A Little Fairness may Induce
a Lot of Redistribution in Democracy

Jean-Robert Tyran and Rupert Sausgruber

December 2002 Discussion Paper no. 2002-30

Department of Economics   University of St. Gallen

A Little Fairness may Induce a Lot of Redistribution in Democracy[1]

Jean-Robert Tyran and Rupert Sausgruber

Author addresses: Dr. Jean-Robert Tyran
                           Volkswirtschaftliche Abteilung
                           Universität St. Gallen
                           Bodanstrasse 1
                           CH-9000 St. Gallen
                           Tel.   +41 71 22423 15
                           Fax   +41 71 224 28 16
                           Email   jean-robert.tyran@unisg.ch
                           Website http://www.vwa.unisg.ch/tyran/

                           Dr. Rupert Sausgruber
                           Institut für Finanzwissenschaft
                           Universität Innsbruck
                           Universitätsstr. 15
                           A-6020 Innsbruck
                           Tel.   +43 512 507 7176
                           Fax   +43 512 507 2877
                           Email   rupert.sausgruber@uibk.ac.at
                           Website http://homepage.uibk.ac.at/homepage/c404/c40422/

Abstract

We use a model of self-centered inequality aversion suggested by Fehr and Schmidt (1999) to study voting on redistribution. We theoretically identify two classes of conditions when an empirically plausible amount of fairness preferences induces redistribution through referenda. We test the predictions of the adapted inequality aversion model in a simple redistribution experiment, and find that it predicts voting outcomes far better than the standard model of voting assuming rationality and strict self-interest.

Keywords

Fairness, Voting, Redistribution.

JEL Classification

D31, D63, D72.

1 Introduction

In his authoritative textbook on public choice, Dennis Mueller (1989: 456) con-
cludes a chapter on redistribution by stating that "the narrow self-interest model
of voting does not explain well the voting behavior of many individuals. Nor does
it explain all redistribution activity." What Mueller calls "the narrow self-interest
model" in fact is the standard approach to the economic analysis of politics today.
This approach assumes that all voters are rational and egoistic. While leading
scholars have noted for quite some time that this approach produces predictions
that are apparently at odds with observed outcomes, most researchers have been
reluctant to allow for non-egoistic preferences in the analysis of voting on redistri-
bution for methodological reasons. Without independent empirical discipline, it
is argued, allowing for non-egoistic preferences is like opening Pandora's box. In
fact, virtually any redistributive outcome can be "explained" by ad hoc invoking
"fairness-minded" voters.

Over the last two decades, however, a vast literature on the measurement
of social preferences has accumulated, amply providing the requested empirical
discipline. This literature has shown that people are heterogeneous with respect
to fairness preferences (see Camerer 2001: Ch. 3 for a survey). That is, some
people are more altruistic than others, some are more spiteful than others, and
some are just plainly egoistic. Recently, the literature has started to address the
important issue of how to model economic interaction when players have hetero-
geneous fairness preferences (see, for example, Rabin 1993, Falk and Fischbacher
2000, Levine 1998, Fehr and Schmidt 1999, 2001, Bolton and Ockenfels 2000,
Dufwenberg and Kirchsteiger 2001).

This paper adapts the framework suggested by Fehr and Schmidt (henceforth
FS 1999) to analyze voting on redistribution in a population with heterogeneous
fairness preferences. We use the FS-approach for two main reasons. First, the
FS-framework can easily be adapted to analyze redistribution through voting
because it is parsimonious and tractable. Second, the FS-framework lends itself
to experimental testing because it generates clear-cut predictions which in some
instances sharply differ from the standard theory of voting.

In the first part of the paper, we identify cases in which the standard model

predicts no redistribution, but the adapted FS-model predicts a lot of redistribution if we assume only "a little fairness", i.e., if we assume that people are at most as fairness-minded as is empirically plausible. In the second part of the paper, we use estimates of distributions of fairness preferences from other (non-voting) experimental studies having measured social preferences to predict voting on redistribution. These predictions are applied to a particular redistribution proposal, which we implement in a laboratory experiment. The purpose of this simple experiment is to test the predictions of the adapted FS-model against the standard model. However, the experiment neither provides a full-fledged test of all implications of the FS-model nor a comparative evaluation of alternative fairness models. We find that the parameterized predictions of the adapted FS-framework perform much better than the ones from the standard model, assuming rationality and strict self-interest.

The main contribution of this paper is twofold. On a theoretical level, we identify conditions when the predictions of the standard theory are non-robust with respect to fairness preferences. On an empirical level, we show that incorporating heterogeneity of fairness preferences may generate much better predictions on voting on redistribution than the standard model. Our paper strives to contribute to the discussion of when and how empirically plausible deviations from the standard assumptions of economics can have large aggregate-level effects. While this discussion has considerably progressed with respect to relaxing rationality assumptions (e.g., Akerlof and Yellen 1985, Haltiwanger and Waldman 1989), much less work seems to be available with respect to relaxing the strict self-interest assumption. In fact, Alesina and Angeletos (henceforth AA 2002) to the best of our knowledge is the only other contribution investigating the effect of fairness preferences on democratic redistribution. However, the AA-model differs in scope and approach from ours. AA aim at explaining differences in redistribution in the US and Europe while we aim at testing the predictions of the FS-approach against the standard theory. The AA-model assumes that agents are heterogeneous with respect to abilities and willingness to work which induces heterogenous evaluations of the fairness of the income distribution. In contrast, we directly assume heterogeneity of fairness preferences and test whether this assumption is appropriate in voting on redistribution.

2

We proceed as follows. Section 2 explains how the FS-framework can be adapted to analyze voting on redistribution. Section 3 explains when a small amount of fairness can induce large-scale redistribution. Section 4 reports experimental results, and section 5 provides some concluding remarks.

2   Fairness and voting on redistribution

Empirical research suggests that some democratic redistribution is hard to reconcile with the assumption of rational and strictly egoistic voters (e.g., Bütler 2000, van der Heijden et al. 1997, Boeri et al. 2001). To be able to better account for voting on redistribution, we suggest to take into account that voters are heterogenous with respect to fairness-preferences.

While this paper focuses on preference-based explanations for voting, we are well aware that there are several alternative ways to explain existing large-scale redistribution in developed economies (e.g., Persson and Tabellini 2000, van Winden 2002, Boadway and Keen 2000). Obviously, our experiment provides no assessment of the relative importance of these alternative explanations. Instead, we isolate the effect of fairness preferences on referendum outcomes under controlled conditions.

