimate a multinomial logit model using maximum likelihood techniques for the data from the first four rounds. The dependent variable takes on a value of zero if the subject voted for the less equal distribution, one if the subject chose greater equality, and two if the subject abstained. The models are estimated separately for the two regimes, behind the veil and positions known.

The variable LEAK is defined as the percent of efficiency loss in the round and takes on the values of 0, 0.5, 0.67 and 1. GAIN is the gain or loss of points which the subject will experience if the transfer passes and is only relevant when positions are known. GAIN is included to capture the role of self interest in voting behavior. In contrast, the variable EQUAL is the proportion of times[13] a subject votes for transfers in the last five rounds which involve zero efficiency losses when positions are known and payoffs are unaffected by the distribution chosen by the majority. EQUAL serves to isolate a subject's preference for equality because considerations of efficiency and self interest do not come into play in these rounds. The variable HIGH is a dichotomous variable that takes the value of one if the round occurred under high pay and zero if low pay. Location dummies (e.g. DENVER, TUSCALOOSA and TIANJIN) were included to test for differences relative to Guizhou in China. The variable PART2 tests whether the order of the regime (whether rounds from behind the veil or positions known were administered first) makes a difference. We entered GAIN, EQUAL and LEAK as quadratic forms to capture nonlinearities and tested for various interactions between LEAK, GAIN and EQUAL. In all specifications, a number of these variables were found to be insignificant, and, as a result, were dropped from subsequent estimations.

Table 3 reports results from three estimations for the rounds conducted behind the veil. Examining column I, LEAK, TUSCALOOSA and TIANJIN are significant at the 1% level. LEAK possesses a negative coefficient, indicating that an increase in efficiency loss lowers the probability of voting for the transfer.[14] The coefficients on TUSC and TIANJIN are significant and negative; voters in Tuscaloosa and Tianjin are less likely than voters in Guizhou to vote for a redistribution. All other variables were insignificant.

Columns II and III report results with nonlinearity and interaction terms included. None of these turn out to be significant. The three rows after the variables report log likelihood ratios, and significance tests on whether additional variables should be included. The indication is that additional variables should not be included at conventional levels of significance. Finally, from column I, the model predicts 488 of 632 total cases[15] correctly (77.2%).

Table 3. Multinomial Logit Analysis.

Independent Variables	Behind the Veil of Ignorance		
	I	II	III
Constant	1.998 (5.873)**	1.882 (5.399)**	1.492 (3.596)**
Leak	−3.896 (–11.846)**	−2.899 (−3.296)**	−2.108 (−2.085)*
Leak^2		−1.106 (−1.178)	−1.118 (−1.184)
Equal	0.1378 (0.444)	0.136 (0.439)	0.919 (1.576)
Equal × Leak			−1.580 (−1.588)
Denver	−0.147 (–0.490)	−0.1508 (−0.491)	−0.139 (−0.451)
Tuscaloosa	−1.463 (–4.451)**	−1.455 (−4.444)**	−1.452 (−4.442)**
Tianjin	−0.986 (–3.506)**	–0.983 (–3.503)**	−0.977 (−3.481)**
High	0.0726 (0.264)	0.073 (0.792)	0.072 (0.263)
Part 2	0.232 (1.166)	0.232 (1.164)	0.228 (1.141)
Log likelihood	−334.787	−331.951	−330.553
Add variables?			
χ^2		1.387	3.826
Significance		0.239	0.148
Cases correct	485	486	486
Cases incorrect	147	146	146

Note: Coefficients refer to probability of voting for the more equal distribution. Coefficients for the probability of abstaining are not reported – all are statistically insignificant. There are 251 votes for the more equal distribution, 375 for the less equal and 6 abstentions.

Source: *T*-values are in parentheses.

*Indicates significance at the 5% level.

**Indicates significance at the 1% level.

Thus, behind the veil, voters are sensitive to efficiency losses with support diminishing as efficiency losses rise. There is no evidence to suggest that concern for equality affects voting behavior behind the veil. There are differences, however, by location. Voters in Tuscaloosa and Tianjin are more likely to vote against redistributions than voters in Guizhou[16] and Denver, suggesting that regional differences within a country may be more important than differences between nations. This is all the more interesting because one country is officially communistic while the other is widely regarded in the world as one of the more capitalistic.

Table 4 summarizes estimations on rounds conducted when positions are known. In contrast to the previous estimations, tests to determine whether the nonlinear and interaction terms should be included in the estimations are significant at the 1% level. Focusing on column III for this reason, we find the first and second degree quadratic terms on LEAK and GAIN are significant. Figure 5 shows the relationship, for both regimes, between the percentage of leakage

Table 4. Multinomial Logit Analysis.

Independent Variables	Positions Known		
	I	II	III
Constant	1.285 (3.609)**	1.056 (2.780)**	0.114 (0.237)
Leak	−0.509 (−1.747)	2.530 (2.527)**	4.078 (3.574)**
Leak^2		−3.217 (−3.331)**	−3.128 (−3.191)**
Equal	0.221 (0.657)	0.246 (0.704)	1.598 (2.453)**
Equal × Leak			−2.360 (−2.495)**
Gain	0.025 (8.690)**	0.043 (6.189)**	0.048 (6.207)**
Gain^2		−0.0001 (−3.726)**	−0.0001 (−4.040)**
Denver	0.617 (1.740)	0.626 (1.727)	1.992 (3.104)**
Denver × Leak			−2.203 (−2.693)**
Tuscaloosa	−0.241 (−0.716)	−0.283 (−0.814)	−0.221 (−0.632)
Tianjin	−0.813 (−2.825)**	−0.865 (−2.910)**	−0.865 (−2.986)**
High	0.398 (1.401)	0.424 (1.446)	0.454 (1.539)
Part 2	−0.785 (−3.589)**	−0.829 (−3.670)**	−0.850 (−3.711)**
Log likelihood	−374.135	−357.532	349.59
Add variables?			
χ^2		27.828	14.332
Significance		0.000	0.000
Cases correct	460	462	470
Cases incorrect	172	170	162

Note: Coefficients refer to probability of voting for the more equal distribution. Coefficients for the probability of abstaining are not reported. Abstentions are less common if the subject stands to gain or lose from the outcome and fade with experience, all other effects are statistically insignificant. There are 333 votes for the more equal distribution, 251 for the less equal and 48 abstentions.

and the probability of voting for the transfer, estimated at the extreme values of equal, and the mean values of other variables. Behind the veil, the relationship is inversely monotonic. For positions known, the negative quadratic term begins to dominate the positive linear term at roughly 40%; up to this point voters are more likely to vote for the transfer the greater the leakage. After 40%, the probability of supporting the transfer falls with increases in the leakage.

The logit results also allow us to estimate where on average Okun's line will be drawn. In Fig. 5, as long as a logit function lies above (below) the 0.5 probability of voting for a transfer, a majority of voters will on average pass (reject) the redistribution. In short Okun's line is drawn where the function crosses 0.5. Behind the veil, the line is drawn at 40% efficiency losses for the average subject. Subjects with a preference for inequality would support only a 20% leak while a strong equality preference lead to toleration of a 60% leak. When positions are known however

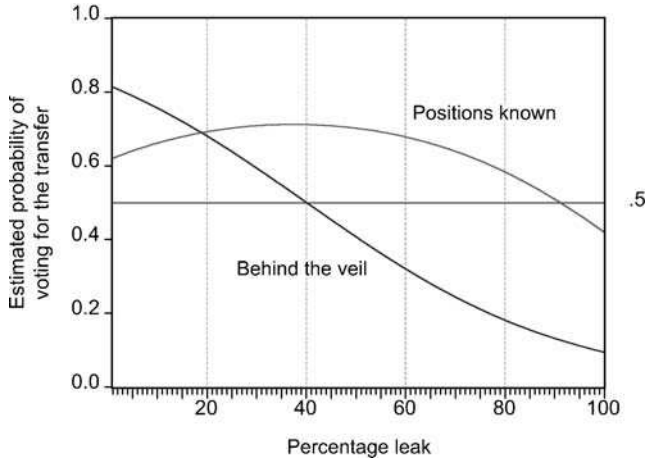


Fig. 5. Effect of Leaky Bucket on Voting Behavior.

transfers pass with efficiency losses as high as 92% – more than double the rate of leakage compared to behind the veil, and regardless of the equality preference.

The relationship between the points gained or lost by the transfer (i.e. GAIN) and the probability of voting for the transfer is shown in Fig. 6. The point to

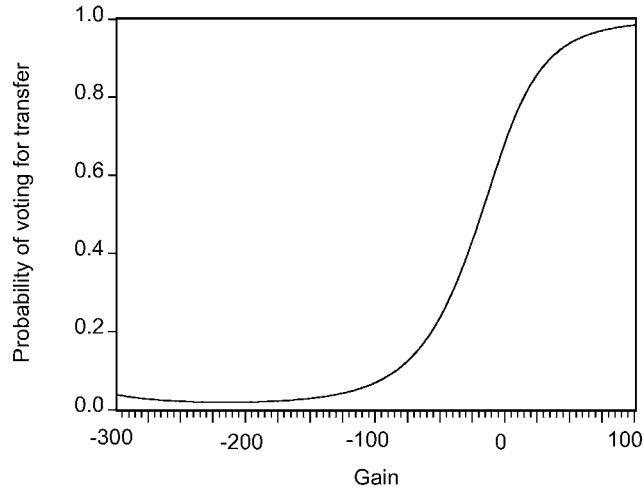


Fig. 6. Effect of Gain on Voting Behavior.

be emphasized is that if a one hundred point or greater loss is involved (i.e. the amount of the transfer) the probability of supporting the redistribution always falls below 4%. As the loss decreases, the probability of support rises rapidly until at a gain of 100 points the probability of supporting the transfer reaches 98%.[17] This is not to suggest that preferences for equality do not have an effect. To the contrary, EQUAL is highly significant indicating that a demonstrated preference for equality increases the probability of voting for a transfer, ceteris paribus.

The interaction term between EQUAL and LEAK is highly significant as well, and as the percentage of leakage rises the effect of a preference for equality diminishes. A number of other variables, which were not significant behind the veil, now are.[18] In fact one is struck by the number of other variables, locational and interaction terms as well as the order of the experiment, which turn out to be significant. The suggestion is that when positions are known and fixed a much more complex process of determining voting behavior is at work with many more considerations coming into play. And we would emphasize the result of these additional considerations is less efficiency. This we believe is important for reasons pointed out in the next section.

5. MOBILITY, EFFICIENCY AND MAJORITARIAN DEMOCRACY

The experimental evidence presented above suggests that income mobility may play a more important role in a majoritarian democracy than has been previously recognized. As is well known, greater mobility in an income distribution implies greater equality in lifetime incomes. What has not been recognized, however, is the effect of mobility on efficiency.

At the outset it is important to note that the two regimes, behind the veil and positions known and fixed, represent the opposite ends of a continuum defined on mobility in an income distribution. In the literature on income distribution, mobility refers to the degree to which individuals are able to move from one position in a distribution of income to another and, in the most fundamental sense, incorporates the prospects of holding various positions in the distribution in the future.[19]

The abstraction of “behind the veil” corresponds to perfect mobility because, regardless of current position, all subjects possess equal chances at every position in the distribution. As we have seen, in these cases, voters weigh efficiency losses against more equality and reduce their support for equality as losses rise.

In contrast, the regime of “position known and fixed” corresponds to perfect immobility. In this case, a subject is assigned his/her relative position and has

no chance of changing it. Under immobility, gains to one's position can only be effected by redistribution between classes. If no gains are possible in the position in which the subject finds him/herself, behavior founded upon a preference for equality, indifference or less socially acceptable motivations such as envy surface as the only options remaining. In these circumstances, voters, in general, support transfers with larger efficiency losses than from behind the veil even at times going to the extreme of 100% leakage.

The fact that a preference for equality is not in evidence behind the veil is thus not surprising because all subjects possess equal expected payoffs, just as with perfect mobility all individuals possess equal expected lifetime incomes, regardless of the static level of inequality currently extant. The level of efficiency, affecting the average level of payoffs, then comes to the fore.

The evidence thus suggests that societies with greater mobility in the distribution of income will support greater efficiency and less mandated redistribution. Beyond avoiding class divisiveness, such societies stand to gain a higher standard of living while at the same time enjoying more equal chances of enjoying it.

6. CONCLUSION

Nearly two decades ago Okun asked how much efficiency leakage individuals would be willing to accept to transfer income from the rich to the poor. It is to the question which Okun posed in 1975 that this paper is addressed. We use the methods of experimental science to examine the preferences of subjects in four locations in China and the U.S. with regard to transfers involving varying efficiency losses. Subjects vote and are paid under two different regimes: (1) behind the veil of ignorance under which votes are cast before positions in the distribution are known; and (2) when positions are known before voting takes place.

Behind the veil of ignorance, transfers with zero efficiency losses are approved in all locations by majorities ranging from 90% in Denver to 67.5% in Tuscaloosa with the Chinese results intermediate between the two. As efficiency losses rise to 50%, however, only voters in Denver continue to support the transfer with Tuscaloosa, China high pay and China low pay rejecting it by majorities of 75%, 53% and 55% respectively. Leakages of 67% and 100% are rejected by large majorities in all locations.

These results markedly contrast with those when positions are known. Transfers entailing 67% leakage now pass in all locations and large numbers of voters, 33–48%, support the imposition of a 100% deadweight loss on the top group. This compares with 0–13% behind the veil. When we consider only voters whose

payoffs are not affected by the transfer we find that the 100% loss actually garners majority support both in Tuscaloosa and high pay China and substantial support in Denver and low pay China.

Support for redistribution however is strongly conditioned by whether or not subjects have to pay for it. When we introduce a second round with zero efficiency losses but in which all subjects, other than the poor, are asked to contribute according to their income, the measure is resoundingly defeated in all locations with majorities ranging from 80% in China to 67.5% in Denver – although the same transfer in which only the rich were required to pay was passed by these same voters with majorities ranging from 90 to 67.5%. "Altruism" among subjects thus runs thin when it comes at their own expense rather than another's.

Estimations from a logit model allow us to further narrow in on where Okun's line would be drawn. Behind the veil of ignorance, we estimate that transfers entailing as large as 40% losses will pass. When positions are known, however, this rises to 92% – more than double that from behind the veil.

This leads to an interesting implication. Our results suggest that, in democracies where voters perceive opportunities to occupy upper income positions, less mandated redistribution, substantially greater efficiency levels and higher payoffs (standards of living) are likely to be observed. In contrast, in those democracies in which little or no mobility exists, large scale, mandated redistributions primarily focused on takings from the upper class are the likely result. In the latter case, the cost of transfers in terms of efficiency losses could be very large. Thus, mobility plays a critical dual role in democracies. It not only contributes to more equal distributions of lifetime incomes but promotes higher standards of living as well.

NOTES

1. See Bishop, Formby and Smith (1991) who emphasize this explanation.
2. Okun refers to this as the Tax and Transfer Equalization Act.
3. See Sen (1973, p. 27) for a discussion of the Pigou-Dalton condition.
4. The instructions in the appendix differ in one respect from those actually used in the experiment. To facilitate reporting the results in this paper, rounds 1–10 in the appendix have been reordered. In the original experiment round 1 was actually round 8, round 2 was round 9, round 3 was round 10 and round 4 was round 7 and round 5 was round 6. Rounds 6 through 10 were originally 1 through 5. We refer to these rounds throughout the paper as they appear in the appendix, not as they appeared in the original experiment.
5. In the U.S., 200 points = $1.
6. Abstentions created the possibility of a tie vote. In these cases the instructions provides a tie-breaker, which involves a toss of a coin.

7. There is a marked contrast between rounds 3 and 4 which is highly significant in regimes as can be confirmed by applying χ^2 tests to total votes for and against the transfer by round. These are reported below.

	Behind Veil		Position Known – All Votes		Position Known – Fixed Pay	
	Transfer	No Transfer	Transfer	No Transfer	Transfer	No Transfer
round 3	44	112	94	54	60	23
round 4	13	146	54	87	56	53
	$\chi^2 = 21$	$p = 4 \times 10^{-6*}$	$\chi^2 = 18$	$p = 0.00002$	$\chi^2 = 7.9$	$p = 0.005$

Note: p is the probability of a larger χ^2 value.

8. All are individuals in the position with the highest payoff who vote to transfer funds from themselves to the person in the lowest position. Five occur when the efficiency loss is 0%, 2 when the loss is 50% and 2 when the loss is 67% with no votes against self interest when the efficiency loss reaches 100%.

9. Voters whose payoffs are not affected by which distribution is chosen are in positions 2, 3 and 4 in rounds 1 through 3 and positions 1, 2, 3, and 4 in round 4.

10. The weak influence of efficiency on voting when positions are known is due almost exclusively to the comparison of the 3rd and 4th rounds. Page's tests for the first three rounds when positions are known have probabilities of larger L of 0.48 and 0.56.

11. Rounds 6 through 10 were used in another context to determine the extent to which the principle of transfers was supported.

12. Other possible and plausible explanations certainly could exist.

13. Equal is constructed by identifying the set of votes in the last five rounds with own pay unaffected, creating a new variable, vote2, which counts abstentions as half a vote, summing vote2 and dividing by the number of observations. The number of obsevations varies subject by subject as assignment to positions with pay affected or unaffected is random. For two individuals, there are no observations.

14. We also reestimated the model with three dummies for the treatments (50% leak, 67% and 100%) in the place of leak and leak^2. This of course allows freedom of functional form. The improvement in log-likelihood for column III positions known from -228.266 to -227.596 is not statistically significant, and the coefficients of other variables do not change in any appreciable manner. The same pattern as the quadratic formulation emerges and the same conclusions follow. The results for column I behind the veil of ignorance are similar. Results from these estimations are available from the authors upon request.

15. The variable equal is unavailable for two subjects, both in the U.S. Since the multinomial logit uses four rounds per U.S. subject, 8 observations are unavailable.

16. The experimenter was the same in Guizhou and Tianjin suggesting the difference in behavior is due to subject pool effects.

17. The functional form indicates a turning point at about -240 points but the positive effect from -300 to -240 is very marginal and due to the functional form used.

18. The interpretation of these results are straightforward and are not discussed in detail for this reason.

19. See Dardanoni (1993) on this point.

ACKNOWLEDGMENTS

This paper was presented at the Conference on Growth, Equity and the Distribution of Income, East-West Center, Honolulu, Hawaii, July, 1994. The authors wish to thank the participants for helpful comments on an earlier version with specific thanks to John Bishop, McKinley Blackburn, John C. H. Fei, James E. Foster, Chao-Cheng Mai, Daniel J. Slottje, Paul Thistle and Buhong Zheng. We would also like to thank Myung Kim, Naci Mocan and Dan Rees. Any errors are of course ours.

REFERENCES

Amiel, Y., & Cowell, F. (1992). Inequality measurement: An experimental approach. *Journal of Public Economics*, *69*(February), 3–26.

Amiel, Y., & Cowell, F. (1993). Inequality changes and income growth. In: W. Eichorn (Ed.), *Models and Measurement of Inequality*. Berlin: Springer-Verlag.

Amiel, Y., Creedy, J., & Hurn, S. (1999). Measuring attitudes to inequality. *Scandinavian Journal of Economics*, *101*, 83–96.

Amiel, Y., & Cowell, F. (1994). Income inequality and social welfare. In: J. Creedy (Ed.), *Taxation, Poverty and Income Distribution*. Aldershot: Edward Elgar.

Beckman, S., Formby, J. P., Smith, W. J., & Zheng, B. (2002). Envy, malice and pareto efficiency: An experimental examination. *Social Choice and Welfare*, *19*, 349–367.

Bishop, J., Formby, J. P., & Smith, W. J. (1991). Incomplete information, income redistribution and risk averse median voter behavior. *Public Choice*, *68*(January), 41–55.

Buchanan, J. (1984). The political economy of the welfare state. In: J. M. Buchanan & R. D. Tollison (Eds), *The Theory of Public Choice-II*. Ann Arbor, MI: University of Michigan Press.

Dardanoni, V. (1993). Measuring social mobility. *Journal of Economic Theory*, *61*, 372–394.

Harrison, E., & Seidl, C. (1994). Perceptional inequality and preferential judgements: An empirical examination of distributional axioms. *Public Choice*, *79*(April), 69–81.

Kachelmeier, S., & Shehata, M. (1992). Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China. *American Economic Review*, *82*(December), 1120–1141.

Okun, A. (1975). *Equality and efficiency*. Washington: Brookings Institution.

Page, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association* (February), 216–230.

Rawls, J. (1971). *A theory of justice*. Cambridge: Harvard University Press.

Sen, A. (1973). *On economic inequality*. New York: W. W. Norton & Co.

Tullock, G. (1983). *Economics of income redistribution*. Boston: Kluwer.

APPENDIX

Instructions Part 1

Version 1

In this experiment, you will be a member of a five person committee. There may be more than one committee but you will always be a member of the same committee. The committee does not meet and the votes are anonymous. At the end of the experiment your pay is determined by your choice, the majority vote of your committee and by chance. We will go through one round for practice but all following rounds are for real money – like the money the experimenter is showing to you now.

If less than 10 subjects show up for the experiment there will be only one committee. Ten or more subjects but less than 15 leads to two committees. Due to the nature of the experiment we can have no more than two committees. Those not assigned to a committee will receive a $3 payment and are excused.

In this session of the experiment there will be _____ committees and the first thing we will do is assign people to their committee. This will be done as follows. Each of you received a numbered folder as you entered the room. The folders are numbered 1, 2, 3 . . . and up to 15. These are your subject numbers and will be used to assign committees.

- Those with numbers 1, 3, 5, 7 and 9 are assigned to committee 1.
- Those with numbers 2, 4, 6, 8 and 10 are assigned to committee 2.

There are 10 rounds in this part of the experiment. All the rounds begin in the same way. The experimenter will shuffle five playing cards, an ace, a two, a three, a four and a five. He will then fan these out in his left hand and walk before each of you with the cards face down. Each of you will point to one card. The experimenter will turn card face up and place it on the table in front of you. He will then walk to the next subject and they will point to one of the four remaining cards. In this way all of you will be randomly assigned to one of five different positions. Once you know your position mark your record sheet accordingly, but please do not pick up or mark the card in any way.

Now that positions are assigned you will vote for one of two possible alternative outcomes, denoted as A and B. The majority vote will determine which alternative is chosen. The alternatives are payment schedules for five outcomes or position numbers and are always presented in the same format as alternatives A and B below.

	A: Average # of points = 120					B: Average # of points = 126				
Position	1	2	3	4	5	1	2	3	4	5
Points	40	80	120	160	200	40	115	120	140	200

In alternative A there are five possible outcomes: 40, 80, 120, 160 and 200 points. In alternative B the possible outcomes are 40, 115, 120, 140 and 200 points. Alternative B will change each round and in some rounds alternative A will change as well.

In most rounds the "average number of points" expected from A and B are the same, but, as in the example above, they are different in some rounds. It is up to you to choose A or B depending on the alternative you think is best from your perspective. To make informed choices it may be helpful to know how to interpret the "average number of points". If alternative B prevails and if there were many random draws to determine the outcome, then 126 is the average number of points that can be expected. This is much like tossing a coin many times. After many tosses the average expected outcome is 50% heads and 50% tails. Similarly, in the example above the average outcome from alternative A and B assuming many random draws is 120 and 126. You should recognize however that while the average expected payoff is higher for B than A, only one draw will be used to determine the outcome, not many. Just as heads or tails are equally likely in a single coin toss, outcome 1, 2, 3, 4, and 5 are equally likely in this experiment. Further the individual payments associated with each outcome may be different in A and B. You will want to carefully compare and think about any differences in the payments (points) associated with each outcome in A as compared to B. For example, in the choice above if the choice is A and outcome 2 prevails you will receive 80 points, but if B is chosen outcome 2 yields 115 points. However, if outcome 4 prevails, B yields 140 points, which is fewer than the 160 points received if A is chosen.

Points are Converted to Cash at the End of the Experiment at a Rate of 200 Points = $1

The particular outcome that prevails is determined by which alternative the committee selects by majority vote and your position number. The first outcome goes to the person in the first position, the second to the person in the second position. If the committee votes for alternative A and you are in position 4 then your payment is 160 points. However, if the committee votes for alternative B then you will earn 140 points regardless of your individual vote. Since we are interested in individual votes, we ask that you not show your vote to anyone and that you be silent once the experiment is underway. However questions are encouraged in the instruction phase.

In each round you will vote by writing an A, B or X on your record sheet. An entry of A or B shows your preference for alternatives A or B while an X means you abstain. Once the votes have been cast the experimenter will walk by and tally the results for each committee. In the event of a tie vote within a committee, a coin will be tossed to determine the outcome. The experimenter will announce the

overall vote of each committee and you will mark the outcome under the heading "majority vote" on your record sheet.

Points earned will be determined by your position number and the alternative chosen by your committee. The experimenter will check that you have entered the points earned on the record sheet and then highlight this as well.

Let's Try a Round for Practice

First, record your subject number on your record sheet.

The experimenter will show you the five cards, shuffle them, fan them out face down and walk before you. Please point to one card.

This determines your position number, which will be entered on your record sheet and highlighted by the experimenter.

Next enter your vote on the record sheet by writing an A, B or X.

The experimenter will now walk by and tally the results for each committee and announce the overall vote for each committee.

You will mark the committee's choice under the heading "majority vote" on the record sheet.

Based upon your position number and the majority vote you will then enter the points earned on the record sheet. For example, if you occupy the fifth position and voted for B but the majority voted for A, then your record sheet would look as follows:

	Position Number	Majority Vote	My Vote	Points Earned
round 0	5	A	B	200

Now that the practice round is over, are there any final questions?

This completes the instruction phase for part 1.

We now conduct the experiment for money.

Instructions for Part 2

This part of the experiment is exactly the same as Part 1, except each person must vote without knowing their position number. That is, you will vote for A or B first and then find out which of the positions you will occupy.

The record sheets and outcome sheets for this part of the experiment are the same as those used in part 1. New record sheets and outcome sheets are being

passed out now and old ones collected. Please enter your subject number on the new record sheet.

Lets try a practice round.

First, record your subject number on the record sheet.

Next, look at the alternatives A and B for round zero, and vote by writing an A, B or X on the record sheet.

The experimenter will now walk by, tabulate the majority vote, and announce the outcome.

You will mark the majority's decision on your record sheet.

The experimenter will now show you the five cards, shuffle them, fan them out face down and walk before you. Please point to one card, which determines your position number.

You will enter your position number on the record sheet and the experimenter will highlight it.

Based on the majority vote and the position number you will now enter the points earned on your record sheet.

Now that the practice round is over, are there any questions?

This completes the instruction phase for part 2

We now conduct the experiment for money.

Outcome Sheet

		A					B				
Round 0 – A practice round											
(0)	Position	1	2	3	4	5	1	2	3	4	5
	Points	40	80	120	160	200	40	115	120	140	200
		Average # of points = 120					Average # of points = 126				
Rounds 1 through 10 – for money											
(1)	Position	1	2	3	4	5	1	2	3	4	5
	Points	10	110	120	160	500	110	110	120	140	400
		Average # of points = 180					Average # of points = 180				
(2)	Position	1	2	3	4	5	1	2	3	4	5
	Points	10	110	120	160	500	110	110	120	160	300
		Average # of points = 180					Average # of points = 160				
(3)	Position	1	2	3	4	5	1	2	3	4	5
	Points	10	110	120	160	500	110	110	120	160	200
		Average # of points = 180					Average # of points = 140				

		A					B				
(4)	Position	1	2	3	4	5	1	2	3	4	5
	Points	10	110	120	160	500	10	110	120	160	200
		Average # of points = 180					Average # of points = 120				
(5)	Position	1	2	3	4	5	1	2	3	4	5
	Points	10	110	120	160	200	90	100	105	140	165
		Average # of points = 120					Average # of points = 120				
(6)	Position	1	2	3	4	5	1	2	3	4	5
	Points	10	40	70	100	130	10	50	60	100	130
		Average # of points = 70					Average # of points = 70				
(7)	Position	1	2	3	4	5	1	2	3	4	5
	Points	40	80	120	160	200	60	60	120	160	200
		Average # of points = 120					Average # of points = 120				
(8)	Position	1	2	3	4	5	1	2	3	4	5
	Points	40	80	120	160	200	40	100	100	160	200
		Average # of points = 120					Average # of points = 120				
(9)	Position	1	2	3	4	5	1	2	3	4	5
	Points	40	80	120	160	200	40	80	140	140	200
		Average # of points = 120					Average # of points = 120				
(10)	Position	1	2	3	4	5	1	2	3	4	5
	Points	40	80	120	160	200	40	80	120	180	180
		Average # of points = 120					Average # of points = 120				

HOW MANIPULABLE ARE FAIRNESS PERCEPTIONS? THE EFFECT OF ADDITIONAL ALTERNATIVES

Yoella Bereby-Meyer and Brit Grosskopf

ABSTRACT

In customer or labor markets raising prices or cutting wages is perceived as unfair if it results from the exploitation of shifts in demands. In a series of manipulations we show that adding an alternative to the original choice set alters the perception of fairness of the final outcome. Adding a worse alternative lowers the perception of unfairness, whereas adding a better alternative raises the perception of unfairness. These findings supplemented with existing experimental evidence cast doubt on purely outcome-based theories of fairness and suggest that fairness perceptions are highly manipulable.

1. INTRODUCTION

With the increasing use and recognition of experimental methods in economics, the traditional economic assumption that people are self-interested and seek to solely maximize their own monetary payoffs in social interactions seems not only stark but non-descriptive of how human beings actually behave. Ever since the first experiments on the ultimatum game (e.g. Güth et al., 1982) researchers are aware that considerations for fairness and factors such as trust and reciprocity affect behavior. In recent years attempts to model social preferences by augmenting

Inequality, Welfare and Income Distribution: Experimental Approaches
Research on Economic Inequality, Volume 11, 43–53

ISSN: 1049-2585/doi:10.1016/S1049-2585(04)11003-X

agents' utility functions with, for example, preferences for equity, reciprocity and relative payoffs have abound.[1]

The current research aims to show that the perception of fairness is affected by the perceived intentions behind agents' actions. These perceived intentions are prone to cognitive biases and can be easily manipulated by altering the set of alternatives an agent originally chooses from. We find that while keeping the outcome unchanged, it is sufficient to *mention* different sets of alternatives to shape perceived intentions and accordingly change perceived fairness to the worse or to the better. These results are strengthened by incorporating evidence from existing experiments with real stakes that similarly show that changing the set of alternatives, while keeping the outcome constant, affects economic behavior. Taken together, these findings suggest that the same economic situation might not only be perceived differently but can lead to very different economic behavior.

Our results have implications to models of social preferences. We will discuss them at the end of the paper, where we will also allude to the potential dangers that might result from the high manipulability of fairness perceptions.

Empirical evidence regarding the importance and robustness of fairness consideration for economic behavior is plentiful. One game that has been studied extensively in this context is the ultimatum game (see Roth, 1995, for a survey). The ultimatum game is a two-player game. One player (the proposer), who is randomly chosen, receives a fixed amount of money that she needs to divide between herself and the other player. The other player (the responder) has to decide whether to accept or reject the proposed division. If he accepts the money is divided accordingly, if he rejects both players do not receive anything.

According to standard economic theory, the responder should accept any proposal greater than zero, and therefore the unique subgame perfect equilibrium prediction is that the proposer offers the smallest possible amount. Experimental evidence does not support this prediction. It rather shows that individuals incorporate fairness considerations into their offers and acceptance or rejection decisions. The average offer to the responder is usually more than 30% of the available pie. In addition, responders usually reject profitable but "unfair" offers (e.g. Bolton & Zwick, 1995; Güth et al., 1982).

To account for this deviation from standard economic theory, Rabin (1993) suggested the concept of a fairness equilibrium that is based on the premise that people are motivated to help those who help them and hurt those who hurt them. Models of this type, known also as reciprocity-based models, assume that people are motivated not only by their final outcomes, but also by the way the outcome has been achieved. Players care about intentions behind actions and may be willing to sacrifice material payoff to reciprocate, i.e. reward fair behavior and punish unfair

behavior (see for example Bereby-Meyer & Niederle, forthcoming; Dufwenberg & Kirchsteiger, 1998; Falk & Fischbacher, 1999).[2]

Fairness considerations are important in many domains. For example, customers suspecting a supplier to treat them unfairly might feel angry and start searching for other alternatives. Anticipating this behavior causes firms not to raise prices, if this raise will be perceived as unfair. Similarly, fairness considerations inhibit employers from cutting wages during periods of high unemployment (Akerlof, 1982; Akerlof & Yellen, 1990; Bewley, 1998; Solow, 1980). The susceptibility of economic behavior to fairness issues makes the understanding of fairness norms an important issue. Kahneman et al. (1986) conducted a telephone survey that aimed to elicit community standards of fairness for the setting of prices and wages. They showed that the reference transaction, a relevant precedent that is characterized by a reference price or wage and by a positive reference profit to the firm, affects fairness perceptions. The behavior of a firm will be perceived as unfair if the firm will increase its profit by arbitrarily changing the reference price or wage (Bazerman, 1985). Similarly, consumers seem to grant special status to the manufacturer's list price, even if they do not expect to pay that amount. Exceeding that amount is perceived as unfair (Bazerman, 2002).

The reference transaction and the list price act as a reference point for assessments of fairness. Outcomes above the reference point are perceived as fairer and outcomes below the reference point are perceived as less fair.

In the current research we suggest that the reference point according to which people evaluate fairness, i.e. the reference transaction, can be manipulated by enlarging the set of possible alternatives that describe what *could have* been a possible outcome.

In line with intentionality-based models we suggest that the way the outcome has been achieved forms perceived intentions and affects the perception of fairness. If the additional ex-ante feasible outcome is worse than the current outcome, the current outcome will be perceived as a gain and consequently as a kind behavior. If the additional ex-ante feasible outcome is better than the current outcome, it will be perceived as a loss and consequently as a mean behavior even if the final outcome does not differ in the two situations.

In the experiment reported below we gave participants hypothetical scenarios similar to Kahneman et al. (1986). We asked the participants to evaluate the fairness of the different actions that were described. Without changing the final outcomes we mentioned an action that could have been taken: for example, a 3% reduction in wages that could have been a reduction of 5%. We found an increase or a decrease in the perception of fairness by just mentioning a worse outcome that could have been obtained. This feasible but yet not obtained outcome seems to have served as the reference point relative to which fairness and the perceived

kindness of the firm were evaluated. Our findings are additional evidence that people do not evaluate the utility of alternatives based on final outcomes, as was expected by once standard economic theory, but rather are affected by the way the outcome is presented to them.

2. THE EXPERIMENT

2.1. Participants

Two hundred and forty-nine undergraduate students participated in this study. One hundred and ninety-two were recruited from the Boston area (Harvard, MIT, BU) and fifty-seven from Ben-Gurion University, Israel. We ran the experimental sessions in Boston and at Ben-Gurion University respectively.

2.2. Apparatus and Procedure

The experiment was conducted in a classroom setting. Participants were told that they were about to participate in a study on decision-making, and were asked to give their fairness evaluation to scenarios like the one that is described below (for all other scenarios see the Appendix):

> A small company employs several workers and has been paying them the average wage. There is severe unemployment in the area and the company could easily replace its current employees with equally skilled workers at a lower wage. The company has been making money. The owners considered reducing current hourly wages by 5%. Finally it was decided to reduce the hourly wages by 3%. How do you judge the decision of the company? Please indicate your judgment on a scale from 1 to 7, where 1 refers to 'not fair at all' and 7 means 'extremely fair'.

Two factors were manipulated:

(1) *Outcome*: *the direction of change that was either negative or positive*. When the direction of change was negative, the change resulted in an outcome reduction (e.g. wage cut) and when the direction of change was positive, the change resulted in an outcome growth (e.g. wage increase).
(2) *Additional alternative*: *an additional alternative that was either mentioned or not*. When an additional alternative was mentioned, it was stated that the decrease (or increase) could have been 5% but eventually was 3%. When the additional alternative was not mentioned, only the decrease (or increase) of 3% was mentioned.

In addition, two types of scenarios were given in each condition.

Scenario A: The scenario described a decision concerning wages (see Appendix, examples A1 and A2).

Scenario B: The scenario described a decision concerning consumer prices (see Appendix, examples B1 and B2).

Participants of each subject pool were randomly assigned to the 8 experimental conditions of the 2 × 2 × 2 design (2 "directions of change" × 2 "with or without an additional alternative" × 2 "types of scenarios"). The number of participants in each condition ranged from 12 to 16.

3. EXPERIMENTAL RESULTS

Figure 1 presents the mean of the observed fairness perception as a function of the outcome and the addition of an alternative. The presence of an additional alternative raises perceived fairness in the case of negative outcomes and lowers perceived fairness in the case of positive outcomes.

In order to test the significance of this effect, we ran a linear regression on the perception of fairness as the dependent variable and the outcome, the mentioning of an additional alternative and the interaction between the outcome and the mentioning of an additional alternative as independent variables.[3,4] *Outcome*

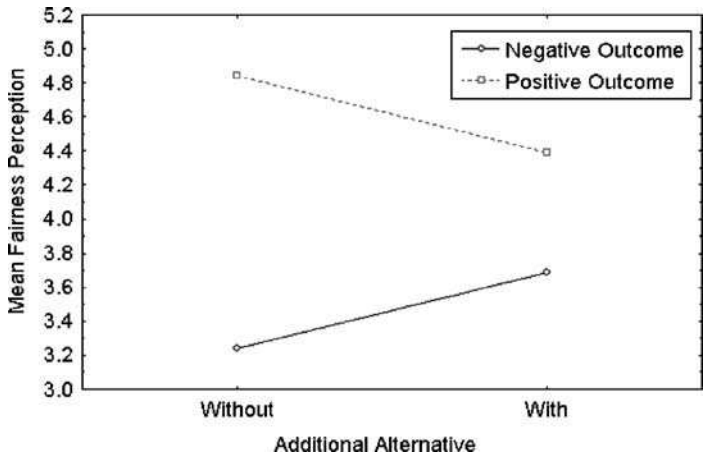


Fig. 1. Observed Mean Fairness Perception as a Function of the Outcome and the Presence of an Additional Alternative.

Table 1. Linear Regression Results.

Variables	Coefficients ($t_{(244)}$, p-Value[a])
Constant	2.07 (3.9, 0.005)
Outcome positive for evaluator	1.6 (5.74, 0.001)
With additional alternative	0.45 (1.58, 0.05)
Outcome × Additional Alternative	−0.9 (−2.27, 0.01)
# Observations	249
$R^2 = 0.14$, $F(3,245) = 13.07$	$p < 0.0001$

[a] The p-value refers to a one-tailed test.

positive for evaluator equals 1 for a positive change and 0 for a negative change. *With additional alternative* equals 1 when mentioned and 0 when not. Table 1 presents the results of the linear regression.

As can be seen in Table 1, we find a significant effect for the outcome, i.e. the direction of change. The coefficient of 1.6 indicates that positive outcomes are perceived as fairer than negative outcomes. However, more importantly, we observe a significant interaction between the outcome and the addition of an alternative. Perceived fairness increases when a worse alternative is mentioned ($t(125) = 1.62$, $p = 0.05$) and decreases when a better alternative is mentioned ($t(118) = -1.71$, $p < 0.044$). These findings illustrate that the same outcome can be judged as fair or unfair depending on the set of available alternatives it was chosen from.

One possible limitation of our study is its hypothetical nature – we did not test choice behavior with performance-based monetary payments. Yet there is evidence that responses to hypothetical questions are often consistent with actual behavior (e.g. Lichtenstein & Slovic, 1973 on preference reversal; Grether, 1980 on probabilistic reasoning).[5]

For an inference about consequences of fairness perceptions and resulting choice behavior we enrich our study with existing experimental evidence that used performance-based incentive pay and combine the results of the two approaches in a later section.

3.1. How Does the Perception of Fairness Affect Actual Choice Behavior?

The following section briefly reviews experimental evidence of choices in games. These experiments show that the set of alternatives that are available to the proposer in ultimatum games affects responder's choice behavior.

Mini-ultimatum games have been quite popular in studying rejections of profitable but "unfair" offers in the ultimatum game. In a mini-ultimatum game the choice set of the proposer is artificially constrained. Its structure allows for a clearer separation of outcomes of different types and the control over players' perceptions and expectations is increased. But clearly, these studies only examine subjects' choices and not their thought processes, i.e. their judgment.

The study that best complements our hypothetical scenarios on players' judgment of fairness is Brandts and Solà (2001). They analyze whether a benchmark split of (320,80) – with (proposer, responder) payoffs – is more or less acceptable depending on what the only other possible choice of the proposer was. They find that the benchmark split of (320,80) has a higher rate of acceptance (hardly ever was rejected), when the only other choice was (350,50), i.e. if the proposer chose the "fairer" of the two "unfair" options, than when the proposer could have split the pie equally, i.e. (200,200). If the other option available to the proposer was a split of (100,300) then the rejection rates of the benchmark split (320,80) increase even further.[6] Similarly, Nelson (2002) finds that an offer of \$4 in a \$20 ultimatum game has a higher probability of being accepted if \$4 is the highest possible offer (mini-ultimatum game) compared to when \$20 is the highest possible offer (full-blown ultimatum game). Falk et al. (2003) show in mini-ultimatum games that the unequal offer of (8,2) has a higher probability of being rejected if the proposer could have proposed an equal offer (5,5) than if the proposer could have proposed only an even more unequal offer, e.g. (10,0).[7]

All these experimental findings suggest that the acceptability of an offer is affected by the set of available offers. Depending on the available set of alternatives for the proposer, identical offers signal different intentions of the proposer and consequently are being accepted or rejected differently.

4. DISCUSSION

The findings of this study suggest that fairness perceptions are affected by the intentions of the agents and are manipulable. Complemented by the existing experimental evidence that fairness perceptions severely affect choice behavior, this is of great economic interest.

Consistent with intentionality-based models we have shown that the way the outcome has been achieved forms perceived intentions and consequently affects fairness perception. Mentioning a worse possible outcome causes the current outcome to be perceived as a gain and consequently as a kind action, while mentioning a better possible outcome causes the current outcome to be perceived

as a loss and consequently as a mean action. In both cases the ex-ante feasible alternative does not affect the final outcome, but nevertheless affects the fairness perception. This shows how highly manipulable the fairness perceptions of the final outcomes are.

Our results are not restricted to changes in fairness perception only. Experimental research on bargaining has shown that agents are likely to act upon their perception of fairness. Responders in a an ultimatum game accept or reject similar offers differently depending on what the proposer could have done, i.e. what alternatives the proposer was choosing from.

Taken together these results suggest that fairness models should take into account that people value the intentions behind actions besides having preferences over final outcomes and equitable distributions. Identical actions – depending on the alternatives available – are likely to signal different intentions of the other party and consequently may lead to a different choice. Therefore, more complex models of social preferences have to take the set of available actions into account.

The fact that choices are dependent on the set of possible alternatives gives also rise to abuses and arbitrary manipulations. Justifications as to how a certain decision came about seem to play an important role in evaluating a decision. Consequently, it might be in the interest of one party to distort the set of available alternatives in order to make the final outcome seem fairer, especially when the available choice set is not directly observable. For example, universities that are currently suffering severe budget cuts may find much more support from staff and faculty if they exaggerated the budget situation, i.e. announce intended salary cuts, and then decide to cut costs in a slightly more moderate way. Sellers may have no incentive to update list prices when prices are falling, since customers will perceive a bigger discount as a kinder act.

NOTES

1. We will discuss models that incorporate social preferences in more detail later in the paper.

2. Another class of fairness models, known as outcome-based models, is concerned with the distribution of payoffs. Fair is defined not only in terms of absolute income but also in terms of the relative share. In order to reduce payoff inequality a player may reduce her payoff if this leads to a greater reduction in the other players' payoff. A player would, however, never sacrifice to increase payoff inequality (see Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999). However, Charness and Grosskopf (2001) and Charness and Rabin (2002) find that people are willing to sacrifice money to achieve efficiency, even when this increases payoff inequality.

3. The effect of the subject pool and its interaction with the other predictors were found insignificant. Therefore we analyze the data pooled from both subject pools.

4. In order to examine whether there are differences in the way the additional alternative affects perceived fairness in the different scenarios, we ran for each type of outcome (decrease/increase) a linear regression on the fairness perception as the dependent variable and the type of scenarios, the mentioning of an additional alternative and the interaction between them as the predicting variables. In both regressions no significant interaction was found. This indicates that the type of scenario did not alter the influence of the additional alternative. As a consequence, the analysis was done with the pooled data of both scenarios.

5. For surveys on experimental procedures and on whether actually paying subjects in experiments alters their behavior see Camerer and Hogarth (1999) and Hertwig and Ortmann (2003).

6. The observed rejection rates are 0.0333, 0.2183 and 0.3492 respectively.

7. Güth et al. (2001) show that almost equal splits instead of equal splits evoke very different behavior on both, the proposers' and the responders' side.

REFERENCES

Akerlof, G. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, *97*, 543–569.

Akerlof, G., & Yellen, J. (1990). The fair wage-effort hypothesis and unemployment. *Quarterly Journal of Economics*, *105*, 255–283.

Bazerman, M. H. (1985). Norms of distributive justice in interest arbitration. *Industrial and Labor Relations Review*, *38*, 558–570.

Bazerman, M. H. (2002). *Judgment in managerial decision making* (5th ed., 1998). Wiley.

Bereby-Meyer, Y., & Niederle, M. (forthcoming). Fairness in bargaining. *Journal of Economic Behaviour and Organization*.

Bewley, T. (1998). Why not cut pay? *European Economic Review*, *42*, 459–490.

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

Bolton, G., & Zwick, R. (1995). Anonymity vs. punishment in ultimatum bargaining. *Games and Economic Behavior*, *10*, 95–121.

Brandts, J., & Solà, C. (2001). Reference points and negative reciprocity in simple sequential games. *Games and Economic Behavior*, *36*, 138–157.

Camerer, C., & Hogarth, R. M. (1999). The effects of financial incentives in experiments. *Journal of Risk and Uncertainty*, *19*, 7–49.

Charness, G., & Grosskopf, B. (2001). Relative payoffs and happiness: An experimental study. *Journal of Economic Behavior and Organization*, *45*, 301–328.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, *117*, 817–869.

Dufwenberg, M., & Kirchsteiger, G. (1998). A theory of sequential reciprocity. Mimeo.

Falk, A., Fehr, E., & Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry*, *41*(1), 20–26.

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

Grether, D. M. (1980). Bayes rule as a descriptive model: The representativeness heuristics. *Quarterly Journal of Economic*, *95*, 537–557.

Güth, W., Huck, S., & Müller, W. (2001). The relevance of equal splits in ultimatum games. *Games and Economic Behavior*, *37*, 161–169.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, *3*, 367–388.

Hertwig, R., & Ortmann, A. (2003). Economists' and psychologists' experimental practices: How they differ, why they differ, and how they could converge. In: I. Brocas & J. D. Carrillo (Eds), *The Psychology of Economic Decisions, Vol. 1: Rationality and Well-Being* (pp. 253–272). Oxford University Press.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review*, *76*, 728–741.

Lichtenstein, S., & Slovic, P. (1973). Reversal of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *101*, 16–20.

Nelson, W. R. (2002). Equity or intention: It is the thought that counts. *Journal of Economic Behavior & Organization*, *48*, 423–430.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, *83*, 1281–1302.

Roth, A. E. (1995). Bargaining experiments. In: J. E. Kagel & A. E. Roth (Eds), *Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press.

Solow, R. M. (1980). On theories of unemployment. *American Economic Review*, *70*, 1–11.

APPENDIX

The wording in parentheses refers to the version with the additional alternative. The two first scenarios refer to a positive outcome and the two last scenarios refer to a negative outcome.

A1. A computer company has been making moderate profit. Recently there has been inflation of 3%. As a consequence the company [considered raising the salary by 5% but in the end] decided to raise the salary of its workers by 3%.

B1. A cosmetic company has been making profit. Recently the price of raw materials used in the cosmetic production decreased by 3%. As a consequence the company [considered decreasing the price of its products by 5%. In the end they] decided to decrease the price of its products by 3%.

A2. A small company employs several workers and has been paying them the average wage. There is severe unemployment in the area and the company could easily replace its current employees with equally skilled workers at a lower wage. The company has been making money. As a consequence the company [considered reducing the current hourly wages by 5%. In the end they] decided to reduce the current hourly wages by 3%.

B2. A grocery store has several months supply of peanut butter in stock that it has on the shelves and in the storeroom. The current price of a peanut butter jar is $7. The owner hears that the wholesale price of peanut butter has increased and [she considered increasing the price on the current stock by 5% immediately. In the end] she immediately raised the price on the current stock by 3%.

AN EXPERIMENTAL ANALYSIS OF SOCIAL MOBILITY COMPARISONS

Michele Bernasconi and Valentino Dardanoni

ABSTRACT

Social mobility is an issue at the crossroad of various disciplines: sociology, statistics, political science and economics. We review alternative approaches to the analysis of intergenerational income mobility, and conduct a questionnaire aimed to reveal students' opinions on some basic principles developed in the literature. The questionnaire includes questions focussed on: (a) the difference between structural and exchange mobility; (b) the decomposition of mobility tables into parameters linked to structural mobility and parameters linked to exchange mobility; (c) the effects of transformations of the status variables (incomes) on mobility comparisons. These issues have been formalized as hypotheses that can be formally tested by the questionnaire. We find various regularities in the data, but also some rejections of basic principles that require further scrutiny.

1. INTRODUCTION

Social mobility is the process through which, under the effect of different transition mechanisms, the distribution of some relevant measure of individual status changes over time. It may be viewed from an *intergenerational* perspective or *intragenerational.* The former notion relates to the transitions of family lines from one generation to the next, usually traced through the male line. The latter

Inequality, Welfare and Income Distribution: Experimental Approaches
Research on Economic Inequality, Volume 11, 55–84

ISSN: 1049-2585/doi:10.1016/S1049-2585(04)11004-1

refers to changes of position within the lifespan of the individual. Social mobility is an issue at the crossroad of various disciplines, including sociology, statistics, and economics (see e.g. Bartholomew, 1996; Prais, 1955, for classical references to the sociological and statistical literature; see Atkinson, 1981; Shorrocks, 1978, for pioneering works on social mobility by economists).

There are, indeed, many reasons for social scientists to be interested in social mobility (see Fields & Ok, 1999, for a review of the literature). First of all, in an ethical perspective, social mobility is often interpreted as a factor determining equality of opportunity. In fact, some scholars even argue that equality of opportunity should be the main, if not exclusive, point of political concern in a society (see e.g. Stokey, 1998), and that in more socially mobile communities, the support for redistribution might be lower (Alesina & La Ferrara, 2001; Benabou & Ok, 2001). In addition, as argued by some, social mobility may promote economic efficiency and stimulate economic growth (see Breen, 1997; Yishay & Moav, 1999, for recent examples and positions within this literature). Social mobility may further directly improve social welfare, when one considers the distributions of incomes of different generations (Atkinson, 1981).

Obviously, given the various reasons one may be interested in social mobility, it is perhaps not surprising that different disciplines and scholars have developed different approaches to social mobility measurement. This is witnessed by the great number of synthetic measures of social mobility (mobility indices) which have been developed by different scholars and which, in many situations, provide quite divergent indications when applied to a given set of real world mobility data (Checchi & Dardanoni, 2003a).

In this paper we focus on intergenerational mobility. We use responses from about 350 university students in Italy, to test the extent to which they agree on a few basic concepts underlying social mobility studies. To the best of our knowledge, this is the first attempt to design a questionnaire to find out people's opinions on basic principles of social mobility measurement. The method of using students' responses for testing basic principles in issues concerning social measurement and ethics has been pioneered by Amiel and Cowell (1992) for the theory of income inequality. Since then, a growing literature has mounted on other aspects of distributive justice (see Amiel, 1999, for a survey; and see Moyes et al., 2002, for a collection of recent papers in the area).

There may be objections to this approach. One may for example argue that scholarly introspection, deduction from basic premises, and academic consensus should be the only validation criteria for the progress of knowledge in the field. However, while we do not believe that people's views should become a substitute for academic confrontation, we also think that it would be presumptuous to proceed without any tests of what people may actually think about the ideas developed by

the scientific community. Amiel and Cowell (1992) are in particular convincing to argue that, if anything, "this might avoid becoming hostage of the conventions that accompany any academic specialism" (p. 4). In the context of the literature on social mobility, this may be even more important since very few conventions are firmly established in the area and much theoretical work is still undergoing.

Some may also object to the use of students as subjects to be targeted in the questionnaire. Perhaps, one may be more interested in experts' view or judgments by the layman. As pointed out by Amiel and Cowell (1992), the former people may be more affected by prejudice, while the latter may fail to express their opinions coherently, due to the lack of experience to think about logical propositions. On the other hand, students may have greater practice to think in abstract terms on general principles and working with simple numerical examples. This attitude may be especially important in the context of social mobility comparisons, which by their multidimensional nature are intrinsically more problematic than inequality comparisons.

In the next section, we give a formal treatment of the various dimensions of comparisons involved in the present experiment. In particular, we focus on the sharp distinction drawn by sociologists between the so called dimensions of structural mobility and exchange mobility and provide formal hypotheses to test them by the questionnaire. We also discuss various invariance hypotheses on the effect that the location and the dispersion of the socio economic indicator, both in the fathers' and the sons' generations, may have on the perception of intergenerational mobility. Next we present the questionnaire and the results. We find various regularities across subjects' responses. Not always, however, these are in the direction of the theoretical predictions. Among others, perhaps the most surprising is that subjects seem to consistently fail to recognise social mobility along the dimension of what we will define as exchange mobility.

2. THE COMPARISON OF MOBILITY TABLES

The intergenerational mobility of a society can be described by the joint distribution of a pair of random variables X and Y which represent, respectively, fathers' and sons' *socio-economic status*. Henceforth, we will simply use income as the relevant indicator of status. The joint distribution $H(x,y)$ contains all the relevant information to study the intergenerational mobility of a society. In particular, H contains information on different aspects of the society's intergenerational mobility. The marginal distributions determine both the average level of status and its dispersion within the fathers and sons generations. One could say then that the marginal distributions contain information of a *static* nature. However, the joint

distribution determines also the conditional distributions of Y for each given level of fathers' income, which indicate how the chances of a son to achieve a given income depend on level of income of his father, and thus provide a truly *dynamic* view of social change.

When analyzing the intergenerational mobility of a society, the interplay between the distributions of X and Y can be described by two quite different concepts: *structural mobility* and *exchange mobility.* Structural mobility refers to, and is measured by, the difference between the fathers' and sons' marginal distributions of income. For example, if a country is experimenting a substantial economic growth, there will be a greater number of high-income positions available to the sons compared to the fathers. The opposite may occur in the case of an economic decline. In either case, there will be some kind of social change. However, there are many ways in which a marginal distribution of Y can be obtained from a given marginal distribution of X. In fact, two hypothetical societies could display the same amount of structural mobility because they have the same marginal distributions, but they could differ in how families interchange their relative positions. In particular, for any given marginal distributions of fathers and sons income, it is equally well possible to imagine an hypothetical society where X and Y are independent, and one where Y is a deterministic monotone function of X. So, while in the first situation any son has the same chances as any other, in the other society the income of a son is uniquely determined by the income of his father. The investigation of the question: *when does a society offer individuals better chances of determining their own income irrespective of the income of their father*s? is generally referred to as the study of exchange mobility.

The importance of the distinction between structural and exchange mobility when analyzing a mobility structure is well known. Most sociological research on social mobility, at least since Rogoff (1953), has fruitfully employed this distinction in virtually all explanations of observed mobility data (see among others Duncan, 1966; Featherman et al., 1975; Goldthorpe, 1980; Hauser & Grusky, 1988; Hope, 1982; Sobel et al., 1998; Wong, 1990, for an overview). On the other hand, the economics literature has been much slower in appraising the distinction between structural and exchange mobility. In particular, the ubiquitous Galton's (1886) model of regression to the mean, which seems to be the workhorse of virtually all empirical studies of economic mobility (see e.g. Mulligan, 1997; Solon, 1999), does not allow an easy and explicit distinction between structural and exchange mobility.[1]

A convenient way to represent an intergenerational mobility structure (a joint distribution of fathers' and sons' income) is by means of so called *mobility tables*. As an example, consider the simplest case where there are only two values that fathers' income can take in a society, that is x_l (a given low level of income) and

x_h (a given high level of income). Similarly, let y_l and y_h be two levels of income (low and high), for the sons. Social mobility may then be described by means of a 2×2 mobility table (Example 1).

Sons → Fathers↓	y_l	y_h	fathers' marginal distribution
x_l	n_{ll}	n_{lh}	$n_{ll} + n_{lh}$
x_h	n_{hl}	n_{hh}	$n_{hl} + n_{hh}$
sons' marginal distribution	$n_{ll} + n_{hl}$	$n_{lh} + n_{hh}$	

Example 1. A General 2×2 Mobility Table.