One possible interpretation of fairness is inequity-aversion (see Konow 2002 for a survey). However, what people consider as inequitable may depend on many (e.g., contextual) factors. In some situations, and in particular in laboratory settings, inequity-aversion can be reasonably approximated by *inequality*-aversion (see FS 1999: 820ff. for explanations. Survey-based evidence that people care about relative incomes is provided by Clark and Oswald 1996, Solnick and Hemenway 1998, Amiel and Cowell 1992, 2001, Carlsson et al. 2001). While this (over-) simplification has the drawback of ignoring potentially important aspects of voting (see section 4.3), it has the advantage of being parsimonious and tractable.

We now provide intuitions for why a *small* amount of heterogeneous fairness preferences may have large effects on redistribution outcomes. Consider for concreteness a referendum to tax the rich and to redistribute the money to the poor. Suppose that the rich slightly outnumber the poor. Therefore, if all voters were

3

exclusively motivated by material self-interest they would reject the redistribution policy by a narrow margin. The first explanation for why "a little" fairness may induce a lot of redistribution prevails if *few* voters are *strongly* inequality averse. For example, suppose that 51% of the voters are rich, and 49% are poor, but that 1% of the rich voters have such strong fairness-preferences that they vote against material self-interest. In this rather extreme case, the presence of these few voters is sufficient to tip the balance, and to induce redistribution.

A second explanation for why fairness preferences may induce a lot of redistribution is that some voters are materially unaffected by redistribution. These "onlookers" may take sides and vote for fair distribution because it is not costly to them. Suppose, for example, that 40% of voters are rich, 40% are middle-class, and 20% are poor. Suppose again that the proposal is to redistribute money from the rich to the poor, leaving the middle-class monetarily unaffected. A rational and strictly self-interested middle-class voter is indifferent and may abstain or flip a coin to decide on how to vote. In contrast, a weakly inequality-averse middle-class voter votes for redistribution. Therefore, *many weakly* inequality-averse voters can tip the balance in favor of redistribution. In this example, the proposal is accepted if more than half of the middle-class voters are slightly inequality averse.

An important advantage of the FS-approach is that it allows to model the effect of fairness preferences on voters who are materially unaffected by redistribution. While the approach by Bolton and Ockenfels (BO 2000) is similar to the FS-approach in many respects, the BO-approach assumes that individuals compare themselves to the average individual in the population. Hence, a change in the income distribution that leaves the average income and voter i's income unaffected will not affect voter i's utility. However, it should be noted that our experiment was designed to provide a simple test of the FS-approach against the standard model, and not to test how alternative fairness theories perform in voting (see Charness and Rabin 2002 and Engelmann and Strobel 2002 for such comparisons in contexts other than voting).

A third explanation for why an empirically plausible extent of fairness preferences induces a lot of redistribution is provided by strategic low-cost consid-

erations. According to this explanation voters consider the fact that they are not pivotal most of the time (Mulligan and Hunter 2001). If an individual voter does not affect the outcome, he or she may just as well cast a fair vote to feel good (see Kirchgässner 1992, 2000: Ch. 5; Brennan and Lomasky 1993). We do not formally analyze this third explanation because it is much more difficult to model. In particular, one would need to model expectations which are not part of the FS-framework (see Alesina and Angeletos 2002 on the potential importance of expectations). However, we do test the expectations-based low-cost theory in our experiment (see section 4.3).

We use the adapted FS-model to calculate the minimum extent of fairness preferences necessary to tip the outcome of a referendum over a given redistribution policy R. For example, we calculate that at least $\overline{\lambda}$ percent of voters have to have fairness preferences of a particular type to tip the balance. If $\overline{\lambda}$ is small compared to the extent of fairness preferences that have been measured in various experimental studies, we conclude that *a little* fairness is sufficient to induce redistribution R. Alternatively, we could have calculated the maximum redistribution that would just be accepted by an electorate with an empirically plausible distribution of fairness preferences. For example, we could have calculated the maximum redistribution $\overline{R}$ that voters are willing to accept by a margin of, say, 50 percent plus 1 vote. Obviously, $\overline{\lambda}$ and $\overline{R}$ are flip sides of the same coin, and, as a consequence, $\overline{\lambda}$ is an indirect measure of the maximum amount of redistribution that is accepted. Therefore, identifying conditions when $\overline{\lambda}$ is small is equivalent to identifying conditions, as suggested in the title of the paper, when a little fairness induces *a lot* of redistribution.

2.1 Adapting the Fehr and Schmidt-framework

Equation (1) shows how FS incorporate inequality aversion into the utility function of individual i. We henceforth assume that all individuals are also voters. Voter i's utility has two components. The first component depends only on i's absolute income x_i, while the second depends on i's income relative to other voters j.

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

The parameter α_i captures how much voter i dislikes to be worse off than other voters j, and β_i captures how much voter i dislikes to be better off than others. The authors refer to extensive research in social psychology and experimental economics to justify the following parameter restrictions: $\alpha_i \geq \beta_i \geq 0$ and $\beta_i < 1$. This means that some voters dislike favorable inequality $\beta_i \geq 0$. But those who do, dislike unfavorable inequality more than favorable inequality ($\alpha_i \geq \beta_i$, see FS 1999 for a detailed discussion). FS implicitly assume that players have different incomes x_i, and dividing by $N-1$ normalizes with respect to all players other than i. In the FS-approach inequality aversion is self-centered. A reduction of inequality increases the utility of voter i irrespective of how this reduction affects the utility of other voters. This assumption distinguishes the FS-approach from the related approach of Hochman and Rodgers (1969).

FS use (1) to discuss how distributions of α_i and β_i in the population may lead to particular aggregate-level outcomes under various institutional frameworks. For example, they explain that a particular distribution of inequality aversion may result in relatively large deviations from the predictions of the standard theory in bilateral bargaining but not in a competitive market. We now extend their analysis to the most important non-market institution: democracy. To do so, we analyze voting in referenda which constitute a simple and natural form of democratic choice (Butler and Ranney 1994).

To be able to use the FS-approach to analyze voting in referenda, we assume that each voter $i = 1, ..., N$ is a member of one of three income classes. In particular, we assume that n_r voters are "rich", n_m are in the "middle class", and n_p are "poor" ($n_r + n_m + n_p = N$). To simplify the analysis, we assume that all voters within income class $k = r, m, p$ have the same income and that all redistribution is between (but not within) income classes. We denote by the vector $\mathbf{x}_0 = (x_r, x_m, x_p)$ the initial income distribution where $x_r > x_m > x_p$. The assumption that voters are members of income classes necessitates a class-specific

normalization. We denote by $w_{kj} \equiv \frac{n_j}{N-n_k}$ the weight a voter in income class k attaches on inequality with respect to class j. It holds that $\sum_{j \neq k} w_{kj} = 1$.

These assumptions allow us to considerably simplify (1). Voter i's utility in class k when the initial income distribution prevails is

$$U_i(\mathbf{x_0} \,|\, r) \;=\; x_r - \beta_i \left[w_{rm}(x_r - x_m) + w_{rp}(x_r - x_p) \right], \qquad (2)$$

$$U_i(\mathbf{x_0} \,|\, m) \;=\; x_m - \alpha_i w_{mr}(x_r - x_m) - \beta_i w_{mp}(x_m - x_p), \qquad (3)$$

$$U_i(\mathbf{x_0} \,|\, p) \;=\; x_p - \alpha_i \left[w_{pr}(x_r - x_p) + w_{pm}(x_m - x_p) \right]. \qquad (4)$$

The focus on three income classes simplifies the analysis because there are no voters with endowments higher than x_r, and, as a consequence, there is no α-term in $U_i(\mathbf{x_0} \,|\, r)$. Similarly, because there are no voters with endowments below x_p, there is no β-term in $U_i(\mathbf{x_0} \,|\, p)$. The assumption that all voters in a particular income class have the same income allows us to dispense with the max-operators and the summation.

2.2   Voting for redistribution from rich to poor

In the following, we provide a detailed account of the redistribution policy R_{rp}. If this policy is accepted, each rich voter has to pay a tax of $t_r > 0$ and each poor voter receives a benefit of $b_p > 0$. We focus our discussion on policy R_{rp} to keep the discussion simple and to provide a basis for our experiment (see section 4). Note that policy R_{rp} reduces inequality, and is, so to speak, a "fair redistribution". However, our framework can easily be used to analyze other types of redistribution (e.g., from poor to rich).

We analyze "non-revolutionary" redistribution proposals that preserve the prevailing ranking of income classes. In particular, we assume throughout that income classes k are distinct in the sense that if the order of pre-redistribution income is $x_r > x_m > x_p$, then post-redistribution incomes must have the same ordering. Hence, in the discussion below we assume that $x_r - t_r > x_m > x_p + b_p$. The vector $\mathbf{x_1}$ denotes post-redistribution income distribution. In the example

we are discussing, $\mathbf{x_1} = (x_r - t_r, \; x_m, \; x_p + b_p)$. The utility of a voter i in income class k in case of acceptance of R_{rp} is

$$U_i(\mathbf{x}_1 \,|\, r) \;=\; x_r - t_r - \beta_i \left[w_{rm}(x_r - t_r - x_m) + w_{rp}(x_r - t_r - x_p - b_p) \right], \quad (5)$$

$$U_i(\mathbf{x}_1 \,|\, m) \;=\; x_m - \alpha_i w_{mr}(x_r - t_r - x_m) - \beta_i w_{mp} \left(x_m - x_p - b_p \right), \quad (6)$$

$$U_i(\mathbf{x}_1 \,|\, p) \;=\; x_p + b_p - \alpha_i \left[w_{pr}(x_r - t_r - x_p - b_p) + w_{pm}(x_m - x_p - b_p) \right]. \quad (7)$$

To derive predictions for individual voting behavior, we compare voter i's utility in case of acceptance and in case of rejection. In line with standard economics models of voting (Downs 1957), we assume that voters are rational expected utility maximizers, that the opportunity cost of voting is zero and that voters behave as if they were pivotal. In particular, a voter for whom $U_i(\mathbf{x}_1 \,|\, k) > U_i(\mathbf{x}_0 \,|\, k)$ votes for the proposal with probability $\pi = 1$. A voter for whom $U_i(\mathbf{x}_1 \,|\, k) = U_i(\mathbf{x}_0 \,|\, k)$ is assumed to randomize between voting yes and no with probability $\pi^{ind.} = 0.5$. A voter for whom $U_i(\mathbf{x}_1 \,|\, k) < U_i(\mathbf{x}_0 \,|\, k)$ is assumed to vote yes with probability $1 - \pi = 0$. These assumptions imply the following for individual voting behavior in the three income classes.

a) Rich voters. A rich voter i approves of policy R_{rp} if $U_i(\mathbf{x}_1 \,|\, r) > U_i(\mathbf{x}_0 \,|\, r)$, i.e., if

$$\beta_i > \frac{t_r}{w_{rm} t_r + w_{rp}(t_r + b_p)} \equiv \overline{\beta}(R_{rp}). \quad (8)$$

Equation (8) shows that a rich voter votes for redistribution if he is sufficiently averse to favorable income inequality, i.e., if his β_i exceeds the critical value $\overline{\beta}(R_{rp})$. Note that only the favorable inequality aversion β_i matters for a rich voter while the unfavorable inequality aversion α_i is irrelevant.

We can now calculate for any given distribution of β_i in the population the percentage of rich voters with $\beta_i > \overline{\beta}(R_{rp})$. We denote by $\lambda \in [0, 1]$ the percentage of rich voters with a sufficiently pronounced favorable inequality aversion to induce them to vote against their material self-interest. The percentage of rich voters voting for R_{rp} is a decreasing function of $\overline{\beta}(R_{rp})$:

$$\lambda = \lambda \left[\overline{\beta}(R_{rp}) \right], \quad \partial \lambda / \partial \overline{\beta} < 0. \tag{9}$$

From (8) and (9) it can be seen that the percentage of yes-voters among the rich λ is cet. par. the higher,

- the lower the monetary cost of redistribution, i.e., the lower the tax t_r a rich voter has to pay $(\partial \lambda / \partial t_r < 0)$,

- the higher the per capita benefit b_p a poor voter receives according the proposal $R_{rp}, (\partial \lambda / \partial b_p > 0)$,