In the table, n_{ij}, with $i, j = h, l$, denotes the number of families in the society with father belonging to category i and son to category j, with $\sum_i \sum_j n_{ij} = n$. Dividing n_{ij} by n, one obtains the relative frequency of the transition between the class i and j, which is an estimate of the probability of that transition. Mobility tables are sometimes expressed in terms of relative frequencies, rather than absolute. But for consistency with the display format used in the experiment, we maintain throughout the notation based on absolute frequencies.

Special cases of mobility tables, as described in the literature, are those of: (a) perfect immobility, where the elements outside the main diagonal are all zeros; (b) origin independence, sometimes also referred to as the case of equality of opportunities, where sons' position is statistically independent from fathers', that is $n_{ll} \times n_{hh} = n_{hl} \times n_{lh}$; and (c) perfect negative dependence, where all the elements outside the counter diagonal are zero.

Mobility data can be considered by various different perspectives and, as anticipated in the introduction, no simple or unanimous answer would in general be given to questions like "when a society is more mobile than another?" or "what would make a society more mobile than another?"

In this paper we follow the route which is most commonly used when analysing mobility tables, and is now the standard paradigm in the sociological literature on social mobility comparisons (see, e.g. Sobel et al., 1998): structural mobility is analyzed by the comparison of fathers' and sons' marginal distribution, while exchange mobility is analyzed by looking at the association structure of the mobility table. In particular, for a generic 2×2 mobility table as in Example 1, define the *odds ratio* as $or = (n_{ll}/n_{lh})/(n_{hl}/n_{hh})$. The odds ratio then describes the ratio between the odds of a son with a father with low income of remaining with low income rather than moving upwards, with respect to the odds that a son with a father with a high income has of becoming poor, rather than remaining rich. The odds ratio can thus be considered as a measure of association between individuals

of different social origin and is therefore an index of the rigidity in the society. Thus, for instance, a society where sons' position is statistically independent from fathers', the case of origin independence referred to above (case b), has an *or* equal to 1. We define this as the case of absence of association. On the other hand, rigid societies where fathers' and sons' incomes are positively associated have odds ratios grater than 1; while societies with negative association between fathers' and sons' incomes have odds ratio below one. Clearly, the case (a) above of perfect immobility is the situation of maximum positive association (highest possible *or*); while the case (c) of perfect negative dependence is the situation of maximum negative association (lowest possible *or*).

We emphasize that in the remaining of the paper we restrict attention to tables displaying non-negative association. The main reason for this restriction is that real world mobility data *never* display negative association between fathers' and sons' status.

Some qualifications regarding this restriction are nevertheless necessary. In particular, there is an important stream of economic literature that stresses the importance in assessing mobility of the degree to which fathers' and sons' ranks are reversed over time. For example, Atkinson (1981) and Atkinson and Bourguignon (1982) consider a framework where welfare is maximized by complete reversal, where all rich become poor and all poor become rich. In general, it is well known that there is a tension between the concept of mobility as origin independence and that of rank reversal (see Shorrocks, 1978, for a clear early axiomatic analysis, and Gottschalk & Spolaore, 2002, for a more recent treatment). On the other hand, within the tables displaying positive association it is easy to show that welfare is maximized by origin independence. This restriction will simplify our analysis by assigning maximum exchange mobility to structures with perfect origin independence ($or = 1$), ignoring the issues of reversal and negative association.

To better illustrate now the concepts of structural mobility and exchange mobility followed in this paper, and the role of odds ratios in assessing the latter, consider the three hypothetical societies S, T and U, of Example 2.[2]

Society **S**

	y_l	y_h
x_l	32	16
x_h	16	32

Society **T**

	y_l	y_h
x_l	32	30
x_h	8	30

Society **U**

	y_l	y_h
x_l	32	8
x_h	30	30

Example 2. 2×2 Tables with Different Structural Mobility and Same Exchange Mobility.

In comparing these tables, we assume for the time being that the four values of x_l, x_h, y_l and y_h, are equal across the tables. Society T is characterised by strong economic growth between generations (62% of the fathers, but only 40% of the sons, have low income). Society U, on the other hand, shows a general

impoverishment in the generational passage (60% of fathers, but only 38% of sons have high incomes). Society S, finally, is an example in which there is no change in the marginal distributions between generations. Thus, structural mobility is greater in both T and U than in society S. Which society, however, can be considered as having more exchange mobility?

Consider again society S: a son with a father with low income has twice as much the probability of remaining with low income than moving to high income. On the other hand, a son with the father with high income has half the probability of having low, rather than high, income.

Thus, its *or* is 4. Note that *or*s of societies U and T are also equal to 4. Therefore, one could argue that the three societies, while differing in terms of structural mobility,[3] are characterised by similar social rigidity in terms of equal positive association between the income of the fathers and the sons.

One can also have situations in which the opposite holds. It is easy to check that the following tables, namely S′ and T′, which display the same structural mobility (since they have the same marginal distributions), differ in terms of exchange mobility since T′ is an example of independence ($or = 1$), while in S′ there is positive association between incomes in the two generations ($or = 16$) (Example 3).

Socicty **S'**	y_l	y_h
x_l	40	10
x_h	10	40

Socicty **T'**	y_l	y_h
x_l	25	25
x_h	25	25

Example 3. 2×2 Tables with Same Structural Mobility and Different Exchange Mobility.

Thus, following the above discussion, we formulate hypotheses on structural mobility by looking at the distance between the fathers' and sons' marginal distributions; and formulate hypotheses on exchange mobility by looking at the odds ratios. In particular, the first two hypotheses that we will control in the experiment are:

H1. In pairwise comparisons between 2×2 tables having the same structural mobility, the table showing lower *odds ratio* has greater exchange mobility.

H2. In pairwise comparisons between 2×2 tables characterised by the same *odds ratios*, the table showing the greater difference in the margins has greater structural mobility.

Obviously, note that the two hypotheses introduce only a partial ordering for comparisons between mobility tables. Indeed, we emphasize that, while the hypotheses apply to more general comparisons than those involving the simplest cases of independence and perfect immobility, they are still quite restrictive in terms of the number of situations in which they can be used to obtain firms prediction. Notice, in particular, that H2 is silent even in comparisons between

tables like U and T, with the same *odds ratios*, and reversed marginal distributions between fathers' and sons' between the two tables (see also Section 3.1, below). Nevertheless, as emphasized above, we believe that the two hypotheses are at the cornerstone of the literature on social mobility measurement, and therefore they deserve particular attention. In line with this approach of considering the most basic principles underlying the theory of mobility measurement, we now focus on the two most obvious limitations of the discussion so far conducted.

The first limitation requiring attention concerns the effect on social mobility comparisons of a transformation of the fathers' and/or sons' incomes. Recall that we have so far discussed comparisons in which the values of x_l, x_h, y_l and y_h, were always equal across the tables. In this regard, we first of all note that for some scholars (see e.g. Fields, 2001, and references therein), as long as the values of incomes are expressed in the same unit across all tables, no transformations are allowed in mobility comparisons, because mobility is an absolute concept which refers to fathers' and sons' actual incomes.

Alternatively, different transformations may be proposed. In the experiment, we specifically test four well-known hypotheses of transformations that according to various literature (see again Fields, 2001) would not affect mobility comparisons. They are (also note the nested nature of the hypotheses):

H3. *Weakly relative invariance:* (x, y) has the same mobility (interpreted either as structural or as exchange mobility) as $(\alpha x, \alpha y)$ for all $\alpha > 0$;

H4. *Weakly affine invariance:* (x, y) has the same mobility as $(\alpha x + \beta, \alpha y + \beta)$ for all $\alpha > 0$ and all β;

H5. *Strongly relative invariance:* (x, y) has the same mobility as $(\alpha x, \gamma y)$ for all $\alpha > 0$ and $\gamma > 0$;

H6. *Strongly affine invariance:* (x, y) has the same mobility as $(\alpha x + \beta, \gamma y + \delta)$ for all $\alpha > 0$ and $\gamma > 0$ and all β and δ.

The second obvious limitation of the discussion underlying hypotheses H1 and H2 concerns the analysis of mobility tables of an order greater than 2×2. In this respect, we note that the study of the decomposition of mobility tables into parameters of structural mobility and parameters of exchange mobility, is an area of quite active research, stemming from the seminal work of McCullagh and Nelder (1989) on generalised linear models (see, e.g. Bartolucci et al., 2001; Sobel et al., 1998). In particular, Dardanoni and Forcina (2002) have recently argued that the relevant extension of the idea of *odds ratios* in mobility tables of an order greater than 2×2 is the so called *generalised odds ratios* (*gor*) (see Agresti, 1990, for a definition). *Gor*s essentially are odds ratios calculated for all possible 2×2 tables

that can be constructed from a $k \times k$ table by summing up families in adjacent categories. The following example shows how to construct, from a 3×3 mobility table (with the three levels of fathers' and sons' incomes, denoted as x_l, x_m and x_h, and y_l, y_m and y_h), four "2×2" tables by adjacent dichotomisations (Example 4).

	y_l	y_m	y_h
x_l	n_{ll}	n_{lm}	n_{lh}
x_m	n_{ml}	n_{mm}	n_{mh}
x_h	n_{hl}	n_{hm}	n_{hh}

T_1

n_{ll}	$n_{lm} + n_{lh}$
$n_{ml} + n_{hl}$	$n_{mm} + n_{mh} + n_{hm} + n_{hh}$

T_2

$n_{ll} + n_{lm}$	n_{lh}
$n_{ml} + n_{mm} + n_{hl} + n_{hm}$	$n_{mh} + n_{hh}$

T_3

$n_{ll} + n_{ml}$	$n_{lm} + n_{lh} + n_{mm} + n_{mh}$
n_{hl}	$n_{hm} + n_{hh}$

T_4

$n_{ll} + n_{lm} + n_{ml} + n_{mm}$	$n_{lh} + n_{mh}$
$n_{hl} + n_{hm}$	n_{hh}

Example 4. A General 3×3 Mobility Table with the Four 2×2 Tables Constructed by Dichotomisation.

*Gor*s are the standard *or*s calculated for the four tables. Intuitively, when comparing two tables of the same order, one can argue that if a table has all the *gor*s lower than the corresponding *gor*s of the other table, then in the former table there is more exchange mobility. Dardanoni and Forcina (2002) show that this criterion is in fact equivalent to the so called *positive quadrant dependence* discussed in the statistical literature and applied to social mobility measurement by Atkinson (1981) and Dardanoni (1993).

The decomposition naturally leads to reformulate H1 and H2 for cases of tables of an order greater than 2 as:

H1′. In pairwise comparisons between tables of the same order and having the same structural mobility, the table showing lower *generalised odds ratios* has greater exchange mobility.

H2′. In pairwise comparisons between tables of the same order and characterised by the same *generalised odds ratios*, the table showing the greater difference in the margins has greater structural mobility.

3. QUESTIONNAIRE DESIGN

Three versions of the questionnaire were prepared. For reasons explained below, we refer to the three versions as: the "2×2 relative" questionnaire, the "2×2 affine" questionnaire, and the "3×3" questionnaire. Participants to the three

questionnaires were, respectively, 148, 78, and 113 students, and they came from three different classes in economics at two Italian universities: the "2 × 2 relative" and the "3 × 3" questionnaires were conducted at the University of Varese, the "2 × 2 affine" questionnaire was administered at the University of Pavia.

All the three questionnaires contained pairwise comparisons between mobility tables. Both the "2 × 2 relative" and "2 × 2 affine" questionnaires presented nine questions involving two-by-two tables; the "3 × 3" version contained eight questions with three-by-three tables.

All questions asked to state which table in each comparison displayed, according to the respondent's view, a greater degree of social mobility. The format for a typical question is shown in Fig. 1.

Respondents could give four types of answers to each question: (a) they could say that social mobility was greater in the one society called Alphaland; (b) in the other society called Betaland; (c) they could say that the two societies had the same amount of social mobility; and (d) they could say that the two societies were not comparable.

The questionnaires were given to students with a statement of instructions, which was also read aloud at the start. This was a delicate step. On the one hand, social mobility is not something respondents may have immediate ideas or intuitions about: in fact, as emphasized throughout, even theorists do not provide a unified discourse of analysis. Therefore, instructions had to explain what the questionnaire was really about. On the other hand, our worry was that, if the instructions explained too much, we could have guided respondents towards certain answers, which was not the purpose.

We first gave a brief definition of what the questionnaire meant by "social mobility." We in particular defined social mobility as *the transition of socio-economic class within a family line, from the fathers' generation to the sons'*

Question 5. Imagine two societies, Alphaland and Betaland, with associated the following tables of social mobility

Alphaland

Sons' income → Fathers' income ↓	100	200	300
100	6	6	6
200	11	19	3
300	16	8	24

Betaland

Sons' income → Fathers' income ↓	100	200	300
100	6	6	6
200	11	11	11
300	16	16	16

In which society do you think there is more social mobility (answer by circling your opinion)?

(a) Alphaland
(b) Betaland
(c) The two societies have the same social mobility
(d) The social mobility of the two societies cannot be compared

Fig. 1. A Typical Question from the Questionnaire.

generation. Secondly, we gave a short explanation about how to read the numbers of a typical social mobility table. We insisted that the incomes in the two societies being compared were expressed in the same unit of account, and that there was no inflation moving from the fathers' generation to the sons' generation in either societies. We also noticed that the questionnaire was personal and anonymous. (The full set of instructions is reported in the Appendix.)

For the same purpose of trying not to guide answers, we also decided to use only numerical examples and to not include verbal versions of the principles to be tested, as it is sometimes done in areas characterised by more firmly established conventions (as for example in the field of income inequality measurement, e.g. Amiel & Cowell, 1992).

In the following section we will present the actual questions and the results from the 2×2 "relative" and "affine" questionnaires, which had a similar structure; we then move on to consider the "3×3" questionnaire.

3.1. The "2 × 2" Questionnaires

3.1.1. Design

Both the "2×2" questionnaires were divided into three parts: the first part compared mobility tables with identical supports for the income distributions of the fathers' and sons' generations in both societies; the second part presented tables with different supports for the income distributions for the fathers' and sons' generations and in both societies; the third part showed tables in which the supports of the income distributions were different between the fathers' and sons' generations, but was the same across the two societies. Participants to the questionnaire knew of this division since it was made explicit by a brief statement at the beginning of each part on the questionnaire.

Table 1 shows the actual mobility tables presented on the "2×2" questionnaires. The first part contained four questions. The questions were identical on the two questionnaires.[4]

Questions 1, 2 and 3 focus on exchange mobility. In Question 1 the marginal distributions of fathers' and sons' incomes are identical within each table. Structural mobility is thus silent in regard to this question. On the other hand, considering exchange mobility, Alphaland gives perhaps the clearer instance of a case of independence ($or = 1$) between the fathers' and sons' classes;[5] whereas Betaland shows the case of a strong association between the two classes ($or = 5.44$). The theoretical prediction of H1 is therefore that in Alphaland there is more exchange mobility.

A similar prediction holds for Question 2. Here Alphaland is an example of stochastic independence, though not a bistochastic table; Betaland is again a

Table 1. "2 × 2" Questionnaires.

Alphaland			Betaland			Theoretical Predictions
Fathers' Income	Sons' Income		Fathers' Income	Sons' Income		
Q.1						
	50	100		50	100	H1: Same structural
50	25	25	50	35	15	mobility, greater
100	25	25	100	15	35	exchange mobility in Alphaland
Q.2						
	50	100		50	100	H1: Same structural
50	21	49	50	27	43	mobility, greater
100	9	21	100	3	27	exchange mobility in Alphaland
Q.3						
	50	100		50	100	Indeterminate
50	25	25	50	9	21	
100	25	25	100	21	49	
Q.4						
	50	100		50	100	H2: Same exchange
50	16	14	50	27	43	mobility, greater
100	14	56	100	3	27	structural mobility in Betaland
Q.5 "relative"/"affine"						
	60	90		30/130	45/145	H3/H4: Same mobility
40	25	25	20/120	25	25	because of weak
80	25	25	40/140	25	25	"rel./aff." invariance
Q.6 "relative"/"affine"						
	120	300		240/190	600/550	H5/H6: Same mobility
100	25	25	150/200	25	25	because of strong
200	25	25	300/350	25	25	"rel./aff." invariance
Q.7 "relative"/"affine"						
	80	120		40/10	60/30	H5/H6: Same mobility
40	25	25	80/110	25	25	because of strong
60	25	25	120/150	25	25	"rel./aff." invariance
Q.8 "relative"/"affine"						
	25/75	50/100		25/75	50/100	H1: Greater exchange
100/50	21	49	100/50	27	43	mobility in Alphaland
200/150	9	21	200/150	3	27	H5/H6: Same mobility as in Q.2
Q.9 "relative"/"affine"						
	250/150	500/400		250/150	500/400	H5/H6: Same mobility as in Q.3
200/100	25	25	200/100	9	21	
400/300	25	25	400/300	21	49	

case in which there is positive association ($or = 5.65$) between the fathers' and sons' income classes. A further difference from the previous example is that in Question 2 there is an upward movement from the fathers' to the sons' generation, which is the same in the two societies (so that structural mobility is equal in the two tables).

Question 3 compares the two different forms of independence: Alphaland is the bistochastic example already met in Question 1; whereas Betaland is a different case of independence, in which both the fathers' and sons' generation are better off than in the bistochastic case. Therefore, the theoretical prediction is that the comparison is indeterminate.

Question 4 looks at structural mobility, testing H2. Specifically, in both societies there is a similar positive association between the fathers' and sons' income classes (the *or*s being 4.57 and 5.65, in Betaland and in Alphaland, respectively).[6] The marginal distributions are, however, different in the two societies: in Betaland there is an upward movement of the sons' income from the fathers', whereas in Alphaland the sons' and fathers' marginal distributions are equal.

The next three questions (Questions 5–7) look at invariance: the "2 × 2 relative" questionnaire to relative invariance; the "2 × 2 affine" questionnaire to affine invariance. Question 5 tests on both questionnaires the weak form of invariance (H3 and H4, respectively). In particular, the tables of Betaland are obtained as transformations of both fathers' and sons' incomes from the same Alphaland society. In the "relative" questionnaire, the transformation is $Y_{\beta\text{-land}} = 0.5 \times Y_{\alpha\text{-land}}$; in the "affine" questionnaire the transformation is $Y_{\beta\text{-land}} = 0.5 \times Y_{\alpha\text{-land}} + 100$. (In Table 1, a slash distinguishes the different transformations operated in the two questionnaires from the same table, namely Alphaland.)

Questions 6 and 7 focus on strong invariance (H5 and H6). Alphaland is again the same in the two questionnaires, whereas different transformations are used to obtain the tables for Betaland. On the "relative" questionnaire, the transformations in Question 6 are: $Y_{\beta\text{-land}} = 1.5 \times Y_{\alpha\text{-land}}$ and $Y_{\beta\text{-land}} = 2 \times Y_{\alpha\text{-land}}$ for the fathers' incomes and the sons' incomes, respectively; and the transformations for Question 7 are: $Y_{\beta\text{-land}} = 2 \times Y_{\alpha\text{-land}}$ and $Y_{\beta\text{-land}} = 0.5 \times Y_{\alpha\text{-land}}$ for fathers and sons, respectively. On the "affine" questionnaire, the transformations for fathers' and sons' are, in the order: $Y_{\beta\text{-land}} = 1.5 \times Y_{\alpha\text{-land}} + 50$ and $Y_{\beta\text{-land}} = 2 \times Y_{\alpha\text{-land}} - 50$, in Question 6; and $Y_{\beta\text{-land}} = 2 \times Y_{\alpha\text{-land}} + 30$ and $Y_{\beta\text{-land}} = 0.5 \times Y_{\alpha\text{-land}} - 30$, in Question 7. Notice that the design implies that in Question 6 both fathers' and sons' are better off moving from Alphaland to Betaland, while in Question 7 fathers are better off while sons are worse off.

The last two questions also test strong invariance, but indirectly. Question 8 replicates Question 2, where incomes in both societies have been transformed as follows: in the "relative" questionnaire, fathers' incomes are $Y_{Q8} = 2 \cdot Y_{Q2}$

Table 2. Results from the "2 × 2" Questionnaires.

	Answers	(a) Alphaland	(b) Betaland	(c) Same Mobility	(d) Not Comparable	H_0: $p(a) = p(b) = p(c + d)$	H_0: $p(a) = p(b)$	Hom. Test
Q. 1	"2 × 2 Relative"							
	n. 148	44	73	30	1	18.74^{****}	-2.58^{**}	0.80
	p.	0.30	0.49	0.20	0.01			
	"2 × 2 Affine"							
	n. 78	20	38	16	4	8.31^{**}	-2.23^{**}	
	p.	0.26	0.49	0.21	0.05			
Q. 2	"2 × 2 Relative"							
	n. 147	60	26	59	2	16.20^{****}	3.56^{***}	3.22
	p.	0.41	0.18	0.40	0.01			
	"2 × 2 Affine"							
	n. 78	34	20	22	2	4.00	1.77^{*}	
	p.	0.44	0.26	0.28	0.03			
Q. 3	"2 × 2 Relative"							
	n. 148	46	89	7	6	58.88^{****}	-3.61^{***}	0.75
	p.	0.31	0.60	0.05	0.04			
	"2 × 2 Affine"							
	n. 78	20	51	5	2	39.31^{****}	-3.56^{***}	
	p.	0.26	0.65	0.07	0.03			
Q. 4	"2 × 2 Relative"							
	n. 147	43	73	18	13	19.10^{****}	-2.69^{***}	1.63
	p.	0.29	0.50	0.12	0.09			
	"2 × 2 Affine"							
	n. 77	26	40	5	6	16.39^{****}	−1.60	
	p.	0.34	0.52	0.06	0.08			
Q. 5	"2 × 2 Relative"							
	n. 147	70	17	49	11	32.37^{****}	5.58^{****}	
	p.	0.48	0.12	0.33	0.07			

	"2 × 2 Affine"						
	n. 77	35	21	16	5	5.09*	1.74*
	p.	0.45	0.27	0.21	0.06		
Q. 6	"2 × 2 Relative"						
	n. 147	18	74	38	17	33.10****	−5.73****
	p.	0.12	0.50	0.26	0.12		
	"2 × 2 Affine"						
	n. 78	10	53	9	6	43.10****	−5.29****
	p.	0.13	0.69	0.12	0.08		
Q. 7	"2 × 2 Relative"						
	n. 147	38	15	76	18	67.39****	3.02***
	p.	0.26	0.10	0.52	0.12		
	"2 × 2 Affine"						
	n. 78	34	30	6	8	8.74**	0.38
	p.	0.44	0.39	0.08	0.10		
Q. 8	"2 × 2 Relative"						
	n. 144	50	29	57	8	13.63***	2.25**
	p.	0.35	0.20	0.40	0.06		
	"2 × 2 Affine"						
	n. 77	24	20	27	6	3.45	0.45
	p.	0.31	0.26	0.35	0.08		
Q. 9	"2 × 2 Relative"						
	n. 144	39	87	12	6	52.13****	−4.37****
	p.	0.27	0.61	0.08	0.04		
	"2 × 2 Affine"						
	n. 77	29	37	4	7	13.82***	−0.86
	p.	0.38	0.48	0.05	0.09		

*Rejection at 10% significance level.
**Rejection at 5% significance level.
***Rejection at 1% significance level.
****Rejection at 0.1% significance level.

and sons' income are $Y_{Q8} = 0.5 \times Y_{Q2}$; in the "affine" questionnaire the transformations are $Y_{Q8} = 2 \times Y_{Q2} - 50$ and $Y_{Q8} = 0.5 \times Y_{Q2} + 50$, for fathers and sons, respectively. Question 9 is a similar replication of Question 3, where the transformations in the two questionnaires are: $Y_{Q9} = 4 \times Y_{Q3}$ and $Y_{Q9} = 5 \times Y_{Q3}$, for fathers and sons, respectively, in the "relative" questionnaire; and $Y_{Q9} = 4 \times Y_{Q3} - 100$ and $Y_{Q9} = 5 \times Y_{Q3} - 100$ in the "affine" questionnaire. It is easy to show that H5 and H6 imply that participants should give in Questions 8 and 9 the same answers as in Questions 2 and 3, respectively.

3.2. Results

Table 2 shows the results from the two questionnaires. The first part of the Table presents the distributions of answers to each question; the second part reports the tests we have conducted. To explain the latter, we first of all note that given the novelty and the rather difficult task involved in the questionnaire, the primary purpose of the tests conducted was to look for some form of regularity or systematic patterns in the answers, and only then compare the consistency possibly found with the theoretical predictions.

In this perspective, we conducted two main tests. The first is a standard χ^2 test controlling whether the distribution of answers on each question might be considered as a purely random, as for example might be the case when the respondents either do not comprehend the concept of social mobility, or do not understand how to read a mobility table, or perhaps understand but make too much confusion and too many mistakes. In constructing the test, we in addition thought that for "confused" people an answer of type (d), namely that the tables are "not comparable," might in fact be viewed as a good substitute for an answer of type (c), that the tables have the "same mobility." We thus summed in the test the answers of the two types. Therefore, our first test assesses the degree of correspondence between the number of observed and expected responses in each of the category (a), (b) and (c + d), under the null hypothesis that all are equally likely.[7]

One may, however, consider this a rather weak test of randomness, since one may object that most people, even if "confused," would anyhow attempt a definite answer either for (a) or for (b). Therefore, as a second test, we check by a standard normal approximation of the binomial test whether the two are equally likely.[8]

On the first four questions, which are common to both the "relative" and "affine" questionnaires, we also conducted a test for the homogeneity of the distributions of the responses across the two questionnaires.[9] As argued below, this may also add to the evidence on the randomness vs. the regularity of the responses.

Overall, we believe that despite the difficulty of the task involved in the questionnaire, the evidence rejects the hypothesis that responses are given randomly. In fact, notwithstanding the variation in participants' opinions that the questionnaires document, there are various regularities across subjects' responses which we consider particularly interesting and to which we now address attention.

Looking at the answers to Question 1 we find that most subjects, in both questionnaires, consider that Betaland is the society showing greater mobility. This is a violation of H1, which is certainly most surprising; nevertheless, we anticipate that this is evidence consistently found throughout the whole questionnaire, as in fact also documented by the great homogeneity in the distributions of responses to the two questionnaires.[10]

In particular, it seems that Alphaland is not perceived as reflecting high social mobility even though it implies independence between the fathers' and sons' economic status, while Betaland implies a strong positive association between fathers' and sons' statuses. Also notice that this evidence is in contrast with other simpler rules that participants may use to assess mobility. Suppose, as for example suggested by one referee, that respondents simply compute the percentages of the sons that, having a low or a high-income father, end up with a high or a low income, respectively. Even in this case, they should have ranked Alphaland as more mobile than Betaland, as the percentages of sons who change position are 50% in Alphaland and 30% in Betaland. In general any existing concept of greater social mobility would rank Alphaland as the most mobile society. This is so because, since they are both bistochastic matrices, they have the same degree of structural mobility, but in Betaland there is a much greater tendency for sons to stay in the same social class as their fathers than in Alphaland.

Responses to Question 2 are more consistent with H1, since most subjects regard Alphaland (which is an example of independence, but not a bistochastic table), as showing grater mobility than Betaland. Notice, however, that the evidence is not very strong, as also a substantial proportion of participants rank the mobility in the two societies equally.

The evidence from Question 3 is stronger. Recall that this is an example in which both tables imply independence between the fathers' and sons' classes, the difference being that Alphaland is also bistochastic. As in Question 1, the latter characteristic seems to consistently be regarded contrary to social mobility, as the vast majority of participants rank Betaland more mobile than Alphaland. One possible interpretation here is that, even though in Betaland there is both the same amount of structural mobility as Alphaland (namely none) and the same amount of exchange mobility (since they both display independence), Betaland is a *richer* society than Alphaland in terms of stochastic dominance of the marginal and the

Table 3. Patterns of Responses in Questions 1 and 3.

Question 1	Question 3				Total
	(a) Alphaland	(b) Betaland	(c) Same Mobility	(d) Not Comparable	
"2 × 2 Relative" questionnaire					
(a) Alphaland	33	8	1	2	44
(b) Betaland	5	67	0	1	73
(c) Same mobility	7	14	6	3	30
(d) Not comparable	1	0	0	0	1
Total	46	89	7	6	148
"2 × 2 Affine" questionnaire					
(a) Alphaland	13	7	0	0	20
(b) Betaland	3	32	3	0	38
(c) Same mobility	4	9	2	1	16
(d) Not comparable	0	3	0	1	4
Total	20	51	5	2	78

dynastic income distributions. Respondents may be reacting to this asymmetry by declaring the latter society as more mobile.

To further understand the robustness of this evidence, and more generally the consistency of responses among the various categories, Table 3 shows the distributions of answers over Questions 1 and 3 in the two questionnaires. In particular, since the main regularity in the answers to the two questions when they are considered individually is the opposition to the bistochastic table Alphaland, we want to check whether the persons who oppose Alphaland in Question 1 also answer Betaland in Question 3. The results show that this is in fact the case in both questionnaires. In fact, more generally, the patterns show a consistency in the responses over the two questions, which we take as evidence confirming that the variation is subjects' responses cannot be simply imputed to randomness, but that it may in fact reflect a genuine difference in patterns of answers.[11]

The results from theses three questions leave open the issue whether H1 is rejected: on the one hand, the hypothesis is strongly rejected in Question 1, but not in Question 2; on the other hand, answers to Question 3 suggest that the anomaly of Question 1 may be due to the special nature of independence for bistochastic tables and its perception by participants. In any event, we comment further on this anomaly when we come to the evidence from the "3 × 3" questionnaire and in the conclusion.

In Question 4 an answer in favour of Betaland is consistent with the prediction of H2: the majority of participants, though not overwhelming, gives indeed such

type of answer. Once again, we note the homogeneity of the distribution of responses in the two questionnaires.

Question 5 focuses on weak invariance. In the "relative" questionnaire, only one third of participants agree with H3, that the two societies have equal mobility. Among the rest, the great majority judges Alphaland more mobile. This might be possibly explained into two ways: the first is that Alphaland is a society in which both the fathers' and sons' generations are richer than in Betaland; the second is that the income differences between the rich and the poor in both generations are greater in Alphaland than in Betaland.

The evidence from Question 5 on the "affine" questionnaire suggests that, at least for the majority of participants, the second explanation is more likely. In particular, recall that in the "affine" questionnaire Betaland is a richer society than Alphaland, though the income differences between the rich and the poor is still greater in it than in Betaland (see Table 1 again). Responses from the questionnaire show that the majority of the violations of invariance (only 21% of participants are consistent with the theoretical prediction of the same mobility) are also in favour of Alphaland (though, in this case, the difference of proportions is significant only at the 10% level).

Responses to Questions 6 and 7 add to the evidence on invariance, looking to strong invariance, namely the hypotheses H5 and H6. In Question 6, the income transformations from Alphaland to Betaland imply that in the "relative" and also in the "affine" questionnaires the latter society is both richer and the one with the greater difference between the poor and the rich. Thus, the two effects noted above operate in the same direction and the majority of violations to the predictions of equal mobility are by far for Betaland.

In Question 7, the transformations work differently. In the "relative" questionnaire, the fathers' and the sons' incomes are reversed from Alphaland to Betaland. In this case, interestingly, the majority of participants answer consistently with the strong relative invariance principle. In the "affine" questionnaire, the effect of the transformation is less clear: very few subjects are consistent with invariance, but the violations are not systematic, because almost an equal amount of subjects respond either Alphaland or Betaland.

Responses to the last two questions test strong invariance indirectly. Question 8 is a replica with transformed incomes of Question 2, in which (we recall), exchange mobility implies that Alphaland is the society with the greater mobility; Question 9 is instead a replica of Question 3, in which the theoretical prediction is indeterminate, but for which the evidence was that the majority of participants answered Betaland. For both pairs of Questions (2, 8) and (3, 9) and in both questionnaires, invariance alone implies the same distributions of responses. The evidence is consistent with invariance. This is also confirmed by Table 4 that

Table 4. Patterns of Responses in Questions (2, 8) and (3, 9).

Question 2	Question 8				Total
	(a) Alphaland	(b) Betaland	(c) Same Mobility	(d) Not Comparable	
Questions (2, 8) – "2 × 2 relative" questionnaire					
(a) Alphaland	32	9	13	4	58
(b) Betaland	4	12	8	1	25
(c) Same mobility	14	8	35	2	59
(d) Not comparable	0	0	1	1	2
Total	50	29	57	8	144
Questions (2, 8) – "2 × 2 affine" questionnaire					
(a) Alphaland	18	7	8	1	34
(b) Betaland	5	8	4	2	19
(c) Same mobility	1	5	15	1	22
(d) Not comparable	0	0	0	2	2
Total	24	20	27	6	77

Question 3	Question 9				Total
	(a) Alphaland	(b) Betaland	(c) Same Mobility	(d) Not Comparable	
Questions (3, 9) – "2 × 2 relative" questionnaire					
(a) Alphaland	29	4	7	2	42
(b) Betaland	10	74	3	2	89
(c) Same mobility	0	3	2	2	7
(d) Not comparable	0	6	0	0	6
Total	39	87	12	6	144
Questions (3, 9) – "2 × 2 affine" questionnaire					
(a) Alphaland	14	5	1	0	20
(b) Betaland	12	31	2	5	50
(c) Same mobility	2	1	1	1	5
(d) Not comparable	1	0	0	1	2
Total	29	37	4	7	77

shows the distributions of answers across the two pairs of questions in the two questionnaires.[12]

Overall, the tests of invariance indicate that invariance axioms should be taken with care, since in various contexts they seem to be rejected. In particular, subjects seem to take into account both the average level and the dispersion of income in the marginal distributions of fathers and sons when comparing the social mobility of two societies.

3.3. The "3 × 3" Questionnaire

3.3.1. Design

The questions posed in the "3 × 3" questionnaire are presented in Table 5. The structure of the questionnaire design is simple. The first three questions rotate three basic mobility tables. In Question 1 and Question 2, Alphaland is a bistochastic table; in the first question it is confronted with a table obtained by rearranging the number of subjects in the lower left corner of the table, while in the second it is confronted with a table obtained by rearranging the number in the lower right corner. These two transformations are well known in the literature (see Atkinson, 1981; Dardanoni, 1993; Shorrocks, 1978; Tchen, 1980) and are referred to as "off-diagonal" and "diagonalising" switches, respectively. Since both types of switches increase *gor*s while keeping marginal distributions unchanged,[13] both Questions 1 and 2 test H1′. Question 3 directly compares the tables with the two different kinds of switches, which by construction have *gor*s not uniformably comparable so that the answer is theoretically indeterminate.

The next three questions replicate the same structure of Questions 1, 2 and 3, in that they rotate three mobility tables: in Question 4 and 5, Alphaland is a table displaying independence (but not bistochastic); in the first question, it is compared with a table obtained by an "off-diagonal" switch, in the second with a table obtained as a "diagonalising" switching. The latter two tables are then directly compared in Question 6. The predictions for the three questions replicate those of the previous triple: more exchange mobility in Alphaland in both Questions 4 and 5, indeterminate in Question 6.

The two tables obtained as "off-diagonal" and "diagonalising" switches of Question 6 are also used in Questions 7 and 8 to test structural mobility, namely H2′. Betaland and Alphaland of Question 6 are compared in Question 7 and Question 8, respectively, with two different mobility tables, which in each case have a richer (in the sense of first order dominance) marginal distribution of sons' income and equal *gor*s. Thus, the latter tables, namely Alphaland in Question 7 and Betaland in Question 8, are in both cases more structurally mobile.

3.4. Results

The results of the "3 × 3" questionnaire are shown in Table 6. Overall, the results confirm some indications obtained in the "2 × 2" questionnaires, but also add new elements of interest. We, first of all, run the same tests to control for pure randomness in the responses. Once again, the general picture is that, despite the variation in the responses, the hypothesis of pure randomness is on the whole rejected.

Table 5. "3 × 3" Questionnaire.

Alphaland				Betaland				Theoretical Predictions
Fathers' Income	Sons' Income			Fathers' Income	Sons' Income			
Q. 1								
	100	200	300		100	200	300	H1′: Same structural
100	11	11	11	100	11	11	11	mobility, greater exchange
200	11	11	11	200	18	4	11	mobility in Alphaland
300	11	11	11	300	4	18	11	
Q. 2								
	100	200	300		100	200	300	H1′: Same structural
100	11	11	11	100	11	11	11	mobility, greater exchange
200	11	11	11	200	11	19	3	mobility in Alphaland
300	11	11	11	300	11	3	19	
Q. 3								
	100	200	300		100	200	300	Indeterminate
100	11	11	11	100	11	11	11	
200	11	19	3	200	18	4	11	
300	11	3	19	300	4	18	11	
Q. 4								
	100	200	300		100	200	300	H1′: Same structural
100	6	6	6	100	6	6	6	mobility, greater exchange
200	11	11	11	200	19	3	11	mobility in Alphaland
300	16	16	16	300	8	24	16	
Q. 5								
	100	200	300		100	200	300	H1′: Same structural
100	6	6	6	100	6	6	6	mobility, greater exchange
200	11	11	11	200	11	19	3	mobility in Alphaland
300	16	16	16	300	16	8	24	
Q. 6								
	100	200	300		100	200	300	Indeterminate
100	6	6	6	100	6	6	6	
200	11	19	3	200	19	3	11	
300	16	8	24	300	8	24	16	
Q. 7								
	100	200	300		100	200	300	H2′: Same exchange
100	4	6	10	100	6	6	6	mobility, greater structural
200	6	18	6	200	19	3	11	mobility in Alphaland
300	10	6	34	300	8	24	16	
Q. 8								
	100	200	300		100	200	300	H2′: Same exchange
100	6	6	6	100	4	6	10	mobility, greater structural
200	11	19	3	200	12	4	14	mobility in Betaland
300	16	8	24	300	4	20	26	

Table 6. Results from the "3 × 3" Questionnaire.

Answers	(a) Alphaland	(b) Betaland	(c) Same Mobility	(d) Not Comparable	H_0: p(a) = p(b) = p(c+d)	H_0: p(a) = p(b)
Q. 1						
n. 115	21	81	11	2	72.07****	−5.84****
p.	0.18	0.70	0.10	0.02		
Q. 2						
n. 115	30	71	12	2	45.10****	3.98****
p.	0.26	0.62	0.10	0.02		
Q. 3						
n. 115	29	40	40	6	3.88	−1.20
p.	0.25	0.35	0.35	0.05		
Q. 4						
n. 113	17	76	14	6	58.64****	−6.01****
p.	0.15	0.67	0.12	0.05		
Q. 5						
n. 113	36	57	14	6	18.28****	−2.07**
p.	0.32	0.50	0.12	0.05		
Q. 6						
n. 113	24	36	41	11	11.27***	−1.42
p.	0.21	0.32	0.37	0.10		
Q. 7						
n. 113	46	42	5	20	6.60*	0.32
p.	0.41	0.37	0.04	0.18		
Q. 8						
n. 113	17	72	4	20	47.59****	−5.72****
p.	0.15	0.64	0.04	0.18		

*Rejection at 10% significance level.
**Rejection at 5% significance level.
***Rejection at 1% significance level.
****Rejection at 0.1% significance level.

Answers to Questions 1 and 2 and to Questions 4 and 5 seem to unambiguously reject H1. In particular, in all cases the independence tables are always considered, by a statistically significant majority of respondents, as less mobile when compared with tables with the same marginal distributions, but displaying positive association. This is most unexpected, and, contrary to the case of the "2 × 2" questionnaire, the evidence against H1 is not limited to the special bistochastic case.

Indeed, taken together, the evidence from the two questionnaires shows anomalies that cannot be lightly dismissed. In particular, given the central role of stochastic independence in measuring exchange mobility, future research may specifically focuses on the conditions that are related to the violation of the two hypotheses.

In this respect, we also notice the evidence against H1 in the "3×3" questionnaire is stronger in Questions 1 and 4, where the comparisons involve the "off-diagonal" switching tables, rather than in Questions 2 and 5, respectively, which use the tables with the "diagonalising" switches.

Questions 3 and 6 provide a direct test whether subjects actually regard "off-diagonal" switches as leading to more social mobility than "diagonalising" switches. The answers show that in both questions there is a substantial proportion of participants who respond that either the two tables have the same mobility, or that are not comparable (the two types of responses (c) and (d) sum up to 40% of answers in Question 3 and to 47% in Question 6). Among those who give a straight answers for either table, we note that a tiny majority judges the "off-diagonal" switch table as more mobile in both questions.

Although in neither question the difference is statistically significant, we believe that the overall tendency of a part of the subjects to consider "off-diagonal" switches as carrying more mobility than "diagonalizing" switches is an evidence due to more than pure chance.

Responses to the last two questions on structural mobility, namely H2, show also elements of interest, which partially confirm the same tendency too. In particular, structural mobility implies that Alphaland in Question 7 should be considered more mobile than the "off-diagonal" switch table Betaland of Question 6; and similarly, in Question 8, Betaland is more structurally mobile than the "diagonalizing" switch table Alphaland of Question 6.

The evidence is consistent with H2 only in the latter case; while in the former, apparently, the attitude of some participants to consider the off-diagonal switching (Betaland) as a sign of social mobility partially offsets the implication of structural mobility, with the result that responses in Question 7 are almost perfectly evenly distributed between answers either for Alphaland or for Betaland.

4. CONCLUSIONS

Mobility measurement is a topic of great theoretical and practical importance. This paper, to the best of our knowledge, is the first attempt to consider mobility measurement from an experimental perspective. We have restricted our attention to intergenerational income mobility. Even in this special case, mobility measurement

is conceptually a very difficult task, because social mobility is really a multifaceted phenomenon that can be analyzed from many different viewpoints.

In this paper we have selected few issues which are crucial to understand social mobility measurement: (1) the difference between structural and exchange mobility; (2) the decomposition of mobility tables into parameters linked to structural mobility and parameters linked to exchange mobility; and (3) the effects of transformations of the status variables (incomes) on mobility comparisons.

These issues have been formalized as hypotheses that can be formally tested by the questionnaire. The results of the experiments we have conducted seem to show that answers cannot be considered as purely random: there are many systematic effects and regularities; but there is also a lot of variation in the data, which makes difficult to come up with definite answers.

In particular, one result from the questionnaire, which certainly needs a deeper scrutiny, is the apparent rejection of the hypothesis that, ceteris paribus, a table with positive association between fathers and sons incomes displays less mobility than a table where the variables are independent. This is certainly surprising; an analogy can be found with the apparent rejection of the Pigou-Dalton principle in inequality analysis recorded by Amiel and Cowell (1992).

Further investigations should reveal whether this anomaly is due to the lack of understanding of the properties of statistical independence,[14] to the refusal of the idea that greater positive association of income implies a more rigid society, or to the fact assessing social mobility exclusively through odd ratios ignores the potential tension between reversal of ranks and origin independence.

The evidence on structural mobility is more consistent with the theory, perhaps given the fact that this concept is easier to understand and more apparently visible in a mobility table. Still, we observed some substantial number of violations against H2: for example, in Question 4, only about 50% of the respondents got the right answers. Similarly, the evidence on the different invariance axioms typically invoked in the literature is not conclusive, and the recommendation for the time being is that caution should be exercised in their use.

Perhaps, one could possibly obtain firmer conclusions including, together with numerical examples, verbal statements of the principle to be tested, as it is sometimes done in experiments on income inequality. In such a case, however, the risk is to artificially induce less variation in the results. Indeed, in issues regarding social mobility, there may be a genuine large difference of opinions inherent to the multidimensional nature of the concept.

An alternative way to control for these differences could be obtained conducting a questionnaire using a “preference” frame, rather than a “measuring” frame. For example, one could use a question format asking: “which society would you

choose to live in behind a veil of ignorance?" This might also help the evaluation of the welfare implications of social mobility.

Ultimately, we believe that only more experimental research can answer some of the above questions and enhance our understanding of the various factors that influence social mobility comparisons and in its perception by individuals.

NOTES

1. A significant exception is the economics literature on mobility measurement, typically axiomatically based, which, since Markandya's (1982) initial use of the structural/exchange mobility distinction, now routinely uses this paradigm for the social mobility analysis. See for example Fields and Ok (1999) for an excellent overview of this literature.

2. The examples are from Checchi and Dardanoni (2003b).

3. Note that the societies also differ in terms of number of families (in particular $n = 100$ in T and U, while $n = 96$ in S). Since, however, are relative frequencies which matter both in assessing marginal distributions and *or*s, even if tables are expressed in terms of absolute frequencies, the difference in n is irrelevant in assessing mobility in the various societies.

4. The actual order in which the questions were presented on the questionnaire was the same as in Table 1. The position and the name of the societies were instead randomized.

5. In particular, in this table one need not to calculate the conditional probability to immediately ascertain statistical independence. This is always the case for bistochastic mobility tables.

6. All the tables in this experiment have been constructed by a MATLAB program which takes as inputs the marginal distributions and the association parameters (*or* in the "2×2" questionnaires and *gor*s in the "3×3" questionnaire) and gives as output a mobility table. The tables actually included in the questionnaire are obtained by rounding the output tables to the nearest integer, and this explains the slight difference in the *or*s used in this question.

7. We also conducted tests in which answers of the types (c) and (d) are not added up, with similar overall results. In fact, using 4 categories rather than 3, given that very few respondents choose (d), we obtain even stronger evidence against the hypothesis that answers are randomly distributed (see below).

8. Notice that the values of the test reported in Table 2 are corrected for continuity (see e.g. Siegel & Castellan, 1988). Significance levels are for two-tailed test.

9. The homogeneity test is based on a standard chi-square test statistics (see e.g. Harrison & Seidl, 1994; Siegel & Castellan, 1988, for a discussion of problems of non-homogeneity which may sometime arise in questionnaire experiments). The null hypothesis is that the distributions of responses in the two questionnaires are drawn from the same sample. The larger the value of the test, the less likely the null is to be true.

10. We cannot reject the hypothesis of homogeneity between responses to the two questionnaires at any significance levels; see the last column of Table 2.

11. In this respect, it is perhaps also worthwhile noticing that among those who give different answers over the two Questions (1, 3), very few give the opposite responses (Alphaland, Betaland) or (Betaland, Alphaland), but more often report answers which may be considered to belong to more adjacent categories, like (Same mobility, Alphaland) or (Same mobility, Betaland).

12. The various figures reported in Table 4 refer to the participants who answered both questions of each pair (2, 8) and (3, 9) in both questionnaires.

13. In all this questionnaire, the 4 *gor*s given as input to the MATLAB program to construct the tables with the "off- diagonal" and "diagonalizing" switches are, respectively (1, 1, 5, 1) and (1, 1, 1 5), where the order of the *gor*s is that of the theoretical tables T_1, T_2, T_3 and T_4 in Section 2. (*Gor*s in the actual tables of the questionnaire may be slightly different from the input values due to rounding.)

14. In this respect, we note that there are various experiments showing that, even in simple individual decision making problems, subjects often fail to recognize fundamental statistical concepts and apply basic statistical tools, like for example in computing compound probabilities down of the various branches of a decision tree. (See Conlisk, 1996, for a discussion of the role of bounded rationality in economics.)

15. The instructions for the "3×3" questionnaire were similar.

ACKNOWLEDGMENTS

Financial support from "Ministero dell'Università e della Ricerca Scientifica" (Miur) is gratefully acknowledged by both authors. We thank colleagues Gianluca Colombo, Giovanna Gavana and Francesco Sacco, who let us conduct the questionnaires during their lectures. Thanks to the students who answered the questions. We benefited from comments and suggestions of two referees and the editor.

REFERENCES

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Alesina, A., & La Ferrara, E. (2001). Preferences for redistribution in the land of opportunities. NBER Working Paper No. 8267.

Amiel, Y. (1999). The measurement of income inequality. The subjective approach. In: J. Silber (Ed.), *Handbook of Inequality Measurement*. Dordrecht: Kluwer Academic Publishers.

Amiel, Y., & Cowell, F. A. (1992). Measurement of income inequality. Experimental test by questionnaire. *Journal of Public Economics*, *47*, 3–26.

Atkinson, A. (1981). The measurement of economic mobility. In: A. B. Atkinson (Ed.), *Essays in Honor of Jan Pen*, Reprinted in *Social Justice and Public Policy*. Brighton: Wheatsheaf Books Ltd. (1983).

Atkinson, A., & Bourguignon, F. (1982). The comparison of multidimensional distributions of economic status. *Review of Economic Studies*, *49*, 183–201.

Bartholomew, D. J. (1996). *The statistical approach to social measurement*. New York: Academic Press.

Bartolucci, F., Forcina, A., & Dardanoni, V. (2001). Positive quadrant dependence and marginal modelling in two-way tables with ordered margins. *Journal of the American Statistical Association*, *96*, 1497–1505.

Benabou, R., & Ok, E. (2001). Social mobility and the demand for redistribution. *Quarterly Journal of Economics*, *116*, 447–487.

Breen, R. (1997). Inequality, economic growth and social mobility. *The British Journal of Socioloy*, *48*, 429–449.

Checchi, D., & Dardanoni, V. (2003a). Mobility comparisons: Does using different measures matter? *Research on Economic Inequality*, *9*, 113–145.

Checchi, D., & Dardanoni, V. (2003b). Mobilità sociale. *Enciclopedia Treccani* (forthcoming).

Conlisk, J. (1996). Why bounded rationality? *Journal of Economic Literature*, *36*, 669–700.

Dardanoni, V. (1993). Measuring social mobility. *Journal of Economic Theory*, *61*, 372–394.

Dardanoni, V., & Forcina, A. (2002). Mobility tables comparisons. Mimeo.

Duncan, O. D. (1966). Methodological issues in the analysis of social mobility. In: N. J. Smelser & S. M. Lipset (Eds), *Social Structure and Mobility in Economic Development*. New York: Aldine.

Featherman, D., Jones, L., & Hauser, R. (1975). Assumptions of social mobility in the U.S.: The case of occupational status. *Social Science Research*, *4*, 329–360.

Fields, G. S. (2001). *Distribution and development*. Cambridge, US: MIT Press.

Fields, G. S., & Ok, E. (1999). The measurement of Income mobility: An introduction to the literature. In: J. Silber (Ed.), *Handbook of Inequality Measurement*. Dordrecht: Kluwer Academic Publishers.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, *15*, 246–263.

Goldthorpe, J. (1980). *Social mobility and class structure in modern Britain*. Oxford: Oxford University Press.

Gottschalk, P., & Spolaore, E. (2002). On the evaluation of economic mobility. *Review of Economic Studies*, *69*, 191–208.

Harrison, E., & Seidl, C. (1994). Acceptance of distributional axioms: Experimental findings. In: W. Eichorn (Ed.), *Models and Measurement of Welfare and Inequality*. Berlin: Springer-Verlag.

Hauser, R. M., & Grusky, D. B. (1988). Cross-national variation in occupational distributions, relative mobility chances, and intergenerational shifts in occupational distributions. *American Sociological Review*, *53*, 723–741.

Hope, K. (1982). Vertical and nonvertical class mobility in three countries. *American Sociological Review*, *47*, 99–113.

Markandya, A. (1982). Intergenerational exchange mobility and economic welfare. *European Economic Review*, *17*, 307–324.

McCullagh, P., & Nelder, J. (1989). *Generalised linear models* (2nd ed.). New York: Chapman & Hall.

Moyes, P., Seidl, C., & Shorrocks, A. (2002). Inequalities: Theories, experiments and applications. *Journal of Economics*. New York: Spriner-Verlag.

Mulligan, C. (1997). *Parental priorities and economic inequality*. Chicago: University of Chicago Press.

Prais, S. J. (1955). Measuring social mobility. *Journal of the Royal Statistical Society*, *118*, 56–66.

Rogoff, N. (1953). *Recent trends in occupational mobility*. Glencoe: Free Press.

Shorrocks, A. F. (1978). Measurement of mobility. *Econometrica*, *46*, 1013–1024.

Siegel, S., & Castellan, N. J. (1988). *Non parametric statistics for behavioral sciences*. New York: McGraw-Hill.

Sobel, M., Becker, M., & Minick, S. (1998). Origins, destinations and association in occupational mobility. *American Journal of Sociology*, *104*, 687–721.

Solon, G. (1999). Intergenerational mobility in the labor market. In: O. Ashenfelter & D. Card (Eds), *Handbook of Labor Economics* (Vol. 3c). Amsterdam: North Holland.

Stokey, N. (1998). Shirtsleeves to shirtsleeves: The economic of social mobility. In: N. L. Schwartz, D. Jacobs & E. Kalai (Eds), *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures 1983–1997*. Cambridge: Econometric Society Monographs.

Tchen, A. (1980). Inequalities for distributions with given marginals. *Annals of Probability*, *8*, 814–827.
Wong, R. S. K. (1990). Understanding cross-national variation in occupational mobility. *American Journal of Sociology*, *55*, 560–573.
Yishay, M., & Moav, O. (1999). Intergenerational mobility and the process of development. *The Economic Journal*, *109*, 677–697.

APPENDIX: THE INSTRUCTIONS FOR THE "2 × 2" QUESTIONNAIRE[15]

A Questionnaire on "Social Mobility"

Thank you, Professor (class instructor's name) and ladies and gentlemen for participating to the following questionnaire. The questionnaire is about social mobility. Social mobility is defined as the transition of socio-economic class within a family line, from the fathers' generation to the sons' generation.

The social mobility of a given society can be represented by a simple table, called *social mobility table.* For example, imagine a hypothetical society with a given currency. Imagine that there is no inflation and that the currency has constant purchasing power. In the society there are 100 people, the fathers, who give birth to an equal number of individuals, the sons.

Assume that fathers' incomes can take one of two alternative values: either 60 or 150; and that sons' income can also take two possible values: either 40 or 170. In a table of social mobility, rows are for fathers' incomes and columns are for sons' incomes. The number in each cell refers to the number of families whose father has the income reported on the corresponding row and whose son has the income shown on the corresponding column. The table below is an example of a society where: 35 families have father with an income of 60 and son with an income of 40, 10 families have father with an income of 150 and son with an income of 40, 25 families have father with an income of 60 and son with an income of 170, and 30 families have father with an income of 150 and son with an income of 170.

An example of a table of social mobility

Fathers' Income	Sons' Income	
	40	170
60	35	25
150	10	30

Social mobility is an active area of research in economics, sociology, political science and statistics. Various scholars have, however, different views about how to measure social mobility.

With the questionnaire, we intend to know people's view on the issue.

In the questionnaire you will face few comparisons of pairs of mobility tables of hypothetical societies. In each pairwise comparison you are asked to state which society, according your view, has the greater degree of social mobility. If you think the two societies have the same degree of social mobility, or if you think that the social mobility of the two societies cannot be compared, you can give such answer at the bottom of each question.

Please answer by your own, without discussion with friends or neighbours; please do not give your name; and please remain quiet when you are done so that others can concentrate. In a later class, you will be offered a handout, which explains the motivation for the individual questions and the pattern of your answers. Thanks again for participation.

SOCIAL WELFARE, THE VEIL OF IGNORANCE AND PURELY INDIVIDUAL RISK: AN EMPIRICAL EXAMINATION

Kristof Bosmans and Erik Schokkaert

ABSTRACT

We present the results of a questionnaire study with Belgian undergraduate students as respondents. We consider the relationship between people's direct ethical preferences, their preferences behind a veil of ignorance, and their purely individual risk preferences over income distributions. The results reveal that, although there are important similarities between the three types of preferences, the first and third types form two extremes, while the second type lies in between the other two. Consistency of response patterns with the expected utility (EU) and rank-dependent expected utility (RDEU) models – natural analogues of the social welfare functions most frequently used in the literature on inequality and social welfare – is tested as well. For all three types of preferences the results reveal that, in the considered context, the RDEU model does not add explanatory power to the EU model. However, preferences appear to be relatively well described by some of the basic concepts from non-expected utility theory not usually considered in the income distribution literature.

Inequality, Welfare and Income Distribution: Experimental Approaches
Research on Economic Inequality, Volume 11, 85–114

ISSN: 1049-2585/doi:10.1016/S1049-2585(04)11005-3

1. INTRODUCTION

The central problem of distributive justice is finding an ethical ranking of income distributions. It is generally accepted that such an ethical ranking should reflect in a certain sense the preferences of an impartial and sympathetic observer (henceforth referred to as "ISO preferences") – "... a person taking a positive sympathetic interest in the welfare of *each* participant but having no partial bias in favor of *any* participant" (Harsanyi, 1977, p. 49). ISO preferences have been analysed in the literature in many different ways, but a particularly influential approach has been the exploration of the formal links between inequality and risk (Cowell & Schokkaert, 2001). This link has been put forward in its most explicit form in Harsanyi's (1953, 1955) approach of the veil of ignorance.[1]

Harsanyi rephrases the problem of distributive justice as a problem of individual decision making under risk: income distributions should be ranked according to the preferences of a rational individual behind the veil of ignorance (henceforth, "VOI preferences"). VOI preferences are the preferences over income distributions of a rational individual who does not know her own position in each income distribution (nor the position of the other members of society) and has (like these other members), for each income distribution, an equal probability of ending up with the income of any of the members of society. Harsanyi argues that rationality requires that VOI preferences be consistent with expected utility (EU) theory. By consequence, the social welfare function, which represents ethical preferences, inherits the formal properties of the EU model and is of the mean utilitarian type.[2] This approach is often seen as providing a justification for the most frequently used social welfare function in the income distribution literature, which is of the mean utilitarian form with utility a function exclusively of own income and an identical utility function for each individual.[3] However, this approach raises two sets of questions.

First, it is not obvious that VOI preferences and ISO preferences indeed coincide. The idea of the veil of ignorance is only one among many proposed approaches to the problem of finding an ethical ranking of income distributions. Moreover, the assumption that utility is a function exclusively of own income does not follow directly from Harsanyi's conditions. Indeed, VOI preferences are preferences over lotteries that have complete income distributions as outcomes, not preferences over lotteries with individual incomes as outcomes. We refer to the latter type of preferences as purely individual risk preferences (henceforth, "PIR preferences"). The assumption that utility is a function exclusively of own income can be justified if VOI preferences are identical to PIR preferences. Differences between VOI preferences and PIR preferences can result from the fact that the individuals do not care only about their own incomes, but also for instance about overall

equality or about their own relative income position. A comparison of ISO and VOI preferences with PIR preferences therefore could give some insight into the importance of externalities. What is the exact relationship between ISO, VOI and PIR preferences?

Second, the risk literature has provided ample empirical evidence of systematic violations of EU theory (the Allais paradox being the most famous example). A theoretical literature on non-expected utility (non-EU) models has developed mainly to accommodate these empirical violations.[4] It seems interesting to check whether these violations of EU theory for PIR preferences are also relevant for the ethical ranking of income distributions, that is, for ISO and VOI preferences. In fact, one of the most popular concepts from the non-EU literature, i.e. rank-dependent expected utility (RDEU), has in its simplified form (Yaari, 1987) received considerable attention in the income distribution literature because it provides a normative basis for an important subclass of the class of generalized Gini inequality indices.[5] Recent contributions have explored further links between the RDEU model in its general form and the measurement of inequality (Gajdos, 2001). How attractive are these non-EU approaches from an ethical point of view?

The present paper examines both issues through a questionnaire approach with Belgian students. We want to check whether their intuitions coincide with the formal approaches used by economists. In order to benefit from the accumulated knowledge in the risk literature, the set-up of our questionnaire will be analogous to the conventional approach used in that literature. We put respondents into three different choice contexts allowing revelation of ISO, VOI and PIR preferences, respectively. In each of these cases we test whether we discover any violations of the standard properties of the EU model. Such violations can also raise doubts about some of the standard assumptions in the literature on income distribution. Moreover, we will also check the empirical relevancy of the Yaari and RDEU models as well as that of some more basic non-EU concepts.

The questionnaire approach has recently become more popular in the economic literature on distributive justice. It has been used extensively for testing the acceptance of the crucial axioms from the literature on income distribution.[6] Recent work has explicitly compared the acceptability of these axioms for the income inequality and the risk setting (Amiel & Cowell, 2002; Amiel, Cowell & Polovin, 2001). Traub, Seidl and Schmidt (2003) and Camacho-Cuena, Seidl and Morone (2003) have run experiments in which subjects get material incentives to rank either income distributions or risky prospects. The close relationship between social welfare judgements and choice under risk and the theoretical suppositions of the EU approach are far from evident for large groups of respondents. Closest related to our work is a questionnaire study by Bernasconi (2002). He also checks the relevance of EU axioms for ISO, VOI and PIR preferences. The formulation

of our questions is very different, however, and we go further in testing explicitly some non-EU alternatives. Despite these differences, some of our results turn out to be similar to those of Bernasconi.

The paper is organized as follows. Section 2 gives an overview of relevant findings from EU theory and non-EU theory and links these to the evaluation of income distributions. The actual questionnaire study is presented in Section 3. In Section 4, we present the results. Section 5 concludes.

2. (NON-)EXPECTED UTILITY THEORY AND THE EVALUATION OF INCOME DISTRIBUTIONS

We first consider EU theory (Section 2.1) and some basic concepts from non-EU theory (Section 2.2). In Section 2.3, we summarize the basic characteristics of the RDEU model and of Yaari's theory. Finally, in Section 2.4, we return to the evaluation of income distributions.

We use the following notation. The set of incomes is $X = \{x_1, x_2, \ldots, x_n\}$, where the incomes are indexed such that $x_1 \leq \cdots \leq x_n$. An income distribution is a vector $\mathbf{p} = (p_1, \ldots, p_n)$ with $p_i \in [0, 1]$ for all i and $\sum_{i=1}^{n} p_i = 1$, where p_i is the proportion of the population with income x_i. In the case of individual decision under risk, income distributions have to be interpreted as lotteries, where p_i is the probability of outcome x_i. Preferences over alternatives, either income distributions or lotteries, are captured by a binary relation $\succeq$ ("is at least as good as"). The relation has an asymmetric factor $\succ$ ("is better than"), and a symmetric factor $\sim$ ("is equally good as"). Under certain conditions, a function, F, can be used to represent preferences. The function F has to be interpreted either as a social welfare function or as an individual utility function, depending on the given choice situation.

A convenient representation to compare the implications of EU theory with the implications of various non-EU theories is the so-called Marschak-Machina triangle[7] (see Fig. 1). Focusing on lotteries with only three possible outcomes (or income distributions with only three income levels) $x_1 < x_2 < x_3$, each alternative can be written as a pair (p_1, p_3), with p_2 determined implicitly as $p_2 = 1 - p_1 - p_3$. Since, furthermore, for $i = 1, 2, 3$, it holds that $p_i \in [0, 1]$, all these alternatives are points in the triangle $\{(p_1, p_3) \in \mathbb{R}^2_+ \mid p_1 + p_3 \leq 1\}$. In the Marschak-Machina triangle of Fig. 1, the different points represent thirteen possible alternatives. Our questionnaire study will focus on eight pairwise choices: each choice problem, $j = 1, \ldots, 8$, involves a choice among a pair of alternative lotteries or income distributions $(\mathbf{a}_j, \mathbf{b}_j)$. Note that the dotted lines connecting each of these pairs of

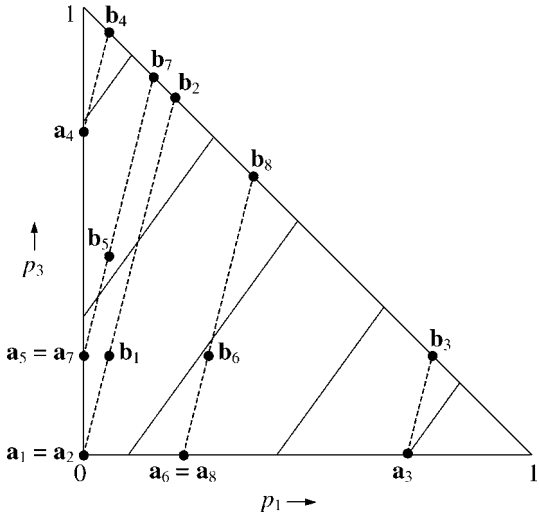


Fig. 1. EU Indifference Curves in the Marschak-Machina Triangle.

alternatives have the same slope equal to four. The probabilities corresponding to the specific options represented in Fig. 1 are shown in Table 1.

2.1. Expected Utility Theory

Let us first summarize in a loose way the basic idea of expected utility (EU) theory. Suppose that all the alternatives can be ordered (implying that the preference relation is reflexive, transitive and complete) and that this ordering is continuous and monotonic. Suppose moreover that the following condition holds:

Table 1. The Choice Pairs (p_1, p_2, p_3).

Question	**a**	**b**
1	(0, 1, 0)	(0.05, 0.75, 0.2)
2	(0, 1, 0)	(0.2, 0, 0.8)
3	(0.75, 0.25, 0)	(0.8, 0, 0.2)
4	(0, 0.25, 0.75)	(0.05, 0, 0.95)
5	(0, 0.8, 0.2)	(0.05, 0.55, 0.4)
6	(0.2, 0.8, 0)	(0.25, 0.55, 0.2)
7	(0, 0.8, 0.2)	(0.16, 0, 0.84)
8	(0.2, 0.8, 0)	(0.36, 0, 0.64)

Independence. For any alternatives $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$ and any scalar $\alpha \in (0, 1)$, it holds that $\mathbf{p} \succeq \mathbf{q}$ if and only if $\alpha\mathbf{p} + (1 - \alpha)\mathbf{r} \succeq \alpha\mathbf{q} + (1 - \alpha)\mathbf{r}$.

Under these assumptions, preferences over alternatives can be represented by

$$F(\mathbf{p}) = \sum_{i=1}^{n} p_i u(x_i), \tag{1}$$

where u is a strictly increasing function. This condition on u ensures monotonicity, which means that (first order) stochastically dominating alternatives are preferred. Strong risk aversion, implying that mean preserving spreads are disapproved, requires that u be strictly concave.

Expression (1) has very strong implications for alternatives in the triangle diagram. In fact, it is immediately clear that the slope of the implied indifference curves is

$$\left.\frac{\mathrm{d}p_3}{\mathrm{d}p_1}\right|_{F=\bar{F}} = \frac{u(x_2) - u(x_1)}{u(x_3) - u(x_2)}, \tag{2}$$

which is constant (since the incomes x_1, x_2 and x_3 are given for all points in the triangle) and positive (since under monotonicity $u(x_3) > u(x_2) > u(x_1)$).

Positivity of the slope of indifference curves is a general property of preference theories that respect monotonicity. Note that monotonicity also implies that indifference curves lying more to the northwest correspond to higher preference. For any point $\mathbf{p}$ in the triangle, the set of points strictly to the northwest of $\mathbf{p}$ (that is, all points $\mathbf{q}$ such that $q_1 \leq p_1$ and $q_3 \geq p_3$, with at least one of the inequalities strict) constitutes the set of points strictly stochastically dominating $\mathbf{p}$.

The important distinguishing implication of EU theory, however, is the fact that the slope of these indifference curves is constant.[8] Thus, in EU theory, indifference curves are parallel straight lines. The continuous lines in Fig. 1 represent such a set of EU indifference curves. One number, the value of the constant slope, determines the preferences over the entire triangle diagram. The figure shows that this feature severely restricts the number of response patterns allowed. In fact, EU theory implies that respondents choose consistently either **a** or **b** or are indifferent in each of the eight choice pairs. With the indifference curves drawn in the figure, this choice should be **b**. With a larger value for the slope it could be indifference or **a**. Note that in EU theory the slope can be seen as a kind of measure for the degree of risk aversion – in a choice between a certain lottery and a risky one, such as in pairs 1 and 2 in the figure, the certain one is chosen only for sufficiently high values of the slope.[9]

2.2. Some Basic Concepts from Non-Expected Utility Theory

The well-known problems discovered by Allais (1953) offer an important challenge to the restrictive implications of EU theory. The first three choice pairs in Fig. 1 are selected so as to illustrate these problems. Allais' "common consequence effect" (also known as the Allais paradox) suggests a tendency for choosing **a** in choice pair 1 and **b** in choice pair 3, thus violating EU theory. Allais' "common ratio effect" concerns a tendency for choosing **a** and **b**, respectively, in choice pairs such as 2 and 3, again violating EU theory. There is by now overwhelming experimental evidence for the empirical relevancy of both predictions (Camerer, 1995; Starmer, 2000).

One solution for "explaining" the Allais problems is dropping the assumption of parallel indifference curves. In fact, Machina (1982) introduced for that purpose the notion of *fanning-out*. In its pure form, fanning-out represents a monotonic increase in the slope of indifference curves as one moves northwest in the triangle (Fig. 2). More specifically, it says that, given any two points **p** and **q** in the triangle, such that **q** lies to the northwest of **p** (that is, **q** stochastically dominates **p**), the slope in point **q** has to be at least as high as that in **p**. For the choice pairs in the figures, fanning-out has the following implications: given any two choice pairs k and l, if $\mathbf{a}_l$ stochastically dominates $\mathbf{a}_k$ and $\mathbf{b}_l$ stochastically dominates $\mathbf{b}_k$, then the choice of alternative **a** from pair k implies that alternative **a** has to be chosen

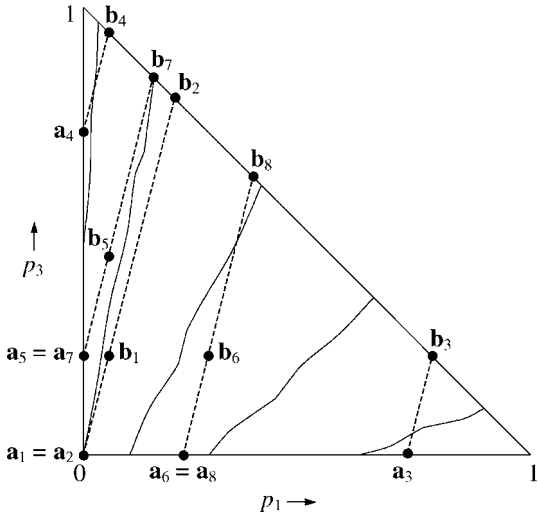


Fig. 2. Fanning-out.

from pair l as well, indifference in pair k implies that either alternative **a** has to be chosen from pair l or that one has to be indifferent between the alternatives of l. Fanning-out therefore accounts for the dominant behaviour in situations such as those suggested by Allais.

Empirical research, however, sometimes reveals the opposite pattern: that of *fanning-in* (see, e.g. Battalio, Kagel & Jiranyakul, 1990). In that case, the slope of the indifference curves becomes smaller as one moves to stochastically dominating alternatives. For the choice pairs in the figures, fanning-in has the following implications: given any two choice pairs k and l, if $\mathbf{a}_l$ stochastically dominates $\mathbf{a}_k$ and $\mathbf{b}_l$ stochastically dominates $\mathbf{b}_k$, then the choice of alternative **b** in k implies that alternative **b** has to be chosen in l as well, indifference in k implies that either alternative **b** has to be chosen in l or that one has to be indifferent between the alternatives of l.

Both fanning-out and fanning-in deal with a change in slope as one moves to different indifference curves (at least when preferences satisfy monotonicity). The research on extensions of EU theory has also focused on the relevancy of the linearity of the indifference curves implied by expressions (1) and (2). Three different assumptions have been proposed:

Betweenness. For any alternatives **p** and **q** and any scalar $\alpha \in (0, 1)$, it holds that $\mathbf{p} \succeq \mathbf{q}$ if and only if $\mathbf{p} \succeq \alpha\mathbf{p} + (1 - \alpha)\mathbf{q} \succeq \mathbf{q}$.

Quasi-convexity. For any alternatives **p** and **q** and any scalar $\alpha \in (0, 1)$, $F(\alpha\mathbf{p} + (1 - \alpha)\mathbf{q}) \leq \max\{F(\mathbf{p}), F(\mathbf{q})\}$.

Quasi-concavity. For any alternatives **p** and **q** and any scalar $\alpha \in (0, 1)$, $F(\alpha\mathbf{p} + (1 - \alpha)\mathbf{q}) \geq \min\{F(\mathbf{p}), F(\mathbf{q})\}$.

Betweenness obviously is an implication of independence. It implies that, if $\mathbf{p} \sim \mathbf{q}$, then for any scalar $\alpha \in (0, 1)$ it holds that $\mathbf{p} \sim \alpha\mathbf{p} + (1 - \alpha)\mathbf{q} \sim \mathbf{q}$, which means that indifference curves are straight lines – but not necessarily parallel. Betweenness implies neutrality to mixtures of alternatives on the same indifference curve. Straightforward extensions are concave indifference curves (corresponding to the assumption of quasi-convexity), describing mixture aversion, and convex indifference curves (corresponding to the assumption of quasi-concavity), describing mixture proneness. The latter case is illustrated in Fig. 3. Betweenness, quasi-convexity and quasi-concavity have implications for the combinations of choice pairs (1, 2), (5, 7), and (6, 8) in the figures. In each of those combinations, the only response patterns consistent with betweenness are **aa**, **bb** and $\sim\sim$. Quasi-convexity allows, in addition to the betweenness patterns, **ab**, $\mathbf{a}\sim$ and $\sim\mathbf{b}$. Quasi-concavity, on the other hand, allows, in addition to the betweenness patterns, **ba**, $\mathbf{b}\sim$ and $\sim\mathbf{a}$.

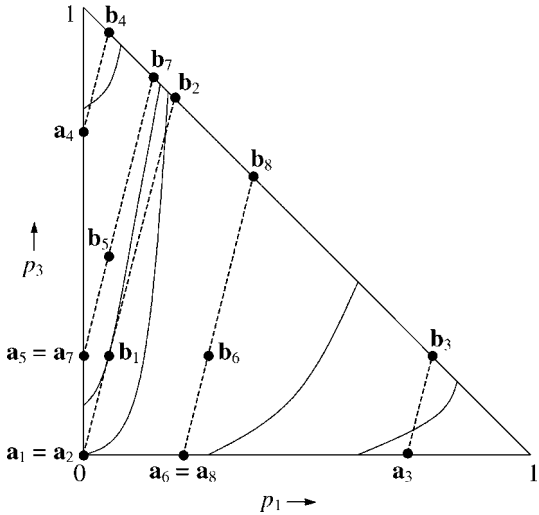


Fig. 3. Quasi-concavity.

2.3. Rank-dependent Expected Utility and Yaari's Dual Theory

The most popular alternative to the EU model, at least for economists (see, e.g. Starmer, 2000), is Quiggin's (1982) rank-dependent expected utility (RDEU) model. Most popular within the income distribution literature is Yaari's (1987) dual theory, which is a special case of the RDEU model. We will first summarize Yaari's model and then return to the more general RDEU approach.

If preferences are consistent with Yaari's theory, they can be represented by

$$F(\mathbf{p}) = \sum_{i=1}^{n} w(p_i, p_1 + \cdots + p_i)x_i, \tag{3}$$

where for any $i \neq n$

$$w(p_i, p_1 + \cdots + p_i) = f(p_i + \cdots + p_n) - f(p_{i+1} + \cdots + p_n),$$

$w(p_n, p_1 + \cdots + p_n) = f(p_n)$ and $f : [0, 1] \to [0, 1]$ is a strictly increasing and continuous function for which $f(0) = 0$ and $f(1) = 1$. Given the conditions on f, preferences are monotonic. Strong risk aversion requires that f be strictly convex (Yaari, 1987). Note that while in the EU approach a change in an income is evaluated in function of the size of the income, in the Yaari approach it is evaluated as a function of its rank position (defined as $p_1 + \cdots + p_i$ for an income x_i).

For the alternatives in the triangle diagram, Yaari's theory implies that

$$F(\mathbf{p}) = [1 - f(1 - p_1)]x_1 + [f(1 - p_1) - f(p_3)]x_2 + f(p_3)x_3, \tag{4}$$

which yields for the slope of the indifference curves

$$\left.\frac{\mathrm{d}p_3}{\mathrm{d}p_1}\right|_{F=\bar{F}} = \frac{f'(1 - p_1)}{f'(p_3)} \frac{x_2 - x_1}{x_3 - x_2}. \tag{5}$$

Again, indifference curves are positively sloped (since $f'(p) > 0$ for all p). If f is strictly convex, the slope decreases as p_1 increases, ceteris paribus, and also as p_3 increases, ceteris paribus. If p_1 decreases and p_3 increases, the slope does not necessarily go up or down. This means that indifference curves strictly fan out horizontally – that is, the slope becomes strictly higher moving horizontally west in the triangle diagram – and strictly fan in vertically – that is, the slope becomes strictly smaller moving vertically north in the triangle diagram. Moving diagonally northwest, however, the slope can go up or down. This pattern is illustrated in Fig. 4. By consequence, for the choices in the figures, fanning-out has to hold for combinations of the choice pairs 1, 3 and 6 (an horizontal move in the triangle), while fanning-in has to hold for combinations of the choice pairs 1, 4 and 5 (a vertical move). There are no implications concerning fanning-out or fanning-in for combinations of choices 2, 3, 4, 7 and 8.

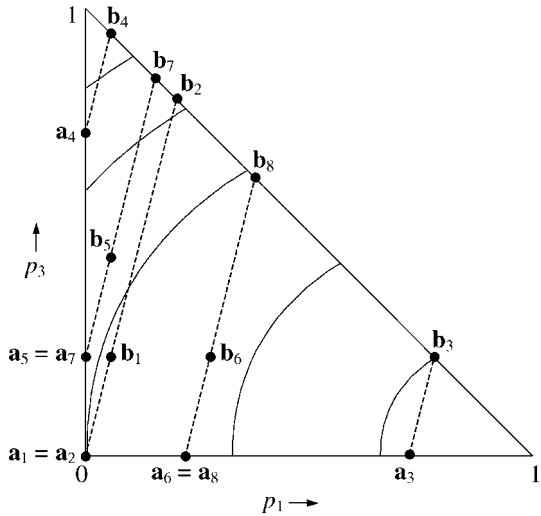


Fig. 4. RDEU (Yaari) Indifference Curves.

Another important property is that, whenever f is strictly convex, the slope of an indifference curve decreases as one moves to the northeast. Therefore, under the assumption of strong risk aversion, indifference curves are strictly concave (preferences are strictly quasi-convex).

The Yaari model is a special case of the RDEU model. The latter model is given by

$$F(\mathbf{p}) = \sum_{i=1}^{n} w(p_i, p_1 + \cdots + p_i)u(x_i), \tag{6}$$

where for any $i \neq n$

$$w(p_i, p_1 + \cdots + p_i) = f(p_i + \cdots + p_n) - f(p_{i+1} + \cdots + p_n),$$

$w(p_n, p_1 + \cdots + p_n) = f(p_n)$, $f : [0, 1] \to [0, 1]$ is a strictly increasing and continuous function for which $f(0) = 0$ and $f(1) = 1$ and u is a strictly increasing function. Again the conditions required for monotonicity are satisfied. Strong risk aversion requires that the function f be convex and that the function u be concave and, furthermore, that either f be strictly convex or u be strictly concave or both (Chew, Karni & Safra, 1987). When u is the identity function, the RDEU model (6) reduces to the Yaari model. When f is the identity function, it reduces to the EU model.

The slope of an indifference curve in the triangle diagram for the RDEU model is

$$\left.\frac{\mathrm{d}p_3}{\mathrm{d}p_1}\right|_{F=\bar{F}} = \frac{f'(1-p_1)}{f'(p_3)}\frac{u(x_2)-u(x_1)}{u(x_3)-u(x_2)}. \tag{7}$$

Clearly, the indifference curves of the RDEU social welfare function have (more or less) the same properties as those of the Yaari model. That is, indifference curves are concave, fan out horizontally and fan in vertically.

2.4. Evaluating Income Distributions

There is a close formal relationship between the literature on income distribution and the theory of decision making under risk. With the Gini index as a prominent exception, the most common inequality measures (including the Atkinson-Kolm and the generalized entropy measures) can all be interpreted in a social welfare framework formally equivalent to the EU model as given in expression (1). This means that they can be interpreted as reflecting VOI-preferences, i.e. the preferences of a rational individual behind the veil of ignorance.[10] Of course one can also

defend EU-type assumptions without explicitly referring to the idea of the veil of ignorance. One then has to justify the independence condition for ISO preferences directly on ethical grounds rather than as a requirement of rationality behind the veil of ignorance.

A strong competitor of the Atkinson-Kolm and the generalized entropy measures is the class of generalized Gini indices. These are based on a social welfare function of the form of the Yaari model (3) (or, at least, an important subclass is) and therefore do not satisfy the independence condition. The most popular social welfare function of the form (3) is the S-Gini social welfare function, where $f(p) = p^{\rho}$ with $\rho > 1$ (Donaldson & Weymark, 1980; Yitzhaki, 1983). The parameter ρ can be seen as a measure for the degree of inequality aversion. Note that the popular Gini index is based on the S-Gini social welfare function with $\rho = 2$. A few studies such as Ebert (1988) and more recently Chateauneuf (1996) and Chateauneuf, Gajdos and Wilthien (2002), have considered the evident extension to the Yaari model which is to base the evaluation of income distributions on the RDEU model.

The idea of strong risk aversion is interpreted within the income distribution literature as the Pigou-Dalton transfer principle, i.e. the notion that a rank preserving transfer from a rich to a poor person increases social welfare. As we have seen, the transfer principle requires in the EU model that the function u be strictly concave. This assumption does not affect the response patterns compatible with EU theory for the choice pairs in the triangle diagram of Fig. 1, however. As can be seen from expression (2) the restriction to linear parallel indifference curves does not depend on the concavity of u and a test of this restriction can be seen as a direct test of the independence assumption without any need to make assumptions about risk aversion.

On the other hand, imposing the transfer principle has stronger consequences for the Yaari and the RDEU models within the triangle. As we have seen, it requires, for instance, in both cases that the indifference curves be strictly concave. Since the transfer principle occupies such a dominant position in the income distribution literature, we will use in the empirical part the terms Yaari model and RDEU model for expressions (3) and (6), respectively, with the assumption of concave indifference curves imposed.

However, we know from previous empirical work that the transfer principle is violated consistently by respondents.[11] Let us therefore define the weaker principle of "weak inequality aversion": given a fixed population, it should always hold that a completely equal income distribution is better than any unequal income distribution with the same total income. This principle seems absolutely essential for an egalitarian social welfare function. It gives additional support for

the transfer principle with an EU social welfare function, i.e. a social welfare function satisfying independence, because such a welfare function will only satisfy weak inequality aversion if it satisfies the transfer principle. However, in the Yaari (and RDEU) framework, weak inequality aversion does not imply the transfer principle. It has been shown (Chateauneuf, 1996) that the Yaari social welfare function (3) satisfies weak inequality aversion if and only if $f(p) < p$ for all $p \in (0, 1)$. This condition is strictly weaker than strict convexity (since $f(0) = 0$ and $f(1) = 1$). The RDEU social welfare function satisfies weak inequality aversion if $f(p) \leq p$ for all $p \in (0, 1)$ and u is concave, with at least one of the conditions holding in its strict version.[12] In our empirical work we will consider these extensions as well and label them Yaari′ and RDEU′.

In the risk literature, forms of the RDEU weighting function f that do not satisfy the condition relating to weak inequality aversion have been considered and sometimes offer a better explanation of observed choice patterns (see, e.g. Gonzalez & Wu, 1999). We do not consider these forms in our empirical analysis because in our view it does not make sense to base the evaluation of income distributions on a welfare function which does not even satisfy the principle of weak inequality aversion.

3. THE SETTING OF THE QUESTIONNAIRE

The target group of the questionnaire consisted of first year business students of the K.U.Leuven (Catholic University Leuven, Belgium), who had not yet been exposed to any lectures on the evaluation of income distributions or on decision making under risk. The questionnaires were distributed and filled-in in the classroom, after the teacher had given a short and non-suggestive oral introduction. The survey was organized twice (with different respondents in two subsequent academic years): in April 2002 and in November 2002. The results were stable over time. In order to test for the differences between ISO, VOI and PIR preferences, there were three different versions of the questionnaire. Accordingly, the group of students was divided into three subgroups. Each subgroup participated in only one version of the questionnaire and respondents did not know that there were three different versions. For the ISO version, the VOI version and the PIR version, there are 93, 92 and 94 respondents, respectively.

Each questionnaire version consists of the same eight questions, where in each question, the respondent is asked to make a choice between two alternatives, which are either income distributions or lotteries, depending on the given choice situation. The eight choice pairs correspond to the alternatives shown in Fig. 1 (with the

probabilities as given in Table 1). Throughout the questionnaire, the same set of three incomes $X = \{x_1 = €500, x_2 = €1500, x_3 = €2500\}$ is used. In line with the Allais problems described earlier, we refer to questions 1, 3, 4, 5 and 6 as the "common consequence questions," and to questions 2, 3, 4, 7 and 8 as the "common ratio questions."

Although the same choice pairs are used, the background stories are different for the three versions of the questionnaire.[13] Each of the three versions deals with recently graduated students that are going to be employed in one of two firms. Each firm offers three types of jobs which are identical in every respect except for the income that is earned: the first job pays €2500, the second €1500 and the third €500. For the ISO and VOI versions, a firm corresponds to an income distribution, for the PIR version it corresponds to a lottery.

In the ISO version, the respondent is asked to consider the situation of 100 recently graduated students that will all be employed in either of two firms, which are different only with respect to the number of positions that are available for each of the jobs. The respondent is then asked to reveal, for the eight cases, which of the two firms he or she thinks offers the largest social welfare.

The VOI version also asks the respondent to consider the situation of 100 recently graduated students, but this time the respondent has to picture himself or herself as being one of them. Again, the firms are different only with respect to the number of positions that are available for each of the jobs. The respondent and the 99 other graduated students will all be employed in the same firm and each has an equal chance of ending up in any of the 100 positions available in the firm. The respondent is then asked to state, for each of the eight cases, which firm he or she prefers.

In the PIR version, the respondent is asked to picture himself or herself as being a recently graduated student who will be employed in either of two firms. The firms are identical except with respect to the probabilities of ending up with each of the jobs. The respondent is then, again, asked to state, for each of the eight cases, which firm he or she prefers.

As mentioned already in the introduction, the setting of our questionnaire is similar to the one used by Bernasconi (2002). There are three main differences, however. First, we use different and more income distributions (and therefore test some axioms which could not be tested by him). Second, he represents the different income distributions in the questionnaires with pie charts, while we simply give the relevant sets of numbers. Third, he formulates the ISO, VOI and PIR cases in a more abstract form, while we tried to formulate a question which was closer to the everyday experience of our respondents. The comparison of his results with ours will therefore give some insight into the importance of framing effects (for which, again, see Camerer, 1995).

4. RESULTS

Our discussion of the results focuses on the two general issues raised before: the comparison of the ISO, VOI and PIR versions of the questionnaire, and the degree of consistency with the preference theories presented in Section 2. In Section 4.1, we have a first look at the question of how the three versions of the questionnaire compare through an analysis of the responses for separate questions. Combining the answers on different questions makes it possible to test also the relevancy of the different basic axioms of choice theory (Section 4.2). In Section 4.3, we conclude the discussion by focusing on the different theories which have been proposed in the income distribution literature.

4.1. A First Look

Table 2 and Fig. 5 give the results for the separate questions. The chi-square test statistics reported in Table 3 test for each question separately the null hypothesis that population proportions for categories **a** and **b**, respectively, are equal for the two versions under comparison (ISO-VOI, VOI-PIR or ISO-PIR) (there is one degree of freedom).[14] To some extent Table 3 suggests that the results for the ISO and PIR versions are furthest removed from each other while the results for the VOI version lie in between. This is exactly what one would expect a priori: ISO preferences deal exclusively with uninvolved common interest, PIR preferences deal exclusively with involved self interest and VOI preferences deal with involved common interest (that is, the common interest is at stake). We will see that this pattern is confirmed in more detailed analyses.

Table 2 shows that, overall, alternative **b** is more popular than the other two alternatives. In the risk literature **b** alternatives are usually seen as more risky

Table 2. Results for Separate Questions (in %).

Question	ISO			VOI			PIR		
	a	**b**	∼	**a**	**b**	∼	**a**	**b**	∼
1	37	59	4	26	72	2	27	69	4
2	50	45	5	45	51	4	58	43	0
3	30	66	4	20	77	3	17	78	5
4	61	37	2	38	57	5	28	69	3
5	44	55	1	29	67	3	35	62	3
6	19	76	4	16	82	2	35	63	2
7	56	42	2	41	53	5	51	43	6
8	30	61	9	36	60	4	48	45	8

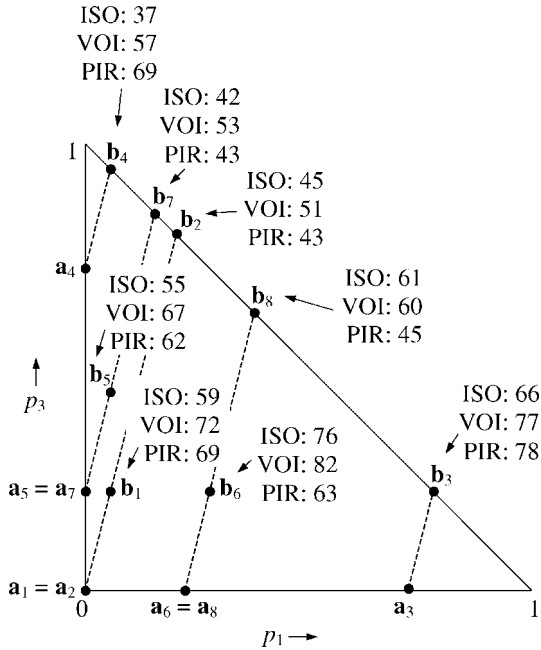


Fig. 5. Overview of the Results, **b** Responses (in %).

than the corresponding **a** alternatives. Analogously, we could say that they are more unequal in the income distribution context. The popularity of the **b** answers can be explained by the choice of the set of incomes in our questionnaire design. Consider as a benchmark the case of a respondent who has preferences consistent with the Atkinson social welfare function: for the given income amounts, such a

Table 3. Chi-square Tests for Homogeneity for Separate Questions.

Question	ISO-VOI	VOI-PIR	ISO-PIR
1	2.72 (0.099)	0.03 (0.867)	2.20 (0.138)
2	0.57 (0.451)	2.15 (0.143)	0.49 (0.483)
3	2.93 (0.087)	0.15 (0.703)	4.35 (0.037)
4	8.94 (0.003)	2.68 (0.101)	21.29 (0.000)
5	3.91 (0.048)	0.71 (0.399)	1.31 (0.253)
6	0.38 (0.539)	8.64 (0.003)	5.47 (0.019)
7	3.23 (0.073)	2.07 (0.151)	0.12 (0.726)
8	0.39 (0.530)	3.58 (0.058)	6.21 (0.013)

Note: p-Values between brackets.

respondent only prefers **a** over **b** if she has a relatively high value of 2.5 or more for the parameter of inequality aversion.

4.2. Testing Some Concrete Hypotheses of Choice Theory

More interesting insights can be gained by analysing the response patterns for different choice pairs together. We will first look at combinations of two questions and then analyse the overall response patterns for the eight questions. We focus, again, on the two main issues. In the first place, we test the empirical relevancy of the concrete hypotheses of choice theory. In the second place, we check for the possible differences between ISO, VOI and PIR preferences.

4.2.1. Pairs of Questions

(a) Tables 4 and 5 show the results for combinations of several pairs of common consequence questions and common ratio questions, respectively. For each

Table 4. Results for Pairs of Common Consequence Questions (in %).

Questions	Version	EU (**aa**, **bb**, ∼∼)	Fanning-out (**ba**, **b**∼, ∼**a**)	Fanning-in (**ab**, **a**∼, ∼**b**)
3, 1	ISO	61	23 (0.203)	16
	VOI	63	22 (0.196)	15
	PIR	63	23 (0.088)	14
1, 4	ISO	53	35 (0.001)	12
	VOI	60	26 (0.049)	14
	PIR	55	22 (0.562)	22
3, 6	ISO	61	14	25 (0.066)
	VOI	73	12	15 (0.345)
	PIR	62	27 (0.014)	12
6, 1	ISO	55	31 (0.010)	14
	VOI	67	22 (0.049)	11
	PIR	63	14	23 (0.088)
1, 5	ISO	67	19 (0.237)	14
	VOI	72	16 (0.279)	12
	PIR	66	21 (0.108)	13
5, 4	ISO	62	28 (0.003)	10
	VOI	54	27 (0.140)	18
	PIR	63	15	22 (0.155)

Note: p-Values between brackets.

Table 5. Results for Pairs of Common Ratio Questions (in %).

Questions	Version	EU (**aa**, **bb**, ∼∼)	Fanning-out (**ba**, **b**∼, ∼**a**)	Fanning-in (**ab**, **a**∼, ∼**b**)
3, 2	ISO	55	33 (0.001)	12
	VOI	57	36 (0.000)	8
	PIR	51	45 (0.000)	4
2, 4	ISO	56	27 (0.106)	17
	VOI	62	15	23 (0.155)
	PIR	57	6	36 (0.000)
3, 8	ISO	62	20 (0.368)	17
	VOI	60	29 (0.004)	11
	PIR	53	41 (0.000)	5
8, 2	ISO	61	29 (0.002)	10
	VOI	52	29 (0.087)	18
	PIR	64	21 (0.196)	15
2, 7	ISO	68	17 (0.428)	15
	VOI	62	17	21 (0.368)
	PIR	67	16	17 (0.500)
7, 4	ISO	60	23 (0.256)	17
	VOI	62	17	21 (0.368)
	PIR	56	9	35 (0.000)

Note: p-Values between brackets.

combination of two choice pairs (described in the first column) we give separately the results for the three versions of the questionnaire. As shown in Section 2.1, only three of the nine possible response patterns are consistent with EU theory for each of the combinations of two questions included in Tables 4 and 5: the respondent can prefer **a** in both choice pairs, she can prefer **b** in both pairs or she can be indifferent (∼) in both choice situations. We call these patterns, (**aa**, **bb**, ∼∼), therefore "EU consistent" and the percentage of respondents with one of these three response patterns is given in the third column of Tables 4 and 5. Analogously we can say that the response patterns (**ba**, **b**∼, ∼**a**) and (**ab**, **a**∼, ∼**b**) are consistent with indifference curves that fan out and fan in, respectively. In both cases, we exclude EU consistent patterns from the categories fanning-out and fanning-in. The percentages of respondents with these patterns are given in the last two columns of the tables.

Clearly, EU consistent responses dominate. One should be aware that this does not necessarily imply that our respondents follow the axioms of EU theory, as it is quite possible for an individual to be consistent with EU theory over two questions

but not over three or more. We will return to this issue in the next section. At this stage it is more interesting to consider whether the violations of EU theory for each of the question pairs are systematic, that is, whether the percentage of observed patterns consistent with fanning-out (or fanning-in) is significantly higher than the percentage that would be observed if the response patterns of the respondents that violate EU theory were completely random. The null hypothesis is that the population frequency of fanning-out (or fanning-in) violations relative to the total population frequency of violations is equal to 50%. The tables report *p*-values for the one sided exact test based on the binomial distribution.

The first combinations of choice pairs in Tables 4 and 5, the combinations (3, 1) and (3, 2), are of particular interest, as they are similar to the original examples used by Allais for introducing the common consequence and common ratio effects, respectively. In both cases the predicted fanning-out patterns are more popular than the fanning-in patterns. The statistical significancy of fanning-out is much weaker for Allais' common consequence effect (questions 3 and 1) than for Allais' common ratio effect (questions 3 and 2).

The overall picture shows some interesting differences between the ISO, VOI and PIR versions of the questionnaire. A mixed pattern of fanning-out and fanning-in is observed in the PIR version. This is in line with the experimental research on decision making under risk. However, with only one exception, fanning-out is always dominating in the ISO version. The VOI version is between the other two, but with a relatively strong presence of fanning-out. Table 6 presents the chi-square test statistics for the hypothesis of homogeneity of two versions with respect to the categories EU, fanning-out and fanning-in between versions

Table 6. Chi-square Tests for Homogeneity for Pairs of Questions.

Questions	ISO-VOI	VOI-PIR	ISO-PIR
3, 1	0.06 (0.969)	0.12 (0.942)	0.20 (0.907)
1, 4	1.93 (0.381)	2.15 (0.342)	5.88 (0.053)
3, 6	3.16 (0.206)	6.43 (0.040)	8.03 (0.018)
6, 1	3.11 (0.211)	6.04 (0.049)	8.99 (0.011)
1, 5	0.56 (0.756)	0.86 (0.650)	0.14 (0.932)
5, 4	3.07 (0.216)	4.25 (0.120)	8.40 (0.015)
3, 2	0.96 (0.620)	2.04 (0.361)	5.01 (0.082)
2, 4	4.00 (0.135)	6.33 (0.042)	18.16 (0.000)
3, 8	2.85 (0.241)	4.07 (0.131)	13.25 (0.001)
8, 2	3.23 (0.199)	2.65 (0.267)	2.20 (0.333)
2, 7	1.05 (0.591)	0.57 (0.753)	0.16 (0.923)
7, 4	0.94 (0.626)	6.56 (0.038)	11.80 (0.003)

Note: *p*-Values between brackets.

(there are two degrees of freedom).[15] The hypothesis formulated on the basis of Table 3 is corroborated by the results in Table 6: the results for the ISO and PIR versions form the extremes while the results for the VOI version are situated in between.

The question pairs in Table 4 also allow to test for some aspects of the Yaari and RDEU models (with the Pigou-Dalton transfer principle imposed). As we have seen (Section 2.3), these models imply that fanning-out holds horizontally, that is, for the question pairs (3, 1), (3, 6) and (6, 1), while fanning-in holds vertically, and thus for the question pairs (1, 4), (1, 5) and (5, 4) (of course, the EU patterns for these pairs are also consistent with the models). This pattern is not supported by the results for the ISO and VOI versions, especially where the Yaari and RDEU models imply fanning-in.

(b) Table 7 gives the results for the question pairs (6,8), (1,2) and (5,7). These combinations allow to test betweenness, i.e. the linearity of indifference curves (which is EU consistent) against quasi-convexity (excluding EU consistent patterns) and quasi-concavity (again, excluding EU consistent patterns). The corresponding response patterns have already been described in Section 2.2. The results in Table 7 are striking. There is a clear and significant domination of quasi-concavity, i.e. convex indifference curves. This mixture proneness is found in all three versions of the questionnaire.[16] Quasi-concavity has also been found in experimental work on decision making under risk (see, e.g. Camerer & Ho, 1994). Its implications for welfare analysis, however, are important. We mentioned already that imposition of the transfer principle in the Yaari and RDEU models implies quasi-convex preferences. We will return to the implications of these findings in Section 4.3.

Table 7. Results for Pairs of Questions (in %).

Questions	Version	EU (**aa**, **bb**, ∼∼)	Quasi-convexity (**ab**, **a**∼, ∼**b**)	Quasi-concavity (**ba**, **b**∼, ∼**a**)
6, 8	ISO	57	15	28 (0.040)
	VOI	61	9	30 (0.001)
	PIR	54	14	32 (0.007)
1, 2	ISO	70	8	23 (0.006)
	VOI	61	9	30 (0.001)
	PIR	55	7	37 (0.000)
5, 7	ISO	63	13	24 (0.061)
	VOI	57	14	29 (0.019)
	PIR	53	14	33 (0.005)

Note: p-Values between brackets.

Table 8. Results for the Combination of All Eight Questions (in %).

	EU	Fanning-out	Fanning-in	Betweenness	Quasi-convexity	Quasi-concavity
Reference	0.8	6.3	6.3	12.5	42.2	42.2
ISO	10	30	13	32	46	68
Test 1	(0.000)	(0.000)	(0.013)	(0.000)	(0.246)	(0.000)
Test 2		(0.000)	(0.889)	(0.002)	(0.854)	(0.001)
VOI	13	29	18	26	40	70
Test 1	(0.000)	(0.000)	(0.000)	(0.000)	(0.686)	(0.000)
Test 2		(0.000)	(0.570)	(0.393)	(0.998)	(0.003)
PIR	11	24	15	21	37	72
Test 1	(0.000)	(0.000)	(0.002)	(0.012)	(0.859)	(0.000)
Test 2		(0.002)	(0.762)	(0.675)	(0.999)	(0.000)

Note: p-Values between brackets.

4.2.2. The Total Pattern of Answers

In Table 8, we summarize the results for a more ambitious approach in which the eight questions are considered jointly. Each column refers to a specific hypothesis of choice theory. We first give, for each hypothesis, as a reference point the proportion of the 256 ($= 2^8$) possible patterns that is actually consistent with the given hypothesis.[17] If individual response patterns were completely random, we would expect to find the "reference" degree of support for the various hypotheses. We then test whether the actual number of consistent response patterns in the data is significantly larger than what would be expected for random responses.[18] This test is labelled "Test 1" in Table 8.

For all three versions, all hypotheses except quasi-convexity pass Test 1. Note that about 10–13% of the observed patterns are consistent with EU theory – which is significantly more than the 0.8% which would be found with a completely random response pattern. An explanation of the success of EU theory could be that respondents use the expected value rule. At the same time it should be mentioned that 10–13% is far from overwhelming considering the focal role of EU theory in the risk and in the income distribution literature.

Since all the other hypotheses generalize EU theory, they all benefit from the relatively good performance of that theory. It is more revealing therefore to test whether they "add" something to EU theory. We do this by removing from the sample all EU consistent patterns. For the remaining (non-EU consistent) responses we follow an analogous procedure as described before. For each hypothesis (each column) we first compute, with respect to the set of all possible patterns excluding the EU consistent patterns, the proportion of consistent responses to be expected if individual response patterns were completely random. We then test whether

the proportion of consistent responses in the (reduced) sample is significantly larger than what would be expected in the random case. The resulting p-values are summarized in Table 8 under the label "Test 2."[19]

For all three versions, fanning-out adds significantly to EU theory, while the fanning-in hypothesis does not. Looking at the shape of the indifference curves, betweenness adds explanatory power to EU theory for the ISO version, but not for the other versions. An approach with linear but non-parallel indifference curves in the Marschak-Machina triangle seems to have some relevance to describe the preferences of an impartial and sympathetic observer. However, more striking is the significance of quasi-concavity for all three versions. The global response patterns therefore confirm what we found already by analysing the combinations of choice pairs two by two.

4.3. The Fate of Different Theories of Income Distribution Evaluation

The importance of quasi-concavity and fanning-out already suggests that the most popular approaches in the income distribution literature will not get much support in our data. Table 9, which is constructed in a similar way as Table 8, summarizes the results in a more structured way. We repeat the results for the EU model as a benchmark. Remember that the EU approach performs significantly better than what would be predicted if the answers were random. As shown by the results for "Test 1," the same is true for the S-Gini, the Yaari, the RDEU, the Yaari′ and the RDEU′ models (for the latter three only in the VOI and PIR versions).

Table 9. Results for the Combination of All Eight Questions (in %).

	EU	S-Gini	Yaari	RDEU	Yaari′	RDEU′
Reference	0.8	2.7	15.2	16.4	65.6	79.3
ISO	10	13	23	23	69	78
Test 1	(0.000)	(0.000)	(0.039)	(0.075)	(0.298)	(0.633)
Test 2		(0.273)	(0.708)	(0.805)	(0.895)	(0.990)
VOI	13	16	25	25	78	88
Test 1	(0.000)	(0.000)	(0.010)	(0.023)	(0.006)	(0.021)
Test 2		(0.268)	(0.794)	(0.869)	(0.518)	(0.829)
PIR	11	15	26	26	74	90
Test 1	(0.000)	(0.000)	(0.007)	(0.016)	(0.042)	(0.003)
Test 2		(0.113)	(0.496)	(0.621)	(0.627)	(0.440)

Note: p-Values between brackets.

However, in our set-up all these alternative theories are less restrictive than EU theory. In fact, each of them can also rationalize each pattern that is EU consistent.[20] We therefore want to test whether any of these theories adds some explanatory power to the EU model. Analogously to the previous section, we therefore computed again the "Test 2" results. For none of the versions, Yaari's theory or the (more restricted) S-Gini model passes this stricter test. Nor does the RDEU model. To repeat: this implies that the proportion of observed response patterns in the subsample of non-EU consistent responses which is consistent with these models is not significantly larger than what would be expected if the answers of the respondents were completely random. It is important to remember that we imposed the Pigou-Dalton transfer principle in the Yaari and the RDEU-model, i.e. convexity of the weighting function f, and that our results can only be seen as a test of this restricted model. Yet relaxation of this convexity condition does not seem to help very much, given the fate of the Yaari′ and the RDEU′ models, which only impose the property of weak inequality aversion. It is difficult to see how one could construct an attractive egalitarian theory of social welfare which does not satisfy this very weak property. Both models (Yaari′ and RDEU′) are quite flexible and it is therefore not surprising that the proportion of response patterns compatible with them is very high. Again, however, the models do not add significantly to EU, in the sense that randomly chosen patterns would have performed equally well.

These results seem to suggest that it is worthwhile to work out alternatives for the EU-type social welfare functions, i.e. to try and elaborate an alternative which does not embody the independence assumption. At the same time, however, the Yaari- and RDEU-type extensions with weak inequality aversion imposed do not seem to be very promising, at least when one wants to rationalize the preferences of our respondents (and they appear to be even less successful for the ISO than for the VOI or PIR version). Comparing Tables 8 and 9 it is striking how much better is the performance of other alternatives to the EU model like fanning-out and quasi-concavity. It remains to be seen whether these ideas can be integrated in an attractive theory of income distribution.

5. CONCLUSION

With our questionnaire study we wanted to test whether the veil of ignorance approach captures in an adequate way the preferences of an impartial and sympathetic observer. Moreover, we wanted to check whether the answers of our respondents satisfy the independence axiom – underlying EU theory and most approaches to inequality measurement – and its most popular alternatives. Both

questions are related but different. One can accept the VOI approach and at the same time argue in favour of a non-EU model behind the veil of ignorance. And one can defend the independence assumption for inequality measurement without the detour of the veil of ignorance.

As to the first question, the results for the three questionnaire versions (ISO, VOI and PIR) are to a certain degree similar: both of Allais' problems are present, there is quite a lot of systematic fanning-out or fanning-in, and quasi-concavity is an important systematic violation of EU (or betweenness). However, there are differences and it appears that the ISO and PIR versions are at both extremes. The identification of ISO preferences with VOI preferences is not evident. Note that the results for the PIR version are reassuringly comparable to the results encountered in empirical studies from the literature on decision under risk: Allais' problems, a complex fanning pattern, systematic violations of betweenness.

The EU model yields a significant contribution to the explanation of the response patterns. At the same time, however, there are clear indications of the relevancy of fanning-out and quasi-concavity, also in the ISO version. Fanning-out and quasi-concavity do not characterize the most popular alternatives to the EU model – the RDEU model with as a special case the Yaari model, which provides the normative basis for an important subclass of the family of generalized Ginis. It is therefore not surprising that they do not add much explanatory power.

These are the results of only one limited study. However, they are in the line of much previous research on the empirical acceptance of the most popular inequality axioms. Moreover, despite the differences in the concrete formulation of the questionnaires and in the general set-up of the empirical study, some of our results are strikingly similar to those of Bernasconi (2002): he also finds that the equivalence of VOI and ISO preferences cannot be taken for granted and that quasi-concavity, i.e. mixture proneness, is important to explain the empirical results.

The conclusion that the traditional inequality literature does not adequately capture the intuitions of our respondents seems clear. Even if we take the Yaari and the EU model together only a quarter of our students shows a response pattern which is in line with one of them. Of course, one can reasonably argue that the normative relevancy of this kind of questionnaire results is limited, as they can never substitute for critical reflection and thorough assessment of the ethical argumentation. We do not go into that debate here. However, a conditional conclusion seems possible. If one wants to construct a theory of income distribution which is more attuned to the intuitions of lay respondents, the RDEU model with imposition of weak inequality aversion does not seem to be the most promising starting point.

NOTES

1. Similar approaches have been proposed by Vickrey (1945, 1960) and Rawls (1958, 1971). The latter coined the term "veil of ignorance." Harsanyi used the approach to justify (mean) utilitarianism while Rawls used it to justify his deontological theory which couples a respect for basic liberties with maximin in "primary goods."

2. Harsanyi's claim that mean utilitarianism follows from his assumptions has been criticized on several accounts. See Mongin (2001) for a thorough overview of the literature.

3. See Cowell (2000) and Lambert (2001) for recent overviews of this literature.

4. For overviews, see Camerer (1995) and Starmer (2000).

5. To be precise, we are referring to the subclass that satisfies Dalton's Population Principle. See, for instance, Gajdos (2001) for details.

6. The most influential work is by Amiel and Cowell, who summarize their most important findings in Amiel and Cowell (1999). See also Harrison and Seidl (1994a, b).

7. This graphical device was introduced into the literature by Marschak (1950) and popularized by Machina (1982). It has since been used in many empirical studies concerning individual decision under risk.

8. Of course, the slope is not required to be equal across different triangles, when different sets X are considered.

9. Indeed, Machina (1982) has shown that the slope of the EU model, given in expression (2), is related to the Arrow-Pratt measure of risk aversion.

10. Dahlby (1987) explicitly works out this interpretation.

11. This is found especially in the context of inequality comparisons (see for instance Amiel & Cowell, 1992, 1998; Ballano & Ruiz-Castillo, 1993; Harrison & Seidl, 1994a, b), but also in the context of social welfare comparisons (Amiel & Cowell, 1994).

12. In fact, Chateauneuf (1996) has shown that these conditions for the Yaari and RDEU models imply consistency with the "absolute differentials ordering," which is a stronger requirement than the one of weak inequality aversion. This stronger principle can be formulated as follows. Suppose that we have two income distributions with the same population and total income, and in the first income distribution the absolute income difference for each income pair is greater than, or equally great as, in the second distribution while for at least one pair the absolute income difference is greater, then the first income distribution is more unequal than the second. It seems natural to extend the principle to the social welfare context by stating that the second income distribution should be evaluated as better than the first.

13. The precise formulation of the background stories in each of the versions is given in Appendix A. For each background story there were two variants of the questionnaire with the questions ordered differently. Since the results show that there is only a slight indication of order effects, we simply pooled the answers for these different variants.

14. We ignore the category of indifference ($\sim$) in the tests because it usually has frequencies lower than five, which would make the chi-square test less appropriate.

15. Note the difference with Table 3, in which we tested for homogeneity of the three versions with respect to the responses (**a** or **b**) for the eight separate questions. Table 6 tests for homogeneity of the three versions with respect to response patterns (EU consistent, fanning-out or fanning-in) for combinations of two questions.

16. Chi-square tests show that the null hypothesis of homogeneity over the versions cannot be rejected.

17. For convenience, we have neglected patterns with indifferences. There are only very few cases of indifference in the answers of our respondents.

18. More specifically, we test the null hypothesis that the population proportion in support of the given hypothesis of choice theory is equal to the population proportion in support of the same hypothesis if choices were completely random against the alternative hypothesis that it is greater.

19. More specifically, "Test 2" considers the null hypothesis that the population proportion in support of a specific non-EU hypothesis, excluding the part of the population that is in support of EU theory as well, is equal to what would be the population proportion in support of that non-EU hypothesis, excluding the part of the population that is in support of EU theory as well, if choices were random. The alternative hypothesis is that the former population proportion is greater than the latter.

20. This is not a general property – but it holds for our set of specific questions within the Marschak-Machina triangle.

ACKNOWLEDGMENTS

We thank two anonymous referees for their valuable comments.

REFERENCES

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole américaine. *Econometrica*, *21*, 503–546.

Amiel, Y., & Cowell, F. A. (1992). Measurement of income inequality: Experimental test by questionnaire. *Journal of Public Economics*, *47*, 3–26.

Amiel, Y., & Cowell, F. A. (1994). Income inequality and social welfare. In: J. Creedy (Ed.), *Taxation, Poverty and Income Distribution* (pp. 193–219). Cheltenham: Edward Elgar.

Amiel, Y., & Cowell, F. A. (1998). Distributional orderings and the transfer principle: A re-examination. *Research on Economic Inequality*, *8*, 195–215.

Amiel, Y., & Cowell, F. A. (1999). *Thinking about inequality*. Cambridge: Cambridge University Press.

Amiel, Y., & Cowell, F. A. (2002). Attitudes towards risk and inequality: A questionnaire-experimental approach. In: F. Andersson & H. J. Holm (Eds), *Experimental Economics: Financial Markets, Auctions and Decision Making* (pp. 85–115). Deventer: Kluwer.

Amiel, Y., Cowell, F. A., & Polovin, A. (2001). Risk perceptions, income transformations and inequality. *European Economic Review*, *45*, 964–976.

Ballano, C., & Ruiz-Castillo, J. (1993). Searching by questionnaire for the meaning of income inequality. *Revista Espanola de Economia*, *10*, 233–259.

Battalio, R. C., Kagel, J., & Jiranyakul, K. (1990). Testing between alternative models of choice under uncertainty: Some initial results. *Journal of Risk and Uncertainty*, *3*, 25–50.

Bernasconi, M. (2002). How should income be divided? Questionnaire evidence from the theory of 'impartial preferences'. In: P. Moyes, C. Seidl & A. Shorrocks (Eds), *Inequalities: Theory, Experiments and Applications, Journal of Economics* (Suppl. 9, pp. 163–195).

Camacho-Cuena, E., Seidl, C., & Morone, A. (2003). Income distributions versus lotteries: Happiness, response-mode effects and preference reversals. Economics Working Paper, 2003-1, Christian-Albrechts-Universität Kiel.

Camerer, C. (1995). Individual decision making. In: J. Kagel & A. E. Roth (Eds), *Handbook of Experimental Economics* (pp. 587–703). Princeton: Princeton University Press.

Camerer, C., & Ho, T.-H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, *8*, 167–196.

Chateauneuf, A. (1996). Decreasing inequality: An approach through non-additive models. *Cahiers Eco & Maths*, 96.58, Université de Paris 1.