- the larger cet. par. the middle class $(\partial \lambda / \partial n_m > 0)$, and the more poor voters benefit from redistribution $(\partial \lambda / \partial n_p > 0)$.

b) Middle-class voters. A middle-class voter votes for policy R_{rp} (compare (3) and (6)) if

$$\alpha_i w_{mr} t_r + \beta_i w_{mp} b_p > 0. \tag{10}$$

Therefore, a middle-class voter votes for the proposal if either $\alpha_i > 0$ or if $\beta_i > 0$. Define $\mu = const.$ as the percentage of voters with $\alpha_i > 0$ and $\beta_i \geq 0$. The number of voters with $\alpha_i = \beta_i = 0$ in income class n_m is $(1 - \mu) n_m$. These voters are indifferent about whether redistribution is implemented. We assume that indifferent voters randomize between voting yes and no with probability $\pi^{ind.} = 0.5$. Hence, the expected number of yes-voters in the middle class is $[(1 + \mu)/2]n_m$.

c) Poor voters. From (4) and (7) a poor voter i votes for R_{rp} if

$$\alpha_i > -\frac{b_p}{w_{pr}(t_r + b_p) + w_{pm} b_p} \ . \tag{11}$$

Since the rhs of (11) is negative by definition, and since $\alpha_i \geq 0$ it follows that all n_p poor voters vote for R_{rp}. In other words, poor voters vote for redistribution policy R_{rp} irrespective of fairness preferences.

d) Summing up. The preceding discussion has formalized the intuition that rich voters vote for the proposal R_{rp} if they are sufficiently strongly averse to favorable inequality $\beta_i > \overline{\beta}$, while middle-class voters vote for R_{rp} even if they are only weakly averse to favorable or unfavorable inequality, and poor voters vote for the policy irrespective of their inequality aversion. Most importantly, we have established the fact that the percentage of rich voters (λ) and the percentage of middle-class voters (μ) voting for redistribution depends on the distributions of α_i and β_i in the population.

From (8) to (11), it follows that the total number of yes-votes $y(R_{rp})$ for R_{rp} is

$$y(R_{rp}) = \lambda n_r + \frac{1 + \mu}{2} n_m + n_p. \tag{12}$$

Policy R_{rp} is adopted if the number of yes-votes is sufficiently high. We denote by $q \in [0, 1]$ the quorum, i.e., the percentage of yes-votes that has to be exceeded for the proposal to be accepted. For example, in majority voting the approval rate requirement is $q = 0.5$, and the number of yes votes that has to be exceeded for R_{rp} to pass is qN. Hence, R_{rp} is accepted if $y(R_{rp}) > qN$.

3   When fairness induces redistribution

In this section we discuss two reasons why a *little* fairness may induce a lot of redistribution in democratic referenda. First, when all voters have material interests at stake, *only few* sufficiently inequality averse voters deciding against material self-interest may tip the balance. For example, if the rich only slightly outnumber the poor, few sufficiently inequality averse rich voters may tip the

balance in favor of R_{rp}. However, as will be shown in section 3.1., these fairness-minded voters are irrelevant if the rich are much more numerous than the poor or if the poor outnumber the rich.

Second, voters which are *only weakly* inequality averse may vote for redistribution if redistribution reduces inequality but is costless to them in monetary terms. For example, a sufficient number of weakly fairness-minded middle-class voters may induce redistribution from rich to poor.

To be able to analytically clearly separate the two accounts, we assume in section 3.1 that there is no middle class $(n_m = 0)$. Section 3.2 discusses how the existence of middle-class voters affects redistribution policy.

3.1 Effects of few strongly inequality averse voters

We now explain under which conditions the presence of few strongly fairness-minded rich voters induces redistribution when the standard model predicts no such redistribution. The main result of this section is that the effects of inequality are asymmetric, with the strength of the effect depending on the initial income distribution. The intuition for this result is that few fairness-minded voters can only tip the referendum outcome if it was close in the first place. In addition, fairness preferences induce voters to vote for fair redistribution (inequality increasing) or against unfair redistribution proposals. Hence, fairness preferences matter most when an unfair distribution would have been closely accepted by strictly self-interested voters. This intuition is restated in result 1.

Result 1 *The effects of few strongly inequality averse voters on referendum outcomes are asymmetric. The effects are disproportionately large when the standard theory predicts a close rejection of fair redistribution or a close acceptance of unfair redistribution.*

To understand result 1, we start by solving the condition for acceptance $y(R_{rp}) > qN$ (see (12)) for λ to obtain

$$\lambda > q + \left[q - \frac{1+\mu}{2}\right] \frac{n_m}{n_r} + (q-1) \frac{n_p}{n_r}. \tag{13}$$

To simplify the analysis, we assume that there are only rich and poor voters ($n_m = 0$). With this simplification, (13) yields the critical percentage of rich voters $\overline{\lambda}$ necessary to tip the balance in favor of acceptance

$$\overline{\lambda}_{|n_m=0} \equiv max \left[0, \; q + (q-1)\frac{n_p}{n_r} \right]. \tag{14}$$

Consider first the case where all agents are exclusively motivated by material self-interest ($\alpha_i = \beta_i = 0$ for all i). In majority voting, policy R_{rp} will be rejected if the rich are in majority, i.e., if $n_r/n_p > 1$. Figure 1 illustrates this prediction of the standard theory. The figure shows for $q = 0.5$ that the standard theory predicts acceptance of R_{rp} for $n_r/n_p < 1$, but rejection for $n_r/n_p > 1$ (note that $(1-q)/q = 1$ for $q = 0.5$). This prediction is a special case of the well-known result that majority voting results in redistribution if the income of the median voter is below the mean income (e.g., Meltzer and Richard 1981).

Incorporating inequality aversion into the analysis changes the identity of the median voter. In standard models of redistribution, the median voter simply is the one with median income. In our framework, however, the existence of fairness preferences ($\beta_i > \overline{\beta}$) induces some rich voters to vote against material self-interest. The heavy line in figure 1 plots the critical percentage $\overline{\lambda}$ of rich voters necessary to induce redistribution (14). We henceforth call the graphical representation of $\overline{\lambda}$ the acceptance frontier. Figure 1 shows that the frontier has a kink. It is constant at 0 for $n_r/n_p < 1$, indicating that the redistribution is accepted irrespective of fairness preferences if the poor are in majority. However, if the rich are in majority ($n_r/n_p > 1$), the critical percentage $\overline{\lambda}$ is positive. This indicates that if the actual percentage of sufficiently fairness-minded rich voters is below the threshold ($\lambda < \overline{\lambda}$), the proposal is rejected (see lightly shaded area in figure 1).