Chateauneuf, A., Gajdos, T., & Wilthien, P.-H. (2002). The principle of strong diminishing transfers. *Journal of Economic Theory*, *103*, 311–333.

Chew, S. H., Karni, E., & Safra, Z. (1987). Risk aversion in the theory of expected utility with rank dependent probabilities. *Journal of Economic Theory*, *42*, 370–381.

Cowell, F. A. (2000). Measurement of inequality. In: A. Atkinson & F. Bourguignon (Eds), *Handbook of Income Distribution* (Vol. 1, pp. 87–166). Amsterdam: Elsevier.

Cowell, F. A., & Schokkaert, E. (2001). Risk perceptions and distributional judgments. *European Economic Review*, *45*, 941–952.

Dahlby, B. (1987). Interpreting inequality measures in a Harsanyi framework. *Theory and Decision*, *22*, 187–202.

Donaldson, D., & Weymark, J. A. (1980). A single-parameter generalization of the Gini indices of inequality. *Journal of Economic Theory*, *22*, 67–86.

Ebert, U. (1988). Measurement of inequality: An attempt at unification and generalization. *Social Choice and Welfare*, *5*, 147–169.

Gajdos, T. (2001). Les fondements axiomatiques de la mesure des inégalités. *Revue d'Economie Politique*, *5*, 683–720.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, *38*, 129–166.

Harrison, E., & Seidl, C. (1994a). Acceptance of distributional axioms: Experimental findings. In: W. Eichhorn (Ed.), *Models and Measurement of Welfare and Inequality* (pp. 67–99). Berlin: Springer.

Harrison, E., & Seidl, C. (1994b). Perceptional inequality and preferential judgments: An empirical examination of distributional axioms. *Public Choice*, *79*, 61–81.

Harsanyi, J. C. (1953). Cardinal utility in welfare economics and in the theory of risk taking. *Journal of Political Economy*, *61*, 434–435.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *63*, 309–321.

Harsanyi, J. C. (1977). *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge: Cambridge University Press.

Lambert, P. J. (2001). *The distribution and redistribution of income* (3rd ed.). Manchester: Manchester University Press.

Machina, M. J. (1982). Expected utility analysis without the independence axiom. *Econometrica*, *50*, 277–323.

Marschak, J. (1950). Rational behavior, uncertain prospects, and measurable utility. *Econometrica*, *18*, 111–141.

Mongin, P. (2001). The impartial observer theorem of social ethics. *Economics and Philosophy*, *17*, 147–179.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behaviour and Organization*, *3*, 323–343.

Rawls, J. (1958). Justice as fairness. *Philosophical Review*, *67*, 164–194.
Rawls, J. (1971). *A theory of justice*. Oxford: Oxford University Press.
Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, *38*, 332–382.
Traub, S., Seidl, C., & Schmidt, U. (2003). Lorenz, Pareto, Pigou: Who scores best? Experimental evidence on dominance relations of income distributions. Economics Working Paper, 2003-4, Christian-Albrechts-Universität Kiel.
Vickrey, W. (1945). Measuring marginal utility by reactions to risk. *Econometrica*, *13*, 319–333.
Vickrey, W. (1960). Utility, strategy, and social decision rules. *Quarterly Journal of Economics*, *74*, 507–535.
Yaari, M. E. (1987). The dual theory of choice under risk. *Journal of Economic Theory*, *44*, 381–397.
Yitzhaki, S. (1983). On an extension of the Gini inequality index. *Quarterly Journal of Economics*, *93*, 321–324.

APPENDIX A.1: ISO VERSION

Consider the situation of two firms, A and B, that each plan to employ 100 recently graduated students. Assume that in each firm there are three types of jobs that are identical in all respects but yield a different monthly net income. The first job yields €2500, the second €1500 and the third €500. The firms differ however with respect to the numbers of positions they have available for each of the three jobs.

Evidently, due to the different distribution of incomes, the global welfare of the 100 employees can be different in the firms A and B. We are interested in your personal judgement of these welfare differences.

Indicate in each of the eight questions below which firm leads to the highest welfare according to you by marking A or B. So, the marked letter corresponds to the firm that you prefer from a welfare perspective. If you consider both firms to be equally good, then mark both letters. Of course each question needs to be treated separately and a different answer can be given in each case.

	A:	B:
Question 1	100 earn €1500 each	20 earn €2500 each 75 earn €1500 each 5 earn €500 each
Question 2	100 earn €1500 each	80 earn €2500 each 20 earn €500 each
Question 3	25 earn €1500 each 75 earn €500 each	20 earn €2500 each 80 earn €500each
Question 4	75 earn €2500 each 25 earn €1500 each	95 earn €2500 each 5 earn €500 each

Question 5	20 earn €2500 each 80 earn €1500 each	40 earn €2500 each 55 earn €1500 each 5 earn €500 each
Question 6	80 earn €1500 each 20 earn €500 each	20 earn €2500 each 55 earn €1500 each 25 earn €500 each
Question 7	20 earn €2500 each 80 earn €1500 each	84 earn €2500 each 16 earn €500 each
Question 8	80 earn €1500 each 20 earn €500 each	64 earn €2500 each 36 earn €500 each

APPENDIX A.2: VOI VERSION

Try to put yourself in the position of a recently graduated student who has to choose, just as 99 other recently graduated students, between accepting a job in firm A or in firm B. Assume that in each firm there are three types of jobs that are identical in all respects but yield a different monthly net income. The first job yields €2500, the second €1500 and the third €500. The firms differ however with respect to the numbers of positions they have available for each of the three jobs.

You and the 99 other recently graduated students either all end up in firm A or all in firm B. Each of the 100 of you has an equal probability of ending up in each of the 100 positions. So, it is unknown beforehand which job you will get.

Indicate in each of the eight questions below which firm you would prefer by marking A or B. So, the marked letter corresponds to the firm that would be preferred by you in this situation. If you consider both firms to be equally good, then mark both letters. Of course each question needs to be treated separately and a different answer can be given in each case.

Note. The formulation of the questions is identical to that of the ISO version in Appendix A.1. The questions are therefore omitted.

APPENDIX A.3: PIR VERSION

Try to put yourself in the position of a recently graduated student who has to choose between accepting a job in firm A or in firm B. Assume that in each firm there are three types of jobs that are identical in all respects but yield a different monthly net income. The first job yields €2500, the second €1500 and the third €500. The firms differ however with respect to the numbers of positions they have available for each of the three jobs. Beforehand it is not known with certainty which of the three possible jobs you will eventually get. Your chances are different in both firms.

Indicate in each of the eight questions below which firm you would prefer by marking A or B. So, the marked letter corresponds to the firm that would be preferred by you in this situation. If you consider both firms to be equally good, then mark both letters. Of course each question needs to be treated separately and a different answer can be given in each case.

	A:	B:
Question 1	100% chance that you earn €1500	20% chance that you earn €2500 75% chance that you earn €1500 5% chance that you earn €500
Question 2	100% chance that you earn €1500	80% chance that you earn €2500 20% chance that you earn €500
Question 3	25% chance that you earn €1500 75% chance that you earn €500	20% chance that you earn €2500 80% chance that you earn €500
Question 4	75% chance that you earn €2500 25% chance that you earn €1500	95% chance that you earn €2500 5% chance that you earn €500
Question 5	20% chance that you earn €2500 80% chance that you earn €1500	40% chance that you earn €2500 55% chance that you earn €1500 5% chance that you earn €500
Question 6	80% chance that you earn €1500 20% chance that you earn €500	20% chance that you earn €2500 55% chance that you earn €1500 25% chance that you earn €500
Question 7	20% chance that you earn €2500 80% chance that you earn €1500	84% chance that you earn €2500 16% chance that you earn €500
Question 8	80% chance that you earn €1500 20% chance that you earn €500	64% chance that you earn €2500 36% chance that you earn €500

AN EXPERIMENTAL STUDY OF THE POUM HYPOTHESIS

Daniele Checchi and Antonio Filippin

ABSTRACT

The "prospect of upward mobility" (POUM) hypothesis formalised by Benabou and Ok (2001a) finds explicit assumptions under which some individuals that are poorer than the average optimally choose to oppose redistribution policies. The underlying intuition is that these individuals rationally expect to be richer than average in the future. This result holds provided the mobility process is concave in expectations, redistribution policies are expected to last for a sufficiently long period and individuals are not too risk averse. This paper tests the POUM hypothesis by means of a within subjects experiment where the concavity of the mobility process, the degree of social mobility, the knowledge of personal income and the degree of inequality are used as treatments. Other determinants of the demand for redistribution, such as risk aversion and inequality aversion are (partially) controlled for via either the experiment design or the information collected during the experiment. We find that the POUM hypothesis holds under alternative specifications, even when we control for individual fixed effects.

1. THEORETICAL INSIGHTS

Benabou and Ok (2001a) have recently suggested a hypothesis to explain why a majority of poor do not fully expropriate the rich. They formally show that a rational

Inequality, Welfare and Income Distribution: Experimental Approaches
Research on Economic Inequality, Volume 11, 115–136

ISSN: 1049-2585/doi:10.1016/S1049-2585(04)11006-5

individual with an income below the average who expects to achieve an income above the average in the near future will reduce her support for redistribution, once she takes into account the income gains she may obtain in the future. They indicate this situation as prospect of upward mobility (POUM) and provide several propositions that characterise the validity of the claim.[1]

The type of redistributive scheme they adopt is a standard one in political economy models and is based on proportional taxation with universal lump-sum redistribution, without dead-weight losses from tax collection or subsidisation.[2] In such a context, the economic agent has to choose a tax rate τ_t in order to maximise the discounted flow of (expected) future income, where the net income accruing in each period is given by

$$\hat{y}_t = y_t + \tau_t(\overline{y}_t - y_t)$$

and where $\hat{y}_t$ indicates income net of taxation and $\overline{y}_t$ represents the average income in the population.[3] This set-up can be simplified by neglecting income growth ($\overline{y}_t = \overline{y}, \forall t$). When income is deterministic, the optimal choice rule, as long as inequality aversion does not matter, is zero redistribution ($\tau_t = 0$) when you are richer than average, and full redistribution ($\tau_t = 100$) when poorer than the average. When income becomes stochastic, additional elements interplay in shaping individual choices: the degree of risk aversion, the time length of validity of the tax rate to be selected and the skewness of future income distribution.

Under the assumptions of risk neutrality and deterministic transition functions, Benabou and Ok (2001a) obtain two main results:

(1) the more concave (i.e. the more skewness-reducing) the transition function, the smaller will be the fraction of population with a below-average current income supporting redistribution (Theorem 1);[4]
(2) the longer the period of validity for the chosen taxation, the smaller will be the fraction of population with a below-average current income supporting redistribution (Theorem 2).

When they consider stochastic transition functions, the concavity property is required in expectations over the following period(s). For discrete income values they also provide a characterisation of 3×3 monotone transition matrices assuring long-term, non-degenerate distributions, which are characterised by a strict majority of population voting for current redistribution that differs from the strict majority voting against future redistribution. Abandoning risk neutrality in favour of risk aversion under income uncertainty, the support for income redistribution increases, creating a trade-off between upward mobility and income insurance.[5]

Their paper provides theoretical insights into the analysis of the demand for redistribution, which has typically been analysed on empirical grounds using social surveys of the population attitude towards the role of government. Using a survey conducted in Russia in the mid-1990s, Ravallion and Lokshin (2000) find that the support for the proposition "Should the government restrict the income of the rich?" declines not only with current socio-economic status (as proxied by current expenditure) but also with the expectation of future welfare improvements. Additional evidence is offered by Corneo and Grüner (2002) using a wider sample of 12 countries surveyed in 1992.[6] In addition to the negative correlation between current status and support for redistribution (that they term "*homo œconomicus effect*"), they show that individuals take into account *societal values* and expectation on relative social position (termed "*social rivalry effect*"). These last two factors have connections with the degree of mobility. In the first case, the authors find support for the argument initially put forward by Piketty (1995), according to which people are less favourable to redistribution if they believe that individual effort is the main determinant of one's social position. On the contrary, they will favour redistribution whenever family background is held to be the major determinant of income. As a consequence, individuals (or regions/countries) who experienced high degrees of mobility will oppose redistribution, whereas we will observe support for it in less immobile societies. The last factor, social rivalry, is expected to affect the opinions of middle income families: if a middle income person fears that redistribution will reduce her social distance from the poor, she might oppose it despite a positive income gain associated to redistribution.[7]

While the POUM idea makes reference to individual prospects of mobility, both Ravallion and Lokshin (2000) and Corneo and Grüner (2002) consider the perception of mobility experienced at community level, on the argument that individuals form their expectations through on observing what happens around them. However, the two concepts can be distinguished, as done by Alesina and LaFerrara (2002). The authors have ingeniously tested the determinants of preferences for redistribution conditioning on expected income mobility. Using data covering two decades of U.S. citizens' samples, they study the determinants of preferences for redistribution,[8] controlling for current income, past experience of mobility, local perception of general (aggregate) mobility and expected future income according to observed transition matrices. While some characteristics play a significant role (minorities, women and young people being more supportive), they rely on proxies to control for risk aversion (self-employment, past experience of unemployment) and altruism (helping others) finding positive and significant effects. They also control for present income, finding a negative impact on the support for redistribution. What is more relevant for our analysis is that

they find a negative impact of the prospect of upward mobility, as measured either by subjective perceptions (by the answer to the question on whether the respondent believes that he and his family "have a good chance of improving their standard of living") or by objective measures (the local probability to go beyond mean income or the individual expected income conditional on current income). Finally they find a mixed effect of individual past experience of mobility (negative impact when measured by occupational prestige, positive impact when measured by years of education). Overall the findings of Alesina and LaFerrara (2002) do not contradict the POUM hypothesis: individuals who expect to achieve an above average income in the near future are less in favour of income redistribution, even if their current income falls below the average. Their results are robust against measured risk aversion and altruism, both factors pushing for more redistribution.

Previous literature suggests that it is almost impossible to provide a convincing test of the POUM hypothesis starting from empirical data,[9] since the demand for redistribution is affected by too many factors: self-interest, risk aversion and individual history of mobility at individual level; altruism, inequality aversion, social rivalry and perception of aggregate mobility at social level. Last, but not least, the POUM hypothesis is obtained in a median voter context, where strategic voting by agents is not considered. Had an agent to consider her direct participation in a redistributive programme, a bargaining over the tax rate could be devised, and clear-cut results are difficult to obtain. Similarly, the empirical test proposed by Alesina and LaFerrara (2002) is based on a question of what a generic government should do, and does not concern individual willingness to participate in this redistribution.

For this reason we undertake a different route, trying to disentangle as many factors as we can in a lab experiment, getting as close as we can to the different assumptions introduced by Benabou and Ok. In this way we are able to isolate (and to control for) the following factors:

- current income position (and therefore self-interest);
- individual prospect of mobility (under rational expectations[10]);
- risk aversion;
- fairness (or inequality aversion);
- length of time horizon;
- imperfect rationality.

The main limit of our strategy is that we neglect the political game. In our experiment we ask participants about their preferred tax rate, irrespective of the actual implementation of the chosen rate. Each participant chose a preferred rate in isolation as if she were the pivotal agent.[11] In this way we neglect the issue of preference aggregation, and concentrate on individual attitude towards

redistribution. These limits are shared by the previous literature: social surveys ask individuals their opinion about redistributing income, not whether they purposely operate to introduce redistributive schemes (like voting for parties promising redistributive actions). A final point is worth discussing: social surveys are based on order preferences measuring the intensity of the internally perceived desire for redistribution; lab experiments provide monetary equivalent of the same desire. We believe monetary equivalent to be a more precise and more trustworthy measure of the intensity of the demand for redistribution. This does not imply that our results actually correspond to behaviour in the real world, where agents do not have direct control over their income nor over fiscal redistribution; in addition, emotions, political ideologies and other factors affect beyond what we can control for. Nevertheless our results are still suggestive of the economic determinants of the attitude towards redistribution.

2. THE EXPERIMENT

2.1. Experimental Procedure

The experiment was run in January 2003, using the zTree software.[12] We recruited subjects from undergraduate courses at the University of Milan. All the subjects were inexperienced. Participants were randomly assigned subjects' numbers and seats. Subjects were told that their physical identity was not associated to their choices during the experiment, the subjects' numbers being their personal identification. They were given written instructions that were also read aloud by the experimenters, stressing that the amount they earned was a function of their decisions. In addition, instructions were also displayed on the screen at the beginning of each treatment (see Appendix A).

After questions were raised, a quiz was run to test their comprehension of the basic redistribution mechanism outlined in Section 1. Then, two treatments were performed. The first asked to choose a tax rate knowing one's income level and the average income. During the second only the set of three income levels and the average income were known. In both cases, the subjects' earnings corresponded in every trial to the level of their final income, net of the chosen tax rate.[13] Another quiz followed, testing subjects' comprehension of a transition matrix. The answers to the first two treatments are used in the sequel to build proxies for fairness and risk aversion, and then used in regression analysis as controls for the actual comprehension of the experiment game.

At this point, six treatments characterized by a common structure but different parameter sets were performed. Subjects, knowing their personal level of initial

income, had to choose their preferred tax rate to be applied to a sequence of unknown future incomes. Every time participants had to choose a tax rate knowing the initial income only. They were supported by both a transition matrix and a "simulator." The latter, without involving real money, allowed the computation of the sequence of expected incomes implied by the proposed transition matrix, net of three different fictitious tax rates chosen by the participant. The simulator was presented as a radio-line along which subjects could move a pointer. Only the extremes of the interval [0,100] were displayed, while the radio-line was not scaled in between (see Appendix B for the instructions submitted to the subjects). After the output of the simulator was shown, participants had to choose the "true" tax rate, i.e. the tax rate that, once applied to the sequence of future incomes, determined their earnings. In more detail, in every round each subject earned the realization of her income, unknown when the tax rate had been chosen, net of the chosen tax rate.

At the end of the experiment a questionnaire was proposed, reminding participants that their physical identity was not associated to their choices and their answers during the experiment. Questions concerned academic as well as personal information.[14] The answers are used as controls in the regressions we show in Section 3. Below (Section 2.3) we present some descriptive statistics of the pool of subjects.

Overall 95 subjects participated in the experiment, which was run in three sessions. The sessions lasted approximately 75 minutes each, and were respectively composed of 28, 30 and 37 subjects. "Points" were the currency used during the experiment, with an exchange rate of 300 points = 1 euro. Final earnings ranged between 8.2 and 10.9 Euro, and were determined by the sum of all the points collected during the experiment, i.e. the sum along periods of the points corresponding to each subject's realized income net of taxation.

One aspect of our experimental procedure needs to be stressed. All the treatments were proposed within each of the three sessions of the experiment. Hence, all the subjects played facing the whole set of parameters. This procedure implies potential carry-over effects from one parameter set to the others, as well as confounding factors arising because of framing, learning and fatigue. However, such effects can be controlled for using an econometric approach to the analysis of the data. The main advantage of this approach is that it allows the control for any observable and/or unobservable individual characteristics that might have affected the choices of the participants during the experiment, as explained in more detail in Section 3. This outcome cannot be obtained under a between-subject experiment where individual characteristics might idiosyncratically affect results across treatments without the possibility of being controlled for. For this reason

we maintain that a within-subject procedure better fits the requirements for estimating the demand for redistribution than a between-subject procedure would do.

2.2. Experimental Design

Given our aim of investigating the demand for redistribution in the lab, with a specific emphasis on the role of the POUM hypothesis, we have purposely excluded any interaction between subjects in the course of the experiment. Each participant decided on her own a tax rate to be applied to her income only, which in turn was not affected by the tax rates chosen by the other participants. Each simply chose her preferred tax rate given a predetermined and constant distribution of incomes within an imaginary population, which nobody else in the lab belonged to. Roughly speaking, every participant belonged to a different population and there were as many populations as participants.

The initial treatments were proposed to collect some information to be used to derive some (admittedly rough) proxies for fairness and risk aversion, two factors that the literature indicates as important determinants of the demand for redistribution. Treatment 1 dealt with fairness. Subjects chose a tax rate to be applied for one period to their known level of income. This choice is repeated three times, and all subjects were both rich and poor at least once. In this case the subjects had all the necessary information to choose the tax rate maximizing their net earnings known with certainty. Calling τ the chosen tax rate in treatment 1, we interpret τ (and indicate this variable as FAIR1) as a proxy for fairness of the individual when rich (i.e. with an income above the average), since a selfish optimal choice would suggest $\tau = 0$. When the individual got an income below average, the optimal choice would have required $\tau = 100$, leading to an egalitarian redistribution of incomes. If the participants held the opinion that full expropriation was excessive, they were expressing a different notion of fairness, where a certain degree of income inequality was held reasonable, or even necessary. Thus we construct a second measure for fairness as $(100 - \tau)$ (and indicate this variable as FAIR2) when the individual obtained an income below the average in Treatment 1. These two measures are positively correlated (0.378) between them.[15] A shortcoming of τ is that it may be confounded by an incomplete comprehension of the game, thus preventing subjects from fully exploiting their opportunities. This possibility is even more likely when we consider that fairness in everyday life means transferring some money to the poor, whereas during the experiment it meant leaving money to the experimenters (given the absence of interaction among the players). However,

the effect of an incomplete comprehension of the game can be separately controlled for by including the number of wrong answers provided during the first of the two quizzes that have been proposed.

Treatment 2 was designed to compute a proxy for risk aversion. Subjects chose a tax rate to be applied to an unknown level of income. They just knew the possible levels of income, the average income, and the relative probability of each occurrence. They chose under two different sets of income levels. Given the rules of the game and the equal probability of each income level, a risk lover should have chosen $\tau = 0$, while a risk averter should have chosen $\tau = 100$.[16] Given that 94 out of 95 subjects chose at least once a tax rate between these two extremes, we thought it would be implausible concluding that these subjects were risk neutral, and we decided to keep the highest tax rate chosen in the two versions of treatment 2 as a proxy for risk aversion (and indicate this variable as RISKAV).[17] There is evidence of a limited degree of risk aversion and inequality aversion, but also excessive equality seems to be rejected by our subjects. Looking at the distribution of our proxies for fairness and risk aversion, there is evidence of a limited degree of risk aversion and inequality aversion, but excessive equality also seems to be rejected by our subjects. In addition, the proxy for risk aversion is not correlated with the proxies for fairness.

The core of the experiment analysed the demand for redistribution under uncertain future incomes, obtained applying a known transition matrix to a known initial income. Six different parameter sets, summarized in Table 1 (rows 3–8), were used in each session. To reduce the role of confounding factors, all the treatments differed from one another by a change in just one parameter.

Within each of these six treatments, participants were initially assigned an initial income, randomly drawn from a uniform distribution. Knowing this initial income, they had the opportunity to simulate what they could expect to get under three fictitious tax rates. After experimenting with these simulations, they had to choose a tax rate to be applied to the unknown income in the following period. The income levels to be taxed were finally assigned conditional on the initial

Table 1. List of Treatments.

1. Tax on known income (proxy for fairness)
2. Tax on unknown income (proxy for risk aversion)
3. Transition matrix: low mobility, high income inequality
4. Transition matrix: high mobility, high income inequality
5. Transition matrix: POUM, high income inequality
6. Transition matrix: low mobility, low income inequality
7. Transition matrix: high mobility, low income inequality
8. Transition matrix: POUM, low income inequality

income and on the probabilities contained in the transition matrix. The following information was displayed at the end of the round: initial income, chosen tax rate, net and gross income obtained in the following period. This procedure was repeated three times within each treatment: the second time a sequence of three income levels was randomly assigned after the choice of the tax rate, while in the third time the chosen tax rate was applied to a sequence of five income levels randomly assigned.[18] The order of these treatments was changed in the three sessions.[19] Three permutations out of the thousands available are clearly insufficient to control for the potential role of carry-over effects, framing, learning and fatigue. Nevertheless, such problems can be dealt with when performing an econometric analysis of the results (see Section 3).

The six parameter sets consisted of three different transition matrices under two different sets of income levels. On the one hand, two different sets of income levels were used to test whether a different degree of income inequality affected the demand for redistribution. The sets of income levels were (30; 40; 65) and (20; 40; 75), respectively. Given that the realizations of initial income were drawn from a uniform distribution, it follows that the average income was always the same, i.e. equal to 45 points in all periods of every treatment. On the other hand, three different transition matrices were aimed at testing two important effects:

(i) how social mobility affects the demand for redistribution;
(ii) the POUM hypothesis.

These two effects are related but they can be tested separately. In fact, looking at the transition matrices that have been proposed during the experiment, we see that:

- matrix A, used as a benchmark, is characterized by a low income mobility and absence of POUM effect: the expected income for the following period of subjects with a median income (40) is equal to 42 in the low inequality case and to 41 in the high inequality case, i.e. lower than the average income (45) in both cases;
- matrix B also, although characterized by higher income mobility,[20] does not satisfy the requirement of the POUM hypothesis: expected future income for the median current income is again below the average income, being either 41 (low inequality) or 39 (high inequality);
- finally, matrix C implies the Prospect of Upward Mobility hypothesis, the expected income for the median current income being above the average income (47 in the low inequality case, 48 in the high inequality case). When looking at Fig. 1 reporting the expected incomes for future period, we notice that middle income still converge to the mean but from above, and not from below as in the no-POUM transition matrices.[21]

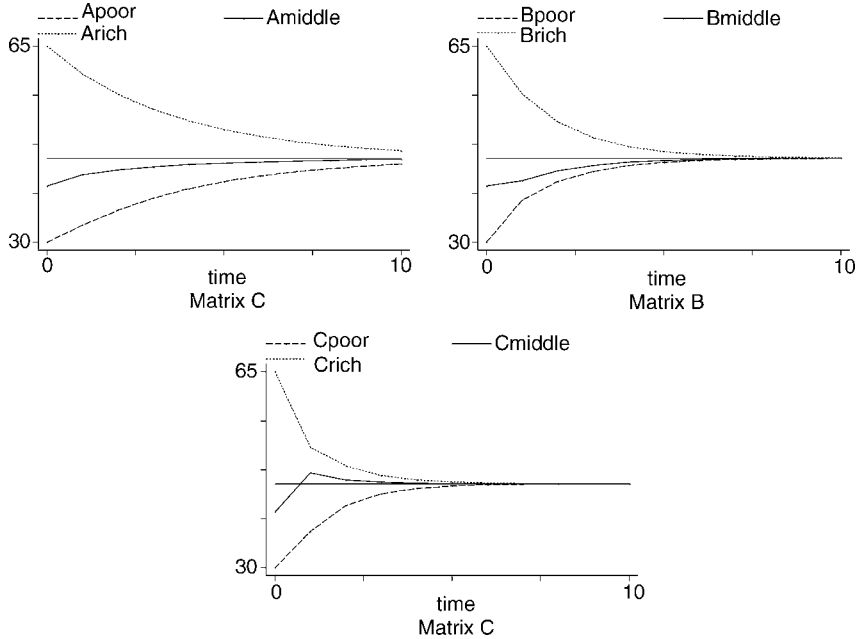


Fig. 1. Expected Incomes According to Different Transition Matrices.

In all cases, note that the different transition matrices share the feature that they leave the distribution of incomes unchanged.[22]

Matrix A

From/to	LOW	MED.	HIGH
LOW	0.7	0.3	0
MED.	0.3	0.5	0.2
HIGH	0	0.2	0.8

Matrix B

From/to	LOW	MED.	HIGH
LOW	0.5	0.4	0.1
MED.	0.4	0.4	0.2
HIGH	0.1	0.2	0.7

Matrix C

From/to	LOW	MED.	HIGH
LOW	0.6	0.3	0.1
MED.	0.3	0.3	0.4
HIGH	0.1	0.4	0.5

The subjects were confronted with all three matrices in different order (see footnote 14), and they were told that the tax they were to choose would last one,

Table 2. Descriptive Statistics of the Subjects – 95 Subjects.

Variable	Mean	Std. Dev.	Min	Max
Gender	0.40	0.49	0	1
Age	23.16	4.07	19	49
Year of enrolment	1.95	2.28	0	10
Ability	0.41	0.29	0	1
Errors 1st quiz	7.12	5.24	0	19
Errors 2nd quiz	0.69	1.81	0	8

three or five periods. Our test of the POUM hypothesis investigates whether, after controlling for all potential observables referred to the subjects and the experiment frame, the tax rate chosen when confronted with matrix C was systematically lower than the rates chosen under matrices A or B.

2.3. Sample Description

From the information collected by means of the final questionnaire, it turns out males were over represented in our sample (60% vs. 40%), and the average age of the pool was 23 years. Other descriptive statistics are reported in Table 2. Most of the participants (89%) came from the School of Political Sciences, and were enrolled in the second year of their degree program. The final mark at the exit of secondary school has been chosen as a proxy for a student's ability, after rescaling the variable in the range [0,1]. Two thirds of the sample came from high schools (*licei*) and one fourth from technical schools (*istituti tecnici*).

Two specific questions concerned the political and religious orientation, given the importance that these two variables are supposed to play in determining the demand for redistribution. An ordered scale from 0 to 5 was used to ask subjects their political orientation (0 = left; 5 = right), without any label on each possible choice. 63% of the subjects reported themselves as being centre-left, i.e. they chose a value from 0 to 2, while 37% as being centre-right. The average was 2.09 while the median choice was 2. With respect to religion, the subjects were asked to locate themselves under three alternatives: "believer and churchgoer," "believer but not churchgoer," "non believer." The proportion of the last occurrence was around one third (see Table 3).

From the quizzes proposed before the core of the experiment, it is possible to infer that subjects had, on average, an imperfect comprehension of the

Table 3. Political and Religious Orientation of the Subjects – 95 Subjects – Self-Assessment.

	Believer and Churchgoer	Believer, Not Churchgoer	Not Believer	Total
Left-wing	2	3	11	16
Left-centre	5	10	10	25
Centre-left	5	9	5	19
Centre-right	4	8	2	14
Right-centre	3	7	1	11
Right-wing	2	5	3	10
Total	21	42	32	95

redistribution mechanism, given that the average number of wrong answers was about 7 out of 21. At first sight, such a result points towards an insufficient comprehension of the basic mechanism. However, it should be taken into account that the problems to be solved within the quiz were more difficult than the problem of choosing a tax rate starting from a given level of income, also because in the latter case a simulator was available for a limited number of trials.[23] On the other hand, the comprehension of the transition matrix was very good, with on average less than 1 mistake out of 8 answers.

3. RESULTS

Our subjects reacted to the different prospect of mobility by adjusting the chosen tax rate. Since each subject was asked to choose a tax rate confronting each transition matrix twice (under reduced and enlarged income variability), under three different time horizons (one, three and five periods), 18 choices are available for each subject, leading to 1710 observations for the "optimal" tax rate. The average tax rate was 46.85, with a median value of 45 and a standard deviation of 36.6. First and third quartiles were 10 and 80 respectively. There is therefore sufficient variability in the data to explore the potential contribution of different factors. When we look at Table 4 we notice that the support for redistribution declined with the time length validity of the choice and with the degree of social mobility. *Prima facie*, the existence of prospects of upward mobility seems to reduce the support for redistribution, and the lengthening of the time horizon works in the same directions. Both results are in line with the predictions of the model proposed by Benabou and Ok (2001a), although the relation is not monotonic when the POUM holds. Since many other factors can confound these

Table 4. Conditional Means of the Preferred Rate – 1710 Choices by 95 Subjects.

Matrix/Period(s)	1	3	5
Matrix A	53.68	51.29	47.63
Matrix B	51.16	47.48	45.30
Matrix C (POUM)	47.62	37.84	39.70

results, we resort to multivariate analysis in order to control for different sources of heterogeneity.

In Table 5 we have estimated a linear model to predict the preferred tax rates by our subjects. The table reports ordinary least square projections, with heteroskedastic robust standard errors estimates.[24] In column 1 we start with the simplest version, accounting for (self reported) personal information, the initial level of income and our proxies for fairness and risk aversion. Contrary to Alesina and LaFerrara (2002) we find that women were less supportive of redistribution, although the coefficient is not significant at 95% confidence level, while similarly to them young people were more willing to redistribute. Religious attitudes do not enter significantly in our results, while political attitude does play a role: right-wing oriented individuals were less in favour of redistribution.[25] A similar attitude is found with respect to individual ability: other things constant, the subjects that experienced greater success in schooling relied less on taxation. We also control for the correct understanding of the game by including the number of errors incurred at the beginning of the experiment. The errors in the 1st quiz indicate the imperfect understanding of the principle of redistribution, and constantly induced an excess of redistribution: the subject with the maximum of errors (19 over 21) on average should have chosen 7.22% points of tax rate in excess of an identical person committing no mistakes. The errors in the 2nd quiz indicate the imperfect understanding of a transition matrix: even if less frequent (the average is 0.7 error per individual), these errors induced less redistribution. Our measures for fairness work in the expected direction: inequality aversion when rich (FAIR1) induced more redistribution (even though of limited amount)[26] whereas equality aversion when poor (FAIR2) reduced the preferred taxation.[27] Our proxy for risk aversion exerted a stronger effect: given a sample mean of 52.1 for RISKAV, we find an additional redistribution induced by this variable of 2.6% points. In column 1 we also control for some conditions potentially affecting the choice, finding that assigned initial income strongly affected the preferred rate. Despite the fact that subjects were informed that subsequent incomes were randomly obtained through the mobility matrices applied to their initial income, they seemed to place strong emphasis on their starting conditions: when initially rich, they chose a sharply

Table 5. Determinants of Preferred Tax Rates – 95 Subjects.

Model	1	2	3
No. obs.	1710	1710	1710
Gender (1 = female)	−2.41	−2.58	
	(−1.57)	(−1.72)	
Age	−0.53**	−0.53**	
	(−2.96)	(−2.95)	
Polit. attitude	−1.35**	−1.32**	
	(−2.69)	(−2.66)	
Relig. attitude	0.65	0.72	
	(0.62)	(0.71)	
Talent	−8.37**	−9.04**	
	(−3.57)	(−3.89)	
Error1	0.38**	0.37**	
	(2.71)	(2.70)	
Error2	−2.09**	−2.08**	
	(−4.84)	(−4.80)	
FAIR1	0.38**	0.33**	
	(6.67)	(5.64)	
FAIR2	−0.07**	−0.036	
	(−2.58)	(−1.36)	
RISKAV	0.05	0.053*	
	(1.85)	(2.10)	
Matrix B	−1.01	−0.54	−0.10
	(−0.66)	(−0.33)	(−0.06)
Matrix C (poum)	−9.60**		
	(−5.96)		
Matrix C × Low income		−7.94**	−6.89**
		(−3.14)	(−2.33)
Matrix C × Middle income		−24.43**	−25.41**
		(−7.76)	(−8.31)
Matrix C × High income		0.25	−2.09
		(0.08)	(0.66)
Length 3 pr		−3.50*	−2.70
		(−1.85)	(−1.47)
Length 5 pr		−6.07**	−5.18**
		(−3.13)	(−2.73)
Initial income	−1.37**	−1.40**	−1.30**
	(−33.14)	(−32.84)	(−34.79)
Initial income dispersion	−1.42	−2.39	−2.10
	(−1.09)	(−1.49)	(−1.34)
Experienced volatility		23.76**	12.08
		(3.11)	(1.34)

Table 5. (*Continued*)

Model	1	2	3
Constant	Yes	Yes	Yes
Year enrl	Yes	Yes	
Secondary	Yes	Yes	
Faculty	Yes	Yes	
Treatment		Yes	Yes
Matrix C × Period		Yes	Yes
Identifier			Yes
R^2	0.472	0.502	0.829

Note: Ols robust standard errors. *t*-Statistics in parentheses.
$^{*}p < 0.05$.
$^{**}p < 0.01$.

lower taxation than when initially poor. This effect is robust against all possible specifications. On the contrary, the extent of initial dispersion did not exert any role, but this does not come as a surprise, since we are already controlling for risk aversion. Without further controls, we find that the presence of matrix C (characterised by the prospect of upward mobility – see Section 2) induced a reduction of 9.15% points in the preferred tax rate. This effect is robust against all successive specifications. It is worth noting that this effect is independent of greater mobility: in fact, matrix B is characterised by higher mobility than matrix A (the excluded case), but its presence was not associated with a significantly lower redistribution (as conditional means of Table 4 would have erroneously suggested).

Column 2 decomposes the POUM effect by interacting it with the type of initial income, and finds that the POUM effect was much more pronounced for individuals that were below (but close to) average income, while there was no effect above the average. Note that these coefficients do not mean that the tax rate chosen by middle-income receivers is lower than the tax chosen by the low-income receivers, because the effect of the initial income must also be taken into account when predicting the optimal demand for redistribution. We also wanted to test whether past experience of mobility (during the course of the experiment) could have affected current choices. Therefore, for each subject we construct a moving dispersion measure over the incomes obtained during previous treatments. Our expectation is that, conditional on perfect partialling out of risk aversion (given the existence of a specific control for it), the volatility experienced by a subject should not affect the optimal choice of redistribution. However, despite the increased significance of the RISKAV coefficient, the experienced volatility bears a positive and significant sign, indicating that our proxy for risk aversion is not perfectly capturing the underlying phenomenon.[28] In fact, when individual effects are introduced in

column 3 to capture all the individual characteristics (including risk aversion), the size and the magnitude of this coefficient changes sharply. Column 2 also adds further controls related to the framing of the experiment: the order in which the treatment appeared, the length of the period of validity of the tax rate alone and interacted with the occurrence of matrix C types. In more details, the number of periods during which the chosen tax rate was effective influenced the preferred tax rate. In line with the model proposed by Benabou and Ok (2001a), the longer the horizon the lower the taxation. However, this effect is not significantly different across matrices, meaning that the same reduction in the preferred tax rate happens regardless of the POUM hypothesis being satisfied or not. The dummies for the order in which treatment appeared turn out not statistically significant, meaning that framing, learning and fatigue did not produce a recognizable pattern in the data. It is worth noting that the changes in the magnitude or in the significance of some coefficients already included in column 1 (e.g. FAIR2) are due to the decomposition of the POUM matrix. In fact, adding only the framing variables to the regression shown in column 1 does not produce significant changes. Despite the additional controls, the POUM effect keeps on holding and the size and the significance of all the coefficients do not vary.

Finally, our strongest check is given by including individual fixed effects in column 3. Even in this case we find that initial income and the type of transition matrix are the main determinants of the extent of preferred redistribution. Other things being constant (and in this specification we are controlling for unobservables as well), the presence of a C-type transition matrix, characterised by prospect of upward mobility, induced a strong reduction in preferred taxation for middle income subjects (in the order of 25% points). Stronger mobility *per se* (captured by B-type transition matrices) did not induce reduction of redistribution.

4. CONCLUSIONS

In this paper we account for a lab experiment where 95 subjects were confronted with the choice of their preferred extent of income redistribution, when different transition matrices stochastically assigned incomes over different time horizons. We find that the preferred taxation declines when the transition matrices are characterised by prospect of upward mobility (i.e. individual below average income have an above average expected income for the next period). This result is robust against alternative specifications, accounting for individual factors (such as inequality and/or risk aversion) and framing effects. It holds even when we introduce individual fixed effects accounting for individual unobservables. Thus

our results support the theoretical predictions of the POUM hypothesis proposed by Benabou and Ok (2001a), showing that under specific transition matrices the support to redistribution is reduced in the vicinity of average income. Also in line with theoretical predictions, we find that longer time horizons call for reduced taxation, because individuals appreciate the freedom of changing the optimal tax when confronted with different incomes in the future. These results hold even when controlling for risk aversion, that on theoretical grounds represent a confounding factor.

In our opinion, the main limit of the present research is the imperfect measures of fairness and risk aversion. However, using fixed effect controls takes care, on econometric grounds, of all these unobservable. We have also avoided to frame social interactions in tax selection, which probably dominates the optimal selection of tax rate in many contexts (for example where tax avoidance is made available to the subjects). This constitutes our next research issue.

NOTES

1. Similarly, Ravallion and Lokshin (2000) borrow the expression of "tunnel effect" to indicate that the attitude towards redistribution depends of expectations over future income trends: "We see that amongst people who expect welfare to fall, there is a very high support for restricting incomes of the rich, and the support is affected little by current level of living. By contrast, support for redistribution is lower than average among those who expect welfare to rise, and is sharply attenuated by higher current levels of living within this group" (p. 97).
2. Linear taxation schedules are discussed in Meltzer and Richards (1981).
3. Notice that we have implicitly assumed that tax rate is expected to last for the future, thus excluding time consistency problems.
4. The analogy between transition function and progressive, balanced budget, redistributive schemes is developed in Benabou and Ok (2001b).
5. In other words, the curvature of transition function and utility work in opposite directions.
6. The crucial question they exploit says "Is it the responsibility of the government to reduce the differences in income between people with high incomes and those with low incomes?"
7. And conversely, a middle-income person with an income above the mean may support redistributive programs despite the associated capital loss if the reform reduces social distance with the elites. This is typical of educational expenditure: see Corneo and Grüner (2000).
8. The support for redistribution is measured by the answers to the following question: "Should the government reduce income differences between the rich and the poor, perhaps by raising the taxes of wealthy families or by giving income assistance to the poor?"
9. Benabou and Ok (2001a) make use of empirically estimated transition matrices from PSID samples to show that under risk neutrality actual data allows for strict majority against redistribution, whereas even small amounts of risk aversion dominate mobility prospects.

10. As it will be made clear in the sequel, we provided the participants with a device allowing them to compute their expected income in the future for any given transition matrix they were facing.

11. Since we provided participants with information about the income distribution, in principle each participant would have been able to assess the likelihood of choice of her preferred rate. But actual gains were obtained using the chosen rate, thus implicitly assuring its implementation.

12. The zTree software has been developed at the University of Zurich, Institute for the Empirical Research in Economics, see Fischbacher (2002).

13. Cowell and Schokkaert (2001) claim that it is unclear that rewards to participants are necessary in the context of social judgements. However, we decided to reward subjects even in this treatment to mimic the fact that fairness is costly.

14. We did not include in the questionnaire verbal questions addressed to find out the attitudes of individuals towards the POUM. Although this would have allowed us to check the numerical answers obtained in the experiment, as Amiel and Cowell (1999) suggest, we did not want to overload the subjects since the design of the experiment required already a large amount of time.

15. Among the determinants of social preferences analysed by Fong (2001), only altruism can contribute to explain FAIR1 and FAIR2. In fact, the setting of the experiment, and in particular the lack of interaction among subjects, prevents equity and reciprocity from playing any role.

16. There is a debate in the literature about whether people are really expected-utility maximizers, and how small-stake gambles in the laboratory, like in Treatment 2 in our experiment, relate with behaviour in the real world (see Rabin, 2000 and a response by Palacio-Huerta et al., 2001).

17. We also tried with an average between the two, finding results that are slightly less significant.

18. The sequences of future income levels were obtained by iterated randomisation based on the probabilities contained in the transition matrix.

19. The sequence of treatments in the three sessions was as follows:

Session 1:	1 2 3 4 5 8 7 6
Session 2:	2 1 8 5 4 3 6 7
Session 3:	2 1 7 6 3 4 5 8

20. The three matrices can be ordered in terms of mobility. If we take the second eigenvalue as a measure of relative immobility, the compute values are the following: $A = 0.764$; $B = 0.564$; $C = 0.464$. See also the speed of mean-regression from Fig. 1.

21. This is equivalent to Fig. 2 in Benabou and Ok (2001a).

22. From the matrices A and B, i.e. when the POUM does not hold, it could be possible in principle to infer something about the risk aversion and the fairness of the subjects. In fact, being the expected income of the non-rich subjects always below the mean, a choice of a tax rate lower than 100% would signal either risk loving or fairness to the rich. Unfortunately, it is not possible to distinguish between the two. Moreover, the correlation between fraction of times in which a subject has chosen a tax rates below 100% when it would have been rational to do so, and any of our proxies of fairness and risk aversion is very low.

23. Given three different tax rates, in the first quiz subjects had to indicate, for each of three different levels of income:

(a) whether a tax or a subsidy emerged;
(b) the amount of the tax or the subsidy;
(c) how much the final income (net of taxation) was.

24. We have also estimated weighed least square, using the inverse of errors as weights, but the results are substantially unaltered. Available from the authors.
25. Constant, Year of enrolment, Secondary school type and Faculty attended are always included as maintained controls.
26. Given a sample mean for FAIR1 of 4.0, the additional redistribution induced by this variable is in the order of 1.5 percentage points.
27. Given a sample mean for FAIR2 of 24.4, the additional redistribution induced by this variable is in the order of 1.7 percentage points.
28. We also counted the number of upward or downward transitions as an alternative measure of individual mobility, without finding statistically significant effects.

ACKNOWLEDGMENTS

Financial support from MIUR-COFIN funds is gratefully acknowledged. We thank Roland Benabou for constructive comments that helped to improve the paper. We also thank Cecilia Garcia-Peñalosa, John Hey, Eliana La Ferrara, Michele Santoni and two anonymous referees for valuable suggestions. All the remaining errors are ours.

REFERENCES

Alesina, A., & La Ferrara, E. (2002). Preferences for redistribution in the land of opportunities. mimeo.
Amiel, Y., & Cowell, F. A. (1999). *Thinking about inequality*. Cambridge: Cambridge University Press.
Benabou, R., & Ok, E. (2001a). Social mobility and the demand for redistribution: The POUM hypothesis. *Quarterly Journal of Economics*, *116*(2), 447–487.
Benabou, R., & Ok, E. (2001b). Mobility as progressivity: Ranking income processes according to equality of opportunity. NBER Working Paper 8431.
Corneo, G., & Grüner, H. P. (2000). Social limits to redistribution. *American Economic Review*, *90*(3), 1491–1507.
Corneo, G., & Grüner, H. P. (2002). Individual preferences for political redistribution. *Journal of Public Economics*, *83*, 83–107.
Cowell, F. A., & Schokkaert, E. (2001). Risk perceptions and distributional judgements. *European Economic Review*, *45*, 941–952.
Fong, C. (2001). Social preferences, self-interest and the demand for redistribution. *Journal of Public Economics*, *82*, 225–246.
Fischbacher, U. (2002). zTree experimenter's manual. University of Zurich, mimeo.

Meltzer, A., & Richards, S. (1981). A rational theory of the size of government. *Journal of Political Economy*, *89*, 914–927.

Palacio-Huerta, I., Serrano, R., & Volij, O. (2001). Rejecting small gambles under expected utility: A comment on Rabin. WP 2001–05, Department of Economics, Brown University.

Piketty, T. (1995). Social mobility and redistributive politics. *Quarterly Journal of Economics*, *110*(3), 551–584.

Rabin, M. (2000). Risk aversion and expected-utility: A calibration theorem. *Econometrica*, *68*, 1281–1292.

Ravallion, M., & Lokshin, M. (2000). Who wants to redistribute? The tunnel effect in 1990s Russia. *Journal of Public Economics*, *76*, 87–104.

APPENDIX A

Instructions Provided Before the Experiment Started

"Points" are the currency during the experiment, with an exchange rate of 300 points = 1 euro.

There are two main concepts behind the experiment.

Taxation

During the experiment you have to choose a tax rate to be applied to your income. This choice will be repeated several times under different rules and conditions.

The choice of every participant affects his/her income only. In a parallel manner, one's income is not affected by the choices of other participants.

In every repetition of the game, you will be assigned an income from three possible levels. You will be asked to choose a tax rate to be applied to that income (in some cases after but usually before knowing it) to determine your earnings in that repetition.

Your earnings in every repetition of the experiment depends upon two factors:

(1) the income level that is assigned to you;
(2) the tax rate that you choose.

What does taxation imply? A tax rate t

(a) determines a decrease of incomes above the average. In particular, a fraction t of the income above the average is collected;
(b) determines an increase of incomes below the average. In particular, it provides a subsidy equal to a fraction t of the difference between the average income and the income that has been assigned.

For instance, if there are three income levels (10, 20, 42) and the average income is 24, the effect of a tax rate equal to 25% is as follows:

Income = 10: taxation determines a subsidy equal to the 25% of (24 – 10), i.e. 3.5 — income after taxation = 13.5

Income = 20: taxation determines a subsidy equal to the 25% of (24 – 10), i.e. 1 — income after taxation = 21

Income = 42: taxation determines a contribution equal to the 25% of (42 – 24), i.e. 4.5 — income after taxation = 37.5

Roughly speaking, taxation redistributes income from the rich individuals to the poor ones, where rich and poor in this context mean with an income above and below the average, respectively.

The Transition Matrix

The transition matrix is a table that summarizes which probability every participant has to reach a given level of income given a starting level of income

	To 10 (%)	To 20 (%)	To 42 (%)
From 10	60	30	10
From 20	30	30	40
From 42	10	40	50

For instance, let's focus on an individual whose starting income is 10. In the following period, his/her income will stay equal to 10 in the 60% of the cases, it will be equal to 20 in the 30% of the cases and it will be equal to 42 in the remaining 10% of the cases. If his/her income in the next period will actually be equal to 42, in the subsequent period his/her income will be equal to10 in the 10% of the cases, it will be equal to 20 in the 40% of the cases and it will stay equal to 42 in the remaining 50% of the cases. The process can be iterated even further in a similar manner.

NB: The transition matrix always refers to INCOMES BEFORE TAXATION. In other words, the matrix refers to gross incomes.

APPENDIX B

Instructions Concerning the Simulator
(Displayed During the Experiment)

The Simulator

The goal of the simulator it to make you familiar with the structure of the game, but it does not directly affect your earnings. You have the possibility to introduce three mock tax rates. The simulator will compute your expected income for next period(s) based on the probabilities displayed by the transition matrix and according to your starting income. This expected income will be modified (deducting taxes if above the average and adding subsidies if below the average) according to each of the mock tax rates you introduced.

ON THE ATTITUDE TOWARDS INEQUALITY

Liema Davidovitz and Yoram Kroll

ABSTRACT

This paper presents an experimental framework for separating the attitude toward inequality from the attitude toward risk. This exploratory experimental study examines the attitude toward inequality while keeping risk constant. The results support the hypothesis of inequality aversion only among middle-income subjects. More interestingly we found that higher equality motivates individuals to take more risk and challenge. This result is a counterpoint to the standard line that inequality is needed to encourage effort.

1. INTRODUCTION

Inequality Aversion (IA) and Risk Aversion (RA) are widely assumed in the economics literature. However, IA is based on ideological and moral beliefs regarding relative incomes among individuals, while RA is related to the individual's preferences for his own income distribution.

Despite this basic difference between the two, researchers analyze them with the same theoretical approaches and empirical evidence. This paper presents a new approach for identifying inequality aversion while avoiding the potential confusion with risk aversion. We propose that the attitude toward inequality be analyzed only as a response to change in inequality among individuals, while maintaining the level of risk.

Inequality, Welfare and Income Distribution: Experimental Approaches
Research on Economic Inequality, Volume 11, 137–147

ISSN: 1049-2585/doi:10.1016/S1049-2585(04)11007-7

Rothschild and Stiglitz (R&S) (1970) provided conditions[1] under which one alternative is riskier than the other. They proved that under these conditions, every risk averter prefers the less risky alternative. Furthermore, by this definition an individual with an unknown utility function, who prefers the less risky alternative according to the integral rules, *must be a risk averter*.

Atkinson (1970), Kolm (1969), Harsanyi (1977) and others applied R&S' definition of more risk to define higher inequality rather than risk. Accordingly, an individual who selects the less risky situation, which is also the more "equal" situation, is then considered to prefer equality (see Amiel & Cowell, 1992, 1994; Cowell, 1985; Gevers et al., 1979; Glejser et al., 1977; Pfingsten, 1988).

In all of the above studies, the means of the two alternatives are the same. However, the level of risk and the level of equality are not identical. Therefore, selecting the more egalitarian economy, which is also the less risky economy, can be due either to RA or to IA. According to our approach, the attitude toward inequality should be analyzed as a response to a change only in inequality among individuals, while maintaining the moments of income distribution constant. Our experiment simulates such a change when an individual moves from a "common gamble" (CG) to an "individual gamble" (IG) environment. In CG the same results, either high or low, apply to all individuals in the reference team. In IG each individual faces a separate gamble and he/she may have results that differ from the results of the other in his team. This cannot happen in a "common gamble," in which all participants receive the same result. The risk in both situations is identical, while only the "common gamble" represents a state of total equality (see Kroll & Davidovitz, 2003).

The next section presents a basic definition of CG and IG. Section 3 describes laboratory experimental tests of IA and the results of the experiment. Conclusions and points for further analysis are in the last section.

2. BASIC DEFINITIONS

Assume n participants in the economy and a random income X with a distribution function $F(X)$. Define an ordered statistic X_i from a sample of size n taken from $F(X)$ by $(X_1, X_2, \ldots, X_n)$. where $X_i \in X \; \forall i$, and $X_1 \leq X_2 \leq \cdots \leq X_n$.

Assume all participants in the economy draw their income randomly from $F(X)$.

Let us define two alternative types of gambles:

Definition I. "Common Gamble" (CG).

All participants receive the same x_i drawn from $F(X)$ by *one mutual draw*.

Definition II. "Individual Gamble" (IG).
Each participant *independently* draws an ordered income x_i from $F(X)$.

In the CG case, each individual has the same realized income x out of $F(X)$ while in the case of IG, each individual is also gambling on his ordered income X_i. CG represents complete ex-ante and ex-post equality whereas IG represents ex-ante inequality with only a small probability of ex-post equality (which occurs only when all n participants draw the same result). The degree of actual ex-post inequality depends on the distribution properties of $F(X)$ and the number of participants n.

According to the traditional Von Neumann and Morgenstern (1947) expected utility approach, the total income (or total wealth) determines preferences between alternatives. On the other extreme, Tversky (1989) claims that the behavior of individuals is more affected by their marginal results rather than by overall or total wealth results. This "mental accounting," in which each decision is treated separately by individuals, recently received the support of Andreoni and Miller (1996).

Our definitions fit situations in which individuals' preferences are determined by marginal distributions. In case total rather than marginal wealth determines preferences, our approach is still appropriate, if marginal prizes are small enough and independent of initial wealth so that the ranking of the incomes cannot be changed due to IG. In such cases CG generates a more equal ex-ante situation than IG.

3. THE EXPERIMENT

The task and reward were selected so that the main goal of the experiment, that of testing attitudes toward inequality, was completely disguised from the students. The students were asked to participate in a portfolio game, and we promised to grant them an additional grade on the last team assignment that they had just handed in, but was not yet graded. The additional percentage grade points were promised to be equal to the percentage return that their selected portfolio would yield.

The subjects were 213 graduate and undergraduate business administration and economics students: 81 graduate students at the Hebrew University of Jerusalem and 132 undergraduate students of the Ruppin Accademic Center. All the students from both universities had taken an investment course and were familiar with risk and portfolio investment theories. The students regularly work in teams of three students.[2]

The formal task of the participants was to select nonnegative proportions (i.e. short sales were not allowed) of two assets portfolio. One of the assets is risk-free with an expected return of 3%; the other is a risky asset whose return is normally distributed with an expected return of 5% and standard deviation of 7%.

In order to ensure full understanding of the risk involved in selecting the risky asset, the complete distribution of the risky asset was plotted for the students. Furthermore, the lower quartiles and standard deviation were calculated for them. (The appendixes include the instructions and explanations which where given to the participants.)

The main goal of the experiment, of testing attitudes toward inequality, was obtained by dividing the participants into two groups, wherein one carried out the portfolio experiments under CG conditions and the other – under IG conditions.

In the CG group, the experimenter told each team that the yield result of the risky asset would apply to all team members. The participants in the CG group were asked to select their own proportions of the risky asset. It was very probable that they would select different proportions (according to their RA). Only if all CG team members select the same portfolio, would they all obtain the same return and would thence be guaranteed complete equality.

Namely, even if the same risky yield is guaranteed to all members in the team they still may have different overall portfolio results due to different proportions of the risky asset. In order to avoid such an inequality possibility, they were also told that the portfolio return will be calculated by using *only one proportion which would be drawn randomly out of the* proportions between the risky and riskless asset which were selected by the three team members. This procedure guranteed that all CG team members would obtain the same return and complete equality would be guaranteed within each team.[3] (Appendix A presents the instructions and explanations which were given to the CG participants.)

The other IG group was informed that each member of the team would carry out a separate gamble and would receive his own specific result (the yield of the risky asset). (Appendix B presents the instructions and explanations that were given to the IG participants.)

In order to eliminate possible deviations in results due to gender, the numbers of females and males in the CG group is identical to their number in IG group.

The students could not consult with their friends and they were guaranteed confidentiality for their answers. Each student participated in only one experimental setting (type of gamble). The students did not know that there were two types of gambles in the experiment.

The experiment was conducted in the classroom during class time. The students had the option not to participate in the experiment. However, since the experiment

provided a risk free bonus option that was significant for the students, all the students decided to participate.[4]

In addition to the main task of portfolio choice, the students had to estimate their grade (before the bonus) in the course. This estimation task was added in order to find out whether there are differences in the attitude toward inequality that are due to the differences in the economic-social level of the participants. This level is simulated by the expected GP of the students. Confidentiality was guaranteed to the students for these estimates along with the assurance that the data would only be used in the research.