Figure 1 illustrates that there are important constellations in which the standard model predicts rejection of R_{rp} but the FS-framework predicts acceptance (see dark shading). Consider, for example, a distribution of preferences implying a small percentage λ_1 of rich voters voting for redistribution. If the rich

12

slightly outnumber the poor, the FS-framework predicts that this small amount of fairness (λ_1) will induce acceptance of the referendum (see point B in figure 1). However, the same percentage λ_1 of sufficiently fairness-minded rich voters does not affect referendum outcomes if the poor outnumber the rich (see point A), or if the rich massively outnumber the poor (see point C).
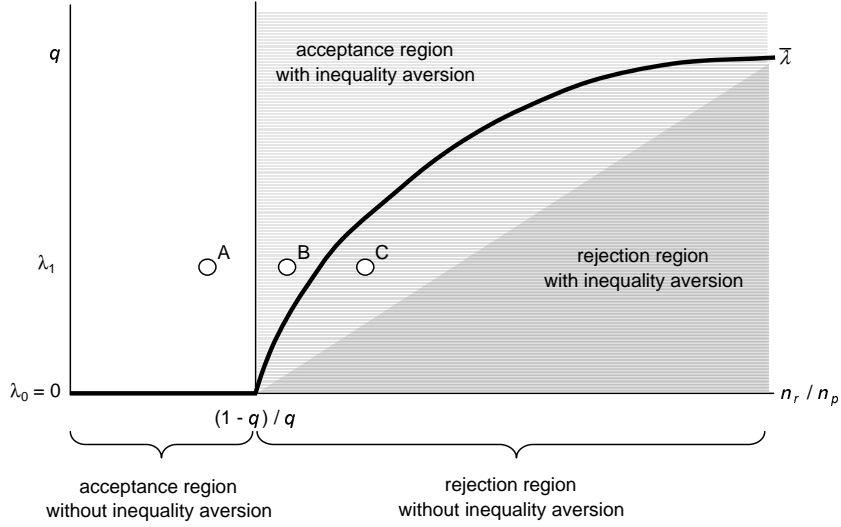

Figure 1: Acceptance frontier $\overline{\lambda}$ for R_{rp} if $n_m = 0$.

This discussion has important implications for the robustness of the predictions of the standard theory for referendum outcomes. Remarkably, the conclusion is not warranted that the standard theory is uniformly unreliable if some voters are fairness-minded. Rather, standard theory's robustness depends on z, the change in proportion of rich and poor voters. In particular, the prediction of the standard economics model is extremely non-robust for small (but positive) z, but robust for $z < 0$, and relatively robust for large $z > 0$ in $n_r/n_p = \frac{(1-q)}{q} + z$.

3.2 Low-cost decisions of monetarily unaffected voters

We now discuss the effects on referendum outcomes of voters who are *only weakly inequality averse* ($\alpha_i > 0$ or $\beta_i > 0$), but who are not monetarily affected by redis-

tribution policy. In particular, middle-class voters are not affected in monetary terms from policy R_{rp}. However, the adoption of policy R_{rp} increases the utility of middle-class voters with $\alpha_i > 0$ or $\beta_i > 0$ because it affects their relative income position. Hence, just minimally inequality averse middle-class voters vote for R_{rp}. To analyze the effects of inequality-averse middle-class voters, we assume $n_m > 0$.

Result 2 *A materially non-affected inequality averse* $(\mu > 0)$ *middle class cet. par. raises expected approval rates in majority voting* $(q = 0.5)$. *For supermajority requirements* $(q > 0.5)$, *the middle class raises expected approval rates if and only if* $\mu > 2q - 1$.

In the presence of middle-class voters $(n_m > 0)$, the acceptance frontier, i.e., the critical percentage of (rich) voters voting against material self-interest is

$$\overline{\lambda} \equiv max \left[0, \ q + \left(q - \frac{1 + \mu}{2} \right) \frac{n_m}{n_r} + (q - 1) \frac{n_p}{n_r} \right]. \tag{15}$$

With slight rearrangements, (15) is seen to be a piecewise linear acceptance frontier

$$\overline{\lambda} \equiv max \left[0, \ A + B\mu \right], \tag{16}$$

where $A = q + \frac{1}{n_r} \left[\left(q - \frac{1}{2} \right) n_m + (q - 1) n_p \right]$, and $B = -n_m/2n_r$. Figure 2 represents (16) in the λ-μ-space. In this figure, A is the intercept and B is the slope of the frontier for $\overline{\lambda} > 0$. In constructing the acceptance region in the λ-μ-space, we also use the restriction on preference parameters suggested by FS (see section 2.1 for explanations) that $\alpha_i \geq \beta_i \geq 0$. This assumption together with the assumption that the distribution of α_i and β_i is the same in all income classes implies that $\mu \geqslant \lambda$. Figure 2 shows the acceptance region of R_{rp} is bounded from below by the frontier $\overline{\lambda}$, and bounded from above by the restriction $\mu \geqslant \lambda$.

Figure 2 shows that μ and $\overline{\lambda}$ are perfect substitutes since the slope of the frontier $B = -n_m/2n_r < 0$ for $\overline{\lambda} > 0$. The negative slope means that the higher the
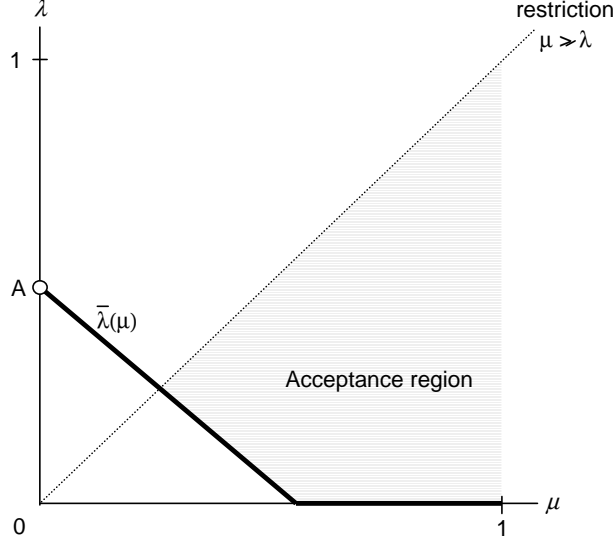
Figure 2: Combinations of λ and μ implying the acceptance of R_{rp} for $n_m > 0$.

percentage of weakly fairness-minded middle-class voters, the smaller the number of sufficiently fairness-minded rich voters necessary to induce redistribution.

a) Size of the middle class. In case of majority voting ($q = 0.5$), a ceteris paribus increase in the number of middle-class voters makes the slope of the frontier steeper, enlarging the acceptance region in figure 2. This effect is the more pronounced, the larger the percentage of fairness-minded middle-class voters μ. To see this, note that $\partial\overline{\lambda}/\partial n_m = -\mu/2n_r < 0$ at $q = 0.5$. If the middle class is at least weakly inequality averse ($\mu > 0$), an increase in n_m reduces the critical $\overline{\lambda}$ and it increases the actual λ for and given distribution of preferences β_i ($\partial\lambda/\partial n_m > 0$).Therefore, an increase in the size of the middle class raises expected approval rates in majority voting (see result 2).

Supermajority requirements ($q > 0.5$) reduce the effect of fairness preferences on two accounts. First, a higher approval rate requirement increases the necessary number of rich voters to tip the outcome of the referendum towards acceptance ($\partial\overline{\lambda}/\partial q|_{\overline{\lambda}>0} > 0$). Second, supermajority requirements also reduce the effect of the middle class. The effect of an increase of the middle class is given by the derivative $\partial\overline{\lambda}/\partial n_m|_{\overline{\lambda}>0} = [2q - 1 - \mu]/2n_r \gtreqless 0$. The ambiguity of this effect can

15

be graphically illustrated in figure 2 by noting that an increase in n_m simultaneously shifts the frontier up and makes it steeper. Hence, for supermajority requirements, an increase in the number of middle-class voters widens the acceptance region for sufficiently prevalent inequality aversion ($\mu > 2q - 1$), but not for small μ.

b) Other redistribution policies. The effects of materially unaffected voters are not necessarily symmetric for symmetric policies. For example, the effects of non-involved poor voters in policy transferring money from the rich to the middle class R_{rm} (with $t_r(R_{rm}) = t_r(R_{rm}), b_p(R_{rp}) = b_m(R_{rm})$) are not the same as the effect of a non-involved middle class in policy R_{rp}. Redistribution involving classes above and below one's income class (e.g., R_{rp}) and otherwise symmetric redistribution involving income classes which are both above (e.g., R_{rm}) or both below (e.g., R_{mp}) one's own income class differ on two accounts. First, inequality-averse voters evaluate inequality with respect to the average income of a class. If redistribution is costless (in the sense that all tax revenues are redistributed), inequality with respect to one class increases by as much as inequality falls with respect to the other class. Hence, costless redistribution of income among voters who are both richer or poorer than voter i does not affect voter i's utility in our framework. For costly ("leaky bucket") redistribution, however, voter i's utility increases (decreases) if redistribution takes place between voters which are all richer (poorer) than i. Second, even if we consider a "leaky bucket" type of redistribution, changes in the relative income position of voters with incomes above are evaluated at α, those below are evaluated at β. In contrast, changes in relative income involving both income classes above and below are evaluated at α and β.

c) Bounded rationality. Above we have identified conditions under which the predictions of the standard model of voting are non-robust with respect to inequality aversion. A second potential source of deviation from the standard model is bounded rationality. A complete analysis of the conditions under which the predictions of the standard model and the FS-approach are non-robust with respect to bounded rationality is clearly beyond the scope of this paper. The

16

reason is simply that there are potentially many ways to be boundedly rational, and the empirical knowledge about the exact form of boundedly rational voting is insufficient to confidently model bounded rationality (see Rubinstein 1998 or Simon 1997 on ways to model bounded rationality in other contexts). Yet, we show in appendix A that the implications of the FS-approach are robust for majority voting with respect to a particular form of bounded rationality (viz., random and uniformly distributed decision errors).

4   An experimental test

We now report results from a simple experiment serving to provide a comparative evaluation of the adapted FS-framework and the standard theory. FS have provided quantitative estimates of the distributions of the preference parameters α_i and β_i which imply numerical predictions for λ and μ, and for individual voting behavior in our experiment. We use the FS-estimates to assess the numerical precision of these predictions.

In our experiment, subjects were endowed with monetary incomes by the experimenter and had to decide on redistribution from rich to poor R_{rp} by majority vote. To be able to comparatively evaluate the standard theory and the adapted FS-framework, we chose the parameters of the experiment such that the predictions of the two accounts sharply differed. In particular, the parameters were chosen such that the standard theory predicts a clear rejection of R_{rp}, while the adapted FS-framework predicts acceptance of R_{rp}.

4.1   Procedures, parameters, and predictions

Subjects were randomly assigned to committees of five voters. Each committee consisted of two rich voters ($n_r = 2$), two middle-class ($n_m = 2$), and one poor voter ($n_p = 1$). The initial endowments were $x_r = 250$, $x_m = 185$, $x_p = 60$ (numbers are Austrian Schillings, ATS 100 $\approx$ US\$ 7 at the time of the experiment). Subjects voted on a proposal R_{rp}, imposing a tax of $t_r = 50$ each on rich voter to

finance a benefit of $b_p = 100$ to the poor voter. The referendum was decided by anonymous majority vote ($q = 0.5$). That is, the policy was adopted if at least three out of the five voters approved.

The experiment was explained to subjects in written instructions (see appendix). We used the neutral labels A, B, and C for rich, middle, and poor voters, respectively. When subjects took their decisions, they did not yet know whether they are in role A, B or C. As a consequence, subjects had to indicate for each of the three roles whether they vote "yes" or "no" on the proposal, abstentions were not possible. Subjects knew that they would be randomly assigned to a committee and a role after they had taken their decisions. Subjects knew that they had to take their decisions only once and that they were paid according to their decisions in the randomly assigned roles and committees.