Hypothesis I. Risk averters (RA), who are also Inequality-Averters (IA), will select a higher proportion of the risky asset if they are facing CG rather than IG.

The reasoning behind this hypothesis is that the equality feature of CG is a compensation for the risk and inequality averse investor. In the IG group, equality may be achieved only by accepting a lower return of the risk-free asset. In the CG group subjects do not have to take a lower proportion of the risky asset in order to lower inequality and their only reason to select less of the risky asset is to reduce the involved risk.

Hypothesis II. The portfolio decisions under CG and IG will also be affected by the self-projected grades of the subjects. Smaller differences between the CG and IG portfolio decisions are expected for students who project either below or above average grades.

The reason for the lower preference of CG among subjects close to the lower and upper bounds of the grade projection is that under a higher inequality regime their deviation from the "norm" is less visible and has a lower probability of potential penalties.

Students with low projected grades may have another reason to prefer IG, because under IG they have some positive chance to change their low ranking. This opportunity does not exist if they select CG.

3.1. Results

In order to test the difference between the portfolios of the two groups (CG vs. IG group) we used Mann-Whitney (MW) test.[5] This procedure examines whether two independent samples come from the same population.

We found that the subjects selected higher proportions of the risky asset under the CG regime than under the IG regime. The average proportion invested in the risky asset is 52.5% under CG compared to only 43.5% for the IG group. This

Table 1. Means and Standard Deviations of Investment in the Risky Asset.

	Common Gamble	Individual Gamble
Number of participants	113	100
Mean of the proportion invested in the risky asset	52.5%	43.5%
Standard deviation	26.1%	26.7%

Note: Results of the test were: $Z = -1.965$, $p < 0.025$, df $= 211$.

result can be interpreted as support for inequality aversion. These results support the first hypothesis with statistical significance.

The results for the first hypothesis are examined in Table 1.

This table presents the main results without taking into consideration the student's grade prediction. Each column contains the number of subjects and the mean and standard deviation of the proportions (in percent terms) invested in the risky asset for each the type of gamble.

In order to test the second hypothesis the students were divided into five categories according to projected final grades. The categories were as follows: far below average, below average, average, above average, and far above average (see Table 2 and the graphic presentation of the table).

Table 2 presents the proportion (in per cent) of investment in the risky asset for each of the five "projected grade" categories. (The number of subjects in each category is in parenthesis.)

The hypothesis is tested by ANOVA. The dependent variable is the investment in the risky asset. The type of gamble (two groups) and the projected grades (five category) are the independent variables.

The main effect is significant with $F = 3.831$, $p < 0.002$. The two partial non-interacted effects are also significant. The differences by types of gamble are significant, $F = 5.516$, $p < 0.02$. The differences according to grade are significant with $F = 3.208$, $p < 0.014$. We did not find significant interaction

Table 2. Average Proportion of Investment in the Risky Asset According to Projected Final Grades.

Mark/ Gamble	FAR ABOVE AVERAGE	ABOVE AVERAGE	AVERAGE	BELOW AVERAGE	FAR BELOW AVERAGE
CG	54.1% (15)	38.5% (13)	50.3% (51)	63.3% (24)	52.5% (10)
IG	59% (15)	35.3% (20)	39.5% (35)	45% (20)	47.5% (10)
	NS	NS	$t = 1.9$, $p < 0.03$	$t = 2.2$, $p < 0.02$	NS

Note: The number of participants is in parenthesis.

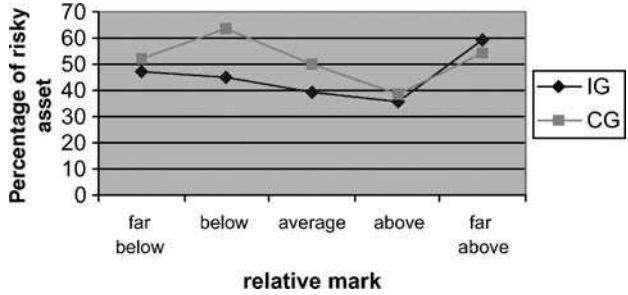


Fig. 1.

between the dependent variables. Therefore, a post hoc *t*-test for differences between the sub-groups is applicable.

The results in most cases support the second hypothesis. In other words, only students with or just below average projected grades invested significantly more in the risky asset under CG (see Table 2).

As anticipated, there were no significant differences in investment in the risky asset between CG and IG for students with projections higher or lower than the average.

The results in Table 2 are depicted in Fig. 1.

Notice that investment in the risky asset is higher for CG only among subjects with projected grades close to the average.

4. CONCLUSION

The paper presents a new approach for identifying inequality aversion while avoiding the potential confusion with risk aversion. The approach is based on comparing decisions under “individual” gambles vs. “common” gambles with the same distribution of outcomes in both.

The results of a laboratory experiment support the main hypothesis that individuals are inequality averters. However, when we grouped them according to their relative mark, this aversion was not evident.

The most important part of our experiment, which separates the impact of more risk from more inequality, is that if risk-aversion as well as inequality-aversion are assumed, then higher equality among participants can motivate risky efforts. This positive impact can be obtained if the risky results of efforts are shared more evenly among the inequality-averse participants.

Note that this positive impact of egalitarian economy on the motivation to risky effort is in contrast to the convention that egalitarian economy reduces motivation toward risky challenges.

Without denying the reasoning behind conventional thought, we claim that our results support a counter point to the convention. Namely, inequality aversion may reduce and even reverse the negative impact of egalitarian economy on the risky efforts motivation (see also Sadrieh & Verbon, 2002).

One should note that our support of egalitarian economy is limited to our data set and experiment's structure. Namely, preference of equality was found within teams of three members who worked closely with each other on the same academic project. Thus, their preference of equal grade on this project for all team members is clear. Further studies should reexamine inequality aversion among larger groups with more anonymous members who take individual tasks. We expect that the larger the group and the more anonymity among its members and the more individual is the task, the aversion to inequality will be lower.

NOTES

1. Rothschild and Stiglitz (1970) provide three coinciding definitions for a random variable Y to be riskier than a random variable X:

(i) Y distributes like X plus a random variable Z which is independent of X and satisfies $E(Z|X) = 0$.
(ii) Every risk averter prefers X to Y.
(iii) Y has the same expected return as X but has more distribution weight in the "tails."

2. We didn't include students who worked on their own in the sample.

3. Note that under this procedure each CG participant knows that there is only 1/3 chances that his actual portfolio selection will be the one selected. The alternative to this procedure is that the actual common proportion will be the average of the selected propotrtion of the three team members.

4. Eckel and Grossman (2000) find that pseudo-volunteers subjects, like in our experiments, are more extreme than pure volunteers and are more affected by nonmonetry rewards. In order to avoid such an effect, as well as other side effects, we completely disguised the main goal of the experiment.

5. The MW test is an a-parametric test that does not require any prior assumptions or specifications related to the distribution of the selected proportions.

ACKNOWLEDGMENTS

The authors acknowledge the financial support of the Kruger Center and a very helpful comment of an anonymous referee.

REFERENCES

Amiel, Y., & Cowell, F. A. (1992). Measurement of income inequality: Experimental test by questionnaire. *Journal of Public Economic, 40,* 3–26.

Amiel, Y., & Cowell, F. A. (1994). Monotonicity, dominance and the pareto principle. *Economics Letters, 45,* 447–450.

Andreoni, J., & Miller, J. H. (1996). *Giving according to GARP: An experimental study of rationality and altruism*. Madison: University of Wisconsin.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory, 2,* 244–263.

Cowell, F. A. (1985). 'A fair suck of the sauce bottle' – Or what do you mean by inequality? *Economic Record, 6,* 567–579.

Eckel, C. C., & Grossman, P. J. (2000). Volunteers and pseudo-volunteers: The effect of recruitment method in dictator experiments. *Experimental Economics, 3,* 107–120.

Gevers, L., Glejser, H., & Rouyer, J. (1979). Professed inequality aversion and its error component. *Scandinavian Journal of Economics, 81,* 238–243.

Glejser, H., Gevers, L., Lambot, Ph., & Morales, J. A. (1977). An econometric study of the variables determining inequality aversion among students. *Economic Review, 10,* 173–188.

Harsanyi, J. C. (1977). *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge: Cambridge University Press.

Kolm, S. Ch. (1969). The optimal production of social justice. In: J. Margolis & H. Guitton (Eds), *Public Economic*. London: Macmillan.

Kroll, Y., & Davidovitz, L. (2003). Inequality aversion vs. risk aversion. *Economica, 70,* 19–29.

Pfingsten, A. (1988). *Empirical investigation of inequality concept: A method and first results*. Karlsruhe University.

Rothschild, M., & Stiglitz, J. E. (1970). Increasing risk: A definition and its economic consequences. *Journal of Economic Theory, 2,* 225–243.

Sadrieh, A., & Verbon, H. A. A. (2002). Inequality, trust and growth: An experimental study. Discussion Paper, 84. Tilburg University, Center for Economic Research.

Tversky, A. (1989). In: W. Sharpe & K. F. Shefered (Eds), *Quantifying the Market Risk Premium Phenomenon for Investment Decision Making*. The Institute of Chartered Financial Analysis (CFA).

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

APPENDIX A

Instructions to the Subjects
(Translated from Hebrew)

This assignment is for research purposes only. You have the right not to participate. We intend to grant you a bonus in terms of points added to the grade of the last

group home assignment that you just handed in. Your bonus will be equal to the yield that you achieve in the following investment portfolio game.

Your task is to construct a portfolio consisting of a risky asset and a risk-free asset. The proportion of each asset in your selected portfolio should be between 0 and 100%, and the total investment in both should be 100%.

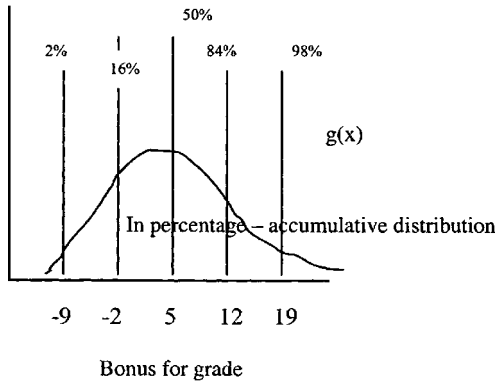
The assests are:

(1) The risk-free return is 3%.
(2) The risky asset is normally distributed with an expected return of 5% and standard deviation of 7%.

(Below is a chart depicting the distribution and its quartiles.)

Following the individual portfolio decisions made by you and the others in your group, a lottery will pick one portfolio for all members in your group. The return on the risky asset will then be determined *by one common drawing* out of normal distribution.
Please note that this procedure will lead to an identical portfolio return for all group members.

Your decision:
Proportion of the risky asset is:
The grade I expect to receive in this course is:
Name:
My group members are: 1._____ 2._____ 3._____.



APPENDIX B

Instructions to the Subjects (Translated from Hebrew)

This assignment is for research purposes only. You have the right not to participate. We intend to grant you a bonus in terms of points added to the grade of the last group home assignment that you just handed in. Your bonus will be equal to the yield that you achieve in the following investment portfolio game.

Your task is to construct a portfolio consisting of a risky asset and a risk-free asset. The proportion of each asset in your selected portfolio should be between 0 and 100%, and the total investment in both should be 100%.

The assests are:

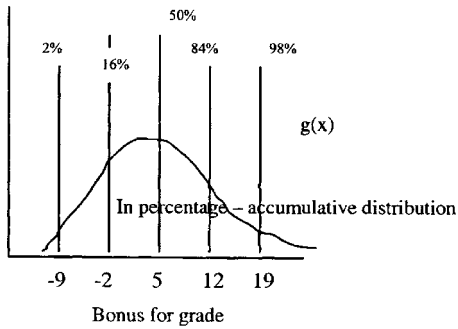
(3) The risk-free return is 3%.
(4) The risky asset is normally distributed with an expected return of 5% and standard deviation of 7%.

(Below is a chart depicting the distribution and its quartiles.)

Following your portfolio decision, the return on the risky asset will be determined by a drawing from the normal distribution.
Please note that the drawing will be separate and independent for each student. Namely even if two students select the same proportion for the risky asset in their portfolios, their bonuses may be different.

Your decision:

Proportion of the risky asset is:
The grade I expect to receive in this course is:
Name:
My group members are: 1._____ 2._____ 3._____.



APPROACHING FAIR BEHAVIOR: DISTRIBUTIONAL AND RECIPROCAL PREFERENCES

Alexander Kritikos and Friedel Bolle

ABSTRACT

This paper suggests to combine different kind of "other-regarding" preferences as an approach to fair behavior which is observed in controlled experiments. We assert that participants in two-person experiments have a good will capital which may be described by altruistic preferences. These preferences guide a large fraction of participants when they have to make distributional choices in one-stage games. We further show that in games with more than one stage the previous action of the other person may cause reciprocal feelings in addition to the altruistic preferences. A friendly (unfriedly) act of the other person may increase (decrease) the good will capital of the participants. Upon these findings, we conclude that a combination of altruism and reciprocity is able to describe the variety of behavior in several experiments despite their differing strategic context.

1. INTRODUCTION

Among experimental economists there is one consensus: the narrow self-interested individual utility function of pay-off or profit maximization is only sometimes apt to explain the behavior of human beings. The neoclassical approach receives support from experimental economics when the assumptions of perfect

Inequality, Welfare and Income Distribution: Experimental Approaches
Research on Economic Inequality, Volume 11, 149–180

ISSN: 1049-2585/doi:10.1016/S1049-2585(04)11008-9

competition are met in market games with standardized goods (cf. Davis & Holt, 1993; Smith, 1962). The results of other experiments in particular with bilateral interaction revealed that often less than 50% of the participants behave like homo oeconomicus even if the games are played as one-shot between anonymous persons. When the participants have the impression that their individual behavior matters for their fellow player(s) the standard assumptions of homo oeconomicus fail to describe the behavior of homo sapiens in a consistent way.

In these experiments it was the emotional ability of homo sapiens and not the self-interested calculation which created social states often superior to those of homo oeconomicus. The participants showed non-selfish behavior in single-stage games like the Dictator Game. Moreover, there is overwhelming evidence that in experiments with more than one stage, the participants trusted in the willingness of their partners to cooperate, and many partners cooperated, although defection had been the "advice" of normative game theory. As the Public Good Experiments with punishment and countless replications of the Ultimatum Game show, participants were also ready to sacrifice material payoffs in order to punish those persons who did not behave in a fair way – then leading to inferior social states which completely selfish players would not have realized.[1]

Thus, motives and emotions of human beings caused but also destroyed fair behavior. These stable results lead to the central question of how to explain the reasons for any kind of behavior beyond selfishness. Experimental economists have suggested to introduce different additional motives into the utility functions such as reciprocity and distributional preferences – but there is no consent about the best way how to *model* behavior beyond selfishness.[2]

Based on the models of Arrow (1975), Becker (1976) and Collard (1978) the first approaches introduced altruistic preferences into the utility function where the utility of the altruistically behaving person depends either on the income or the utility of his fellow player.[3] It is remarkable that general models on altruism include the existence of spite (a negative influence of the other's well-being) or inequity aversion (where the sign of the influence depends on the relation of income). In games with more than one stage, models of altruism are to a certain extent capable to express reciprocal choices.

Parallel to the models on altruism, approaches were developed claiming that it is the intention of each action which drives the decision to choose a strategy beyond egoistic preferences. More specifically, it is assumed that participants are guided by positive or negative concerns for their fellow participants.[4] These models introduced as intentional variable reciprocal motives (cf. Dufwenberg & Kirchsteiger, 1998; Falk & Fischbacher, 1999; Rabin, 1993) into the utility function.[5]

The objections raised by Fehr and Schmidt (1999) against the introduction of intentional motives are that from a theoretical point of view the modelling of

intentions requires the adoption of psychological game theory which is difficult to be applied (see e.g. Geanakoplos et al., 1989). They further pointed out that in all models on reciprocity reference points have to be chosen which separate friendly from unfriendly action. In the course of time, dynamic processes may transform behavior of the participants and, thus, the reference points. And there is a plethora of equilibria depending on these dynamic processes.

From an experimental point of view, Bolton et al. (1998b), Bolton and Ockenfels (BO) (2000) and Fehr and Schmidt (FS) (1999) made clear that reciprocity or altruism are only sometimes apt to explain behavior in experiments. They suggested that decisions can be captured in a better way if the focus is put on the payoffs of each action. Inspired by Loewenstein et al. (1989), they assert that participants are "concerned with securing what they consider to be their fair share" (Bolton et al., 1998b, p. 295). In response to the crucial question of how a fair share is established, BO and FS introduce the equal outcome as reference point. Participants are supposed to behave "inequity averse." Therefore, next to egoism it is inequity aversion which should be introduced into the utility function.[6]

However, doubts were raised about the appropriate choice of the reference point and whether the models of FS and BO are able to consistently capture intentional motives of participants. With respect to the reference point of an equal split, Konow (2001) notes that "equality is not a principle of fairness, at best it is a special case of the principles, when members are equally accountable, efficient or needy." Konow shows that equal splits represent the observer's best estimate of fair allocations when "information about relevant differences (of the subjects) . . . are assumed away." Kahneman et al. (1986) made clear that these differences may have a strong impact: an unequal split can be viewed as fair and an equal split as unfair outcome.

This paper will show from existing as well as from new experiments that, *firstly*, subjects have distributional and reciprocal concerns being different from equity preferences, and that, *secondly*, models on inequity aversion are not able to capture all impacts of the strategic setting of a game. This is why it is not possible to describe behavior of the participants in the variety of experiments by a *single* additional variable, irrespective which of the above mentioned variables – altruistic, reciprocal or inequity averse preferences – is introduced.

Focusing on the principles behind the existing approaches, it shows that in the models of FS and BO participants are supposed to be interested in their own fair share; their behavior can be described as "self-regarding." Approaches based on models of altruism and reciprocity analyze motives why participants may be willing to sacrifice part of their endowment for the well-being of others or for mean actions against others. Other-regarding preferences are the main principle. Accordingly as a *third* point, the present paper shows that there is

no consistent relation between the fairness of an action and the fairness of an income distribution.

Finally, we compare the distributional preferences of these persons with their preferences when choices between the same payoff distributions are embedded in games with a strategic context. To explain these observations, as a *fourth* point the present paper argues to combine altruistic with reciprocal preferences. We assert that a large fraction of participants have an altruistic good-will capital guiding them when they make distributional choices in one-stage games. In games with more than one stage the previously friendly (unfriendly) action of the other person may cause reciprocal feelings and may increase (decrease) their good-will.

In this context, it should be emphasized that parallel to this paper Charness and Rabin (CR) (2002) suggested to insert more than one variable in addition to egoism into the utility function. In their approach, they combine so-called social welfare preferences with reciprocal concerns. Their model is able to explain most but not all of the results presented in this paper. It differs to our suggestions insofar, as they assume that reciprocal motives dominate behavior only when participants in experiments are confronted with mean actions by their fellow players.

The paper is organized as follows. Section 2 sketches the major features of models on inequity aversion. Section 3 explains the variables of altruism and reciprocity which we regard as the main variables to describe deviations from egoistic choices. Section 4 tests the impact of all variables in simple Distribution games and in games with a strategic setting. Section 5 concludes and suggests a formal approach combining altruistic with reciprocal preferences which also could take care for other distributional choices.

2. THE WORLD ACCORDING TO SCIA

Assume that there are n agents $i = 1, \ldots, n$. A general utility function of agent i with self-centered inequity aversion (SCIA) depends on i's income x_i and the differential incomes $\Delta_j = x_j - x_i$ with respect to all other agents.

$$U_i(x) = V_i(x_i, \Delta_1, \ldots, \Delta_{i-1}, \Delta_{i+1}, \ldots, \Delta_n) \tag{1}$$

with[7]

$$V_i(x_i + \varepsilon, \Delta_1, \ldots, \Delta_n) > V_i(x_i, \Delta_1, \ldots, \Delta_n) \quad \text{for } \varepsilon > 0, \tag{2}$$

and the attribute that for given x_i and given $x_j V_i$ is maximized by $\Delta_k = 0$, i.e.

$$x_k = x_i \quad \text{for all } j \neq i, k, \tag{3}$$

In particular, such a utility function implies that i would not like to leave a state of equal incomes to enter a state where he gets less and the others get more, i.e.

$$U_1(y, \ldots, y) > U_1(z, x, \ldots, x) \quad \text{if } z < y, x > y. \tag{i}$$

More general, we can conclude that a person will never give a present to a richer person, no matter how efficient[8] the present is.[9] The model of FS uses a special version of (1),[10]

$$U_i(x) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq 1} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq 1} \max\{x_i - x_j, 0\}, \tag{4}$$

where it is assumed that $\beta_i \leq \alpha_i$ and $0 \leq \beta_i < 1$. In the two-player case $i \in \{1, 2\}$ the simplified utility function is given by

$$U_i(x) = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\}, \quad i \neq j. \tag{4'}$$

The second terms measure the utility loss if i is worse off than others (or than j in the two-player case), while the third terms measure the loss if i is better off. $\beta_i = \alpha_i$ means that the utility loss from a disadvantageous inequity is at least as large as from an advantageous inequity. If we neglect the case $\beta_i = \alpha_i$ then (4) implies that i does not like the others to have larger payoffs than he has, in particular

$$U_1(0, y, \ldots, y) \geq U_1(0, x, \ldots, x) \quad \text{if } y < x. \tag{ii}$$

FS suggest the following distribution of α and β: α can be either 0 or 0.5 or 1 with probability 0.3 each, and α is equal to 4 with probability 0.1. The values for β are 0 or 0.25 each with a probability of 30% and 0.6 with a probability of 40%. $\alpha = 0$ and $\beta = 0$ resemble to egoistic players so that FS expect 30% of the participants to make egoistic and 70% of the participants to make inequity averse choices. FS suggest no distribution for the combination of α and β for the same person. It is not possible to calculate the utility of a person who is confronted with two payoffs, one of them to his advantage and one to his disadvantage.[11]

How are the basic experiments explained by FS? (a) In the Ultimatum Game the utility function (4) says that 70% of the responders reject offers providing them with less than 25% of the pie because they are not satisfied with more unequal splits. (b) At the final stage of the Gift Exchange Game[12] inequity-averse workers raise their efforts after having received higher than competitive wages because, as FS put it, these "workers can move in the direction of more equitable outcomes" in relation to their employer. (c) In the Centipede Game no inequity averse person is willing to give up a higher advantageous payoff in exchange for a lower disadvantageous payoff. This is contradicted by the results of McKelvey and Palfrey (1992). (d) For the dictator game the utility function of FS is not

a good predictor. FS suggest that if the utility function of (4′) is assumed to be concave (instead of piecewise linear), their model captures the behavior of "dictators" who donate on average one fourth of their pie.[13]

In the next two sections, we discuss from a theoretic and from an experimental point of view to what extent these kind of distributional preferences and their generalization, presented in the Eqs (1), (2), and (3) are consistent with behavior not yet explored. We contrast this discussion by suggesting an alternative approach combining altruism with reciprocal behavior and we compare our suggestions to those of CR.

3. AN ALTERNATIVE CONCEPTION

In this section we will introduce the two variables, altruism and reciprocal behavior, which may explain in a different way the behavior of non-selfish human beings. We will discuss the different notions of the two variables and how they may be connected with each other.

3.1. Altruism

It is almost common sense that participants have distributional preferences in experiments as much as outside of the lab. It is not common sense whether these preferences refer to oneself or to another person. We consider altruism (and greed as its negative pendant) as an important human motive to describe distributional preferences. We define altruism as the intrinsic motivation to give material payoffs to an anonymous other if the increase in income of the other person who profits from the altruistic decision satisfies the donor more than the outcome under the non-altruistic move. Thus, we view altruism as an external effect where

$$\begin{aligned} U_i &= V_i(x_1, \ldots, x_i, \ldots, x_n), i = 1, \ldots, n, \quad \text{with} \quad x_j \\ &= \text{income (or consumption) of person } j. \end{aligned} \tag{5}$$

Since (5) is a generalisation of (1) and therefore also of (4) and (4′) there is no direct history related to the person who will profit from the altruistic move. We suggest that intentions do not play a role.[14] This kind of decision can be experimentally tested by distributional choices in one-stage games without strategic context. The motive of an altruistic choice is to *increase or decrease* the income of *another* person. Transfers may be extended beyond the level proposed by inequity aversion where these choices *decrease* the *donor's* relative payoff compared to the payoff of

persons with lower payoffs. The approach of CR differ from ours insofar, as social welfare preferences put emphasis only on the positive side of altruism where the income of another person is increased.

We are in particular interested in making transparent the gradual differences of altruistic choices among which some are compatible with SCIA or with selfish preferences. Let us start with the first degree of altruism which may be called *payoff-increasing altruism*. A decision resembling to this kind of altruism is one which increases the payoff of the decision maker (to be called "oneself") and of the other person which is affected by the decision (to be called "the other"). A person having the choice between the following two payoffs (4, 0) and (5, 5) (the first amount indicating the payoff of oneself, the second indicating the payoff of the other) will choose (5, 5) irrespective whether he is an inequity averse, altruistic or egoistic persons. This choice increases both payoffs and leads to equal payoffs, as well. Matters change if there is a choice between (0, 0) and (1, 4): persons with a sufficient degree of inequity aversion, in the FS approach all persons with inequality averse preferences, will decide for the payoff (0, 0). Egoistic altruistic persons similar to social welfare maximizers will choose (1, 4).[15]

The next degree of altruism will be described as *inexpensive altruism* which increases the payoffs of the other while the payoff of oneself remains unchanged. Consider the following choices: (50, 50) and (50, 60). Inequity averse persons will prefer the equal payoff, altruistic persons will choose to increase the payoff of the other person, and egoistic persons will be indifferent. Thus, the choice of (50, 60) is also compatible with egoistic preferences.

The third degree is called *expensive altruism* lowering one's own and increasing the other's payoff, a setting which is mostly covered by the Dictator game: Altruistic and inequity averse persons will e.g. prefer a payoff of (4, 1) to a payoff of (5, 0). Different to this are efficient donations, for example a payoff of (4, 10) in comparison to (5, 5). Sufficiently altruistic persons would choose (4, 10), inequity averse persons and egoists will prefer the equal split.

For all kinds of choices, we found that SCIA and "sufficiently altruistic preferences" predict contradictory choices. Typically, this is the case if a person has to leave a situation with equal incomes in order to make efficient or social welfare maximizing donations.

3.2. Reciprocity

The exchange of goods is central for economies with highly specialized production. Any two parties exchanging products without an enforceable contract may carry out the exchange for reasons of reciprocity if the welfare of both trading partners is

increased. When one party defects in an exchange process, the desire for revenge seems to be ubiquitous, and revenge – as the other side of reciprocity – appears to be "usual" behavior. Revenge, and the positive side of reciprocity, friendly behavior, are strong intentions.[16] Gouldner (1960) and most other sociologists and economists consider reciprocity as a cornerstone of social behavior.

Reciprocity is the intrinsic motivation to respond to the previous behavior of a related person. *Purely* reciprocal decisions can be found at the final stage of games with more than one stage.[17] People who behave reciprocally, will reward the cooperator from the previous stage of a game and punish a person who behaved uncooperatively in the previous stage even if reward and punishment would not be suggested by normative game theory. Both decisions, reward and punishment, may reduce a person's payoff, but will raise his utility compared to the non-reciprocal choice, while the payoff of the rewarded (punished) person will increase (decrease). Reciprocal choices are, thus, based on the history of the game.

A second crucial point of a reciprocal move is its consequence for the outcome of the two parties. Their main reason for positive reciprocity in an exchange process is that they aim to realize an outcome where *both* players' payoffs are increased when *both* cooperated, compared to the outcomes of both players when they *both* do not cooperate. The same holds vice versa for negative reciprocity. Choices where a first player's outcome remains constant or is reduced when both players cooperated are degenerated reciprocity games. Some tests of CR and of Bolton et al. (1998a) with respect to positive reciprocity refer to this kind of games.

Distributional concerns cannot be the driving force of an exchange process because the relatively poorer person is not willing to reward the trust he is given by the relatively richer one. Therefore, an aversion against an *increase of inequity* might cause the systematic non-compliance to an unenforceable exchange contract. Inequity averse persons would not be true reciprocators. Players who behaved cooperatively in previous stages may get exploited.

While the utility of the rewarded (punished) person will increase (decrease), one has again to distinguish – similar to the grades of altruism – three different grades of reciprocity: payoff increasing, inexpensive and expensive reciprocity. Payoff-increasing positive reciprocity is a move which increases one's own as well as the other outcome. The games to test this kind of reciprocity can be found among the class of coordination games. Payoff increasing negative reciprocity raises the monetary payoff of oneself while the other player's payoff is reduced after his non-cooperative move in the previous stage. The most prominent example for such a decision is the sequential Prisoners' Dilemma. Egoistic and reciprocal players will do a non-cooperative move if the first mover defected, while inequity averse players will make a cooperative move if inequity is sufficiently reduced by that choice.

The second degree – inexpensive reciprocity – is the ability to reward (punish) the other player after a (non)cooperative move while the rewarding (punishing) player's payoff is not reduced by this choice in the last stage. One way of testing this kind of reciprocity is the Indenture Game (described in the next section). Reciprocal players will choose to reciprocate while inequity averse players will make their choice dependent on the payoffs of each move. Egoists are indifferent in their choice. They will either decide upon their taste for fairness (since it is "free of charge") or they will decide upon other variables, as the equilibrium path of the game.

The third degree is expensive reciprocity. Players at the final stage of a game are willing to invest material payoff in order to reward (punish) the (non)cooperative choice of their fellow players in the previous stage. There is an asymmetry between reward and punishment. Reward is an efficient re-action: If a move of a person 1 increased the payoff of a person 2, person 2 is willing to spend material payoff in order to increase the payoff of the person 1 in return. In contrast to this revenge is inefficient: A person 2 needs to sacrifice a certain amount of money in order to decrease the utility (or the payoff) of a person 1 who's non-cooperative move has already lowered the payoff of person 2 below the expected level. Nevertheless, there is even more experimental evidence on revenge behavior than on positive reciprocity (see also CR). Egoists will never choose such reciprocal move and inequity averse players only if it reduces inequity. Reciprocators will make a such a choice if the utility loss of their lower payoff is overcompensated by the utility gain from having reduced their fellow player's payoff. Therefore, from a theoretical point of view, there is no consistent relationship between the fairness of an action and the fairness of the income distribution after the action. Reciprocal choices are only sometimes compatible with distributional concerns.

3.3. Approaching Fair Behavior

We aim to combine the two motives of reciprocity and altruism in order to approach fair behavior. We consider the degree of altruism connected with each player as a kind of each player's individual benchmark for his behavior in strategic games. Any reciprocal move, either friendly in return to a cooperative choice or retaliatory in return to a defective choice of the other player, is related to this benchmark.[18] Such a combination takes care of the fact that the amount of money a final stage player is willing to spend in order to increase (or decrease) the other person's utility may depend on the strategic context.

We also need to touch the question of how to model reciprocal motives. The main problem is how to interpret a friendly and a hostile action. When players

can choose between a small number of discrete strategies, as in the Prisoner's Dilemma Game or the Centipede Game, it is easy to indicate reciprocal choices. It is common knowledge what cooperative and defective moves are because – as we indicated in the subsection on reciprocity – the payoffs of both players are increased if and only if both players choose cooperative moves which resembles to positive reciprocity (and vice versa). Therefore, in this paper we will restrict to games where the choice of each action has a unique 'label' and where it is common knowledge that a friendly (mean) action is recognized by both players as friendly (mean) action.[19]

4. TESTING ALTRUISTIC AND RECIPROCAL PREFERENCES

In this section we make an initial qualitative test of the combination of altruism and reciprocity by comparing different kind of Dictator Games with Ultimatum, Indenture, Sequential Prisoners' Dilemma and Trust Games. This test will be rather basic because we aim to find out the following things: Having conducted several one-stage games without any strategic context our first aim is to discriminate between altruistic and other distributional choices and to find out what kind of distributional preferences appear in simple one stage games. Our second aim is to find out whether choices in games with a strategic setting are different from these distributional preferences observed in one-stage games.

Combining altruism with reciprocity we hypothesize that depending on his degree of altruism a person 2 is willing to behave altruistically to person 1 by giving a certain amount x_{21} to person 1. This altruistic choice in a one-stage-setting can be seen as a baseline because it shows to what extent person 2 is willing to increase the utility of 1. It can be interpreted as person 2's good-will capital in relation to person 1. This good-will capital can be influenced by person 1. x_{21} is reciprocally increased (decreased) when person 1 chose a friendly (hostile) act towards person 2 in the previous stage. The counter-hypothesis of models focusing on distributional concerns, is that the action chosen by player 2 is constant, irrespective of the stages of the game, game structure or intention of player 1 in the previous stage.

We will make a piecewise test.[20] We will compare the behavior of participants in experiments with one stage where we can observe their distributional concerns with the behavior of participants in games with more than one stage where we can observe their behavior after a friendly or hostile move of their fellow player.

4.1. The Ultimatum Game

The first test compares distributional choices in a non-strategic context (the Distribution game) with the same distributional choices in the strategic context of the Ultimatum Game.

In the Ultimatum game, two persons bargain about the distribution of a pie of size x. Person 1, the proposer, offers x_{12} to person 2, the responder, and $x–x_{12}$ to himself. If Person 2 accepts the offer, both persons receive payoffs according to the offer of person 1. If person 2 rejects the offer, both persons end up with a zero payoff. There exists strong evidence about the behavior of the participants.[21] In this paper we will not repeat the results of the experiments, but present an overview in Fig. 1 over the distribution of the acceptance thresholds of responders in the Ultimatum Game (according to the parameter values of FS) with a total pie of 5 units.[22] This game is central for one reason: FS criticize approaches based on altruistic choices as being "inconsistent with the rejection of offers in the Ultimatum Game." From our perspective, the Ultimatum Game, as a two-stage game, is not apt to test for the existence of altruism. We assert that altruistic choices can be found in one-stage-games and that these are trumped by reciprocal motives depending on the behavior at the previous stage.

The Ultimatum Game allows to test for the existence of expensive negative reciprocity when its results are compared with one-stage games having the same

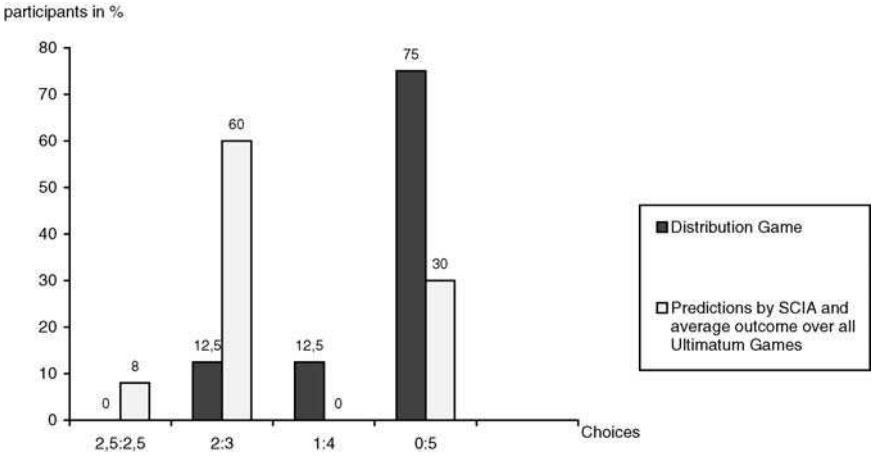


Fig. 1. Acceptance Thresholds of Participants in the Distribution Game and in the Ultimatum Game (Which is Equal to the Predictions of Fehr & Schmidt (FS), 1999).

payoff structure. In order to test what kind of distributional concerns guide the participants, we conducted the following Distribution Experiment (see Kritikos & Bolle, 2001): Persons had to choose between different income distributions which were the same as in the second stage of the Ultimatum Game. Neglecting history, the participants were in the same situation as responders in the Ultimatum Game. Instructions differed only with respect to the fact that there was no previous stage.

Participants being motivated by altruism or egoism will accept any division in the Distribution game. Compatibility with inequity aversion requires that they will reject those income distributions which are rejected in the Ultimatum Game. Similar to CR we predict higher acceptance rates in the Distribution than in the Ultimatum Game.

Result 1. Figure 2 shows that 70 out of 80 participants (87.5%) preferred a payoff of DM 1 for themselves and a payoff of DM 4 for an anonymous other person to a zero payoff to both persons. 10 persons (12.5%) preferred a zero payoff for both players.

Results of the Ultimatum Game would allow for the prediction that 24 persons would choose according to the model of FS (1, 4) and 56 persons would choose (0, 0). A χ^2-test shows that the rejection rate is in the single-stage game (without history) significantly lower than in the Ultimatum Game.

The further results of the experiment – shown in Fig. 1 – support this view. A majority of 60 participants (75%) chose (0, 5), although their payoff remained zero while somebody else's payoff was increased to DM 5.[23] No participant had a preference for equal payoffs only, while FS predicted that there should be 8 players. And instead of its prediction of 48 there were only 10 players who accepted an income split of (2, 3) but none with a lower payoff for themselves. A χ^2-test reveals that the acceptance threshold was significantly lower in the distribution game over one stage compared to the Ultimatum Game.[24]

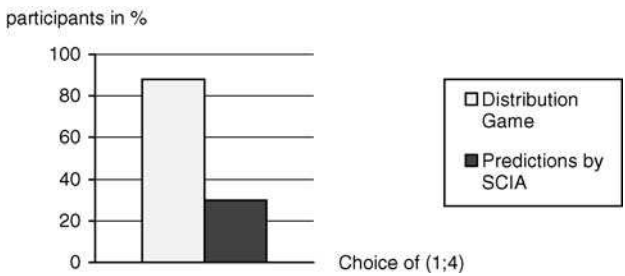


Fig. 2. Share of Participants Who Prefer a Choice of DM 1 for Themselves and DM 4 for an Anonymous Other to a Choice of a Zero Payoff for Both.

Other studies found the same diverging behavior. Similar to our results, in Offerman (1999) 17% of final stage players are willing to sacrifice 1 unit to reduce the other player's payoff by 4 units when the unequal choice of (6, 11) in favour of the other player appeared randomly.[25] If this distribution of (6, 11) was intentionally chosen by the other player, 83% of the participants punished the other player so that both players payoffs were reduced to (5, 7).[26] And in CR all 36 participants of an experiment preferred a payoff of (200, 800) to (0, 0).[27]

This outcome supports our view that participants have a basic good-will capital (or as CR put it that participants have distributional concerns for the social welfare) and that rejections in the Ultimatum Game are not a result of distributional concerns but that the good-will capital of the participants was over-compensated by retaliatory preferences.

4.2. Inexpensive Choices in One-Stage and Two-Stage Games

In the second part of the piecewise test we confront participants with an inexpensive choice in a one-stage game and in a strategic setting after an uncooperative move of the other player.

In the Indenture Game (for description see Fig. 3), the willingness for inexpensive negative (as well as positive)[28] reciprocity can be tested. Two persons sequentially exchange a commodity against a payment. Under the transaction structure used in this game, the seller of the good (player 1) may deliver the item to the buyer (player 2) after player 2 has offered a hostage to him – the indenture. In the final node of the game player 2 is indifferent between transferring and keeping the second part of the indenture, irrespective whether player 1 was cooperative and delivered the item or not. Therefore, player 2 is able to react in the final stage of the game with a friendly (hostile) move after an (un)-cooperative move of player 1 in the previous stage without having to sacrifice any amount of money for his choice.

In Experiment 1 (see Table 1), we asked persons to choose between two income distributions (0, 15) and (0, 35). This is the same situation as for a buyer (person 2) in an Indenture Game after an uncooperative move of player 1. Altruistically motivated persons should prefer (0, 35) in the Distribution Game, Egoists should be indifferent and inequality averse should prefer (0, 15). In the Indenture Game, reciprocators and inequity averse participants will choose (0, 15) and egoists are expected to follow the equilibrium choice which is also (0, 15).

Result 2. As Fig. 4 shows, 37 persons (74%) selected (0, 35) as their favorite move, 13 persons (26%) preferred (0, 15). In the next-to-last stage of the

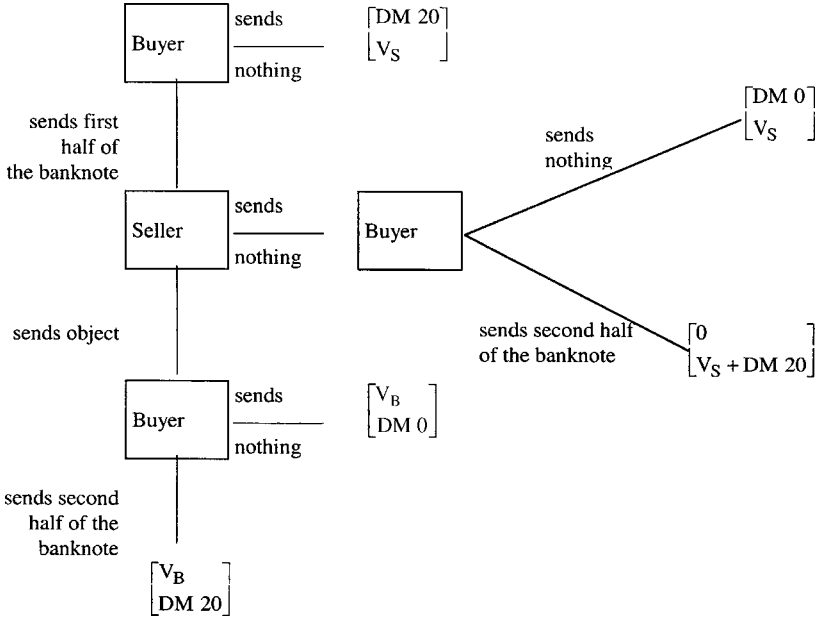


Fig. 3. The Indenture Game. The Structure of Values is V_S = DM 15 and V_B = DM 25.

Table 1. Experiment 1: Experimental Procedure (for Instructions see Appendix B.

Game	Choice 1	Choice 2	Number of Participants
Income distributions			
(1)	(0, 35)	(0, 15)	50
(2)	(10, 40)	(20, 10)	50
(3)	(0, 75)	(10, 10)	50

Note: In a classroom experiment, participants were confronted with an income distribution described by ($y_{\text{for self}}$; $y_{\text{for another person}}$), with the amount of money in German Marks. They were asked: Which distribution do you prefer? Choices were sequentially presented to the participants. They wrote down their decisions together with a pseudonym and password. The decision forms were then collected and new ones distributed. All Games (1) to (3) were conducted with the same participants. After the decisions were done 8 decision forms were selected for payoffs. The amount $y_{\text{for self}}$ was paid to the person who made the decision; the amount $y_{\text{for another person}}$ was paid to another randomly selected anonymous person, from a parallel classroom. Participants received the amounts from a third party not involved in the experiment. The pseudonyms of the winners were named openly, and the winners were required to reveal their password to the third party in order to get paid.

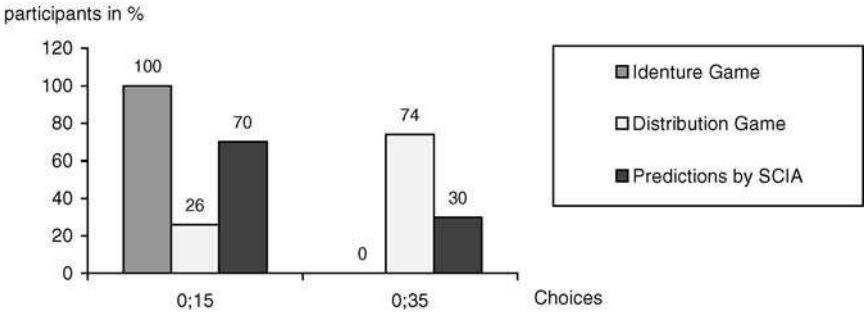


Fig. 4. Predictions of FS and Share of Participants Who had the Choice Between the Payoffs of (0, 15) and (0, 35) in the Distribution Game and in the Indenture Game.

Indenture Game, there had been 35 observations of uncooperative behavior. As Fig. 4 also shows, all 35 participants confronted with non-cooperation, reacted with negative reciprocity by selecting (0, 15).[29]

Behavior in the Distribution Game is significantly different from the predictions of FS ($\chi^2 = 21.6; p < 0.0001$) that 70% of the participants will choose (0, 15) and 30% will be indifferent. Thus, the altruistic preferences in the Distribution Game are significantly reduced when players face a previously uncooperative move.

Moreover, CR find evidence for inexpensive choices of positive reciprocity. In a two stage game, player 2 was asked, after a cooperative move of player 1, to choose between (400, 400) and (400, 750). 94% of the participants cooperated although this choice increased inequality.

4.3. Expensive Choices in One-Stage and Two-Stage Games

In the third part of the piecewise test it is aimed to find out how participants react when they are confronted with an expensive choice in different kind of dictator games and strategic games after a cooperative move of the other player.

The usual Dictator Game to test distributional preferences is as follows: Player 2 is given an amount x and is asked whether he is willing to transfer any x_{21} to an anonymous player 1. Player 1 can do nothing but accept the share x_{21} while player 2 will keep the rest $x–x_{21}$. Since player 2 has a higher endowment than player 1, altruistic and inequity averse persons will share their endowment. There is evidence (e.g. Andreoni & Miller, 2002; Andreoni & Vesterlund, 2001; Bolton & Zwick, 1995; Camerer & Thaler, 1995; Eckel & Grossman, 1996; Forsythe et al., 1994; Hoffman et al., 1994; Kritikos & Bolle, 2002) that on average 70% of the

participants are willing to spend a substantial amount to their anonymous partners, while egoistic participants (again about 30%) pocket the whole amount of money.

This game is the baseline for two tests. In the first test we compare the Dictator Game with the Trust Game (Bolle, 1998): His experiment was designed as follows. The first mover was endowed with DM 80. He could keep his money which left him with DM 80 and nothing for the second mover. Alternatively, he could give the full amount to player 2. Player 2 was able to make a profitable investment and to double the amount *he received by player 1* to DM 160, before he decides upon the amount he returns to player 1. Player 2 could choose any amount x_{21} between DM 0 and DM 160 which he wanted to return to Player 1 and could keep DM 160 – x_{21} for himself. As in the Dictator Game he was not obliged to give anything. Player 2 plays a Dictator Game, after Player 1 has put him into that position.

In this experiment, there were 50 observations of second movers. Comparing the Dictator Game with the Second-mover of a Trust Game reveals the relevance of explicit modelling of positive reciprocity. Theories of inequity aversion predict that player 2 will transfer the same share to player 1 irrespective of the game structure. We predict that player 2 will transfer a significantly higher amount to player 1 in the Trust Game than in the Dictator Game.

Result 3. In Dictator Games, the average share the dictator passes on to the other player is 25%, with the following rough distribution (Fig. 5): 30% of the players transferred nothing, about 50% shared something between 20 and 50% of the pie, and about 20% made an equal split. In the Trust Game, on average 50% of the final stake were transferred with the following distribution (see also Fig. 5): 10% chose the zero share, and 5% shared less than half the pie. Another 25% of the players returned half of the pie (the initial amount) of DM 80 while 60% of the second movers transferred to the first mover more than 50% of

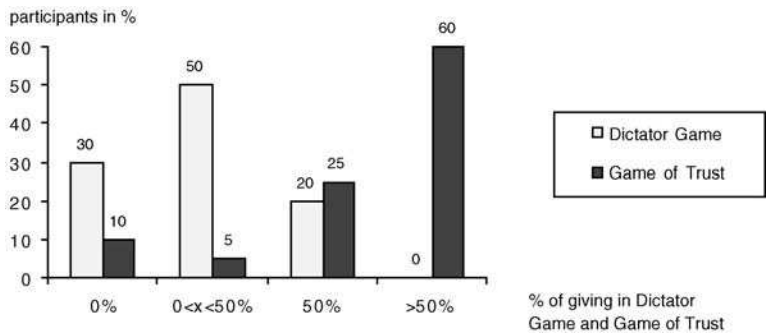


Fig. 5. Share of Participants in the Dictator Game and in the Game of Trust Who Gave (i) Nothing, (ii) Something Between 0 and 50%, (iii) Exactly 50%, (iv) More than 50%.

the pie.[30] Testing the two distributions in the Dictator Game and in the Game of Trust shows that the differences are highly significant ($\chi^2 = 54.6$, df $= 3$, $p < 0.0001$).[31]

The differences coincide with positive reciprocity beyond inequity aversion: 60% of the participants returned more than the initial endowment to player 1. Four persons (about 10%) decided upon equity considerations and stated in the post-experimental questionnaire that an equal share of the joint profit of DM 160–80 was just. The remaining 30% of observations can be explained by egoistic motives, even if some of them returned half of the pie. These stated in the questionnaire that they were guided by the norm that the first player should not face a loss and should be returned the initial endowment.[32]

The second test is provided within our experiment 1 (the one-stage game): The participants were asked to choose between the two income distributions (20, 10) and (10, 40) (cf. Experiment 1 in Table 1). The choice allows for testing distributional preferences beyond inequity reducing choices. Person 2 has to reduce his relative and absolute payoff in order to increase person 1's payoff at the exchange rate 3 to 1.[33] Altruistic persons and social welfare maximizers are expected to prefer (10, 40), inequity averse persons and egoistic (20, 10).

Result 4. Figure 6 shows, that 29 persons (58%) preferred the choice which paid DM 10 to themselves and DM 40 to an anonymous other person. 21 participants preferred to allocate DM 20 to themselves and DM 10 to the other player.[34]

The acceptance rate of the altruistic choice is significantly higher than the prediction of FS.

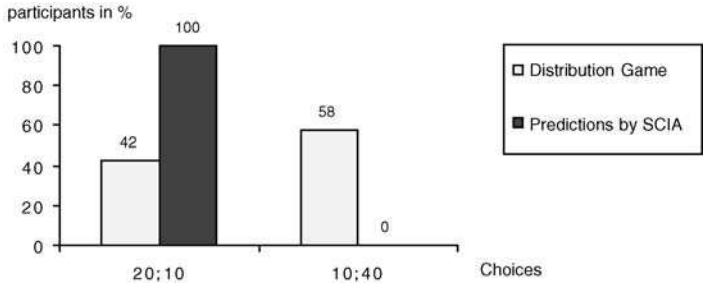


Fig. 6. Predictions of FS and Share of Participants Who had the Choice Between the Two Payoffs of (20, 10) and (10, 40).

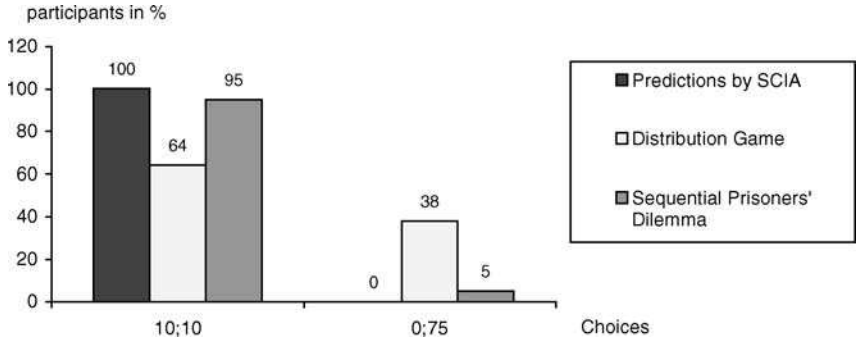


Fig. 7. Predictions of FS and Share of Participants Who had the Choice Between the Payoff of (10, 10) and (0, 75) in Distribution Game and in the Sequential Prisoner's Dilemma Game.

4.4. The Sequential Prisoners' Dilemma

The aim of the final part of the piecewise test is to find out how players react when they have to give up their complete endowment for an altruistic choice – again in a strategic and a non-strategic setting. In the sequential Prisoners' Dilemma, the second mover can decide how to react to a decision of the first mover. Here it is focused on second movers being confronted with a *non-cooperative* choice of first movers. In Bolle and Ockenfels (1990), these second movers had the choice between a defective move inducing a payoff of (10, 10) and a cooperative move leading to (0, 75). It is shown in Fig. 7 that 95% of the second movers (58 out of 61) preferred the

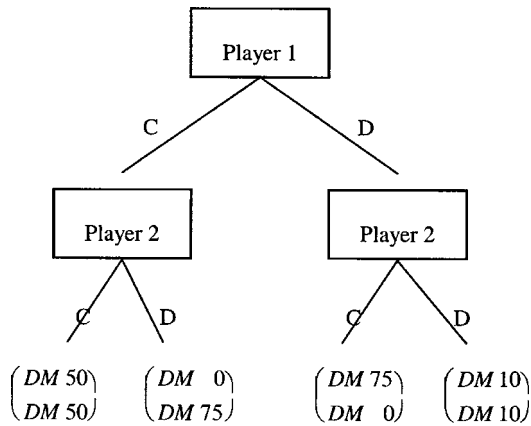


Fig. 8. The Sequential Prisoners' Dilemma in the Experiment of Bolle and Ockenfels (1990).

Table 2. Choices Resembling to Altruism Beyond Inequality Aversion.

	Choices Resembling To		
	Payoff – Increasing Altruism	Inexpensive Altruism	Expensive Altruism
Kritikos and Bolle (2001)	(1, 4) vs. (0, 0) 87.5%	(0.5) vs. (0.0) 75%	(10, 40) vs. (20, 10) 58%
Charness and Grosskopf (2001)	(625, 1200) vs. (600, 600) 88%	(600, 1200) vs. (600, 600) 74%	(600, 1200) vs. (625, 625) 67%
Charness and Rabin (2001)	(200, 800) vs. (0, 0) 100%	(400, 750) vs. (400, 400) 69%	(375, 750) vs. (400, 400) 50%
Offerman (1999)	(6, 11) vs. (5, 7) 83%		
Present paper		(0, 35) vs. (0, 15) 74%	(0, 75) vs. (10, 10) 38%

defective choice (10, 10). This choice is compatible with egoism, egoistic negative reciprocity and inequity aversion (The game is also described in Fig. 8).

In order to compare this outcome with behavior in a non-strategic setting, the Distribution Game was continued in Experiment 1 (Table 1), giving to 50 subjects the choice between two income distributions (10, 10) or (0, 75).

Result 5. As Fig. 7 shows, 19 persons (38%) chose a distribution which paid nothing to themselves and DM 75 to an anonymous other person. 31 participants preferred to allocate DM 10 to themselves and DM 10 to the other player.

A Fisher exact probability test shows that the rate of the "altruistic choice" is significantly higher ($p = 0.000001$) than predicted by FS. This outcome indicates that for the majority of the participants it is either egoism and/or negative reciprocity, driving their non-cooperative behavior in the Prisoners' Dilemma. Nevertheless, it has to be emphasized that the altruistic choices in this part of the distribution game were less than in any other game (see Table 2). It remains open whether the lower willingness for altruistic choices is owed to the high exchange rate or to the fact that the proposer had to give up his complete endowment.[35]

4.5. Discussion of Further Experiments

We gave evidence that in strategic games positive (negative) reciprocity increased (destroyed) the altruistic attitudes of a high share of participants after a cooperative (defective) move of their fellow players in the previous stage – even beyond inequity aversion. Yet, in this context it has to be emphasized that in all experiments (also in those mentioned in the introduction) the reciprocal attitudes

were indirectly activated by the experimental setting. The instructions indicated that in the first stage of those games the action was "named" – mostly a transfer of money (e.g. in the investment game or in the gift exchange game) or of an item (e.g. in the Indenture Game), a move which both players could understand as co-operative or defective.

More recent experiments – conducted by Bolton et al. (1998a, 2000) – indicated that in the domain of positive reciprocal behavior distributional concerns might be more important than intentional reciprocity. However, it is necessary to clarify that these experiments by Bolton et al. (1998a, 2000) were designed in a different way, since the participants received instructions where only the payoffs at the end of the game were revealed. Actions at each stage of the game (e.g. the transfer of money or of an item) were not labelled, at all. Since the activation of intentions is very sensitive to the level of information given to the participants through the instructions, we assert that the players who were making a certain choice did not recognize the possible intention which they transport with a certain action. Their choice of action might have become different. The same holds for the final stage players. Since they were not able to recognize any intention behind the choice at the previous stage of their fellow players, their choice might have changed, as well. Thus, we assert that intentional choices are done by players, if a certain intention is activated through a certain amount of information.[36]

Therefore, in Experiment 2 we aim to test whether intentional reciprocity indeed needs to be activated by offering a minimum amount of information for each action in a game. For this test, we use a Prisoner's Dilemma where we presented only the payoffs at the end of the game and where we did not provide any "background story" for each action.

4.5.1. Procedure of the Sequential Prisoners' Dilemma (for Instructions see Appendix B)

Participants were divided into two groups in two different classrooms and were accordingly assigned either to the role of person I or person II. Participants were anonymously matched and played one-shot games. Those participants being assigned to person 1 were asked to decide between A and B in Fig. 9 and between G and H in Fig. 10. Participants being assigned to person 2 were asked what choice they would do if person 1 had chosen A, then what choice they would do if Person 1 had chosen B, G or H. After a random matching of Persons 1 with Persons 2 by the experimenter, participants received their payments as indicated in the Figs 9 and 10. Participants received the amounts from a third party not involved in the experiment. The pseudonyms of the winners were named openly, and the winners to get paid were required to reveal their password to the third party.

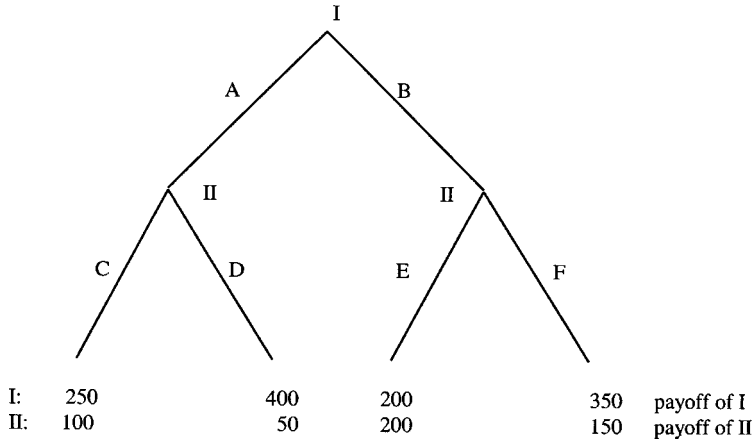


Fig. 9. A Sequential Prisoners' Dilemma Game/Part One of Experiment 2. *Note:* Which choice (between I and K) will you make if player I had chosen G. Which choice (between L and M) will you make if player I had chosen H.

Thus, it was not clear to the participants that in Game 1, player 1 would make a cooperative move by choosing B and a defective move by choosing A, that player 2 would behave positively reciprocal if he would choose F, etc. We further designed the outcomes of the two Prisoners Dilemma Games in a way (see Figs 9 and 10)

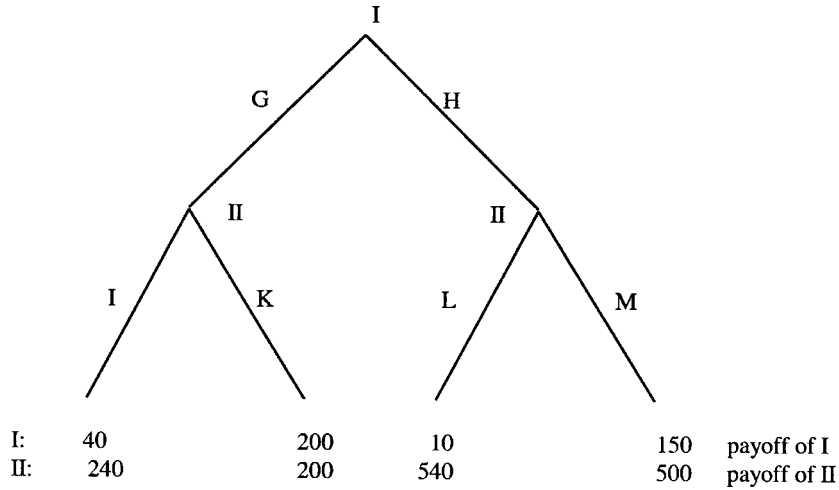


Fig. 10. A Sequential Prisoners' Dilemma Game/Part Two of Experiment 2.

that in the first game inequity averse persons will not cooperate although the first mover did so and that in the second game inequity averse persons will cooperate unconditionally even if the first mover defected.[37]

In this experiment 104 persons participated where 52 of the subjects were assigned to be Person I and 52 to be Person II. We classified five different types of players, the Egoist always playing defection (i.e. CE in Game 1 and IL in Game 2), the Utilitarian always playing cooperation (i.e. DF in Game 1 and KM in Game 2), the Reciprocator playing defection if player I defected and cooperation if player I cooperated, the Inequity Averse choosing always defection in Game 1 and always cooperation in Game 2, the Perverse Player choosing to defect after a cooperative move of player I and to cooperate after a defective move of player 1.

Result 6. As Table 3 indicates we found 19 Egoists, 12 Inequity Averse Persons, 8 Reciprocators, 2 Perverse; 1 Utilitarian Player and 10 person with different combinations.

The outcome of this experiment indicates that there are slightly more inequity averse players than reciprocators – seemingly in contrast to earlier findings. However, we note a clear distinction between the designs of the experiment. Similar to Bolton et al. (1998, 2000) we revealed in Experiment 2 only the consequences of the choices, i.e. the payoffs. Since we did not label any action in the game we hid the intentions of each action.

We argue that in order to distinguish between the consequential approaches (as those of BO and FS) and the intentional approaches the same level of information about each action and about the consequence of each action should be offered to the participants of experiments.[38] Experiment 2 made clear that the experimental results we are comparing depend to a certain extent on the level of information given to the participants through the instructions. Approaches towards reciprocal behavior do better fit with observed behavior when participants are also informed about the actions at each stage of the game.

Table 3. Results of the Sequential Prisoners' Dilemma Game of Experiment 2.

Choices			
Game 1	Game 2		
C E	I L	Egoistic	19
D E	K L	Perverse (partly)	2
C F	I M	Reciprocal (partly)	8
D F	K M	Utilitarian	1
C E	K M	Inequality Averse	12
		Other combinations	10

5. DISCUSSION AND CONCLUSION

Being confronted with the great variety of experimental outcomes economists search for appropriate approaches capturing the diverging results and allowing for correct predictions.

(i) Putting together the results of the present Experiment I and of similar studies there is evidence (see Table 2) that in Distribution Games up to 60% of the participants make choices of expensive altruism, more than 70% of the participants make choices of inexpensive altruistic choices, and up to 100% of the participants make altruistic choices increasing their own payoff. These altruistic choices increased inequity in favor of the other player,[39] supporting our view that participants have a good will capital irrespective of the final payoff distribution and it also supports the view of CR about social welfare preferences.

(ii) The above experiments – mentioned in Section 4 and in the introduction – showed that reciprocal choices exist, even if they increased inequity in favor of the other player. The share of reciprocally behaving persons was mostly above 40% and in some cases even around 60%. A comparison of the experimental results of one-stage with multiple-stage-games yields the following regularities: In multiple stage games, behavior at the final stage was significantly different from the behavior in the corresponding one-stage games with the same payoffs.

(iii) The consequence of (i) and (ii) is that it is not possible to explain the variety of observed behavior in experiments with a single second variable (besides egoism), irrespective whether the second variable is reciprocity, altruism, or inequity aversion.

After this test we suggest a rough hypothesis about the different kind of behavior of the participants (see also Table 2). We distinguish between three types: egoists, reciprocators with an altruistic good-will capital and inequity averse individuals. The data and parameter distributions support the assumption of FS that about 30% of the participants were pure egoists (see in Table 2 the differences between the expensive and payoff-increasing altruistic choices). We speculate that among the remaining 70% of participants, a majority had a good-will capital and will behave like reciprocators in strategic games and a minority will behave like inequity averse individuals. Similar to our findings, CR (2002, p. 834) conclude that "social-welfare preferences and even narrow self-interest outperform difference aversion."

Reciprocity in combination with the altruistic good-will capital is apt to explain these differences[40]: The good-will capital of more than half of the final stage

players was increased after a friendly act while retaliatory preferences overcompensated it after an unfriendly act of their fellow players. As a consequence these final stage players rewarded (punished) their fellow players for their friendly (unfriendly) move at the previous stage which increased (decreased) the payoffs of the fellow players.[41]

The specification of a model based on reciprocity remains to be difficult. For such a purpose, we need to return to the question of the correct interpretation of intentions. We assert that for every simple game different behavioral norms exist (e.g. the efficient, the subgame perfect or the just outcome). Each norm can be seen as a benchmark if it is generally accepted and common knowledge. The norm, and thus the expected "neutral action" may deviate from subgame perfect strategies. Choices above (beneath) this norm are perceived by the final stage player as a (non)cooperative action and answered by a reward (punishment). Likewise, in the Game of Trust (Bolle, 1998) or the Gift Exchange Game (Fehr et al., 1997) contributions above zero (the subgame perfect outcome) were seen as a cooperative move, and the players answered by reciprocal cooperation. In the public good game with punishment (cf. Fehr & Gächter, 2000), the norm was – at least to a certain extent – the private provision of the public good (the efficient result) and persons who did not duly contribute were punished.[42]

From our point of view, a first suggestion for the modelling of social man is as follows. We assume that there is a group of n persons. We *describe* and *denominate* an action by the transfer vector

$$T_i = (T_{i1}, \ldots, T_{in})$$

connected with this action. T_{ij} describes the changes of income of j caused by the action T_i of i. If T_i is a choice directly leading to the payoff vector, than it may be interpreted as the vector of evaluation minus the values of the "neutral action" described above. An action in period t is then selected on the basis of an interdependent utility function:

$$U_i^t = \sum_j a_{ij}^t T_{ij}^t \tag{6}$$

with $a_{ii}^t = 1$ and positive or negative coefficients a_{ij}^t.

All individuals i can observe T_i and, thus, conclude on T_{ij}. From period to period i's incentive a_{ij} to choose an action with transfers to j is discounted. Positive or negative transfers, in addition, contribute to the next period's incentives:

$$a_{ij}^{t+1} = \delta_i a_{ij}^t + \gamma_i T_{ji}^t, \quad 0 < \delta_i < 1, \quad \gamma_i > 0. \tag{7}$$

Some theoretical implications of this approach have been investigated by Bolle and Kritikos (2002). Further, it might be appropriate to supplement (6) by a quadratic altruistic term and/or by a term capturing inequity aversion.

Having shown that it is not possible to describe behavior of the participants in the variety of experiments by a single variable in addition to egoism and having shown that participants have altruistic preferences in simple one-stage games, this paper makes also clear that the question about the appropriate utility function is far from being answered. We still do not know how participants will decide in simple distribution games if they do not have to choose between two but between a continuum of payoffs. Accordingly, we do not know whether it is sufficient to model distributional preferences as concerns for social welfare (as CR suggest) or whether we have to introduce a spite component, the negative side of altruism (as we argued earlier in this paper). Leaving aside the missing answer of a clear definition of positive reciprocity (are games of degenerated reciprocity a useful test for the existence of reciprocity), we also have no clear idea to what extent we may observe choices in the domain of positive reciprocity. We need more experiments with the focus on these two points.

NOTES

1. See the experiments on the Game of Trust (cf. Berg et al., 1995; Bolle, 1998; Jacobsen & Sadrieh, 1996), the Prisoners' Dilemma (cf. Andreoni & Miller, 1993; Bohnet & Frey, 1995; Bolle & Ockenfels, 1990; Cooper et al., 1996), the Gift Exchange Game (cf. Fehr et al., 1997), the Centipede Game (cf. McKelvey & Palfrey, 1992), the Ultimatum Game (cf. Güth, 1995a), the Public Good Game with Punishment (cf. Fehr & Gächter, 2000), or on the Dictator Game (cf. Forsythe et al., 1994).
2. Since Simon's (1957) bounded rationality approach (see in particular also Selten, 1990), economists are looking for pieces of a "theory of real behavior."
3. Cf. inter alia Bolle (1991) and Andreoni and Miller (2002).
4. Both reciprocity and altruism proved to be evolutionary stable (cf. Bester & Güth, 1998; Güth, 1995b).
5. In the models of Levine (1998) and Kritikos and Bolle (2003), intention is substituted by inclination, i.e. by the question whether A has positive feelings towards B. If inclination is extracted from past behavior, we have again a model of indirect reciprocity. A different distinction between altruism and reciprocal altruism was already suggested by Andreoni and Miller (1993).
6. In this context, it should be made clear that BO and FS (similar to Bolton et al., 1998b) expect a close relationship between the perceived fairness of an action and the perceived fairness of the distribution of income generated by an action.
7. This means: Increasing i's income as well as all other agents' incomes by ε makes i better off.

8. A present is efficient if the loss in income of the donor is smaller than the increase in income of the donee.

9. As we will show in the next section, strong enough *altruism* would imply the opposite because the other's increased income might outweigh the losses of the poorer donor.

10. For reasons of comparison, the approach of Bolton and Ockenfels (2000) is shown in Appendix A.

11. Such payoffs appear in every choice of the centipede game which models multiple exchange processes.

12. Similar explanations hold for the last stage of the Sequential Prisoners' Dilemma and in the Game of Trust.

13. In this context it should be emphasized that the approach of BO has an explanation for the dictator game. The model suggests that the dictator will keep for himself something between the half and the full pie. See App. 1.

14. This is in contrast to Falk and Fischbacher (1999) who suggested that any altruistic choice contains also an implicit "fishing for reciprocity."

15. Thus, decisions which are compatible with purely selfish preferences and "simple" altruism are not necessarily compatible with inequity aversion.

16. See the survey by Roth (1995) for supporting experimental evidence.

17. Reciprocal decisions can also be found in earlier stages of games but might interact with other motives.

18. For a formal approach on the development of such a good-will capital, cf. Kritikos and Meran (1998).

19. See e.g. Rabin (1993) or Falk and Fischbacher (1993) who developed reciprocity models taking care for the problem how an intention behind the choice of an action is interpreted in a consistent way.

20. There exist two experiments, one by Offerman (1999), and one with a relatively small number of observations by Bolton et al. (1998a) where positive and negative reciprocity are tested in one and the same experiment. We will subsequently consider the results of both papers in this section.

21. Its empirical results guided FS in specifying the values of α and β in their model and which they used to predict the behavior of the participants across games.

22. For an overview about the facts of the Ultimatum Game, cf. Table 1 in Fehr and Schmidt (1999).

23. Similar results were found by Charness and Rabin (2002). In their experiments 69% of the participants preferred a choice of (400, 750) in favor of the other player to a choice of (400, 400); and by Charness and Grosskopf (2001) where 74% preferred a payoff of (600, 1200) to (600, 600).

24. Further evidence is given by Blount (1995) who asked for the minimum acceptable offer in Ultimatum Games.

25. This game has the same structure as the first part of our game. It differed only with respect to the endowment.

26. There are more experiments supporting this result. In an experiment on the Ultimatum Game by Kagel and Wolfe (2001) knowingly unequal proposals were rejected at substantially higher rates than unintentional unequal proposals. And in an experiment on the "equal punishment game" second movers were ready to punish non-cooperative proposers even if the punishment worsened the relative payoff of the responder (Ahlert et al., 1999).

27. For similar results see Charness and Grosskopf (2001): 88% preferred (625, 1200) to (600, 600).

28. In here we will restrict our analysis on the negative part of reciprocity because only this piece allows for a discrimination between the three variables we are discussing in this paper. For the experimental results on positive reciprocity, cf. Kritikos (2000). For details on the theoretical background, cf. Kritikos and Bolle (1998).

29. For more details on the experiment cf. Kritikos (2000).

30. Similar results were found by Berg et al. (1995). In their experiment player 2 was able to triple the amount received from player 1. Thus, a discrimination in these experiments between inequity aversion and reciprocity is not possible. However in the video-taped replication of these experiments, Jacobsen and Sadrieh (1996) show that the history of the games mattered and that reciprocity was the main motivation in the decision process.

31. For similar results see Charness (forthcoming) who applied Blount's (1995) framework to the gift exchange game.

32. cf. Bolle (1998, p. 91). It would be certainly interesting to further investigate to what extent norm-oriented and pure egoists exist. At least in experimental settings like the investment game their behavior was different.

33. Fehr and Schmidt's (1999) utility function (4) is inspired by an investigation of Loewenstein et al. (1989) where, however, only exchange rates of 1 to 1 were involved in the choices.

34. We found similar support in Charness and Rabin (2002) where 50% preferred (375, 750) to (400, 400) and in Charness and Grosskopf (2001) where 67% preferred (600, 1200) to (625, 625).

35. Andreoni and Vesterlund (2001) who tested different exchange rates offer a gender-specific explanation.

36. In this context, it is necessary to emphasize that there is no consensus among experimental economists what the optimal amount of information is which should be given to the participants and what kind of information does already insinuate a certain kind of behavior. There is consensus that direct explanations like a certain choice would be a "reciprocal choice" should be avoided in the instructions. However, there are also experimentalists who suggest that already by telling a certain story along with an experiment (like in the investment game) participants are induced to behave e.g. like reciprocators because the story reminds them of similar incidences in reality. Participants, then, do not decide anymore outside of their daily norm where one-shot decisions between anonymous persons might not exist. There are other experimentalists who rather think that only by giving a neutral explanation for each action, e.g. the transfer of money in the investment game, it is possible to transfer the necessary amount of information to the participant. Otherwise the participant is simply not able to recognize the structure of an experiment where a certain action may be suited to activate reciprocal feelings at the fellow player.

37. There is ample evidence that reciprocal choice usually prevail in experimental studies on the Prisoners' Dilemma, i.e. that the second mover chooses cooperation only if the first mover cooperated and if the second mover could recognize the first mover's choice as a cooperative one. See inter alia Rapoport and Chammah (1965), Bolle and Ockenfels (1990), Cooper et al. (1996), Watabe et al. (1996), and Clark and Sefton (2001).

38. Of course it is easier to provide information about the consequence of an action (only the payoffs need to be revealed) than about the intention of an action. Since it cannot be excluded that participants are driven by norms when they receive direct information

about the intention of an action (a certain move if labeled cooperative or defective may induce more participants to choose e.g. cooperation), the label of an action should be kept neutral. We think that the descriptions of actions are kept neutral if each action receives its own payoff as it was done in the trust games of Berg et al. (1995) and Bolle (1998) where the choice to invest some money at the first stage of the game was labeled as transfer of money from person 1 to person 2. For similar reasoning, cf Antonides (1994).

39. In altruistic choices reducing inequity we should accordingly observe for each of the three cases a higher share of moves which then qualify for altruism as well as for inequity aversion.

40. As this paper aims to show that we have to explicitly model a combination of altruism and reciprocity beyond inequity aversion, we would like to emphasize that we do not aim to test for the existence of reciprocity and altruism per se. There is much evidence suggesting that these two variables matter – some of this evidence was mentioned in the introduction.

41. CR who also suggest a combination of distributional with reciprocal preferences, conclude in contrast to us that reciprocal feelings prevail on its negative side. However, the experiments on (non-existent) positive reciprocity, they base their findings on, are mostly games which we would describe as degenerated reciprocity games where the player of the previous stage cannot improve his payoff by a cooperative move.

42. It should be emphasized that it is often the experimental setting which induces the norm. Allowing e.g. a punishment of defectors in a Public Goods Game induces the norm of a private provision of this public good.

ACKNOWLEDGMENTS

We would like to thank Matthew Braham, Gary Charness, Alexander Dyck, Ernst Fehr, Werner Güth and Axel Ockenfels and two anonymous referees for helpful comments and discussions on previous versions of the paper. We further thank the German Science Foundation (Grant Bo 747/5-1) for financial support.

REFERENCES

Ahlert, M., Crueger, A., & Güth, W. (1999). An experimental study of the equal punishment game. Discussion Paper, University of Halle.

Andreoni, J., & Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *Economic Journal*, *103*, 570–585.

Andreoni, J., & Miller, J. H. (2002). Giving according to GARP: An experimental study of rationality and altruism. *Econometrica*, *70*, 737–753.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *Quarterly Journal of Economics*, *116*, 293–312.

Antonides, G. (1994). Mental accounting in sequential prisoners' dilemma game. *Journal of Economic Psychology*, *15*, 351–374.

Arrow, K. J. (1975). Gifts and exchanges. In: E. S. Phelps (Ed.), *Altruism, Morality, and Economic Theory*. New York.

Becker, G. S. (1976). *The economic approach to human behavior*. University of Chicago Press.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity and social histoy. *Games and Economic Behavior, 10*, 122–142.

Bester, H., & Güth, W. (1998). Is altruism evolutionary stable? *Journal of Economic Behavior and Organization, 34*, 193–209.

Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes, 63*, 131–144.

Bohnet, I., & Frey, B. S. (1995). Ist reden silber und schweigen gold? Eine ökonomische analyse. *Zeitschrift für Wirtschafts- und Sozialwissenschaften, 115*, 169–209.

Bolle, F. (1991). On love and altruism. *Rationality and Society, 3*, 197–214.

Bolle, F. (1998). Trust – An experimental study. *Theory and Decision, 45*, 83–98.

Bolle, F., & Kritikos, A. S. (2002). Reciprocity, altruism, solidarity: A dynamic model. Discussion Paper, Frankfurt/Oder.

Bolle, F., & Ockenfels, P. (1990). Prisoners' dilemma as a game with incomplete information. *Journal of Economic Psychology, 11*, 69–84.

Bolton, G. E., Brandts, J., & Katok, H. (2000). How strategy sensitive are contributions? A test of six hypotheses in a two-person dilemma game. *Economic Theory*.

Bolton, G. E., Brandts, J., & Ockenfels, A. (1998a). Measuring motivations for reciprocal responses observed in a simple dilemma game. *Experimental Economics, 1*, 207–219.

Bolton, G. E., Katok, E., & Zwick, R. (1998b). Dictator game giving: Rules of fairness vs. acts of kindness. *International Journal of Game Theory, 27*, 269–299.

Bolton, G., & Ockenfels, A. A. (2000). Theory of equity, reciprocity and competition. *American Economic Review, 90*, 166–1993.

Bolton, G. E., & Zwick, R. (1995). Anonymity vs. punishment in ultimatum bargaining. *Games and Economic Behavior, 10*, 95–121.

Camerer, C. F., & Thaler, R. H. (1995). Ultimatums, dictators and manners. *Journal of Economic Perspectives, 9*, 209–219.

Charness, G. (forthcoming). Attribution and reciprocity in an experimental labour market: An experimental investigation. *Journal of Labor Economics*.

Charness, G., & Grosskopf, B. (2001). Relative payoffs and happiness: An experimental study. *Journal of Economic Behavior and Organization, 45*, 301–328.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple test. *Quarterly Journal of Economics, 117*, 817–869.

Clark, K., & Sefton, M. (2001). The sequential prisoner's dilemma: Evidence on reciprocation. *Economic Journal, 111*, 51–68.

Collard, D. (1978). *Altruism and economy*. Oxford: Robinson.

Cooper, R., DeJong, D. V., Forsythe, R., & Ross, T. (1996). Cooperation without reputation: Experimental evidence from prisoner's dilemma games. *Games and Economic Behavior, 12*, 187–218.

Davis, D., & Holt, C. (1993). *Experimental Economics*. Princeton Univ. Press.

Dufwenberg, M., & Kirchsteiger, G. (1998). A Theory of Sequential reciprocity. Discussion Paper, CentER Tilburg University.

Eckel, C. C., & Grossman, P. (1996). Altruism in anonymous dictator games. *Games and Economic Behavior, 57*(16), 181–191.

Falk, A., & Fischbacher, U. (1999). A theory of reciprocity. Discussion Paper, University of Zürich.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*, 980–994.

Fehr, E., Gächter, S., & Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device. *Econometrica*, *65*, 833–860.

Fehr, E., & Schmidt, K. A. (1999). Theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, *114*, 817–868.

Forsythe, R., Horowith, J., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*, 347–369.

Geanakoplos, J., Pearce, D., & Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, *1*, 60–79.

Gouldner, A. W. (1960). The norm of reciprocity. *American Sociological Review*, *25*, 161–178.

Güth, W. (1995a). On ultimatum bargaining experiments – A personal review. *Journal of Economic Behavior and Organization*, *27*, 329–344.

Güth, W. (1995b). An evolutionary approach to explaining cooperative behavior by reciprocal incentives. *International Journal of Game Theory*, *24*, 323–344.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights and anonymity in bargaining games. *Games and Economic Behavior*, *7*, 346–380.

Jacobsen, E., & Sadrieh, A. (1996). Experimental proof for the motivational importance of reciprocity. Working Paper Nr. B386, SFB 303, Bonn.

Kagel, J. H., & Wolfe, K. (2001). Tests of fairness models based on equity considerations in a three-person ultimatum game. *Experimental Economics*, *4*, 203–220.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review*, *76*, 728–741.

Konow, J. (2001). Fair and square: The four sides of distributive justice. *Journal of Economic Behavior and Organization*, *46*, 137–164.

Kritikos, A. S. (2000). Indenture: A contract for selfish and reciprocal types. IMOP Discussion Papers, Athens.

Kritikos, A. S., & Bolle, F. (1998). Indenturing banknotes as mechanism to induce cooperation. *Journal of Economic Behavior and Organization*, *34*, 279–294.

Kritikos, A. S., & Bolle, F. (2001). Distributional concerns: Equity or efficiency oriented. *Economics Letters*, *73*, 333–338.

Kritikos, A. S., & Bolle, F. (2003). Utility- vs. income-based altruism. Discussion Paper, Frankfurt (Oder).

Kritikos, A. S., & Meran, G. (1998). Social norms, moral commitment and cooperation. *Homo Oeconomicus*, *15*, 71–92.

Levine, D. (1998). Modelling altruism and spitefulness in experiments. *Review of Economic Dynamics*, *1*, 593–622.

Loewenstein, G., Thompson, L., & Bazerman, M. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psych.*, *57*, 426–441.

McKelvey, R. D., & Palfrey, T. R. (1992). An experimental study of the centipede game. *Econometrica*, *60*, 803–836.

Offerman, T. J. (1999). Hurting hurts more than helping helps: The role of self-serving bias. Discussion Paper, University of Amsterdam.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, *83*, 1281–1303.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation*. University of Michigan Press.

Roth, A. E. (1995). Bargaining experiments. In: J. Kagel & A. E. Roth (Eds), *Handbook of Experimental Economics* (pp. 253–348). Princeton University Press.
Selten, R. (1990). Bounded rationality. *Journal of Institutional and Theoretical Economics*, *146*, 649–658.
Simon, H. A. (1957). *Models of man*. New York – London: Wiley.
Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy*, *70*, 111–137.
Watabe, M., Terai, S., Hayashi, N., & Yamagishi, T. (1996). Cooperation in the one-shot prisoner's dilemma based on expectations of reciprocity. *Japanese Journal of Experimental Social Psychology*, *36*, 183–196.

APPENDIX A

The approach by Bolton and Ockenfels (2000) is a special version of (1) insofar as *i* cares only about his relative position towards all others. They assume a utility function

$$U_i = v_i(x_i, \lambda_i) \tag{A.1}$$

with λ_i = standardized share of the social income $c = x_1 + \cdots + x_n$ to be allocated, i.e.

$$\lambda_i = \frac{x_i/c}{1/n} = \frac{nx_i}{c} \quad \text{if } c > 0. \tag{A.2}$$

For $c = 0$, λ_i is set to 1. Bolton and Ockenfels (2000) assume further that v_i is differentiable, that the partial derivative with respect to x_i is positive and that, for every given x_i and variable λ_i, v_i takes its maximum at $\lambda_i = 1$. Moreover

$$\frac{\mathrm{d}^2 v(k\lambda, \lambda)}{\mathrm{d}\lambda^2} \leq 0 \quad \text{for every } k. \tag{A.3}$$

This approach, called ERC theory, also implies (i). Instead of (ii), we get

$$(0, y, \ldots, y) \sim 1 (0, x, \ldots, x) \quad \text{for all } x,\ y. \tag{ii$'$}$$

Like in SCIA, the ERC modell does not care about the distribution of income among others. According to ERC, *i* is indifferent between any distribution of $c - y_i$.

APPENDIX B

Instructions to Experiment 1

This is an experiment in the economics of decision making. You will make decisions in several different situations. Each decision (and outcome) is independent from each of your other decisions. In every case you will be anonymously paired with one other person so that your decision affects the payoffs of others.

There are roles for each decision – generally A or B. Only "A players" will have to make decisions. You are a player A. You will be confronted with several income distributions described by ($y_{\text{for self}}$; $y_{\text{for another person}}$), with the amount of money in German Marks. You will be asked which distribution do you prefer:

(1) (0, 35) or (0, 15)
(2) (10, 40) or (20, 10)
(3) (0, 75) or (10, 10)

Instructions to Experiment 2

This is an experiment in the economics of decision making. If you follow the instructions carefully, you can earn a considerable amount of money. You will be paid in private and in cash at the end of the experiment.

There are two rooms of people in this experiment. The people in the other room are hearing exactly the same instructions. For the following procedure you will be matched with an anonymous person from the other room. You are person II, the person in the other room is person I. You will not be informed about the identity of the other person, nor will the other person will be informed about your identity.

The payoff which you can realize during the following procedure depends on your decision and on the decision of person I. Each round will consist of two steps which will take place in sequence: Person I may choose in the subsequently shown "game 1" between A and B. If person I has chosen A, you will have the choice between C and D. If person I has chosen B, you will have the choice between E and F. Any combination of choices leads to different payoffs. The possible payoffs in Euro-cents are shown in Figs 9 and 10.

Please raise your hand if you have any questions.

If you have no further questions, please answer now the following questions:

Which choice (between C and D) will you make if player I had chosen A.
Which choice (between E and F) will you make if player I had chosen B.

FAIRNESS-BASED ALTRUISM AND REDISTRIBUTION: AN EXPERIMENTAL APPROACH

Luigi Mittone

ABSTRACT

The focus of this paper is on altruism and coordination among agents with different income levels. A special form of altruism (fairness based or ethical altruism) is investigated by means of experiments. The definition of altruism used here follows from A. Sen's concept of obligation, *i.e. behaviour that produces advantage for someone whose welfare is not important at all for the agent's well-being. In this sense, the paper investigates altruism* without reciprocity. *A second hypothesis investigated is that the extent of ethical altruism is influenced by gender and by income differences within the population.*

1. INTRODUCTION: FAIRNESS BASED (ETHICAL) ALTRUISM, AND INCOME REDISTRIBUTION

This paper follows a previous one (Mittone, 2002) which investigated the spontaneous onset of altruistic behaviour within small groups of people. The focus of Mittone (2002) was on a specific definition of altruism suggested by Sen (1970). Sen examined the nature of supportive behaviour to distinguish between that generated by a feeling of sympathy and that caused by a moral duty. Both

Inequality, Welfare and Income Distribution: Experimental Approaches
Research on Economic Inequality, Volume 11, 181–213

ISSN: 1049-2585/doi:10.1016/S1049-2585(04)11009-0

concepts apply whenever an economic agent consciously makes a choice which raises the level of well-being of other agents without the latter (the beneficiaries) having to recognize – either consciously or unconsciously – that they owe? any return favour to the agent that has benefited them.

Following Sen, an action of the type just described is produced by sympathy when the increase in the beneficiary's level of well-being positively influences the level of well-being of the agent, who reflexively enjoys the improvement in the other's circumstances. Instead, an action is motivated by obligation when the agent operates according to a moral principle exogenous to the mechanisms that determine the levels of individual well-being. In other words, when the agent feels sympathy for another person, suffers and feels pleasure with him/her, then the agent's actions undertaken to help or harm the other person are ultimately egoistic, because they are intended to improve the agent's own well-being through modification of the other's well-being. Again following Sen, behaviour can be called non-egoistic only when the agent operates in favour of another agent because s/he believes that one must necessarily behave thus in principle.

Unfortunately, the concept of obligation is much more difficult to specify than that of sympathy. This is because the element of obligatoriness that generates behaviour favourable to other agents should be ideally "net" of the psychological costs caused by the sense of remorse. When an agent feels obligated to help another person because otherwise s/he would incur a psychological cost that would alter her/his level of well-being, then this type of behaviour becomes indistinguishable from that prompted by sympathy. Consequently, Sen concludes, one can talk correctly of obligation, and therefore of non-egoistic choices, only when an agent chooses from two possible actions the one that s/he believes to be right even though it yields him less well-being than the alternative. It also follows that this kind of commitment is closely related to a more general concept of fairness. One could in fact argue that Sen's concept of moral obligation comes very close to a feeling of what it is fair to do and what is not.

Sen's discussion of the concepts of altruism and obligation involves subtle distinctions between what is perceived as well-being at the moment when the choice of action is made and what may yield well-being in the long run. Put otherwise, it could be argued that a choice prompted by obligation does not generate well-being at the moment when it is made, and indeed is perceived by the agent as costly in terms of immediate well-being but nevertheless presages well-being in the long run. An example of this situation is provided by a worker who produces greater work effort than that agreed with the firm – and which is obviously a cost in terms of well-being at the moment when it is produced – not because s/he believes it to be *right* on the basis of some ethical principle – for example that one should always give of one's best in every aspect of life regardless of contracts and formal

rules – but simply with a view to an improvement in her/his future well-being brought about by improving her/his career prospects. To clarify this kind of situation, Sen examines how his definitions of sympathy and obligation fare in a context of intertemporal choice. His discussion of the problem will not be treated here for reasons of space, and because it adds nothing relevant for the topic addressed.

It is important to stress that when Sen's definition of altruism (I shall henceforth call it "ethical altruism" or "fairness based altruism") is transferred to the context of the mechanisms within society or within organizations that generate spontaneously "altruistic" behaviour – that is, personal sacrifice for society and other individuals – the distinction between sympathy and obligation is important – from a normative point of view – only when the sense of obligation is in some way relatable to an ethical system which in turn is in some way determined by the social context in which it has developed. In the absence of a three-way linkage among ethical system, social context and obligation mechanisms, it is pointless to distinguish between obligation and sympathy because it would be anyway impossible to implement a strategy designed to set off the virtuous circle by means of appropriate institutional choices.

The special form of altruism in Sen's sense is also important as a justification for income redistribution policies. Implementing a policy aimed at reducing income disparities entails, in fact, asking richer individuals to reduce their wealth in favour of other people, for whom they cannot reasonably feel sympathy – again using the term in the meaning given to it by Sen – because they are totally unknown to them. The reasons that induce the richer part of the population to accept the sacrifice imposed by an income redistribution policy are several, and they oscillate between strictly selfish motives – like the fear of a revolution fuelled by the sentiment of injustice felt by the poor – and the psychological need for "self-absolution" generated by the feeling that one is luckier than other people. Within this wide range of determinants that justify redistribution policies, however, a special place is occupied by ethical altruism, because is the one most internally coherent with the tools of institutional engineering.

When a society or an organization wants to generate spontaneous virtuous mechanisms, in an attempt to produce better cooperation and solidarity among its members, obviously it can only act on obligation-related motivations. In fact, all inducements founded on sympathy or on selfishness (like the just-mentioned fear of violent insurrection) pertain to the purely psychological and personal domain, and they therefore require individually designed incentives. On the other hand, if the intention is to trigger the onset of an incentivising mechanism based on obligation, this must in some way be relatable to a known context – that is, to an ethical system whose connection with a given social system is known. There are two social contexts that construct the individual ethical system: the one internal to

the groups in which the agents operate (family, firm, other organisations, etc.), and society at large. Here the attention is mainly focused on organisations and society, while no consideration is made of other kinds of small group like the family.

Undertaking changes intended to modify the ethical system at the organization level suffers from the obvious limitation that the organisation can only intervene in itself. On the other hand, one might think that the most efficient level for intervention in the ethical system is the one represented by society as a whole, but the problem with such a generalised level is that society is a highly complex environment, a context in which numerous different ethical systems co-exist, and where it is therefore almost impossible to implement tools which modify those systems in the same direction.

On considering practical examples of relational models intended somehow to produce an ethic of commitment, one finds that many of them refer to firms and rely on the reciprocity mechanism. A classic example of this type of reciprocation mechanism is described in Akerlof's celebrated article (1982) on "cash posters," in which he discusses a case reported by George Homans (1954) and relative to a services firm. In the business studied by Homans, a group of women workers engaged in purely routine tasks displayed a propensity to produce individual levels of work effort that were higher (sometimes much higher) than that contractually required. Akerlof explains this apparently irrational behaviour as resulting from a "gift exchange" relation between the female workers and the firm. Because of this relation, the workers produced more than they were contractually obliged, and they received in return a wage which was slightly above the market rate and – perhaps – greater tolerance of cases of (usually temporary) failure to produce the minimum contractual level of effort. In other words, a reciprocity mechanism had been created of the type: "I the worker produce more that I have to, in return I receive from you a wage above the market rate and the assurance that I will not be dismissed if (for a limited but not rigidly defined period of time) I produce less than the contractual minimum."

The reciprocity relation established between firm and workers in the case described by Homans and discussed by Akerlof means that the virtuous behaviour of the workers depended on the firm's respect for the pact, with the consequence that it was potentially unstable. This point is crucial for my discussion because it helps clarify the difference between behaviour dictated by obligation in Sen's sense, and which is strictly ethical, and behaviour which is instead solely the result of a reciprocation mechanism and has little to do with the ethical dimension of human action. Note that the onset of reciprocation mechanisms has the same effect on altruism as that produced by the sentiment of sympathy in Sen's discussion. A costly action which produces an advantage for another agent, but which is chosen under the logic of reciprocity, is in fact intended to yield a subsequent gain in

terms of well-being just as is the action of an agent who acts out of sympathy and ultimately, and again in Sen's set-up, seeks an improvement in her/his well-being.

Following this line of inquiry many other examples can be found in the literature dealing specifically with the social dimension of altruism. Among such examples of "social altruism" one may mention the "Rotten Kid Theorem" developed by G. Becker (1981). According to Becker, if only one member of a family (typically the head of the family) gives some money to the other members, without asking for a rake-off, a virtuous cooperation mechanism of is activated within the family by the other members. Becker's Theorem demonstrates that, given the first injection of altruism, is then rational for the family members to cooperate because they can thus produce a higher level of collective (and also individual) welfare. The example of the family falls within the category of altruism produced by sympathy because it is reasonable to assume that the head of household gains some form of pleasure from helping his relatives. This does not mean that the mechanism described by the Rotten Kid Theorem is not important in society; on the contrary, it may be of crucial importance in all those situations where social cohesion is poor and individuals tend to assume strongly selfish attitudes.

With regard to the negative effects produced by an atomised society, Putnam (2000) offers a vivid picture of a society where social cohesion is very poor. Furthermore, the Putnam's work yields interesting insights into both the problem of a too low percentage of what I shall later call "altruism bearers" and the role played in past and contemporary American society by philanthropy.

When non-egoist behaviour is caused by reciprocity mechanisms, not only is it no longer definable as altruistic in Sen's sense, but also, and especially, it cannot be related to the ethical dimension, because it depends on a strictly consequentialist process of causation. The ethical dimension, in fact, can only be observed in cases of non-contingent behaviour; or in other words, behaviour which is not actuated by reactions similar to those produced by a cause/effect calculation specific to the context in which it is performed. Actions decided on the basis of a cause/effect calculation of the type implicit to reciprocation mechanisms are typically related to the standard process of maximization of individual utility, and therefore cannot at the same time pertain to the sphere of ethical choices. In fact, using the well-known distinction between ethical preferences and subjective preferences propounded by Harsanyi (1955), one can imagine that agents construct a double system of preference ordering: the system of ethical preferences on the one side, that of subjective preferences on the other.

In Harsanyi's theory, agents order their ethical preferences according to an impersonal representation of a "fair" society, while they structure their subjective preferences solely in function of their own well-being, so that they are contingent. From this it follows that decisions are the result of the intersection of the two

preference systems. In other words, actions are decided by a meta-system of preferences structures along a continuum with decisions caused by purely ethical preferences at the left extreme, and purely subjective choices at the right one.

When distinguishing between behaviours dictated by purely ethical reasons and actions instead produced by reciprocation mechanisms – note that reciprocation-based choices lie close to the left extreme of Harsanyi's meta-preferences system – it is important once again to think from a normative point of view (looking at both organisations and at society) because it is thus possible to single out two strategies with which to actuate non-formalised incentives that differ but are at the same time potentially interconnected or interconnectable. The first strategy is the creation of social conditions such that agents find themselves embedded in a network of reciprocity relations. The second is to trigger virtuous processes by injecting "altruistic values vectors" into the organisation or into the society.

Creating a network of reciprocity relations within an organization requires, for example, organizing workers into teams, introducing shared incentive mechanisms (for instance a bonus shared equally by the members of a team), or rotating workers among jobs so that each of them learns to perform different functions. One among the many possible ways to increase the strength of the reciprocity relationships in society is to focus welfare policies on the family rather than on the individual.

To inject altruistic vectors into an organisation is to import an adequate number of altruism bearing elements from the outside world, or society at large. Altruism bearing elements may be conveyed into the organization through persons ethically committed to altruism – for instance, voluntary workers in non-profit organizations which produce social services – or they may in some way be included among the organization's institutional goals. A typical example of the inclusion of ethical goals in an organization's ends is provided by the for-profit enterprise which decides to allocate some of its profits to socially useful activities, for example by financing medical research.

It is more difficult is to import altruistic bearers into society at large because in this case there is no "outside world" to look at. On the other hand a society can promote – e.g. through fiscal policy – individual or group initiatives inspired by altruistic aims, like non-profit organisations or philanthropic programmes.

Note that the introduction of altruistic goals among an organization's objectives is the only instrument that has a bearing on the question of the relation between ethical system and social context. It will be remembered that when discussing Sen's concept of altruism we assumed that the existence of a causal link between social system and ethical system was the precondition for altruism somehow to promote commitment in organizations. If it is true that the ethical values system of an organization's members is influenced – that is, shaped – by the social context

in which they operate, then it is evident that the purpose of an organization's costly choice of sacrificing part of its profits to finance purely social activities is to establish a social system within its walls that is, at least partly, inspired by principles of pure altruism. Conversely, this assumption is entirely irrelevant if the organization opts for the first of the two strategies just outlined, namely that of self-injecting altruism by introducing altruistic agents. In this case, the process by which the altruistic agents have become such is of little importance; what matters is that they are altruistic by virtue of a moral principle, not because of an emotional state (sympathy) or a utilitarian calculation (reciprocation).

Reflecting on the injection of altruistic bearers raises the question as to how these elements can spread, or in other words, how the virtuous processes of commitment discussed thus far can be set in motion. Thus reaffirmed is the importance of the reciprocation mechanisms mentioned earlier, because it is likely that in order to activate a "virtuous" reciprocity cycle, or to break a vicious one of reciprocal harm, it is necessary to reach a minimum threshold of agents willing to behave in a manner that is individually costly but collectively beneficial, albeit one not driven by non-ethical ends. If this minimum threshold is not reached, reciprocation may collapse into a Nash stable, but simultaneously Pareto inefficient, equilibrium. In other words, pure free-riding behaviour may prevail, so that the entire social system is frozen in a sort of irreversible lock-in process of productive and Paretian inefficiency. In this situation, the injection of bearers of pure altruism – that is, agents indifferent to the reciprocation mechanism and who always and invariably behave altruistically – may break the cycle of negative reciprocation. It may likewise prove useful in situations where the reciprocation mechanism is highly unstable, that is, in the presence of unstable internal equilibria.

With regard to the dynamic between bearers of pure altruism and non-altruistic agents we may once again usefully draw on Harsanyi (1977) and his discussion of a particular type of social cost which arises in situations of interpersonal interaction based on reciprocation mechanisms not sustained by a parallel system of punishment – that is, ones based solely on a system of promises among agents unconstrained by mechanisms of coercion or reciprocal punishability. It is well known, in fact, that interaction models based on reciprocation – as described by games theory for example – which do not comprise the possibility to "punish" defectors (those who do not behave cooperatively) will not lead to Pareto-efficient solutions. The classic example is the prisoner's dilemma, which in its one-shot version collapses onto a non-cooperation equilibrium which is Pareto-dominated by the cooperative equilibrium. On the other hand, the cooperative solution can be "spontaneously" obtained when one moves to the repeated version of the game where the players are able to activate appropriate mechanisms of reciprocal reward and punishment.

To sum up, if the cooperative behaviour triggered by reciprocation is to function, it requires both a system of punishments and rewards and, when a relatively large number of people are involved, an adequate number of agents oriented towards cooperation. Note also that this system of rewards and punishments sustained by the reciprocation mechanism is very similar to the concept of "social rule" described by Elster (1989), who explicitly includes reciprocity in his taxonomy of social rules. It should be emphasised that the social rules defined by Elster are explicitly different from ethical rules in that they are often backed by reinforcement mechanisms based on self-interest.

Investigation of the role of altruistic agents in reciprocation contexts requires clarification of the relations among reciprocating behaviour, free riding, altruism and imitation. The most interesting field for study of these relations is, I believe, that of experimental economics. However, as we shall shortly see, the emergence of altruistic behaviour has been little investigated in the experimental literature.

2. RECIPROCATING BEHAVIOUR AND ALTRUISM

There is a large body of literature on the onset of cooperative behaviour in the absence of incentives, and it has examined the phenomenon from three main points of view: the voluntary supply of public goods (e.g. Andreoni, 1988a, b, 1995a, b; Fischbacher et al., 2001; Keser, 1996), complete information games (e.g. Eckel & Grossman, 1996a, b, c; Hoffman et al., 1996a, b; Kreps et al., 1982), and the alteration of market mechanisms by reciprocation processes (Camerer & Thaler, 1995; Güth & Tietz, 1990; Roth, 1995).

These three strands in the literature provide different yet often overlapping explanations for individually costly cooperative behaviour in the absence of incentives for cooperation. The two most relevant to my purposes here are those of reciprocation and of error (discovered preferences). The feature shared both by explanations based on erroneous choices and by those based on reciprocation is that they derive from theoretical models which prescribe pure "egoistic" behaviour in contexts where the players (the experimental subjects) are able to punish what they deem to be unfair behaviour. In parallel – and consequently – they also share the shortcoming that the context selected almost never permits investigation of the existence of pure altruism.

A classic example of this type of "conditioned" cooperation situation is provided by the repeated prisoner's dilemma. Here the players' ability to punish uncooperative behaviour takes the form of reprisal mechanisms – the best known of them being the tit-for-tat strategy (Axelrod, 1984) – so that it is impossible to

distinguish the onset of altruistic behaviour from cooperative behaviour "forced" by the fear of a reprisal. On the other hand, also experiments based on games which apparently enable assessment of altruistic behaviour, like the "ultimatum game" (Stahl, 1972), are in fact conditioned by the operation of the reprisal effect.

Similarly, also experiments on voluntary provision of public goods with repeated choices may give rise to some form of reciprocation (Fischbacher et al., 2001). Not surprisingly these kinds of experiments are compared to the prisoners' dilemma or to the game of chicken, as emphasised by Ledyard (1995a, p. 144): "*Without a threshold the voluntary contributions mechanism is usually a prisoners' dilemma game; with a threshold it becomes a game of chicken*."

The literature on public goods experiments is very wide-ranging and also very difficult to organise. Put more precisely, and once again quoting Ledyard (1995a, p. 112), "*it is difficult to identify a typical public goods experiment. . . . there are as many variations in procedures and treatments as there are research groups*." The only characteristic shared by the majority of experiments on public goods is that the subjects do not know the others' individual contributions but always know the total contribution. In a repeated choices context, knowing the total contribution, round by round, enables the player? to punish or to reward the group by increasing or decreasing his/her individual contribution over time. Not surprisingly, most public goods experiments with repeated choices display a decreasing rate of contribution (e.g. Isaac et al., 1984, 1985; for an extensive review of this literature see the already mentioned Ledyard, 1995a, b).

The progressive increase of free riding over time (widely discussed by the literature but still not explained in unanimous manner) may be due to the negative reciprocation mechanism that has induced the subjects to react negatively to a collective contribution which is less than expected. Imagine a situation where a subject contributes a sum which is 60% of the maximum in round 1, and then, at the end of the round, discovers that the total contribution by the group has been 40%. It is likely that in round 2 s/he will decide to contribute less, possibly a sum close to 40% of the maximum, thereby "punishing" those who are free riding more than her/him. On the other hand it is quite unlikely that the opposite can happen. In other words, if the subject discovers at the end of round 1 that the group has contributed an amount close to her/his own contribution, there is little chance that s/he will increase her/his contribution, because s/he may reasonably conclude that the others are individually behaving in the same way as s/he is behaving. This obviously happens only in very special cases: most of the time there will be quite high variance among behaviours *and therefore there will be someone who will fall in the situation initially described* (non capisco).[1]

On the other hand, if the subject has contributed less than the group and therefore decides to increase her/his contribution, we are back to the other form

of reciprocity, i.e. positive rather than negative reciprocity. The onset of a positive reciprocation effect, instead of a negative one, should produce virtuous results, but with a lower degree of stability because it only needs someone in the group to start free riding in a marked manner for the positive reciprocation cycle to weaken, giving rise to some form of fluctuation. In fact some experiments (Isaac et al., 1990; Palfrey & Prisbrey, 1993) report that repetition has no effect, i.e. there is neither an increase or a decrease in the average rate of contribution.

The "ultimatum game" describes a situation of the following type. Consider the case of two players, A and B. A is given a sum of money (say € 100) on one condition, namely that s/he give player B a part of it, ranging from € 1 to the entire sum. B cannot communicate with A and can only accept or reject A's offer. If s/he rejects it, both players lose the entire sum. If A decides to offer more than the minimum to B (and this situation arises very frequently in experimental contexts: see e.g. Güth et al., 1982, or Roth et al., 1991), one may suppose that this is a case of altruistic behaviour. However, even in this case it may be that A's decision is conditioned by the risk of reprisal by B. In fact, B may decide to "punish" A for an offer which s/he deems too low, rejecting the offer and inflicting damage on A (who would lose the entire sum) which is much greater than s/he would suffer. Put otherwise, the opportunity cost of punishing A's egoistic behaviour is less for B than the psychological pleasure that s/he derives from the punishment inflicted.

The only games that seem able to isolate the phenomenon of pure altruism unconstrained by the fear of reprisal are the "dictator game" and the "impunity game," which are both variants of the "ultimatum game." In the dictator game, the player who is given the sum of money and must decide how much to give to her/his partner does not risk losing anything because the other player cannot refuse. Note that considering the dictator game to be a "game" is misleading because it involves a purely individual choice problem: one, that is to say, without complications of a strategic nature due to interaction with other agents. The impunity game (Bolton et al., 1998) is very similar to the dictator game, with two differences. The first is that the dictator must choose between two possible ways to split the prize: keep most of it but leave a substantial part for the partner, or divide it into two equal parts. The second difference with respect to the dictator game is that in the impunity game the dictator's partner may decide to reject the offer even if this option – as in the dictator game – does not have consequences for the dictator, who in any case collects her/his prize.

A survey of the main results from experiments using the dictator and the impunity games is contained in the already cited study by Bolton et al. (1998), to which the reader is referred for details. Here I shall merely point out some of the features and shortcomings shared by experiments which have used these games.

The main feature shared by experiments based on the dictator game is that they look for explanations of behaviour "unforeseen" by the theory – which predicts that the dictator will keep as much of the money as possible – solely by examining the role played by anonymity. In fact, anonymity as a potential determinant of purely egoistic behaviour (and therefore coherent with the theory) has been tested both with reference to the players alone – i.e. by ensuring that the dictator does not know who her/his partner is and vice versa (Forsythe et al., 1994) – and by extending anonymity to the experimenters as well – i.e. by ensuring that not even the experimenters were able the reconstruct the identities of the subjects of the experiment (Bolton & Swick, 1995; Hoffman et al., 1994). The idea behind the hypothesis that anonymity is important is that the experimental subjects (obviously when they perform the role of dictator) are loath to appear greedy, either to the partner or to the experimenter, for fear of acquiring a bad reputation and also for purely psychological reasons of self-representation.

The differences between the experiments just mentioned consist of various devices introduced into the experimental design: for example, the contextualization of the game in a market. The two shortcomings shared by these experiments are, first, the use of a game that might be already known to the players – although this was a very remote possibility, at least for the first experiments – and second the static nature of both the dictator and the impunity game. A number of experimenters have sought to remedy this second shortcoming, for example Bolton and Zwick, who repeated the game ten times but with experimental subjects who never met more than once.

The results of experiments using the ultimatum game are rather contradictory, except for the fact that they almost entirely confirm the existence of behaviour inconsistent with the hypothesis of pure selfishness. The existence of such behaviour, however, varies greatly in the dimension of the spontaneous contribution. That is, the sums above the minimum threshold fixed for the game which the dictator forgoes are highly variable, and so too are the percentages of non-egoists in the total of dictators reported by the experiments.

The main conclusion to be drawn from the experimental literature on the dictator game is that the impossibility of designing a truly dynamic version of the game, combined with the lack of real strategic interactions among the players, prevents the transfer of the results from these experiments to contexts – such as the one examined here, namely organizations – typically characterized both by the repetition of choices over time and by strategic interaction among players. That said, it is also interesting to note that some authors emphasise the importance of using ethical factors in explanation of the not perfectly egoistic behaviour found by the experiments. For example Hoffman et al., with reference to over-contribution behaviour by dictators, stress:

> At the very minimum, these results suggest that other-regarding preferences may have an overwhelming social, what-do-others-know, component, and therefore should be *derived* formally from more elementary expectational considerations.

Bolton et al. assert something similar in the conclusions to their study:

> Our procedure suggests that dictators giving arises from a concern for fair distribution on the part of dictators. This is not to say that dictators give in order to improve the welfare of others. In our procedure, concerns for a fair distribution originate from personal and social rules that effectively constrain self-interested behavior – although within these constraints dictators *do* behave in a self-interested manner (they act first to secure what they consider to be their own fair share).

The second of these statements is particularly interesting because it is the one that best fits Sen's definition of non-egoism discussed in the first section.

One type of altruism that is explicitly "impure," in that it springs from the fear of reprisal, is what has been called "altruistic punishment" (Fehr & Gächter, 2002). Experiments on altruistic punishment have investigated the particular type of behaviour observed when an experimental subject decides to assume a personal cost in order to enforce a cooperation rule which operates in favour of the group to which s/he belongs. The context typically described by these experiments is one in which the subjects can decide to cooperate or to defect in the production of a public good. Free-riding behaviour, moreover, can be punished by mechanisms which are costly to the person who inflicts the punishment. In other words, one or more subjects may decide to assume the task and the cost of punishing another subject who has chosen a purely opportunistic strategy. This type of behaviour is called "altruistic punishment" or "altruistic reciprocation" because it is believed that the decision to assume for oneself the cost of enforcing the cooperation rule is contrary to a strictly egoistic logic of maximizing individual utility.

An example of an experiment of this type is described by the already-cited Fehr and Gächter (2002), who used groups consisting of four players who at the beginning of the experiment received 20 monetary units which they could use to finance a common project (investing from 0 to 20 units). The subjects could keep the units that they did not spend on the shared project. For each monetary unit invested in the project, the entire group received 1.6 units in return, and each of the group's members received 0.4 units regardless of her/his contribution to the project. Because the minimum investment was 1 monetary unit, while the individual return was 0.4, no individual economic convenience derived from investing in the project. Choices were made simultaneously and anonymously. At the end of the round, the players were told about the choices that the others had made (although identities were still not revealed), and they could punish the other

players by assigning a punishment score which ranged from 0 to 10 points. Each point received cost the punishee 3 monetary units and the punisher 1 monetary unit.

It will be seen that the punitive altruism described by Fehr and Gächter is difficult to fit with the definition of altruism discussed here because it certainly involves some form of psychological interaction between the subject that inflicts the punishment and the subject that receives it. More specifically, in the case of punitive punishment we have a phenomenon that mirrors Sen's sympathy. It is likely that the subject who decides to assume the cost of the punishment will feel some form of "resentment" or antipathy towards the subject that s/he decides to punish. It should be borne in mind, in fact, that the design of the experiment allowed damage to be caused which was more than proportional to the cost of the punishment. This enabled the punisher to obtain a psychological advantage in terms of revenge on the free rider because the relative cost of the punishment was significantly less for the punisher than for the punished. The context was therefore one that could be related to a calculation of subjective utility, but not to strictly ethical choices.

The second explanation of individually costly cooperative behaviour mentioned above is that of error or decline in the level of spontaneous over-contribution in experiments on repeated public goods. This phenomenon was identified in the early 1990s (Ledyard, 1995a, b; Plott, 1995), and it is typically associated with experiments in which the subjects must make a series of spontaneous choices concerning a public good over time. It has been found that the initial levels of spontaneous over-contribution tend to diminish as the experiment proceeds.

The main reason put forward in explanation of this phenomenon is that the subjects initially make a mistake in fixing their levels of contribution and learn with time to correct this error. The interesting feature of these experiments is that this error-correction process never entirely resolves the phenomenon of over-contribution, which consequently is not entirely eliminated. The persistence of behaviour inconsistent with convergence on the Nash equilibrium therefore seems to indicate the existence of some factor besides error. However, the structure of these experiments, which were constructed in order to leave room in any case for the onset of reciprocating behaviour, does not permit one to establish with certainty whether the over-contribution stems from ethical factors or from something else.

A final consideration concerns an aspect somehow implicit in many of the experiments examined thus far but which was not explicitly discussed in any of them. I refer to the role performed by imitative mechanisms in cooperative choices like the ones described. It is likely that strictly individual motivations for costly co-operation are flanked by processes of a conformist nature; that is to say, behaviour which consists in simple adaptation to the choices prevalent in the group. The

phenomena of conformism and imitation of the group's predominant choice may partly explain behaviour which appears to be altruistic but is in fact motivated by pure imitation.