Predictions. The standard theory makes clear predictions in our experiment. Under the assumption that all voters are rational and egoistic ($\alpha_i = \beta_i = 0$), both rich voters vote against the proposal, and the poor voter votes for the proposal. As explained in section 2.2, a materially uninvolved middle-class voter is predicted to randomly vote for or against the proposal. In this case, the prediction would be that the proposal is on average rejected by a narrow margin of 3:2 votes. To make the comparative evaluation of the two theories statistically more powerful, middle-class voters had to pay a small commission of $t_m = 5$ in case the proposal was accepted, and this was known to all participants at the beginning of the experiment. The commission makes it easier to discriminate the two theories because the standard theory now predicts that both middle-class voters on average vote against redistribution, and, therefore, the proposal is predicted to be rejected in each committee by a margin of 4:1 (see table 1).

To calculate the numerical predictions of the adapted FS-framework we use the distributions of α_i and β_i suggested in Fehr and Schmidt (1999: 844). The authors provide a rough characterization of the distribution of fairness preferences as follows: 30 percent of the voters have $\alpha_i = \beta_i = 0$, 30 percent have $\alpha_i = 0.5$ and $\beta_i = 0.25$, and 40 percent have $\alpha_i \geq 1$ and $\beta_i = 0.6$. We would like to emphasize that this characterization should be considered as a handy approximation of a continuous distribution of fairness preferences.

Using this approximation to calculate $\overline{\beta}(R_{rp})$ as in (8), yields $\overline{\beta}(R_{rp}) = 0.6$, and a corresponding value of $\lambda = 40\%$. However, in taking the commission of $t_m = 5$ into account, the condition (8) for a rich voter to vote in favor of redistribution is slightly modified to yield $\beta_i > t_r/[w_m(t_r - t_m) + w_p(t_r + b_p)] \equiv \overline{\beta}(R_{rp})$. Inserting our parameters now yields $\overline{\beta}(R_{rp}) = 0.625$. Therefore, the adapted FS-framework predicts that at most $\lambda \leq 40$ percent of the rich to vote for redistribution. The modified condition for a middle-class voter to approve is $\alpha_i w_r(t_r - t_m) + \beta_i w_p(b_p + t_m) > t_m$ (compare (10)). With our parameters this yields $30\alpha_i + 35\beta_i > 5$, which implies for the distributions of α_i and β_i that $\mu = 70$ percent of middle-class voters vote for redistribution.

The above discussion can be summarized in two hypotheses to be tested. The adapted FS-framework predicts at most 40 percent of the rich ($\lambda \leq 0.4$), 70 percent of the middle-class ($\mu = 0.7$), and 100 percent of the poor to vote for redistribution (see table 1). On average, the proposal R_{rp} is predicted by FS to be accepted by a margin of $[\lambda n_r + \mu n_m + n_p]/N = 64$ percent. In contrast, the standard theory assuming strictly self-interested voters predicts that 0 percent of the rich, 0 percent of the middle-class, and 100 percent of the poor voters to approve. Hence, the standard theory predicts rejection with an average approval rate of $n_p/N = 20$ percent.

4.2   Experimental results

We conducted one experimental session in a large lecture hall at the University of Innsbruck with 80 undergraduate students from various majors. The average subject earned ATS 245 ($\approx$ US\$ 17) in less than an hour.[1] The main result of our experiment is summarized in the following result.

Result 3 *The adapted FS-framework predicts referendum outcomes much better than the standard theory. In addition, the adapted FS-framework provides strikingly accurate predictions for individual voting behavior in all three income classes.*

---

[1]Earnings include a show-up fee and payments to motivate subjects to correctly report their expectations on voting outcomes. See section 4.3 for details.

Table 1: Percentage shares of yes-votes

	Predictions		Observations (N=80)
	Standard Theory	Fehr and Schmidt (1999)	
rich	0	40	33.7
middle	0	70	70.0
poor	100	100	96.3

Support for result 3 is provided by the fact that the referendum is accepted by 14 out of 16 committees, at a total average approval rate of 61.4 percent yes-votes. This observation is much closer to the prediction of the FS-framework of an average approval rate of 64 percent, than of the standard theory (which predicts an approval rate of 20 percent). The adapted FS-framework not only predicts the overall acceptance rate much better, it is also strikingly precise in predicting voting behavior in the different classes. In particular, the overall approval rate for the rich is 33.7 percent ($= 27/80$), and for the middle-class it is 70.0 percent ($= 56/80$) (see table 1). These observations are far away from the predictions of the standard theory (0 percent in both cases), but strikingly close to the predictions of the adapted FS-framework of $\lambda \leq 40$ percent, and of $\mu = 70$ percent.

Table 1 also reports the finding that 96.3 percent of poor voters voted for redistribution. This is not per se evidence against either theory, since both theories predict that 100 percent of poor voters approve of redistribution. The result is nevertheless important for our analysis because it indicates that bounded rationality played almost no role in our simple laboratory setting. This strongly suggests that the deviations from the standard theory prediction for middle-class and rich voters are not due to bounded rationality, but to inequality aversion.

A more detailed analysis of voting decisions reveals that individual voting behavior is very much in line with the predictions of the adapted FS-model. Since each subject had to take contingent decisions for all three income positions, we have three observations for each subject. We denote by $y_i = \{y_r, y_m, y_p\}$ subject i's conditional voting decisions given i is rich, middle, and poor. We classify voters into types according to their voting decisions. In particular, we refer to

$y_i = \{0, 0, 1\}$ as strictly self-interested, $y_i = \{0, 1, 1\}$ as weakly inequality averse, and $y_i = \{1, 1, 1\}$ as strongly inequality averse. We observe that 25 percent of the subjects take decisions that are consistent with strict self-interest, 37.5 percent are consistent with weak inequality aversion, and 30 percent with strong inequality aversion. Only 7.5 percent of subjects were inconsistent with one of these three types. Again, these figures are strikingly close to the distribution of preferences suggested by FS.