Discriminating between subjects induced to cooperate by an ethical motivation and those who instead adjust passively to the preponderance of altruistic or egoistic behaviour is a very complex undertaking, and probably impossible. As we shall shortly see, the experiment described here is unable to shed light on this aspect, although its results suggest that this type of phenomenon was present in the behaviour observed.

3. THE EXPERIMENTS

Starting from the theoretical premises set out in the previous section, it was decided to design an experiment which enabled the study of altruistic behaviour triggered by ethical motives. Specifically, we conducted six experimental sessions with the same design but the differences shown in Table 1.

The sole difference between the first two experiments concerned the degree of anonymity. In Exp1Ba the subject that decided to cooperate knew that none of the other participants would know anything about her/his "sacrifice" because no information was given about her/his choices on the computer screen. In Exp2An the condition of anonymity was stricter because we used a stringent "double anonymity" procedure[2] that assured the subjects that their choices were totally hidden to the experimenters as well. In other words, a given subject in Exp2An who decided to cooperate knew that nobody would know anything about her/his "sacrifice," which was thus a strictly private matter. The reason for this

Table 1. The Experiments.

Experiment	Anonymity	Redistribution	Subjects' Gender	Number of Subjects
Exp1Ba[a]	Yes	No	Male and female	20
Exp2An	Yes double blind	No	Male and female	24
Exp3Dif	Yes double blind	Yes	Male and female	
Exp4Dif(a)[b]	Yes double blind	No	Male and female	12
Exp4Dif(b)	Yes double blind	Yes	Male and female	12
Exp5Wo	Yes double blind	No	Women only	24
Exp6Me	Yes double blind	No	Men only	24

[a]The results from this experiment have also been utilised in Mittone (2002).

[b]Exp4Dif(a) and Exp4Dif(b) form part of a common experiment with two rounds – (a) and (b) – carried out using the same sample of subjects to test two settings: round (a) the same as Exp2An and round (b) the same as Exp3Dif1.

strengthening of the anonymity condition was to totally eliminate any form of selfish motivation – even of a psychological nature – that could be produced by the desire to appear "a nice altruistic person" in the eyes of the researchers.

The essential anonymity condition of Exp1Ba was intended to remove the component of direct reciprocity in the form of both reward and risk of punishment. When individually costly cooperative behaviour is directed towards subjects whose identities are unknown, and when it is undertaken without reward or in the absence of the risk of being punished, it is likely to be altruistic behaviour in Sen's sense of the term. The high anonymity condition of Exp2An was intended to improve the degree of "cleanliness" of the behaviours observed, i.e. so that it would be certain that those who decided to cooperate were totally unaffected either by feelings of reciprocity or by a wish to enhance their image in the eyes of the experimenters.

The third and fourth experiments were structured exactly like the second one, and therefore with total anonymity. But a difference was introduced in the money endowment at the beginning of the experiment. In both Exp3Dif1 and Exp4Dif2 the experimental subjects were divided into two groups: the first group received an extra money reward – 10 euros – for participation in the experiment, while the second group only received the standard payment. Membership of the first or second group was a matter of chance because the subjects were separated using a random procedure. The difference between the starting endowments was introduced to simulate a situation of income disparity within the micro-society of the participants in the experiment. The idea was that differences in individual income act as a stimulus for cooperation and solidarity, i.e. the awareness of the existence of income disparities would interact with the ethical system of the experimental subjects and promote altruism by the "richer" ones. The difference between Exp3Dif and Exp4Dif(a−b) was that in Exp4Dif(a−b) the same subjects played the game twice: the first time without any income differentiation and the second time with the extra money given to half of them.

Finally the fifth and fourth experiments were once again identical to the second one (Exp2An) but the samples of experimental subjects used consisted respectively only of women and of men – whereas in all the other experiments the samples were made up of both women and men. The hypothesis behind the fourth experiment was that women are more oriented by ethical motivations towards social cooperation (selfless) than are men (selfish). The idea that women are less egoistic than men is by no means a new one, and it has been investigated from many perspectives of analysis. In particular, the differences between the behaviours of women and men have been analysed experimentally by using the same kinds of games as mentioned in the previous section. Among the studies carried out using the experimental approach, here I shall cite only Eckel and Grossman (1998,

2001), who held ten experimental sessions based on the dictator game, using samples of subjects made up only of women and men. The results from these experiments seemed to show that on average women donate twice as much as men.

The finding that women are more generous than men in the dictator game context was important for our experiments because the setting of this game is very similar to the one described here. Both the double-anonymous dictator game and the experiments reported in this paper share the characteristics of risk removal, gender-related subject interactions, and the experimenter effect. As already mentioned, the main difference between the two approaches is that the dictator game involves a one shot decision, while the puzzle game reported here required the performance of a long sequence of decisions and therefore made it possible to observe the emergence of systematic behaviours.

It is important to underline that, despite their design differences, the main and common purpose of all five experiments was to eliminate the component of direct reciprocity in the form of both reward and risk of punishment. When individually costly cooperative behaviour is directed towards subjects whose identities are unknown, and when it is undertaken without reward or in the absence of the risk of being punished, it is likely to be altruistic behaviour in Sen's sense of the term. The experiments described here made it possible to isolate a situation based on anonymity, costly cooperation, and absence of direct punishment.

The altruistic behaviour studied by means of the experiments discussed here is particularly important for analysis of organizational formulas and more in general for the design of income redistribution policies.

4. THE EXPERIMENTAL DESIGN OF THE PUZZLE EXPERIMENT

The experiment analysed in what follows involved the construction of a puzzle by four players who took turns to place the pieces in position. The players had to place just one piece in each round of the game, and they had to comply with the rule that the puzzle was to be completed by starting from the central triangles and working anticlockwise towards the outer parts of the design. The pattern of the puzzle is shown in Fig. 1.

The experiment continued until the entire puzzle had been completed. Different quantities of the various kinds of pieces were distributed among the players, so that it might happen that a player was unable to contribute to the puzzle when it was his/her turn to make a move. Failure to make a move caused a delay in completion of the game and affected the monetary prize (€122) which was distributed in equal parts among all the players at the end of the experiment. More

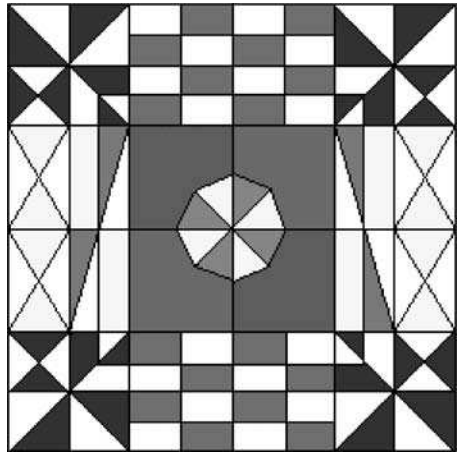


Fig. 1. The Design of the Puzzle.

specifically, every missed turn reduced the amount of the final prize by a fixed sum (€2.06). Because the rule on the division of the final prize stated that the loss was to be equally divided among all the players, it follows that the individual damage caused by failure to insert a piece was one-quarter of the sum subtracted from the collective prize (about half a euro, more precisely €0.515).

The experiment was conducted using computers. Each player saw the pattern of the puzzle and the pieces in his/her possession on the screen but did not know the number and composition of the pieces possessed by his/her partners. For each round, the software showed the space in the puzzle pattern to be filled, and the player whose turn it was had to indicate with the mouse pointer the piece in his/her possession to be inserted in the pattern. If the player did not have the piece required, s/he clicked a button which told the other players that s/he would have to miss his/her turn. The next player could insert one of his/her pieces instead of the missing one, and in this way the final reward was not reduced. However, the player who inserted a piece when it was not his/her turn incurred a penalty (€1.03). Thus, by inserting an optional piece the "altruist" player suffered more damage to his/her individual return than s/he obtained from sharing the common prize without the reduction due to the lost piece (to be precise, s/he lost €0.52). It should also be pointed out that it was impossible to determine whether the choice of behaving altruistically for the whole duration of the experiment would not prove less advantageous – individually – than non-cooperative behaviour, not even in the case in which a situation of close cooperation – choice of altruistic moves – arose among all players. In fact, not knowing what pieces were possessed by the

other players meant that there was a risk of being called upon to cooperate much more frequently than the others, thus generating a result which was collectively better but individually worse than that obtainable by opportunistic behaviour.

The four players sat at an equal number of computer screens together with other experimental subjects (depending on the session, the number of people in the computer room varied from 12 to 20), so that no player knew exactly with whom s/he was playing and could count on remaining anonymous. Because of anonymity and the fact that nobody knew what pieces the other players possessed, it was not possible to enforce cooperative behaviour by punishing free riders. Likewise, if a player decided to cooperate by inserting his/her piece in the place of another player, s/he knew that s/he would incur a cost without being able to count on any form of reciprocity, because s/he did not know the distribution of the pieces among the players. Furthermore, a player who chose to cooperate could not hope to obtain some form of solidaristic recompense for his/her sacrifice from the group as a whole because no one (including the experimenters for Exp2An, Exp3Dif1, Exp4Dif2, Exp5Wo and Exp6Me) ever knew that s/he had cooperated.

The experiment thus made it possible to observe the onset of altruistic behaviour in the absence of both the fear of being punished for non-cooperation and of sympathy for the other players, who remained strictly anonymous.

5. ANALYSIS OF THE RESULTS

The subjects for the experiments were recruited by means of posters put up on the bulletin boards of the Faculty of Economics of the University of Trento. At the beginning of the experiment the subjects were given the list of instruction (see Appendix). The instructions were also read out by a researcher at the beginning of each experimental session. The hypotheses underlying each change to the experimental design were the following:

H1. Exp2An stricter anonymity = lower frequency of altruistic moves: control on Exp1Ba.

H2. Exp3Dif1 and Exp4Dif2(b) differences in the starting endowment = feeling of disparities = greater frequency of altruistic moves by the "richer" subjects: control respectively on Exp2An and on Exp4dif2(a).

H3. Exp5Wo use of only women as experimental subjects = greater frequency of altruistic moves: control on Exp2An and Exp6Me.

The results from the experiment are given in Table 2, which shows the aggregate frequency of the costly cooperation choices made by the participants in each

Table 2. Percentage Frequencies of Cooperative Moves.

Exp1Ba	Grp1.1	Grp2.1	Grp3.1	Grp4.1	Grp5.1	
Grp. Averages	35.75	30	77.75	44.75	39	
Tot. Average	45.45					
Exp2An	Grp1.2	Grp2.2	Grp3.2	Grp4.2	Grp5.2	Grp6.2
Grp. Averages	30	47.25	25	38.5	50.75	59.75
Tot. Average	41.87					
Exp3Dif	Grp1.3	Grp2.3	Grp3.3	Grp4.3	Grp5.3	Grp6.3
Grp. Averages	73	68	92.5	55.5	41.5	51
Tot. Average	63.58					
Exp4Dif(a)	Grp1.4a	Grp2.4a	Grp3.4a			
Grp. Avrgs.	60	75	63.75			
Tot. Avrg.	66.25					
Exp4Dif(b)	Grp1.4b	Grp2.4b	Grp3.4b			
Grp. Averages	55.75	69.5	55.5			
Tot. Average	60.25					
Exp5Wo	Grp1.5	Grp2.5	Grp3.5	Grp4.5	Grp5.5	Grp6.5
Grp. Averages	76	69.75	76.5	50.5	67.75	43.25
Tot. Average	63.96					
Exp6Me	Grp1.6	Grp2.6	Grp3.6	Grp4.6	Grp5.6	Grp6.6
Grp. Averages	76.75	80	89.25	49.25	62.75	67.5
Tot. Average	70.92					

experiment.[3] The values are expressed in percentages: that is, each figure expresses the number of times that a given player in a given group decided to help one of the other players, obviously if s/he was able to do so.

The general average of altruistic moves in Exp1Ba is 45.45, while the same average calculated for Exp2An is 41.87. The difference between the two experimental sessions therefore seems to be quite small. A possible way to check whether the difference between the averages of the two samples of subjects is statistically significant is to compute a normal distribution test.

The test used was a maximum likelihood estimator[4] based on the assumption that a subject's decision to cooperate or to defect was independent of the choices made by the other participants because of the anonymity condition. We then assumed that each individual choice – help; non help – was a Bernoullian, with the values of the parameter θ defined over a λ_1; λ_2 interval:

$$\lambda_{1,2} = (\hat{\theta}_1 - \hat{\theta}_2) \pm 1.96 \times \sqrt{\mathrm{Var}(\hat{\theta}_1 - \hat{\theta}_2)}$$

The results obtained by computing the test are summarised in Table 3.

Table 3. Statistical Significance of Differences Between Groups.

Experiments	θ Values	λ_1	λ_2	Statistical Significance of Differences Between Groups
Exp1Ba – Exp2An	0.459459 0.414414	−0.14177	0.05168	No
Exp3Dif (rich) – Exp3Dif (poor)	0.709091 0.571429	−0.26257	−0.01274	Yes
Exp3Dif – Exp2An	0.63964 0.414414	−0.31571	−0.13473	Yes
Exp5Wo – Exp2An	0.63964 0.414414	−0.31571	−0.13473	Yes
Exp5Me – Exp2An	0.702702 0.414414	−0.37668	−0.19988	Yes
Exp5Wo – Exp5Me	0.63964 0.702702	−0.15026	0.024137	No
Exp1Ba (women) Exp1Ba (men)	0.586956 0.417266	0.33390	0.005473	Yes

The θ and the lambda values for Exp1Ba and for Exp2An show that there is no statistically significant difference between the results from the two experimental sessions, which means that an increase in the anonymity conditions does not change the behaviours of the subjects in any important way.

It is also worth noting that, in both Exp1Ba and Exp2An, cooperating meant deciding to pay a charge without any possibility of obtaining some form of reciprocal aid, neither from the subject directly helped nor from the group as a whole, because no player would ever know if someone had helped and who it had been. In spite of this rather severe contextualisation, almost half of the possible altruistic moves were actually performed by the subjects in both the experiments. The attitude towards ethical altruism – i.e. altruism without reciprocation – therefore seemed to be quite pronounced.

The average frequency of altruistic moves recorded in Exp3Dif is 63.58, which seemingly confirms the first part of hypothesis H2: in fact, it is 21.7 points higher than the average calculated for experiment Exp2An. By contrast, the average frequency of cooperative moves made by the experimental subjects in Exp4Dif(a) is lower than the average reported in the second round of this experimental session, i.e. the round with a different initial endowment – Exp4Dif(b).

This apparent incoherence between the results of the two experimental sessions with different initial endowments – Exp3Dif and Exp4Dif2(a−b) – is a consequence of the decision to use the same sample of subjects in Exp4Dif(a−b). Using the same sample of subjects, and assigning them the higher initial money

endowment at random, incurred the risk of giving the extra money to the subjects who showed themselves to be the most altruistic in the first round. Giving more money to those subjects who had already chosen always to cooperate meant reducing the possibility of testing the role played by income disparities as an incentive for altruism. In fact, it is evident from Table 2 that in Exp4Dif(b) 3 out of a total of 6 "rich" subjects had always cooperated in Exp4Dif(a) and therefore could not improve their degree of altruism. By contrast, a further 3 subjects who had always cooperated in the first round were not chosen for the extra money in the second round. It follows that the average number of altruistic moves decreased in the second round because half of the "rich" subjects could not improve their degree of altruism while all the "poor" ones could reduce their level of cooperation – and in fact 50% of them did so.

For the sake of precision, it should also be pointed out that two "rich" subjects – who produced a low level of cooperation in Exp4Dif(a) – decided even to reduce their effort in the second round, despite the fact that they received the extra money.

The average percentage of altruistic moves, respectively for the "rich" and for the "poor" subjects in Exp3Dif, are 70.5 and 56.67; values which are coherent with the hypotheses assumed. Furthermore, the difference is significant. In fact, the θ values and the acceptance interval for the two sub-samples – the sub-sample of the "rich" subjects and the sub-sample of the "poor" ones – show that the difference between the propensities of the poor subjects and the rich ones to cooperate is statistically significant. An analogous result is also obtained from comparison between the values of λ computed respectively for the whole sample of subjects that participated in Exp3Dif and for the sample of subjects in Exp2An. In this case too, the difference between the averages is statistically significant.

It should be stressed that, although the experimental subjects seemed to be influenced by the initial money endowment, this did not completely eliminate altruistic behaviour from the poor group. Table 2 – Exp3Dif and Exp4Dif(b) – shows that 12 out of the total of 16 participants who had not received the initial money endowment cooperated in more than 50% of the moves, and 5 of them always cooperated. The existence of cooperative behaviour in the sub-sample of the disadvantaged subjects confirms the "ethical" nature of the altruistic behaviour observed. When a participant, even though s/he had been discriminated against, still decided to cooperate, it meant that s/he was behaving in accordance with some psychological-ethical "built-in" mechanism which was context independent.

Checking for differences in the attitude towards altruism of women and men requires comparison between the results from Exp5Wo and Exp6Me. From preliminary analysis of Table 2 it seems that women are less altruistic than men because the percentage of cooperative moves in the women sample is 63.96, while the same average computed for the sample made up by men is 70.92. More

in general, it seems that the samples made up of subjects of the same sex display a more marked solidaristic attitude than do the mixed samples. This impression is confirmed when the θ values and the acceptance interval are computed for Exp5Wo and Exp2An. Given the lambda values, there is a statistically significant difference (95%) in the propensities to cooperate when the sample consisting only of women is compared with the mixed sample. Women make 22% more altruistic moves than do the subjects belonging to the mixed sample.

Similarly, also the theta values computed for Exp6Me and Exp2An and the corresponding acceptance interval $\lambda_1 - \lambda_2$ point to the conclusion that the difference between the average propensities to make altruistic moves is statistically significant. Conversely, the differences between Exp5Wo and Exp6Me are not statistically significant, which may lead indicate that what matters is not the difference between sexes but the fact that the subjects in the sample are of the same sex.

Summarising, it seems that some form of solidarity arises in groups comprising subjects of the same sex. On the other hand, women, when compared with men in a mixed group, display a greater propensity to cooperate. In fact, analysis of the results from experiment Exp1Ba (which was the only mixed experiment with low anonymity and therefore the only one that allowed identification of the sex of the players) shows that the female participants cooperate on average in 58% of moves while the male participants choose to cooperate only in 41.27% of total moves. On computing the theta values respectively for the females and for the males and defining the confidence interval, it emerges that the difference is statistically significant.

More in general, and with regard only to the directly comparable experiments – i.e. excluding Exp4Dif, which had too few subjects and was the only one that used the same subjects to test two experimental settings – one notes that the average frequency of altruistic moves made in the various experimental sessions is consistent with the hypotheses incorporated into the experimental design. The highest values were reported in the experiments with subjects of the same sex and without initial income differentiation, closely followed by the experiment that introduced income differentiation. It is also worth noting that the number of subjects that decided always to cooperate is quite high (28 out of a total of 116 participants excluding Exp4Dif for the reasons just mentioned) and therefore allows one to conclude that the general thesis put forward here is valid: namely that altruism without reciprocation exists not only in a one-shot setting (like the one used by the dictator game) but also in a repeated choices context.

Awareness of the existence of inequalities in the individual initial endowments induced 7 “rich” subjects – plus another one who cooperated in 80% of the moves – out of a total of 12, always to cooperate, while only 2 of the “poor” subjects decided always to do so. These differences (which are statistically

significant) apparently demonstrate that it should be easier to implement an income redistribution policy – and less politically "expensive" – when people are strongly aware that there are inequalities in society.

Analogously, the same awareness should act as an incentive for cooperation within organisations. Consciousness that some members of the organisation are disadvantaged – for example because they suffer from some form of physical or mental handicap – can produce a positive feeling of solidarity which pushes the other members towards altruistic behaviour. Note that this consideration is based on a repeated choices experiment that allows one to state that altruistic behaviour may be "stable," i.e. that it is not circumscribed to a one shot decision.

6. THE OPINIONS OF THE EXPERIMENTAL SUBJECTS

The results from the experiments discussed in the previous section seemingly show the existence of individually costly cooperative behaviour sustained, not by reciprocation mechanisms but by some form of psychological-ethical mechanism. In other words, they seem to prove the existence of "pure altruists," or of non-egoistic behaviour in Sen's sense.

The importance of ethical motives in determining altruistic behaviour is confirmed by another empirical finding, this one based on the opinions of the participants in the experiments.

In order to facilitate interpretation of the results of the experiments, after each session the participants were asked to complete a questionnaire intended to gather their opinions. The questionnaire consisted of a small set of questions mainly focused on the problem of reciprocity: that is, it was designed to verify whether the participants thought they could have somehow punished free-riding behaviour. At the end of the four experiments, 60% of the participants thought it was impossible to punish free-riding behaviour even if they had been informed about the identity of the subject who decided not to cooperate. Consequently, the majority of participants were convinced of the substantial unpunishability of egoistic behaviour even if the anonymity condition were totally relaxed. It should also be stressed that practically none of the subjects who thought that free riders could be punished was able to explain how this could be accomplished in practice. Moreover, the overwhelming majority of them also said that the punishment would in any case affect the other players.

Regarding the specific motivations reported by the subjects who participated in Exp3Dif, the majority (58%) of the "poor" ones declared that their decision to cooperate had been influenced by the feeling that they were disadvantaged with

respect to the others. On the other hand, the subjects in the "rich" sub-sample were equally distributed between those who declared that they had not been pushed into greater cooperation by the feeling that they were luckier than the others, and those who declared that they felt somehow encouraged to cooperate by their higher initial endowment.

The large majority (83%) of Exp5Wo declared that their membership of a group consisting only of women did not affect their attitude towards cooperation at all. The same reply was given by the majority (79%) of the men who participated in Exp6Me. The opinions expressed by the participants in both the experimental sessions with only one sex contrast with the behaviours observed. The most plausible reason for this discrepancy is that people do not like to appear influenced by sex differences because there is some sort of generalised cultural agreement that sex discrimination is "bad."

In the majority of cases, the opinions gathered by the questionnaires seemingly support the conclusions drawn from the experiments; in particular they support the hypothesis that altruistic choices were not influenced by the fear of being punished or by any positive return from the other participants.

7. CONCLUSIONS

Summarising the results from the puzzle experiment we may say that altruism without reciprocation is a quite common form of behaviour, not only in one shot decision games (like the dictator game) but also in a repeated choices context. The extent of the effects produced by ethical altruism may be influenced by socio-economic variables like income inequalities and gender. Disparities in income seem to generate a higher altruistic response from the richest subjects and a lower level of cooperation from the poor ones. This means that the solidarity propensity due to altruism – without reciprocation – is influenced by some broad form of psychological sensitivity to fairness in wealth distribution.

The effects of gender are less clear, but two considerations can be made: the first is that women in mixed groups are more altruistic than men; the second is that awareness of belonging to a group made up of subjects of the same sex increases cooperation both for women and for men.

On the other hand it is worth emphasising that the tendency of women to be more altruistic than men is well known and has been analysed by the Gender Psychology literature. Among the explanations suggested for this greater propensity of women towards altruism one of the more interesting is the evolutionary one. On this line of analysis, women should be more oriented towards altruism because they are less pushed to compete for mates than men are. Another evolutionary based explanation is that women perform a crucial role in the care of children. The more altruistic a

mother (towards her children) the better chances of survival her babies will have, with the consequence that she will have the maximum number of offspring.

The results of the puzzle experiments yield normative indications both when they are applied to organisations and when they are related to society as a whole. The effects produced by ethical altruism on the degree of efficiency of an organisation are quite obvious. Since this determinant of cooperative behaviour is substantially independent from the specific context, it can be "injected" into the organisation without a specific formal contract being necessary. Injecting altruistic values into an organisation means looking for members – who may be either individuals or other organisations as partners – that have ethical aims as their main objective – e.g. volunteers, non profit foundations and organisations, etc. Involving ethically motivated partners may increase the level of cooperation within the organisation, and at the same time it may activate virtuous mechanisms of imitation among the other members. The advantage of injecting bearers of altruism, compared with more traditional instruments to promote productivity and efficiency like money incentives defined by a formal contract, is that these "altruistic holders" are substantially unaffected by the organisational context and therefore are very useful in all cases where a organisation's milieu has deteriorated.

Similarly, at the society level it might be interesting to consider the implications of education in solidarity and cooperation for the political cost – degree of legitimisation – of redistributive policies.

The recent evolution of national economic systems towards globalisation has widened income gaps both within countries and among them. In many countries the impact of globalisation and consumerism has caused a dramatic growth of new forms of poverty characterised by a total lack of solidarity. A mechanical interpretation of the law of market competition has increased the distances among countries and among individuals. Where there were once bonds of solidarity among the residents of the poorer countries, which helped prevent isolation and social deprivation, there is now a sort of desert produced by uncritical imitation of the lifestyles of the richer countries. From the experimental results discussed here is possible to conclude that ethically grounded altruism exists, and that its bearers may push the governments to produce new and more effective policies against inequality and discrimination.

NOTES

1. The situation depicted here has some analogies with the discussion on the effects produced by modifications of the marginal per capita return: see Ledyard (1995a, b) for a review.

2. The double blind procedure adopted has similarities with the procedures developed in Hoffman et al. (1994). It is described in the instructions given in the appendix.

3. A more detailed table with the individual data is given in Appendix B.
4. A detailed description of the test used is given in Appendix C.

ACKNOWLEDGMENTS

The first results from a pilot experiment originally developed by the present writer were used by Barbara Ferro for her degree thesis, of which I was the supervisor. Assistance in the construction of the pilot experiment was provided by Barbara Ferro herself, who designed the puzzle, and by the staff of the Experimental Economics Laboratory of the University of Trento, and in particular by Marco Tecilla, who subsequently wrote the software used to conduct the final version of the experiment to which this paper refers. My especial thanks go to Dominique Cappelletti and Ivan Soraperra for their helpful support both in the practical conduct of the experiments and in the statistical analysis of the results. Finally, I wish to thank two anonymous Referees for their very useful comments, which I have widely used to improve the quality of the final version of this paper. Of course, responsibility for any error or omission is entirely mine.

REFERENCES

Akerlof, G. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics*.

Andreoni, J. (1988a). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics*, *3*(37), 291–304.

Andreoni, J. (1988b). Privately provided public goods in a large economy: The limits of altruism. *Journal of Public Economics*, *1*(35), 57–73.

Andreoni, J. (1995a). Cooperation in public goods experiments: Kindness or confusion? *American Economic Review*, *4*(85), 891–904.

Andreoni, J. (1995b). Warm glow vs. cold-prickle: The effects of positive and negative framing on cooperation in experiments. *Quarterly Journal of Economics*, *1*(110), 1–21.

Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.

Becker, G. (1981). *A treatise on the family*. Harvard: Harvard University Press.

Bolton, G. E., Katok, E., & Swick, R. (1998). Dictator game giving: Rules of fairness vs. acts of kindness. *International Journal of Game Theory*, *27*, 269–299.

Bolton, G. E., & Swick, R. (1995). Anonymity vs. punishment in ultimatum bargaining. *Games and Economics Behavior*, *1*(10), 95–121.

Camerer, C. F., & Thaler, R. H. (1995). Ultimatums, dictators and manners. *Journal of Economic Perspectives*, *2*(9), 209–219.

Eckel, C., & Grossman, P. (1996a). Altruism in anonymous dictator games. *Games and Economic Behavior*, *2*(16), 181–191.

Eckel, C., & Grossman, P. (1996b). The relative price of fairnees: Gender differences in a punishment game. *Journal of Economic Behavior and Organization*, *2*(30), 143–158.

Eckel, C., & Grossman, P. (1996c). Equity and fairnees in economic decisions: Evidence from bargaining experiments. In: G. Antonides, W. F. Raaij & S. Maital (Eds), *Advances in Economic Psychology*. Chichester: Wiley.

Eckel, C., & Grossman, P. (1998). Are woman less selfish than men? Evidence from dictator experiments. *Economic Journal*, *448*(108), 726–735.

Eckel, C., & Grossman, P. (2001). Chivalry and solidarity in ultimatum games. *Economic Inquiry*, *2*(39), 171–188.

Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives*, *4*(3), 99–117.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*, 397–404.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *3*(6), 347–369.

Güth, W., Shmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behaviour in Organization*, *3*, 367–388.

Güth, W., & Tietz, R. (1990). Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology*, *3*(11), 417–449.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *4*(63), 309–321.

Harsanyi, J. C. (1977). Morality and the theory of rational behavior. *Social Research*, *4*(44), 623–656.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). References, property rights, and anonymity in bargaining games. *Games and Economics Behavior*, *3*(7), 346–380.

Hoffman, E., Mccabe, K. A., & Smith, V. L. (1996a). On expectations and the monetary stakes in ultimatum games. *International Journal of Game Theory*, *3*(25), 289–301.

Hoffman, E., Mccabe, K. A., & Smith, V. L. (1996b). Social distance and other regarding behavior in dictator game. *American Economic Rewiev*, *3*(86), 653–660.

Homans, G. C. (1954). The cash posters. *American Journal of Sociology*, *19*, 724–733.

Isaac, R. M., McCue, K., & Plott, C. (1985). Public goods provision in an experimental environment. *Journal of Public Economics*, *26*, 51–74.

Isaac, R. M., Walker, J., & Thomas, S. (1984). Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice*, *43*, 113–149.

Keser, C. (1996). Voluntary contributions to a public goods when partial contribution is a dominant strategy. *Economics Letters*, *3*(50), 359–366.

Kreps, D., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, *2*(27), 245–252.

Ledyard, J. O. (1995a). Public goods: A survey of experimental research. In: J. H. Kagel & A. E. Roth (Eds), *The Handbook of Experimental Economics* (pp. 111–194). Princeton: Princeton University Press.

Ledyard, J. O. (1995b). *The economics of informational decentralization: Complexity, efficiency, and stability*. Boston, Dordrecht and London: Kluwer Academic, pp. xii, 494.

Mittone, L. (2002). Altruismo etico e coordinamento nelle organizzazioni: Un approccio sperimentale. In: C. Borzaga & M. Musella (Eds), *Produttività ed Efficienza Nelle Organizzazioni Nonprofit: Analisi Teoriche e Verifiche Empiriche*, edizioni 31, Trento.

Plott, C. R. (1995). Rational individual behaviour in markets and social choice processes: The discovered preference hypothesis. In: K. J. Arrow, E. Colombatto, M. Perlman & C. Schmidt (Eds), *The Rational Foundations of Economic Behaviour*. London: Macmillan.

Putnam, R. D. (2000). *Bowling alone*. New York: Simon & Schuster.

Roth, A., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An experimental study. *American Economic Review*, *81*, 1068–1095.

Roth, A. E. (1995). Bargaining experiments. In: J. H. Kagel & A. E. Roth (Eds), *The Handbook of Experimental Economics* (pp. 253–348). Princeton: Princeton University.

Sen, A. K. (1970). *Collective choice and social welfare*. San Francisco: Holden-Day.
Stahl, I. (1972). *Bargaining theory*. Stockholm: Economic Research Institute.

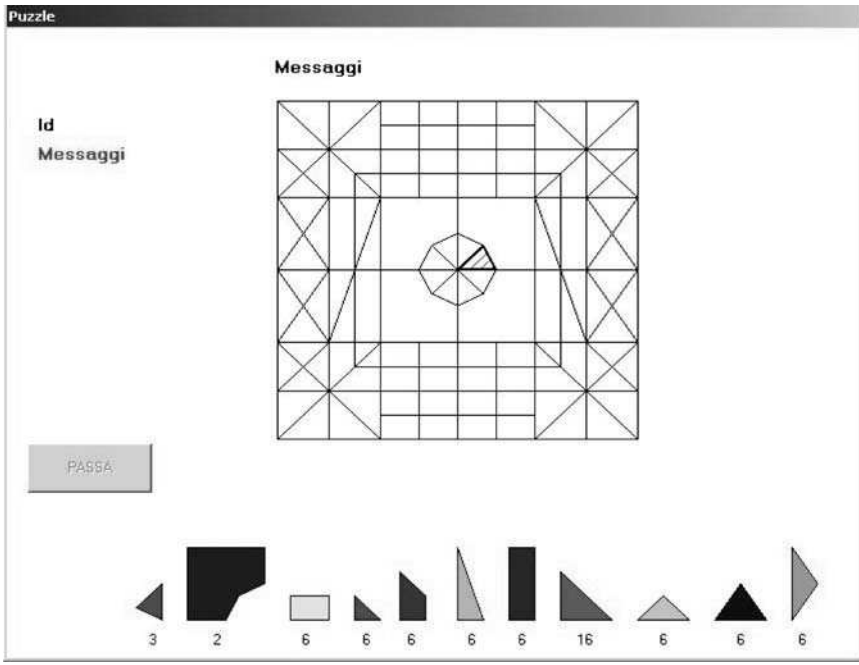
APPENDIX A

Instructions Given to the Participants

You are about to take part in an experiment intended to study the behaviour of people when they cooperate.

Each of you will belong to a group of 4 players chosen at random by computer. No player will know who the other members of his/her group are. Communication will take place solely via your computer screen.

The experiment consists of completing a puzzle of 120 pieces in as few moves as possible. The pieces divide into 11 types which differ by colour and shape. The number of pieces allocated to each player are shown at the bottom of the screen. The numbers written beneath each piece state how many pieces of that type are available to each player.



The puzzle must be assembled following a fixed sequence. The puzzle can be divided into 5 sub-puzzles and it must be assembled in the following order:

- the central octagon;
- the square with the octagon at its centre;
- the first frame of the square;
- the frame of the first frame;
- the frame of the second frame.

Each sub-puzzle is assembled anticlockwise by means of the computer.

Each group has an initial account of €120. Every move in excess of the minimum of 120 moves necessary to complete the puzzle will entail the deduction of €2 from the group's account, which will be divided equally among the 4 players at the end of the experiment.

Each player in turn inserts the right piece in the corresponding area of the puzzle grid. If the player does not have this piece, s/he misses his/her turn and the next player receives a request for help. At this point s/he may choose between the two following alternatives:

- Help the previous player and put the piece in its place on the grid. In this case the person that has helped must pay a penalty of €1 which will be deduced from his/her final personal reward. This action counts as 1 move, so that the player prevents €2 being deducted from the group's account.
- Not help the previous player. In this case the player inserts the requisite piece in the puzzle grid, only when is her/his time to play. This behaviour counts as 2 moves. This means that the minimum number of moves necessary to complete the puzzle increases. €2 are deducted from the group's account.

It may happen that the next player does not have the requisite piece either. In this case, s/he passes the turn to the next player by clicking on the "pass" button.

To sum up, the final amount of the group's account is calculating by deducting €2 from the initial €120 every time a player does not help the previous player who does not have the piece required. This account is then divided equally among the four players in the group. Deducted from the individual accounts are €1 for every time that a player has helped the previous player.

Payment Rules

Before beginning the experiment, you should randomly take a sealed envelope containing your code and your game number. These you should write in the boxes

and then click on the "begin the experiment" button. In this way total anonymity is guaranteed also with respect to the experimenters.

You should keep your code and number sheet with care, without showing it to anyone.

At the end of the experiment, you will go in a room where there are a number of sealed envelopes equal to the number of participants in the experiment. These envelopes are labelled with the codes and numbers of the participants, and they contain the rewards.

At this point, each of you will enter the room one by one and take her/his envelope. You will go back to the common room and wait until all the participants have taken their envelopes.

If someone does not find her/his envelope, or if any kind of irregularity occurs, the experimenters will ask you to show your number-code sheet in order to verify that everybody has taken the right envelope.

APPENDIX B

The Individual Patterns of Co-operation

Table B.1, shows the frequency of the costly cooperation choices made by the participants in each experiment. More specifically, for each experiment the identities of the players are given in the rows, while the figures in the columns refer to the groups of four players which made up each experimental sample. As in Table 2 the values are expressed in percentages.

APPENDIX C

The Statistical Test of Independence

The experimental subjects' decisions have the following distribution:

$$f(x) = \theta^x(1-\theta)^{1-x} \quad \theta \in (0,1) \quad x \in \{0,1\} \tag{C.1}$$

From (C.1) it follows that each experimental session is a random sample with n samplings from a Bernoullian. Therefore the maximum likelihood estimator is given by:

$$\hat{\theta} = \sum_{i=1}^{n} x_i \tag{C.2}$$

Table B.1. Percentage Frequencies of Cooperative Moves.

Exp1Ba											
Player	Grp1.1	Player	Grp2.1	Player	Grp3.1	Player	Grp4.1	Player	Grp5.1		
G1.1.1	89	G1.2.1	100	G1.3.1	100	G1.4.1	78	G1.5.1	0		
G2.1.1	44	G2.2.1	0	G2.3.1	33	G2.4.1	0	G2.5.1	0		
G3.1.1	10	G3.2.1	20	G3.3.1	100	G3.4.1	90	G3.5.1	100		
G4.1.1	0	G4.2.1	0	G4.3.1	78	G4.4.1	11	G4.5.1	56		
Grp. Avrgs.	35.75		30		77.75		44.75		39		
Tot. Avrg.	45.45										

Exp2An											
Player	Grp1.2	Player	Grp2.2	Player	Grp3.2	Player	Grp4.2	Player	Grp5.2	Player	Grp6.2
G1.1.2	44	G1.2.2	11	G1.3.2	67	G1.4.2	0	G1.5.2	0	G1.6.2	0
G2.1.2	33	G2.2.2	100	G2.3.2	0	G2.4.2	44	G2.5.2	67	G2.6.2	100
G3.1.2	10	G3.2.2	0	G3.3.2	0	G3.4.2	10	G3.5.2	80	G3.6.2	50
G4.1.2	33	G4.2.2	78	G4.3.2	33	G4.4.2	100	G4.5.2	56	G4.6.2	89
Grp. Avrgs.	30		47.25		25		38.5		50.75		59.75
Tot. Avrg.	41.87										

Exp3Dif											
Player	Grp1.3	Player	Grp2.3	Player	Grp3.3	Player	Grp4.3	Player	Grp5.3	Player	Grp6.3
G1.1.3(R)	100	G1.2.3(R)	100	G1.3.3(R)	100	G1.4.3(R)	11	G1.5.3(P)	33	G1.6.3(R)	0
G2.1.3(P)	56	G2.2.3(P)	100	G2.3.3(R)	100	G2.4.3(P)	11	G2.5.3(P)	0	G2.6.3(P)	44
G3.1.3(R)	80	G3.2.3(P)	50	G3.3.3(P)	70	G3.4.3(P)	100	G3.5.3(R)	100	G3.6.3(P)	60
G4.1.3(P)	56	G4.2.3(R)	22	G4.3.3(P)	100	G4.4.3(R)	100	G4.5.3(R)	33	G4.6.3(R)	100
Grp. Averg.	73		68		92,5		55,5		41,5		51
Tot. Avrg.	63,58										

R = player with extra reward P = player without extra reward

Exp4Dif(a)					
Player	Grp1.4a	Player	Grp2.4a	Player	Grp3.4a
G1.1.4a	100	G1.2.4a	22	G1.3.4a	33
G2.1.4a	40	G2.2.4a	78	G2.3.4a	22

Table B.1. (*Continued*)

G3.1.4a	100	G3.2.4a	100	G3.3.4a	100
G4.1.4a	0	G4.2.4a	100	G4.3.4a	100
Grp. Avrgs.	60		75		63.75
Tot. Avrg.	66.25				

Exp4Dif(b)

Player	Grp1.4b	Player	Grp2.4b	Player	Grp3.4b
G1.1.4b(R)	100	G1.2.4b(R)	0	G1.3.4b(R)	22
G2.1.4b(P)	22	G2.2.4b(P)	78	G2.3.4b(P)	0
G3.1.4b(P)	90	G3.2.4b(P)	100	G3.3.4b(P)	100
G4.1.4b(R)	11	G4.2.4b(R)	100	G4.3.4b(R)	100
Grp. Avrgs.	55.75		69.5		55.5
Tot. Avrg.	60.25				

R = player with extra reward P = player without extra reward

Exp5Wo

Player	Grp1.5	Player	Grp2.5	Player	Grp3.5	Player	Grp4.5	Player	Grp5.5	Player	Grp6.5
G1.1.5	78	G1.2.5	78	G1.3.5	100	G1.4.5	56	G1.5.5	89	G1.6.5	100
G2.1.5	67	G2.2.5	11	G2.3.5	78	G2.4.5	0	G2.5.5	56	G2.6.5	0
G3.1.5	70	G3.2.5	90	G3.3.5	50	G3.4.5	90	G3.5.5	70	G3.6.5	40
G4.1.5	89	G4.2.5	100	G4.3.5	78	G4.4.5	56	G4.5.5	56	G4.6.5	33
Grp. Avrgs.	76		69.75		76.5		50.5		67.75		43.25
Tot. Avrg.	63.96										

Exp6Me

Player	Grp1.6	Player	Grp2.6	Player	Grp3.6	Player	Grp4.6	Player	Grp5.6	Player	Grp6.6
G1.1.6	67	G1.2.6	100	G1.3.6	100	G1.4.6	56	G1.5.6	100	G1.6.6	100
G2.1.6	100	G2.2.6	100	G2.3.6	67	G2.4.6	100	G2.5.6	11	G2.6.6	56
G3.1.6	40	G3.2.6	20	G3.3.6	90	G3.4.6	30	G3.5.6	40	G3.6.6	70
G4.1.6	100	G4.2.6	100	G4.3.6	100	G4.4.6	11	G4.5.6	100	G4.6.6	44
Grp. Avrgs.	76,75		80		89,25		49,25		62,75		67,5
Tot. Avrg.	70,92										

The maximum likelihood estimator of a Bernoullian random variable, of parameter θ, has a distribution which can be approximated by a normal distribution. Therefore the difference between the parameters computed for different groups has the following distribution:

$$\hat{\theta}_1 - \hat{\theta}_2 \overset{n\to\infty}{\to} N\left(\theta_1 - \theta_2 \frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}\right) \tag{C.3}$$

From (C.3) is possible to compute the 95% confidence interval:

$$\lambda_{1,2} = (\hat{\theta}_1 - \hat{\theta}_2) \pm 1.96 \times \sqrt{\text{Var}(\hat{\theta}_1 - \hat{\theta}_2)} \tag{C.4}$$

Falling within the interval defined by (C.4) are 95% of the values of the difference between the estimators. Therefore, if this interval does not include zero, one can state that the attitude towards cooperation within each couple of groups is different with 95% significance. This means defining an acceptance-rejection zone for the following system of hypotheses:

$$\begin{cases} H_0 : \theta_1 = \theta_2 \\ H_1 : \theta_1 \neq \theta_2 \end{cases}$$

Which means that the null hypothesis is rejected if $0 \notin (\lambda_1; \lambda_2)$.

INEQUALITY AND PROCEDURAL FAIRNESS IN A MONEY-BURNING AND STEALING EXPERIMENT

Daniel John Zizzo

ABSTRACT

This paper presents the results of an experiment where an unequal wealth distribution was created and then subjects could act to change this wealth distribution. Subjects received money by betting and possibly by arbitrary ("undeserved") gifts; they could then pay to reduce, redistribute and, in half of the sessions, steal money from others. The experimental results are incompatible with some standard models of interdependent preferences. Over 80% of redistributors were rank egalitarian, but how subjects perceived the problem significantly affected their redistribution activity: perceptions of fairness were not simply a matter of relative payoff, and changed according to whether a subject was undeservedly advantaged or otherwise.

1. INTRODUCTION

This paper describes the results of an experiment on the economics of inequality and procedural fairness. Models with interdependent preferences (such as altruism, envy and inequality aversion) make predictions on what distributional outcomes agents prefer for themselves and other agents. Interdependent preferences have been incorporated in rational choice models to explain a variety of empirical

Inequality, Welfare and Income Distribution: Experimental Approaches
Research on Economic Inequality, Volume 11, 215–247

ISSN: 1049-2585/doi:10.1016/S1049-2585(04)11010-7

anomalies, for example in relation to consumption and wage determination (Frank, 1985), and public goods contribution (Offerman et al., 1996). Various policy implications derive from the idea of interdependent preferences, for example for tax determination (Frank, 1997) or the relative importance of unemployment and growth as policy goals (Oswald, 1997).

There are many different theories of preferences over distributional outcomes. In addition, evidence exists suggesting that how a decision problem is perceived (categorized) matters in determining behaviour apparently driven by interdependent preferences. For example, in deciding how much to contribute to a public good, it matters whether a "cooperative" or "entrepreneur business-strategy frame" has been induced (Elliott et al., 1998). It matters whether an agent is perceived as deserving a bigger share of the cake being bargained (e.g. Hoffman et al., 1994): agents care about the procedure by which earnings are obtained, i.e. they care about procedural fairness. The evidence from dictator games (where the dictator chooses how to split the cake and the other "player" has no say) shows that the dictator will give more if she believes the recipient is deserving a gift (Eckel & Grossman, 1996) and less if she believes she has "earned" the money she has (Todd, 2001): procedural fairness cuts both ways.

In the experiment presented in this paper, subjects received money by betting and possibly by an arbitrary allocation procedure that induced changes in the perceptions of procedural fairness. By paying a price, they could then eliminate ("burn") and redistribute money (including their own) and, in about half of the sessions, steal money from others. Only one decision was made, and it was the final decision in the experiment, to avoid reputational considerations. Therefore, strictly speaking, only short run behavior was under study (although practice was provided). With this qualification, the experiment made two contributions. First, it provided a new setting to verify that agents care about distributional outcomes and procedural fairness against the null hypothesis of self-interest. Second, it verified the explanatory power of competing hypotheses concerning interdependent preferences and the distribution of interdependent preferences in the population.

We found that the observed redistribution patterns were incompatible with self-interest, pure or impure envy or altruism, and Levine's (1999), Offerman et al.'s (1996) and Charness and Rabin's (2002) distributions of preference types. Over 80% of the subjects engaging in redistribution activity were rank egalitarian: they cared about reducing the scores of richer subjects at least as much or more than the poorer ones. This result supports models of distributional fairness that make subjects care about the others' individual payoffs (e.g. Charness & Rabin, 2002, or a non-linear version of Fehr & Schmidt, 1999, but not Bolton & Ockenfels, 2000),[1] although these also are not without problems.

How subjects perceived (categorized) the problem, especially in relationship to the fairness of the procedure, significantly affected their redistribution activity. For example, almost half of the advantaged subjects in the Stealing condition could be classified as self-interested when the procedure was unfair, but none otherwise.

Section 1 presents the experimental hypotheses and design. Section 2 analyzes the results. Section 3 discusses some possible limitations of the design. Section 4 concludes.

2. THE EXPERIMENT: HYPOTHESES AND DESCRIPTION

2.1. Introduction and Hypotheses

In each session (typically) four subjects participated first to a betting and then to a redistribution stage. The betting stage was instrumental to the creation of an unequal wealth distribution. In addition, both during the betting stage and at the start of the redistribution stage, in half of the conditions (the "Non Desert," nD, condition) additional money was publicly given to some subjects according to some arbitrary criterion, discussed below. In the D ("Desert") condition prizes based on performance were assigned to make the wealth distribution roughly as unequal, or at least as an unequal, as in the nD conditions; this is motivated below.

At a fixed cost of 10% of one's own initial gains, the redistribution stage allowed both for redistribution (also to oneself in the "Stealing," S, condition) and for "burning" (elimination) of anyone's earnings. Practice took place before both stages, and a short questionnaire was administered in the end, before payment to subjects. Apart from the questionnaire, the experiment was fully computerized. Strict anonymity was preserved throughout, and the final decision was one-shot (so no issue of reputation was involved).

The experiment used a 2×2 factorial design crossing the arbitrary assignment of additional money (the Desert factor) with the possibility of stealing (the Stealing factor). However, it is also useful to consider whether a subject was "advantaged" (whether by prizes or arbitrary additional endowments) or not (A/nA). So there were eight possible combinations of Advantage (A/nA), Stealing (S if allowed, nS if not allowed) and Desert (D if Desert, nD if Non Desert). The experimental design enables to differentiate clearly between advantaged and disadvantaged subjects, and to state predictions about the effect of wealth as predictions about the effect of advantage.[2]

The experimental instructions are reproduced in the appendix: they are similar to those published in Zizzo and Oswald (2001). The differences between that

experiment and the one described in this paper concerned: (1) a different pricing system; (2) the constant presence of arbitrary assignments of additional money in the Zizzo and Oswald experiment (i.e. procedural fairness was not manipulated across conditions); (3) the possibility only of burning in that experiment.

We can now formulate hypotheses concerning the outcome of the redistribution stage;[3] a compact summary is presented in Table 1.

H0 *(Pure Self-Interest)*. Self-interested subjects should do nothing in the nS condition (since it is costly) and steal everything from everybody else in the S condition. Even if there are "trembles" out of these dominant strategies, they should be statistically of the same magnitude across conditions. The Desert manipulation should not matter.

We now analyze specific predictions of standard hypotheses on interdependent preferences, summarized under H2 through H5. A common prediction will then be described and contrasted to H6.

H1 *(Pure Envy)*. Since stealing brings about a greater relative advantage than burning, in the S conditions envious subjects should steal everything and burn nothing, exactly as for self-interested subjects. In the nS conditions, either subjects are not envious enough to incur the cost of burning or, if they are, they are best off burning everything of everybody else.

H2 *(Pure Altruism)*. Purely altruistic subjects should not steal and burn. They might redistribute some of their own gains, though this is costly more than one-to-one due to the fixed price of activity and, in addition, there is a free riding problem. Assuming that altruistic giving is a normal good, we would expect more redistribution from wealthy, typically advantaged subjects.

H3 *(Distributional Preferences)*. While the experimental design allows sharp predictions for pure altruism and envy, this is less so for distributional preferences such as inequality aversion. Here expectations are crucial, as subjects would like to tailor their activity on the basis of what others will do. We may be tempted to think that poor, typically disadvantaged subjects should engage in proportionally larger burning and stealing than the rich, typically advantaged subjects.[4] However, if advantaged subjects think that the others will make them poor, it is unclear that they should engage in less activity. Testable predictions exist:

(3.1) in Bolton and Ockenfels (2000), the agent cares only about preserving her relative standing, i.e. having an equal share of payoff herself – not, directly, on the other players' payoffs;

Table 1. Experimental Hypotheses.

Experimental Hypotheses	Some General Predictions	Desert		Rank Egalitarianism
		Main Effect	Interaction Effect	
H0: Pure self-interest	Steal all when feasible, 0 AR when not feasible; SB = 1	No	No	Not consistent[a]
H1: Pure envy	Steal all when feasible (SB = 1), otherwise burn all (SB = 0; this assumes fixed price low enough, else burn 0)	No	No	Not consistent[a]
H2: Pure altruism	Give, do not steal or burn; if giving is a normal good, wealthier subjects give more	No	No	Not consistent
H3: Distributional preferences				
Bolton and Ockenfels (1999)	May steal some when feasible	No	No	Not consistent[a]
Non-linear version of Fehr and Schmidt (1999)	May steal some when feasible	No	No	Predicted
H4: Warm glow or cold frisson	Give or not steal a fixed *k*% across condition (SB constant); never burn	No	No	Not consistent
H5: Distributions of preferences in population				
Andreoni and Miller (1998)	34–35% inequality averse, 43–44% self-interested (SB = 0.43–0.44), 10% fully altruistic	No	No	Not consistent[a]
Charness and Rabin (2002)	70% quasi-Leontief, 20% inequality averse, 10% envious (SB = 0–0.1)	No	No	Predicted for inequality averse subjects
Levine (1998)	About 70% self-interested or envious, should steal all when feasible (SB $\geq$ 0.7)	No	No	Predicted for quasi-Leontief and inequality averse subjects
Offerman et al. (1996)	65% self-interested, 27% altruists, 1% envious (SB = 0.65–0.66)	No	No	Not consistent[a]
H6: Categorization effects				
Desert (and Reciprocity)	May steal some when feasible, SB may be variable	Yes	Yes	No specific prediction

Note: SB: Self-interest Boundary.

[a] A "ceiling effect" bias may predict a spurious rank egalitarian correlation (wealthier people have more wealth to be stolen or burnt).

(3.2) in Charness and Rabin (2002) and a non-linear version of Fehr and Schmidt's (1999) inequality version model, agents care about the distribution of earnings among all players:[5] as a result, we would expect that subjects who are ranked higher in terms of their wealth are made poorer to a greater, or at least to a non-smaller, extent. In other words, we should expect *rank egalitarian* behavior.

H4 *(Warm Glow or Cold Frisson)*. In the S condition, the average amounts (of the sum of the other players' scores) left unstolen should be the same across (D/nD) conditions, and equal to some *k*%, out of a "warm glow" or "cold frisson" from not stealing. We would also expect that higher ranked subjects be not made object of more burning and stealing. We would obviously not expect any burning out of altruism, and, once again, the D manipulation should not matter.

H5 *(Distributions of interdependent preferences in the population)*. Different people may have different preferences: there may be preference heterogeneity in the population. It is easy to think of between-subjects combinations of the above factors allowing fitted to explain intermediate patterns in the data. Under H5 we catalogue a few of the distributions of interdependent preferences that have been estimated in the recent literature:

(5.1) Levine (1998) includes a role for intentions in determining how nice an agent is (the nicer you are, the nicer I am), and has a heterogeneous preference distribution that makes the model consistent with its data. According to his distribution, some 70% of the population is self-interested or envious, and even more (some 86%) behave as such. Being conservative, we expect stealing of everything (when possible) by at least 70–75% of the population. Furthermore, no non-trivial rank egalitarian behavior is predicted.[6]

(5.2) On the basis of an allocation task, Andreoni and Miller (1998) suggested a different type of distribution: they classified 34–35% of their sample as having a "Leontief utility function" (min [own, other's utility]). 43–44% was considered "selfish," and 21–22% purely altruistic with weight 1 on the other's utility. Non-trivial rank egalitarian behavior can be predicted in relation to the Leontief subjects.

(5.3) Charness and Rabin (2002) conjectured that about 70% of the population have preferences similar to "Leontief utility functions" (what they label "quasi-maximin preferences"), 20% are characterized by inequality aversion (their "difference aversion") and 10% are envious (their "competitive preferences"). It is an interesting feature of their distribution that there are no purely self-interested agents. Quasi-Leontief and inequality

averse subjects can be predicted to display non-trivial rank egalitarian behavior.

(5.4) Offerman et al. (1996) hypothesized still another, and rather different, distribution of types. According to their computations, 65% of the subjects are self-interested, 27% altruists, only 1% envious, and 6% choose at random. No non-trivial systematic rank egalitarian behavior is to be expected.

H2–H5 *(The Role of Desert)*. A prediction concerning standard hypotheses on interdependent preferences, such as H2 through H5, is that the Desert manipulation should not matter to them, or at most should act in the opposite direction to that of the concerns about procedural justice which will be discussed in the context of H6. This is because we made sure that in the D condition earnings were (at least) as unequal as in the nD condition. The point value for the variance in the nD condition was higher than that in the D condition, suggesting that, if anything, more H2–H5 redistribution activity should take place in the D condition relative to the nD condition. As we shall see, however, this is strongly falsified by the data. In practice, we have no reason to believe that inequality was significantly different between the D and nD condition: in an F test for the equality of the variances of the scores at the start of the redistribution stage in the D vs. the nD condition is insignificant ($F = 1.336$, $P > 0.1$). This being the case, a common prediction of H2 through H5 is that Desert should not matter, nor should it interact with other factors.[7]

H1 through H5 summarize various predictions from rational choice models. As such they differ from:

H6 *(Categorization Effects)*. When we talk about how agents perceive a decision problem, we are talking about how they *categorize* it. In general, categorization of X is how an agent represents X (Smith, 1995). In relation to interdependent behavior, Zizzo and Oswald (2001) distinguished three logical steps in the categorization process that produces interdependent behavior[8] (or otherwise). These are: (a) the *perception* of the decision problem, such as the definition of the material payoff structure or of one's own position in the game; an implication of this may be considered (b) the *priming* (i.e. activation) of one or more categories specifically relevant to address decision problems that may involve interdependent preferences (let us label them as "social categories" as a shortcut); the outcome will be (c) the *activation* of interdependent preferences and production of behavior.

In the case of our experiment, subjects may, because of the existence of the advantage, perceive the game differently according to the experimental condition. This different game perception implies that subjects may prime differently two

social categories, one based on desert and one on reciprocity. Desert may imply greater burning/stealing/redistribution activity in the nD than in the D condition; reciprocity might imply that, in the nD condition, A subjects may feel entitled to reciprocate and burn/steal more, out of the fear of being burnt/robbed more. Thus, H6 predicts that we should expect greater redistribution activity in nD compared to D conditions, and that interactions between different factors may be present.

Let SB (= "Self-interest Boundary") be the proportion of subjects doing nothing in the nS conditions (i.e. not sacrificing 10% of their own initial earnings), and refraining from stealing an amount equal to at least 10% of their own earnings in the S condition. We call this proportion "Self-interest Boundary," because it yields an upper bound to the proportion of subjects who are self-interested, even assuming some limited trembling. Some tight predictions can be made based on the values of SB. According to H0, $SB = 1$ across conditions. If H1 (pure envy), we should have $SB = 1$ in the S condition and $SB = 0$ in the nS condition. Levine's (1998) distribution would predict $SB \geq 0.7$ in the S condition, and Charness and Rabin (2002) $SB = 0.1$ (0) depending on whether stealing is allowed (or not). Furthermore, according to H1–H5, SB should not change according to the Desert factor or any interaction of this with other factors.

2.2. Detailed Description

The experiment was performed in Oxford between June 23 and July 1, 1998. 32 sessions of 4 subjects were planned, but (due to no show-ups and a computer breakdown at the start of a session) five sessions were run with three subjects.[9] Since three of these sessions were in a particular experimental condition (Desert and Stealing allowed), an extra session (with 4 subjects) was run in this condition. Therefore, the final number of sessions was 33, and the sample size was of 127 subjects. Subjects were mostly students. They could participate in one session only.

The experimental currency was the "doblon." Each doblon was convertible at the end of the experiment in U.K. pounds at the rate of 0.6 pence for doblon. Not considering the redistribution stage, where gains could only be reduced, the average gains were designed to be between 1000 and 1800 doblons (i.e. between 6 and 10.8 pounds). However, as we shall see below, in about half of the sessions the possibility of "stealing" (redistributing other people's money to oneself) provided a chance to increase one's own earnings substantially in the redistribution stage, by an average 22 pounds or more. Subjects got 3 pounds for participation, in addition to any other earnings. The overall experiment lasted 45 minutes on average.

Every effort was made to ensure anonymity among players. The possibility of two subjects knowing each other was minimised in a variety of ways (for example, undergraduates from the same Oxford college were not paired). Subjects were seated as soon as they arrived, and screens prevented view among them. A player number (1, 2, 3 or 4) corresponded to each seat, and seats were assigned according to the alphabetical order of the participants.

The experiment presented four stages: a practice stage, a betting stage, a redistribution stage (starting with further practice), and a payment stage. The wording of the instructions was neutral (words such as "burning" or "stealing" were not used). Instructions for each stage were provided at the start, and only at the start, of the stage (so, for example, subjects did not know the instructions for the redistribution stage until they had completed the betting stage).

2.2.1. Practice Stage

In each of the ten rounds of the practice stage, players received 100 doblons, and had to choose how much of the 100 doblons to bet (i.e. a number between 0 and 100). The computer then randomly generated a number between 1 and 3. If a 1 was drawn, subjects kept the original amount (100) and gained twice the amount they had bet. If a 2 or 3 were drawn, they lost the amount they had bet. The amounts gained in the practice stage did not count towards final actual gains.

2.2.2. Betting Stage

The betting stage was identical to the practice stage except for two things: (1) the scores of all players (labelled as 1, 2, 3 and, if any, 4) were displayed on each screen and updated at the end of each round; (2) in the nD condition, players 1 and 2 – chosen as such only because of alphabetical priority – were assigned (and could bet up to) 130 doblons each round rather than 100,[10] and this was common knowledge; in the D condition, subjects were told that the two top earners[11] at the end of the stage (e.g. after the 10th round) would gain a prize of 30% of their current earnings + 500 doblons.

2.2.3. Practice and Redistribution Stage

In the nD condition, players 1 and 2 were given an additional gift of 500 doblons at the start of the following stage. The kind of computer display faced by subjects is illustrated in Fig. 1.

Subjects were shown a grid displaying, from left to right: (a) red cells with the initial scores of all players, and the endowment each player had received (e.g. 1800 for advantaged, non deserving – A, nD – subjects); (b) green cells in which they could put numbers to eliminate earnings of any player; (c) blue cells in which they could put numbers to redistribute earnings from the player on the row of the grid

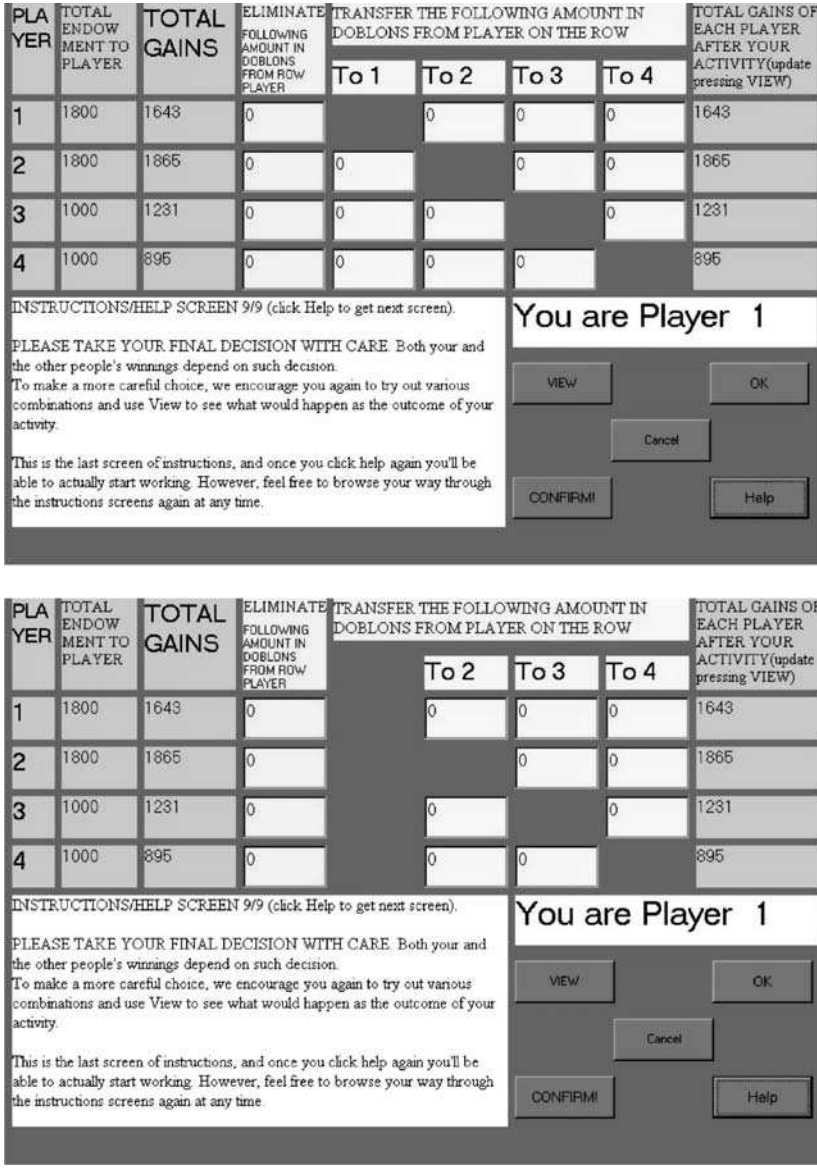


Fig. 1. Screens from the Redistribution Stage. *Note:* The computer displays on the top and the bottom of the page refer to the Stealing and the Non Stealing conditions, respectively.

to the player on the column of the grid (including to oneself in the S conditions); (d) red cells listing the scores of each player after any activity of the subject (but not that of the other subjects). A button called "View" was provided on the screen. By putting numbers in the various cells and clicking View, subjects could make practice. They could see column d updated with what would happen as the aggregate outcome of those numbers, without making any real decision. Subjects were actively encouraged (both in the written instructions and with a verbal reminder) to do practice (for at least ten minutes in the verbal reminder), by putting various combinations of numbers and clicking View, to get a grasp of what they could do. Most subjects appeared to follow the advice. When subjects were happy with their decisions, they could follow a step-by-step procedure to make their final decision.

When everyone had made their decisions for the redistribution stage, a computer calculated the gains of each subject as her initial gains plus the sum of the activities made by each player. If the final balance was negative, it was automatically increased to 0.[12]

Payment Stage. The final score of each player was displayed on her computer screen only. Subjects were asked to fill a short questionnaire that asked for basic questions such as the motivation behind their choices, and which indirectly verified the subjects' understanding of the experiment. They were then asked to sign a pledge of confidence on the content of the experiment plus a receipt, and were paid their earnings, if any, plus the 3 pounds for participation. Players were paid one at a time, in an order designed to ensure that a subject walking out of the room could not see or be seen by the others. They were asked to stay seated until paid. The mean payment in the experiment was 13.35 pounds; payments ranged between 3 and 37.11 pounds.

3. RESULTS

Table 2 contains a compact summary of the experimental results.

3.1. Evaluation of H0

Figure 2 displays the average proportion of redistribution made across conditions, as a fraction of the scores of each player.

Stealing is substantial when it is allowed, but always much lower than 100%. It is unlikely to be motivated only by self-interest, since, when we move from an S to a corresponding nS condition, burning regularly increases: the burning

Table 2. Experimental Results.

Experimental Hypotheses	General Fit with Evidence	Evidence on Desert
H0: Pure self-interest	Steal substantial but lower than 100%; burning an imperfect substitute for stealing, SB < 1 and variable	Unexplained
H1: Pure envy	Steal substantial but lower than 100%; when staling not allowed, only 2 subjects burnt all	Unexplained
H2: Pure altruism	Burning and stealing too high; wealthier subjects do not give more; rank egalitarianism not explained	Unexplained
H3: Distributional preferences		
Bolton and Ockenfels (1999)	Rank egalitarianism not explained	Unexplained
Non-linear version of Fehr and Schmidt (1999)	Rank egalitarianism predicted	Unexplained
H4: Warm glow or cold frisson	Burning and stealing substantial, and variable across conditions; rank egalitarianism not explained	Unexplained
H5: Distributions of preferences in population		
Andreoni and Miller (1998)	Rank egalitarianism predicted for 34–35% of the population; observed SB roughly compatible, though too much variable across conditions	Unexplained
Charness and Rabin (2002)	Rank egalitarianism predicted; observed SB is too high in 7 conditions out of 8	Unexplained
Levine (1998)	Rank egalitarianism not predicted; observed SB is too low	Unexplained
Offerman et al. (1996)	Burning and stealing too high	Unexplained
H6: Categorization effects		
Desert (and Reciprocity)	SB variable, interaction effect exists	Predicted

Note: SB: Self-interest Boundary.

ratio is only 8% when stealing is allowed, but jumps to an average 20.20% in the nS condition. Since burning appears an (imperfect) substitute for stealing, some stealing is likely to be motivated by negative interdependence. Figure 2 appears to show a lack of predictive power by the self-interest hypothesis. Figure 3 displays the SB across conditions and confirms this impression: in one condition the SB is about 46%, in six it is in the 10–25% range, in one it is equal to 0%.

However, in going to more formal statistical testing, we may be wary of data that include the answers of subjects who misunderstood the instructions. Perhaps,

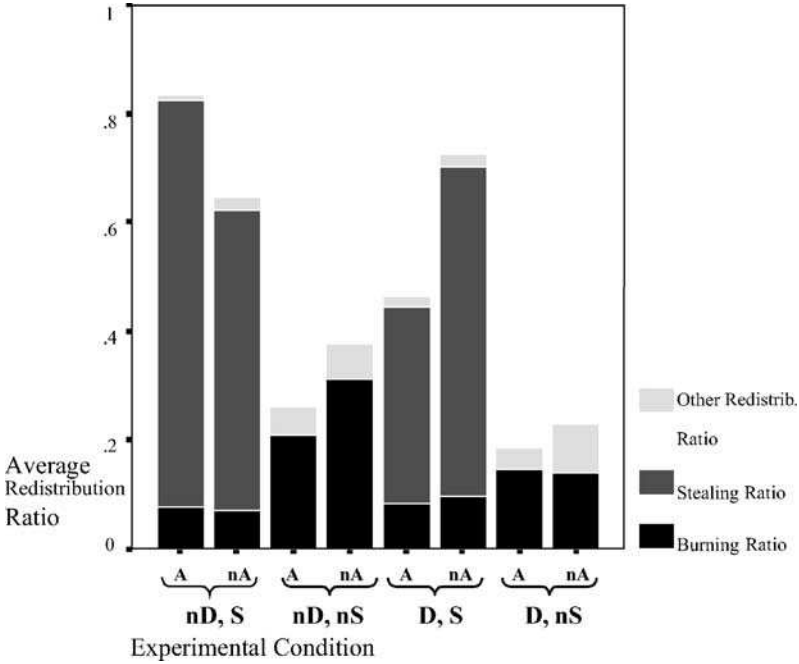


Fig. 2. Redistribution Activity for Each Experimental Condition, Full Sample. *Note:* Average Redistribution Ratio (ARR) for each experimental condition (full sample, $n = 127$). ARR is the sum of the burning, stealing and other redistribution ratio for each subject. The burning ratio is equal to the amount burnt by a player divided by the sum of the scores of the other players in the session, and similarly for the stealing and the other redistribution ratios. Experimental conditions: A, nD, S = Advantaged, Non Desert, Stealing; nA, nD, S = Not Advantaged, Non Desert, Stealing; A, nD, nS = Advantaged, Non Desert, Not Stealing; nA, nD, nS = Not Advantaged, Non Desert, Not Stealing; A, D, S = Advantaged, Desert, Stealing; nA, D, S = Not Advantaged, Desert, Stealing; A, D, nS = Advantaged, Desert, Not Stealing; nA, D, nS = Not Advantaged, Desert, Not Stealing.

eliminating them, the self-interest hypothesis can be rescued. We can use the questionnaires to weed out people whose answers show imperfect understanding of the game they were playing. This removes 19 subjects, so the final testing sample is $n = 108$. Figures 4 and 5 display the data for the restricted sample.

H0 fares no better under $n = 108$ than it does otherwise. A nonparametric sign test rejects the hypothesis that SB = 1 at $P < 0.0005$[13] ($Z = 8.307$). H1 also cannot explain the significance of the Desert factor, as it will be analysed below, or the SB variability shown by Figs 3 and 5.

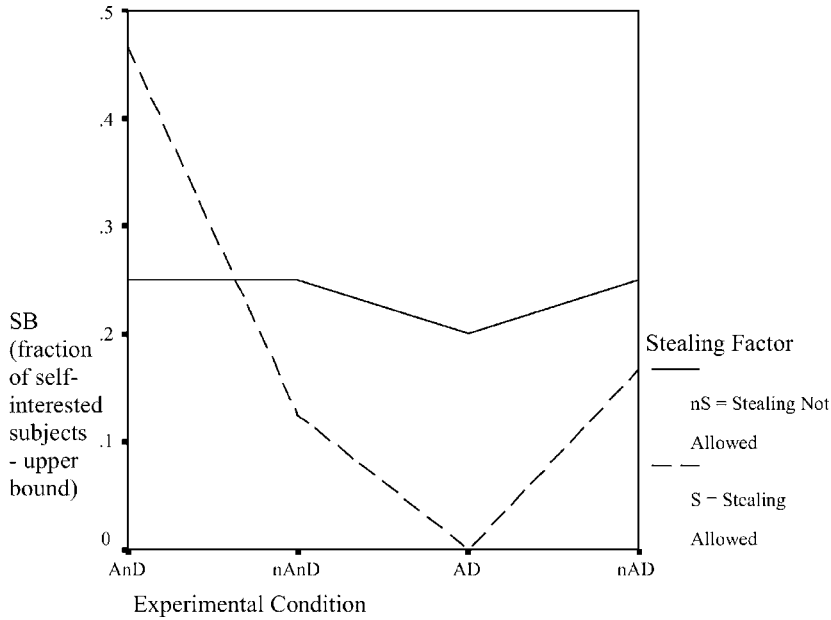


Fig. 3. Upper Bound to Fraction of Self-Interested Subjects in Each Condition; Full Sample. *Note:* Experimental conditions: A, nD = Advantaged, Non Desert; nA, nD = Not Advantaged, Non Desert; A, D = Advantaged, Desert; nA, D = Not Advantaged, Desert.

3.2. Evaluation of H1–H5 vs. H6

One of the most striking features from the figures is the difference in behavior between A, nD, S and A, D, S subjects. Advantaged non-deserving subjects stole some 75% of the gains on average, but advantaged deserving subjects only about 35%. Even more surprising, 45–50% of the subjects in the first group appear to be below the Self-Interest Boundary, but none of the latter is. Since the only difference between the two conditions is the Desert manipulation, this appears evidence against H1–H5.

Since the decision was one-shot, and we are concerned with the evaluation of rational choice models, it may be useful to eliminate the cases in which, from the questionnaires, it appears evident that subjects misunderstood the instructions. Again, this leads us to consider a sample of $n = 108$.

Define ARR as "aggregate redistribution ratio," i.e. the sum of any burning, stealing and other redistribution activity by the subject, divided by the sum of the scores of the other players. An F test on ARR using Desert, Stealing and Advantage

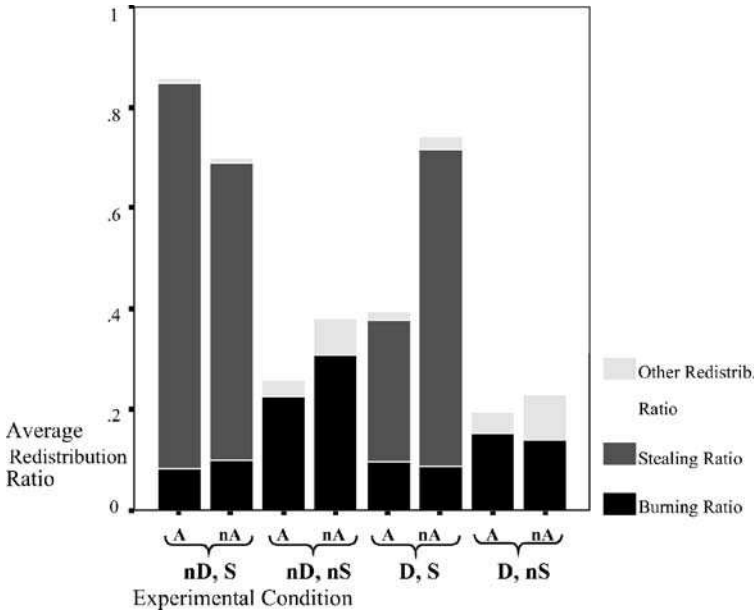


Fig. 4. Redistribution Activity for Each Experimental Condition, "Understanding" Sample. *Note:* Average Redistribution Ratio (ARR) in each experimental condition, for the sample of subjects whose answers in the final questionnaire do not show misunderstandings (Understanding Sample, $n = 108$).

as factors is significant at the 1% level (df $= 7$, $F = 8.901$, $P < 0.0005$). The Stealing factor is significant ($F = 45.6$, $P < 0.0005$), and so is Desert ($F = 5.999$, $P < 0.02$); the interaction term Stealing $\times$ Desert $\times$ Advantage is significant at the 5% level ($F = 4.538$, $P < 0.05$). No other term (including the main Advantage factor) is significant, not even at the 10% level. The significance of Desert and of Stealing $\times$ Desert $\times$ Advantage is replicated if an F test on aggregate redistribution (AR) in absolute terms is performed.

The usage of nonparametrics shows that the significance of Desert is not sensitive to the usage of parametric tests. As predicted by the categorization effects hypothesis H6, the Spearman correlation coefficient between Desert (= 1 in nD conditions, 0 otherwise) and ARR is significantly positive ($\rho = 0.16$, $P < 0.05$). The same nonparametric positive correlation between Desert and AR can be found ($\rho = 0.16$, $P < 0.05$). The results are also robust to the use of the full sample of $n = 127$ (with an F test, Stealing gives $F = 49.035$, $P < 0.0005$; Desert gives $F = 5.043$, $P < 0.05$; Stealing $\times$ Deserving $\times$ Advantage gives $F = 5.201$, $P < 0.05$).

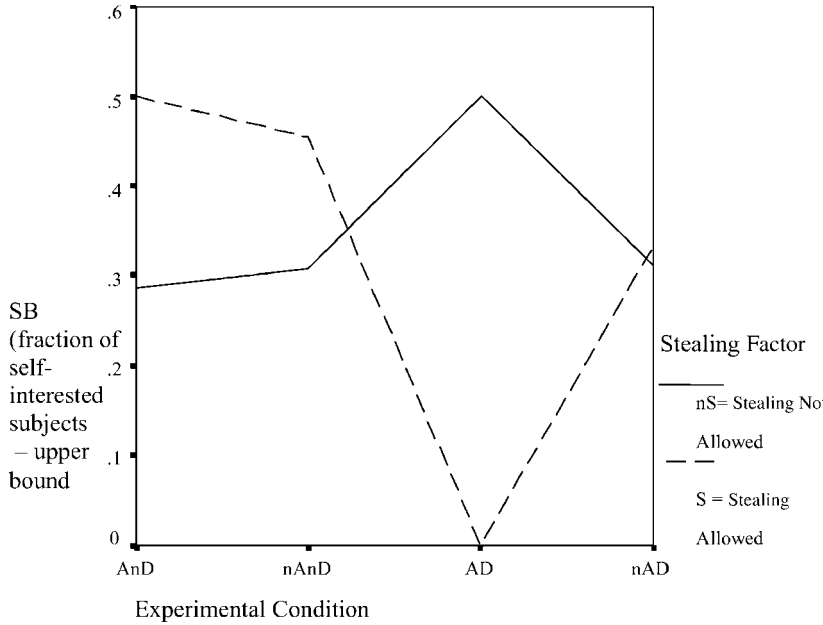


Fig. 5. Upper Bound to Fraction of Self-Interested Subjects in Each Condition; "Understanding" Sample.

In conclusion, the data support the prediction, made only by H6, that the procedure by which subjects earned money matters. It not only triggered a significantly higher aggregate redistribution, but A, nD, S subjects seemed to expect their position to be much more vulnerable than A, D, S subjects because of the unfair source of their advantage. We might conjecture that they reacted by feeling justified to reciprocate and "defend" themselves as much as possible by stealing much more in return.

There is further evidence running against predictions of specific models.

3.2.1. Pure Envy

H1 cannot explain why the stealing ratio is significantly below 1. The average amounts left unstolen per subject varied from 5.12 pounds in the A, nD, S condition to 12.53 in the A, D, S condition. These are obviously large amounts relative to the scale of experimental gains. Moreover, the nS condition prediction that, according to the degree of envy, either a subject should burn nothing or should burn everything is not supported by the data: only 2 out of 32 subjects who did something in the nS condition burnt everything out of everybody else.

3.2.2. Distributional Preferences

We found that the Advantage factor is insignificant in an F test on the aggregate redistribution ratio. Figures 2 or 4 show that such insignificance is not driven by greater gifts by advantaged subjects: such gifts should appear as "other redistribution," and the "other redistribution" by nA subjects is often greater, not smaller, than that provided by A subjects. Nevertheless, as we discussed earlier, it is possible that expectations about others' behavior drives the lack of significance of the Advantage factor, even with inequality averse or Leontief utility subjects.

We can test Bolton and Ockenfels' (2000) claim that subjects care only about their own relative share of the cake, and not on how gains are distributed among the other subjects. Consider a player deciding whether and how much to change the earnings of the other players. The other players can be ranked according to their gains: assign 1 to the top ranked, 2 to the second ranked and 3 to the third ranked, if any.[14] Call this variable Orank. Now let Positive equal 1 if the player increases the gains of another player, and 0 otherwise. According to Bolton and Ockenfels (2000), we should expect no correlation between Orank and Positive. If the subject feels she ought to give some of her share (which may happen if she has a greater share than equitable relative to the size of the cake as a whole), she is indifferent to whom to give. However, if one considers the sample of everyone who made some change in the gains of the other players, there is a significant positive correlation between Orank and Positive ($n = 303$; Spearman's $\rho = 0.335$; $P < 0.0005$).

Now assign 1 to the person whose gains are reduced most by a player, among the other subjects; 2 to the second most reduced; 3 to the least reduced:[15] call this variable Ochange. If Bolton and Ockenfels are right, we may expect a positive correlation between Orank and Ochange because, if players choose randomly how to divide their optimal amount of changes, they might still reduce the amounts of the richer players more on average. Moreover, if they can steal and want to steal a lot, they may be forced to steal more from the rich people, anyway. For example, in the limit (self-interest or pure envy) case of someone stealing everything from everybody, there would be a perfect correspondence between Orank and Ochange, since the richest gets stolen most, the second richest second most, and the poorest the least. This "ceiling effect" may bias the results when stealing is allowed.

In trying to assess whether the correlation between Orank and Ochange was spurious, we drew numbers randomly from a uniform distribution, multiplied them by the score of each player faced by the decision-maker, and then computed a fictional Ochange (call it Ocarlo) based on the Monte Carlo simulation. This procedure was followed 30 times. As expected from the first bias discussed before, there was a significant positive correlation between Orank and Ocarlo: the mean Spearman

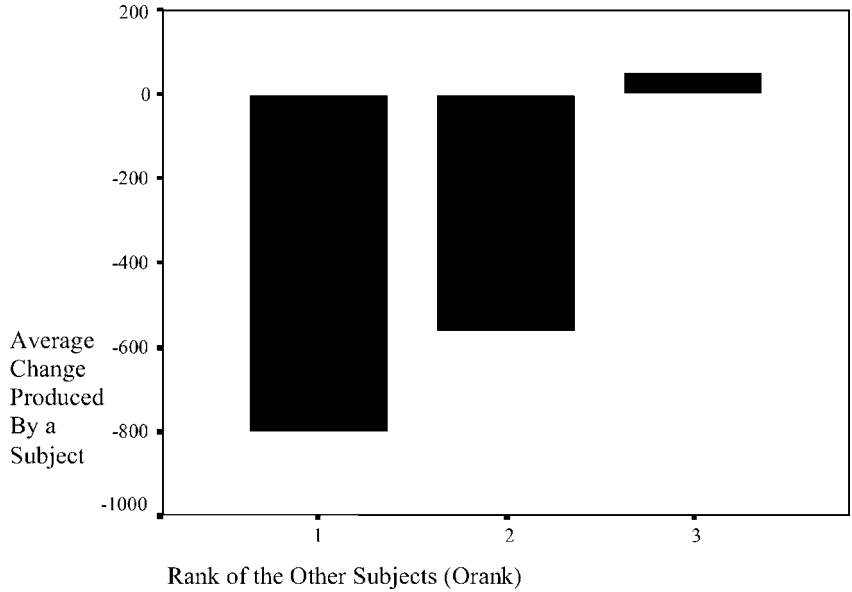


Fig. 6. Score Changes Produced by a Subject in the Scores of the Other Players, According to their Rank, Non Stealing Condition Only. *Note:* Average change produced by a subject on the scores of the other subjects, according to their rank, in the non Stealing condition. Orank assigns a value of 1 to the top ranked among the other subjects, 2 to the second ranked and 3 to the bottom ranked.

correlation was $\rho = 0.38$ (min $= 0.24$; max $= 0.49$; S.D. $= 0.066$). However, this is significantly less than the correlation that we actually find in the data ($\rho = 0.806$).

Even looking at the conditions where stealing was not allowed (see Fig. 6), and so eliminating the "ceiling effect" bias,[16] the correlation was still 0.695 and so significantly different from the Monte Carlo distribution correlation (in a *t*-test, $t = 16.265$, df $= 29$, $P < 0.0005$; in a nonparametric sign test, $Z = 5.295$, $P < 0.0005$). Therefore, a correlation appears to exist between rank and activity of which players are object: subjects seem to care that specifically richer subjects are hit more by their activity. The Bolton and Ockenfels' model fails to take this into account.

Another way to look at the relationship between rank and redistribution is to consider the number of people who satisfy what in Section 2.1 we called a *rank egalitarian* relationship. We consider a subject as satisfying a rank egalitarian relationship if she reduces the score of the richest of the other subjects at least as much as or more than that of the second richest, and that of the second richest at least as much or more than that of the poorest subject. If we just look at the non

Stealing condition, to minimize the "ceiling effect" bias, we find that 83.33% of the subjects who engaged in any activity behave in a rank egalitarian way.[17]

3.2.3. Pure Altruism, Warm Glow, Cold Frisson

Wealthier subjects do not give more. More seriously, there is significant burning, particularly in the nS condition (where on average 20.20% of earnings were burnt): this cannot be explained either by pure altruism or a warm glow. Moreover, the amounts left unstolen show substantial variability across A/nA and D/nD conditions, ranging from $k = 25\%$ of the cake for A, nD subjects, to 42–45% for nA subjects, to 64% for A, D subjects: this runs against the hypotheses of a warm glow or cold frisson. So does our finding of a strong correlation between rank and being victim of redistribution activity.

3.2.4. Distributions of Preferences

Combinations of the above models could be used to try to fit the data better, but they would still be unable to explain the relevance of desert and its interaction with the other factors. There is specific evidence against Levine's (1998) distribution, and Figs 3 and 5 show why: the prediction that 70% (or more) of the subjects would steal everything does not hold. SB is significantly below 0.7, as a binomial test points out (with the full sample, SB = 34.65%; with the "understanding" sample $n = 108$, average SB = 34.26%; either way, Prob(SB = 0.7) < 0.0005). At the same time, though, as the figures show, in seven conditions out of eight SB is above the 0–0.1 range, thus failing to provide support for Charness and Rabin's (2002) suggested distribution.

Offerman et al.'s (1996) distribution of types is also rejected. Giving is much less common than burning or stealing, and the prediction that 65% of the subjects are self-interested fails a binomial test, once again at $P < 0.0005$. Only 1–7% of the subjects should burn in the non Stealing condition, but 49.21% did. In the Stealing condition, no more than 73% should have stolen – we would not expect the altruistic subjects to steal – but 95.31% of the subjects actually stole something. Binomial tests easily show the significance of these differences ($P < 0.0005$). Rank egalitarianism is also not predicted by Offerman et al. (1996) or by Levine (1998).

Andreoni and Miller's (1998) distribution would appear the least off the mark. It allows for a fraction of inequality-averse subjects, and so successfully predicts the aggregate tendency for rank egalitarianism in our data. Their aggregate average SB (43.66%) is not too distant from the value we found, considering that, due to the different structures of our experiments, we employ different criteria to fix the exact boundaries of what to consider self-interested: in a binomial test, equality between our and their SB proportion is rejected at "only" the 5% significance level (e.g. $P = 0.026$ with $n = 127$). Andreoni and Miller's (1998)

model can predict a genuine correlation between rank and having one own score being stolen and burnt, as discussed earlier, at least for about 35% of the subjects. Nevertheless, the Andreoni and Miller's distribution cannot explain the variability in the SB distribution displayed by Figs 3 and 5. The significance of desert and its apparent effect on the expectations formulated by the advantaged subjects in the Stealing condition cannot be accommodated in this framework, at least as long as it conceives desert without paying due consideration to procedural fairness concerns.

In conclusion, models consistent with rank egalitarianism, and Andreoni and Miller (1998) among the type distributions of preferences, present the most adequate fit of our data in a rational choice framework. Nevertheless, all the rational choice models of interdependent preferences we discussed cannot explain certain features in our data. Moreover, Levine's (1998) and Charness and Rabin's (2002) attention to intentions might not be misplaced: what made advantaged subjects in the Stealing condition behave differently according to the source of their advantage might have been what they thought that their disadvantaged counterparts would have done. Expectations were sensitive to the way the problem was perceived.

Parsimonious explanations exist for rank egalitarianism, and we cannot exclude that extensions of the distributional preferences framework to allow for cognitive processing may go a long way explaining other features of the data (e.g. Konow, 2000). All we can say is that consideration of outcomes alone is not enough, and that categorization effects, such as those entailed by the perception of procedural fairness, affected behavior.

3.2.5. Economics Training

It is known from public goods experiments that economists tend to make marginally worse citizens, by contributing less (e.g. Frank et al., 1993). In my experiment, training in economics or game theory was not significantly correlated to aggregate redistribution. However, there was a significant positive correlation between such training and the stealing ratio (Spearman $\rho = 0.191$, $P < 0.02$). Since economics training affects how a subject perceives the decision problem (e.g. modifying the expectations on how the other players will behave), this also possibly reflects a categorization effect.

4. LIMITATIONS

The study presented in this paper has two main limitations. First, the final decision was not repeated many times, and so an opportunity to learn a "more rational"

response was not provided. Many economists would consider rational choice predictions in the short run slightly beside the point. However, this is not the case in general: Andreoni and Miller's (1998) experiment is also static and without repetition or feedback on the given task, and yet the authors stress the conformity of their findings with rational choice. Unlike our experiment, they do not even have a practice stage.

In our experiment repetition is more difficult to implement than in standard bargaining experiments, because of the larger sample size and the importance of having a new wealth distribution at the start of every redistribution stage. Hence, since the design is new, I decided to start from the simplest experimental design possible to avoid reputation effects: a one-shot decision. Undoubtedly, further research must look into repetition.

Nevertheless, it is unclear that repetition would necessarily eliminate an explanatory role for categorization effects (see Cookson, 2000). Moreover, the existence of a practice stage, the statistical analysis with the reduced sample of "surely understanding" subjects and a manipulation made in the money burning experiment described in Zizzo and Oswald (2001) all ensure that the results are not a by-product of misunderstanding of the instructions. In about half of the sessions of the Zizzo and Oswald experiment, we added verbal instructions stressing that any activity was costly and that the decision to be taken (in the Zizzo and Oswald equivalent of the redistribution stage) was the only one and final. We also tried individually to explain subjects exactly what they were doing, whenever they wanted to go on to the final decision, in order to check their full understanding of the consequences of their actions. We found that this "understanding-checking" manipulation was always insignificant.

One might also argue that the study of decisions in the short run is a better mirror of many economic decisions than providing intensive learning incentives across ten or one hundred rounds, which may be unlikely in the real world in many cases. Therefore, how subjects assimilate a decision problem to more familiar ones can be of independent interest. In addition, we found patterns in the data (such as rank egalitarianism or the role of desert) that cannot be explained by random behavior alone: this suggests that at least some subjects took the experiment seriously.

The second limitation of the design is that, due to the role of expectations, we could not test distributional preferences theories as strictly as pure altruism or envy. Zizzo (2003) addressed this concern by having an experiment that was very similar to Zizzo and Oswald's (2001) money burning design, but where the decisions of only one player chosen randomly was implemented, after everyone had made their decisions. Rank egalitarianism carried over when this "random dictator" design was used. But obviously additional research is required: for example, expectations may have mediated the impact of desert on behavior.

Other objections to the experimental design are less serious. For example, one might object that 10% of one's own earnings might be too small a price (typically some 1 or 2 U.K. pounds) to be taken seriously. However, this observation neglects that, in the Stealing condition, the amounts left unstolen were much larger.

That some sessions were with only three subjects is an obvious limitation forced upon us by necessity rather than by desire; since it affected one specific condition particularly (Desert and Stealing allowed), omitting these sessions is not an option as it leaves us too few observations in this condition to get any degree of statistical power. However, variables such as the ARR or the burning or stealing ratio prove to be insignificantly different in sessions with three subjects relative to sessions with four subjects. This suggests that a small difference in the number of subjects did not seriously affect behavior. For example, the ARR is 76.69 (64.60) with three (four) subjects ($t = 1.058$, df $= 62$, $P > 0.1$). Obviously, further work may shed additional light on this.

Another only marginally relevant objection is that the "prize race" of the D conditions may have induced a "competitive frame" that carried out in the redistribution stage, at least for some subjects. Since this bias is in the direction of understating the impact of desert, its elimination may only strengthen the results of this paper. Similarly, the objection that the inducement of desert was too weak in my experiment, based as it was on gains of chance relative to arbitrary distributions, only strengthens the conclusions of this paper: it shows that procedural fairness concerns mattered even if introduced in this minimal way.

Finally, we were forced to increase negative final balances to zero, for obvious ethical and practical constraints – we could not ask subjects to pay us money. This implied, for example, that if two subjects B and C stole all of subject A's total gains, each of them would get the whole of it, and A would simply get 0. It is conceivable that subjects could collude by stealing everything out of everyone else and getting a Pareto superior outcome as a result. In practice, however, this did not happen: only one subject went bankrupt and had her score raised in the Stealing condition.

5. CONCLUSIONS

This paper presents a new experimental test on preferences towards wealth distributions and procedural fairness. I construct a laboratory experiment in which subjects earn money by betting and, in about half of the sessions, by receiving undeservedly assigned gifts. Subjects are then told the experiment is finishing and offered a last decision. They are anonymously allowed to eliminate, redistribute

and, in about half of the sessions, steal other subjects' money. To do this, they have to pay a price (that is, to give up some of their own cash).

A large fraction of subjects was not purely self-interested. Significant amounts of money were left unstolen. When stealing was not allowed, about 20% of the earnings were burnt. Over 80% of the subjects engaging in any activity were rank egalitarian. There was a strong correlation between wealth, or rank, and the amounts by which subjects were burnt: a majority of the subjects was *rank egalitarian*. Bolton and Ockenfels' (2000) model does not predict rank egalitarianism, and is therefore not consistent with our data. Instead, other theories of distributional preferences such as a non-linear version of Fehr and Schmidt (1999) can explain rank egalitarianism.

Apart from testing among different theories of distributional preferences, the experiment allowed a new test of other theories and distributions of interdependent preferences: specific predictions made by pure and impure models of altruism and envy were rejected. We considered the distributions of preference types estimated by Andreoni and Miller (1998), Levine (1998), Offerman et al. (1996) and Charness and Rabin (2002), and found the greatest support for Andreoni and Miller's.

However, the rational choice models considered in this paper cannot easily explain why perceptions of procedural fairness affected behavior, in such a way that the fraction of subjects that can be classified as self-interested varied dramatically across conditions, from 0% to almost 50%. Perceptions of desert mattered, even if introduced in an arguably minimal way, and undeservedly advantaged subjects may have engaged in defensive stealing when this was allowed.

NOTES

1. Note that there is no contradiction between saying that Charness and Rabin's (2002) *estimation* of preference types is rejected and stating that their *model* has some qualified support. Both are in their paper, and are related to one another, but they are not the same thing.

2. The two are very highly correlated ($r = 0.841$), and so it is anyway unfeasible to disentangle the effects of the two with the present sample size. Readers who prefer to reason directly in terms of wealth rather than advantage should feel free to think just in terms of the former: results do not qualitatively change if one replaces one variable with the other when it is possible to do so.

3. The data from the betting stage are analyzed in Zizzo (2001).

4. There are two reasons for this. First, utility depends also on material payoff, so a subject will be less aggressive against herself than against the other subjects: since there are either one or two advantaged subjects per session, it matters significantly whether one

of the two is you, in making your redistribution decision. Second, rich, advantaged subjects may expect their scores to be reduced – perhaps stolen – by the other (say) inequality averse players, and not bother about reducing them themselves – while the poor, disadvantaged subjects do not have this problem.

5. Charness and Rabin (2002) also has a role for intentions and so is not a pure distributional preferences model, but nevertheless does have a role for distributional preferences other things being equal. The basic Fehr and Schmidt (1999, Sections 2–6) model has linear inequality aversion terms, which make irrelevant how wealth is distributed among the other players. However, in the extended version of their model with concave inequality aversion terms (briefly discussed in their Section 7), large deviations from inequality bring greater disutility than small deviations, and so the distribution of wealth among other players matters.

6. A reason for "trivial" rank egalitarian behavior is discussed and evaluated in Section 3: wealthier people may be stolen or burnt more, and this may produce a spurious rank egalitarian relationship.

7. One might reply that it should according to economists, such as Konow (1996), who embrace inequality aversion while at the same time allowing a role for desert. Neverheless, by stressing a key role for the way the problem is perceived (e.g. the defensibility of trying to get a more equitable or inequitable outcome in Konow, 2000), this kind of work is better subsumed under H6 below than under H5. Charness and Rabin (2002) do talk of deservingness, but their notion of deservingness refers to intentions not perceived entitlements.

8. The neurotransmitter serotonin may be involved in this process (Zizzo, 2002).

9. Whether this may have affected results is discussed in Section 3.

10. Only player 1 in sessions with only three subjects.

11. Only the top earner in sessions with three subjects.

12. This may be treated as a potential source of distortions, but in practice it was not, as only one subject went bankrupt (see Section 3).

13. This result is robust to the usage of the full sample ($n = 127$).

14. In case of ties between first and second place, a value of 2 was assigned; in case of tie between second and third, a value of 3.

15. Ties were treated as for Orank (see previous Note).

16. A "ceiling effect" bias would remain in relation to envious subjects who were to burn everything of everyone else. As we mentioned earlier, however, virtually no one did.

17. The fraction increases to 90.48% in the Stealing conditions.

ACKNOWLEDGMENTS

Thanks to the late Michael Bacharach, to Neil Buckley, James Konow, Matthew Mulford, Andrew Oswald, two anonymous referees and the participants to the Economic Science Association meeting held in New Orleans (March 1999) for useful advice, and to Liz Beaumont-Bissell for experimental assistance. Related work was presented in Oxford, Paris and Royal Holloway, enabling me to receive additional helpful feedback. The usual disclaimer applies. Financial

support from the George Webb Medley Fund, University of Oxford, is gratefully acknowledged.

REFERENCES

Andreoni, J., & Miller, J. H. (1998). Analyzing choice with revealed preference: Is altruism rational? In: C. Plott & V. L. Smith (Eds), *The Handbook of Experimental Economics Results*.

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

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics, 117*, 817–869.

Cookson, R. (2000). Framing effects in public goods experiments. *Experimental Economics, 3*, 55–79.

Eckel, C. E., & Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and Economic Behavior, 16*, 181–191.

Elliott, C. S., Hayward, D. M., & Canon, S. (1998). Institutional framing: Some experimental evidence. *Journal of Economic Behavior and Organization, 35*, 455–464.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*, 817–868.

Frank, R. H. (1985). *Choosing the right pond: Human behavior and the quest for status*. New York and Oxford: Oxford University Press.

Frank, R. H. (1997). The frame of reference as a public good. *Economic Journal, 107*, 1832–1847.

Frank, R. H., Gilovich, T. D., & Regan, D. T. (1993). Does studying economics inhibit cooperation? *Journal of Economic Perspectives, 7*, 159–171.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior, 7*, 346–380.

Konow, J. (1996). A positive theory of economic fairness. *Journal of Economic Behavior and Organization, 31*, 13–35.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review, 90*, 1072–1091.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics, 1*, 593–622.

Offerman, T., Somnermans, J., & Schram, A. (1996). Value orientations, expectations and voluntary contributions in public goods. *Economic Journal, 106*, 817–845.

Oswald, A. J. (1997). Happiness and economic performance. *Economic Journal, 107*, 1815–1831.

Smith, E. E. (1995). Concepts and categorization. In: E. E. Smith & D. N. Osherson (Eds), *Thinking: An Invitation to Cognitive Science* (Vol. 3, pp. 3–33). Cambridge, MA and London: MIT Press.

Todd, L. C. (2001). Mental accounting and other-regarding behavior: Evidence from the lab. *Journal of Economic Psychology, 22*, 605–615.

Zizzo, D. J. (2001). Situational determinants of risk-taking behavior in a lottery race game. *Greek Economic Review, 21*, 37–51.

Zizzo, D. J. (2002). Between utility and cognition: The neurobiology of relative position. *Journal of Economic Behavior and Organization, 48*, 71–91.

Zizzo, D. J. (2003). Money burning and rank egalitarianism with random dictators. *Economics Letters*, *81*, 263–266.
Zizzo, D. J., & Oswald, A. J. (2001). Are people willing to pay to reduce others' incomes? *Annales d'Economie et de Statistique*, *63–64*, 39–62.

APPENDIX: EXPERIMENTAL INSTRUCTIONS

There were 4 versions of instructions, according to the experimental condition. Condition 1: Non Desert, Stealing allowed. Condition 2: Non Desert, Non Stealing. Condition 3: Desert, Stealing. Condition 4: Desert, Non Stealing. Small changes had to be made in the three sessions with only three subjects.

Stage 1 Instructions

In this experiment you will use the computer to read information and make decisions.

Typically you will be asked to enter a number in one or more cells – such as that on the bottom-left corner of this screen – and to click some buttons. To input or change numbers, click the mouse pointer in the cell. You will then be able to type or erase numbers in the cell using the keyboard. Please always remember to type numbers as digits (say, 50) rather than as letters (say, fifty). You can give commands to the computer by clicking on the grey buttons at the appropriate times. Examples on the current screen are OK, Confirm, Cancel and Help. Note that only Help is currently highlighted, meaning that you can only click on Help right now (but please wait until you have read these instructions!). To press a button, click on it with the mouse pointer. Always click on Help to pass to the next screen of instructions.

IMPORTANT: please do NOT try to exit the experiment program even temporarily. Do NOT tamper with the computer in any other way (such as turning it off or removing the floppy disk). On various occasions you will be asked to click a button to check whether the other players have made their choices and the computer has made the necessary computations. Please, do NOT click the button continuously. Wait at least 10 seconds between attempts. You are NOT allowed to speak to any other participant in the experiment at any time. Further, if you need to speak to the experimenter, you should do quietly. If you have a query which the instructions are unable to solve, please raise your hand and we'll do our best to solve it – either on a piece of paper or with a low voice.

The above rules are essential for a smooth and speedy completion of the experiment. If you violate them, you may force everyone to lose much additional time, and you may be asked to leave the room and lose ALL gains AND the participation token. Thanks a lot!!!

The experiment is divided into four stages. The first stage is for practice. The second and third are the real experiment. The fourth stage is for the payment. We are going to use an experimental currency, the doblon. Your final doblon gains (except those of the practice stage) will be converted into U.K. pounds in the payment stage, at the rate of 0.6 pence per doblon. Unlike those earned later in the experiment, the doblons earned in the practice stage will NOT count towards your final gains and will NOT be convertable for money – the practice stage is only for practice, not to let you earn money! However, the doblons gained in the real experiment (stages 2 and 3) and which you still have by the end of stage will be converted into U.K. pounds in the payment stage. During the experiment your gains may go down as well as up. However, no player's balance will ever be allowed to fall below zero.

Moreover, whatever your final doblon gains from stage 2 to 3, you will be given an additional payment of 3 pounds for participation in stage 4.

Welcome to the Practice Stage!

There are 10 rounds. Each round you receive 100 doblons for practice and you can choose to bet any amount of them, i.e. you can choose to bet between 0 and 100 doblons each round. Please write your choice in the left-down box of this screen.

To go ahead with your choice, press the OK button of the main screen and then Confirm. If you are not sure about your choice, even after having pressed OK, but before having pressed Confirm, press Cancel. After having pressed OK and Confirm, the computer randomly generates a number between 1 and 3. If you get 2 or 3, you lose the money you bet. If you get a 1, you win: you keep the original amount of money you bet and gain double the amount (for ex., if you bet 100, you get 00 overall).

Example 1: Jill receives 100 doblons. She bets 50 doblons. Assume she wins. Then she retains the 50 doblons she bet (50), plus the money she did not bet (50), plus she earns $2 \times 50 = 100$ doblons more. So she earns a total of 200 doblons from the round. Now assume she loses. Then she is left with only the money she did not bet, that is with 50 doblons.

Example 2: Jamie receives 100 doblons. He bets 0 doblons. He wins 2×0 if a 3 is drawn, and loses 0 otherwise, so, whatever the number, he is left with 100 doblons.

Jane receives 100 doblons. She bets all of them. She wins 2×100 if a 3 is drawn, and loses 100 otherwise. So her overall winning from the round is 300 if she wins, and 0 otherwise.

Click Help to make this screen disappear and the first round start. Click Help another time to make the instructions appear again. Note: while these instructions are in view, you won't be able to take decisions.

Stage 2 Instructions

Welcome to Stage 2 of the Experiment!!!

In this stage you will play bets for real money, and this is why your score is "restarting" from zero.

Non Desert Conditions Only

Players have been assigned a number according to the alphabetical order of their last names. Players 1 and 2 get 130 doblons each round. Players 3 and 4 get 100 doblons each round. Each round you can bet from 0 up to the amount you receive each round (100 or 130). Put the number of doblons you are betting in the box in the bottom-left corner of the screen.

All players are given 100 doblons each round. Each round you can bet from 0 up to the amount you receive each round (100). Put the number of doblons you are betting in the box in the bottom-left corner of the screen.

Desert Conditions Only

The two players who at the end of all ten rounds will have the highest overall winnings, will get a prize equal to 30% of their earnings plus an additional 500 doblons. [If two (or more) players are tied for one prize, who gets the prize between them will be decided entirely randomly.]

To go ahead with your choice, press the OK button and then Confirm. If you are not sure about your choice, even after having pressed OK, but before having pressed Confirm, press Cancel. You can NOT change your choice for the round after having pressed BOTH OK AND Confirm.

After having pressed OK and Confirm, the computer randomly generates a number between 1 and 3. If a 1 is drawn, you win: you keep the money you bet and earn double the amount. If you get 2 or 3, you lose the money you bet.

To pass to the next screen, press the Help button.

There are ten rounds. After having pressed Confirm, and before passing to the following round, the computer will check whether the other players have made

their choices. Once everybody has made her choice, the updated winnings of each player will appear on the screen.

Example: Jill receives 100 doblons. She bets 50 doblons and wins. Therefore she retains the 50 doblons she bet (50), plus the money she did not bet (50), plus she earns $2 \times 50 = 100$ doblons more. So she earns a total of 200 doblons from the round. Now assume she loses. Then she is left with only the money she did not bet, that is with 50 doblons.

Non Desert Conditions Only

In the meanwhile, Jamie receives 130 doblons. He bets 0 doblons. He wins 2×0 if a 1 is drawn, and loses 0 otherwise, so, whatever the number, he is left with 130 doblons.

Jane receives 130 doblons. She bets all of them. She wins 2×130 if a 1 is drawn, and loses 130 otherwise. So her overall winning from the round is 390 if she wins, and 0 otherwise.

Assume that Jill wins and Jane loses. Then, before passing to the following screen, on Jane's screen the new amounts, identified by number, of the other players will appear. For example, if Jamie is Player 1, it will appear that Jamie got 130 doblons more by the end of the round.

Desert Conditions Only

In the meanwhile, Jay made the same bet but lost, so is left with 50 doblons; Jamie bet 0 doblons and so retains his 100 doblons; Jane bets 100 doblons and loses, so she is left with 0 doblons.

Assume now that after the 10 rounds of play, Jill has 1200 doblons, Jamie 1050, Jane 950 and Jay 800. Then Jill wins a further prize equal to the 30% of 1200 (i.e. 360) plus 500 doblons – a total of 860 doblons -, while Jamie gets a prize of 815 doblons.

Click Help to make this screen disappear; a small label reminding your income per round will appear and you'll be able to start. Click Help again to make the instructions appear again. Note: while these instructions are in view, you won't be able to take decisions.

Stage 3 Instructions

Non Stealing Conditions Only

In this stage, you are allowed to eliminate part or all of the winnings of any player – yourself included -, and/or to transfer part or all of them from any player (again, yourself included) to any but NOT to yourself.

Stealing Conditions Only
In this stage, you are allowed to eliminate part or all of the winnings of any player (yourself included), and/or to transfer part or all of them between players (again, yourself included).

Non Desert Conditions Only
Players 1 and 2 get a GIFT of 500 doblons. Our compliments to players 1 and 2. Players 3 and 4 don't get any gift.

To do any activity of elimination or transfer of winnings, you have to pay a price equal to 10% of your total gains.

Non Desert Conditions Only
The total gains are the gains a player had until now, from income we gave her (including gifts) and from winnings.

Desert Conditions Only
The total gains are the gains a player had until now, from income we gave her and from winnings.

Total gains do NOT include the participation token. In other words, the price of elimination and transfer is NOT proportional to the sum of total gains + participation token, but only to total gains. Further, the participation token can NOT be subject to any elimination or transferring activity.

Each row represents a player – the one in the first column from the left. The second column from the left specifies the total amount of doblons we gave each player (=total endowment to the player) in stage 2 and 3.

Non Desert Conditions Only
It includes the 1000 or 1300 doblons each player received in stage 2 – in 10 rounds of 100 or 130 doblons each, plus, if any, the 500 doblons gift previously discussed.

Desert Conditions Only
It includes the 1000 doblons each player received in stage 2. It does not consider winnings dependent on betting choices and outcomes.

The third column from the left has the total gains of the corresponding row player. It may be higher or lower than the endowment, according to the stage 2 performance. The first column from the right displays the total gains after your activity. To update this column, press View (it is also updated automatically when you press OK). All these columns have a RED background. You cannot put any number yourself in any red cell.

You can plug and change numbers in the GREEN and SKY-BLUE cells. To eliminate gains, put the number of doblons gained by a player (and that you want to eliminate) in the green cell of the corresponding row.

To transfer earnings from a player to another, put the number of doblons you want to transfer in the cell which is in the row of the first player and in the column of the other player. (*This grid was made of sky-blue cells.*) You cannot at any time reduce the total gains of any player after your activity to below zero.

Within such limit, once you pay the fixed price for engaging in eliminating and/or transferring activity, you can engage in any amount of elimination and/or transfer you wish, as long as you can pay the price.

Before taking a final decision, you are encouraged to spend some time plugging numbers in the cells and viewing the outcome by pressing View, just to get a better understanding of how things work out.

Once you are happy with your choices, press OK and then Confirm. Press Cancel after OK if you change your mind. Once you press Confirm, you can NOT change your mind anymore.

IMPORTANT: all players have these same instructions in front of them right now.

The final gains of each player are determined as the SUM of the activity of elimination and transfer of winnings made by ALL players. However, if such final gains are below zero, they are automatically raised to zero.

Any activity of transfer and elimination of gains will remain entirely ANONYMOUS both during and after the experiment. After everybody has taken her decisions, a screen with the final winnings (final gains from this stage plus participation token) will appear.

Please stay seated. Payment will be done one at a time and each player will be asked to leave before payment is made to another player. This is to reinforce complete anonymity.

Desert Conditions Only

EXAMPLES: Assume there are two players, Jim (assume player 1) and Joe (assume player 3). Jim starts with 2000 doblons, whereas Joe starts with 1000 doblons.

Non Desert Conditions Only

EXAMPLES: Assume there are two players, Jim (assume player 1) and Joe (assume player 3). Jim receives a 1000 doblons gift and starts with 2000 doblons, whereas Joe starts with 1000 doblons.

Ex. 1: Neither does any activity. Then Jim retains his 2000 doblons and Joe 1000.

Ex. 2: Joe puts 2000 in the green cell in the player 1 row. Jim does nothing. Then Joe gets 900 doblons (for he pays $10\% \times 1000 = 100$) and Jim 0.

Ex. 3: Assume now that there is also Jane, who has 500 initial total gains.

If Joe puts 500 in the green cell of player 1 and 400 in the sky-blue cell corresponding to the player 1 row and Jane's column. (*Due to a typo, the example in brackets was incomplete in the computer instructions, but subjects were told verbally to ignore it*.)

Stealing Conditions Only

Ex. 4: Jim transfers 500 of Joe's doblons to himself; Joe transfers 1000 of Jim's doblons and 250 of Jane's to himself; Jane transfers 1000 of Jim's doblons to herself and 1000 to Joe.

Then Jim's balance is 2000 (inital total gains) -200 (price for activity: 10% of initial total gains) -1000 (transferred away by Joe) -1000 (transferred away by Jane) $+ 500$ (transferred from Jim) $+ 250$ (transferred from Jane) $= 550$ doblons.

Joe's balance is $1000 - 100$ (price for activity) -500 (transferred away by Jim) $+ 1000$ (transferred from Jim by Joe) $+ 1000$ (transferred from Jim by Jane) $=$ 2400 doblons.

Jane's balance is $500-50$ (price for activity) $+ 1000$ (transferred from Jim) -250 (transferred away by Jim) $= 1200$ doblons.

Non Stealing Conditions Only

Ex. 4: Jim eliminates 500 of Joe's doblons and 250 of Jane's; Joe eliminates 1000 of Jim's doblons; Jane eliminates 1000 of Jim's doblons and transfers another 1000 to Joe.

Then Jim's balance is 2000 (inital total gains) -200 (price for activity: 10% of initial total gains) -1000 (eliminated by Joe) -1000 (eliminated by Jane) $=$ -200, hence 0 since a negative balance is not allowed.

Joe's balance is $1000-100$ (price for activity) -500 (eliminated by Jim) $+ 1000$ (transferred by Jane) $= 1400$ doblons.

Jane's balance is $500-50$ (price for activity) -250 (eliminated by Jim) $= 200$ doblons.

PLEASE TAKE YOUR FINAL DECISION WITH CARE. Both your and the other people's winnings depend on such decision.

To make a more careful choice, we encourage you again to try out various combinations and use View to see what would happen as the outcome of your activity.

This is the last screen of instructions, and once you click help again you'll be able to actually start working.

However, feel free to browse your way through the instructions screens again at any time.

PLEASE START WORKING NOW.

First, make some PRACTICE clicking on View to see what happens when you make a choice.

Second, press OK if you are satisfied with your choice and press OK on the message box that will appear.

Third, press Confirm if you are positively sure about your choices. Otherwise press Cancel.

Click Help to get the instructions back on this screen.