To statistically evaluate the predictions of the two models, we construct an exact probability limit around the observed sample approval rates. The intuition is as follows: Can we reject the hypothesis that ρ percent of the population would have voted for the proposal given that we actually observe an approval rate of s percent in our sample? Call ρ the unknown population proportion of voters in class k that would vote in favor of redistribution and s the observed frequency of yes-votes in the sample of $n_k = 80$ bivariate choices. The interval is constructed such that it contains the true parameter with probability $(1 - \alpha)$. That is, we solve $\sum_{x=s}^{n_k} \binom{n_k}{x} \underline{\rho}^x (1 - \underline{\rho})^{n_k - x} = \frac{\alpha}{2}$ and $\sum_{x=0}^{s} \binom{n_k}{x} \overline{\rho}^x (1 - \overline{\rho})^{n_k - x} = \frac{\alpha}{2}$ for the lower and upper exact probability bounds $\underline{\rho}$ and $\overline{\rho}$. For a significance level $\alpha = 1\%$ this yields $[\underline{\rho}, \overline{\rho}] = [0.208, 0.487]$ for the rich voters, $[\underline{\rho}, \overline{\rho}] = [0.553, 0.823]$ for the middle class, and $[\underline{\rho}, \overline{\rho}] = [0.869, 0.996]$ for the poor voters, respectively.

For rich voters, the interval contains the limit prediction of the adapted FS-framework ($\lambda = 0.4$), but is far from containing the prediction of the standard theory ($\lambda = 0.0$). Therefore, we can reject the hypotheses of the benchmark model that $\lambda = 0.0$ but cannot reject $\lambda = 0.4$. For middle-class voters, the interval for the parameter μ again contains the prediction of the adapted FS-framework ($0.553 > 0.7 > 0.823$), but does not contain the prediction $\mu = 0.0$ of the standard theory. Finally, for the poor voters, the upper bound of the confidence limit gets close to 100 percent.

4.3 Discussion of results

We now discuss some potential limitations of the adapted FS-approach and provide some caveats on the interpretation of our experimental results.

Low-cost decisions due to non-pivotality. We have discussed two reasons for why a small amount of fairness may have disproportionately large effects on redistribution in sections 3.1 and 3.2. As briefly mentioned in section 2, a strategic low-cost argument provides a third potential explanation. According to this reasoning, people may vote against their material self-interest because they expect their decision to be irrelevant for the outcome. Indeed, voting against one's material self-interest is costless whenever a voter is non-pivotal. The strategic low-cost hypothesis, therefore, predicts that those who expect to be non-pivotal tend to vote against their material self-interest while those who expect to be pivotal tend to vote according to their material self-interest.

To provide a simple test of this hypothesis, we asked subjects to report the expected number of yes-votes in their committee (excluding their own vote). Subjects had an incentive to correctly predict voting decisions since they received an extra payment of ATS 50 (US$ 3.5, approx.) if their expectation was correct. Subjects believe to be pivotal if they expect exactly two others to approve. We find that 21 of the 80 voters in the role of the rich expected to be pivotal. 42.9 percent ($= 9/21$) of these voters voted for redistribution. In contrast to the prediction of the low-cost hypothesis, the percentage of the non-pivotal rich voters approving of redistribution is lower, at 30.5 percent ($= 18/59$). However, approval rates are not significantly different ($p = 0.304$ according to a χ^2 test). For the middle class too, approval rates are not significantly different between those who expect to be pivotal and those who expect to be non-pivotal ($p = 0.171$). Therefore, we reject the strategic low-cost hypothesis.

In the small electorates of $N = 5$ voters we study, the probability of being pivotal is rather high, and one may suspect that the strategic low-cost hypothesis has stronger predicting power in referenda in larger electorates. Tyran (2002) has experimentally tested the low-cost hypothesis in electorates of 30 voters using more elaborate tests. He finds that even in these relatively large electorates the low-cost hypothesis fails to predict individual voting behavior.

Strategy method vs. veil of ignorance. In our experiment, we apply a technique known as the strategy method to elicit individual voting decisions. Voters have to indicate *conditional* voting decisions in case they are allocated to a committee as a rich voter, as a middle-class or a poor voter. When taking their decisions, voters do not know which income class they will be allocated to. However, all voters in our experiment know that only the decision in the actually allocated income position determines voting outcomes and payoffs. This situation importantly differs from what some authors have called the "veil of ignorance" (Rawls 1971). It has long been recognized that uncertainty about one's future income position is a potential reason for why people agree on fair redistribution out of self interest (Harsanyi 1955). The situation these authors envisage, however, is one in which each voter has to take one *un*conditional voting decision without knowing in which income position he or she is going to be, and this voting decision is binding in all possible states of the world. Hence, purely self-interested voters may (ex ante) vote for (ex post) redistribution out of an insurance motive (see Frohlich and Oppenheimer 1990, 1992 for experimental studies on the issue). In contrast, the strategy method applied in our experiment does not provide incentives to vote fairly. The strategy method seems to be behaviorally equivalent to taking decisions sequentially, i.e., when decision-makers know their position (Brandts and Charness 2000).

Earned income positions and self-serving bias. In adapting the FS-framework, we assumed that the distribution of fairness preferences is the same in all income classes. This seems quite plausible since in our experiment, the initial relative income position was randomly and exogenously determined (by the experimenter). In natural settings, however, the intensity of inequality-aversion may importantly depend on the circumstances that have caused inequality. For example, survey studies suggest that beliefs about self- and exogenous determination of relative income positions strongly affect attitudes towards redistribution (Alesina and LaFerrara 2001, Fong 2001, Corneo and Grüner 2002).

There is also some experimental evidence supporting this view. For example, in ultimatum and in dictator games offers have been found to be less generous if proposers have earned their initial position than if it was randomly allocated (Hoffman et al. 1994). See Gächter and Riedl (2002) and Rutström and Williams (2000) for related studies on entitlements and redistribution.

Some limitations of the adapted FS-approach. It is important to note that our simple experiment served as a test of the prediction of the adapted FS-model against the standard model. The fact that the FS-framework was much more successful in predicting voting behavior should not, however, be interpreted to indicate that the FS-approach is the most useful of all possible approaches or that inequality aversion is the most important of all motives in voting (see Baron 2002 for a broad discussion). These conclusions are not warranted because our experiment was not designed to provide a comparative evaluation of different theories involving inequality aversion or of different non-selfish motives. For a comparative evaluation of alternative models of inequality aversion and a comparative evaluation of efficiency motives in simple redistribution games see Engelmann and Strobel (2002). Charness and Rabin (2000) provide an extended discussion as well as experimental evidence suggesting that maximin-preferences may be more important than inequality aversion in many instances.

We believe that the adapted FS-model is particularly well-suited to analyze pure redistribution. However, the FS-model is probably less appropriate in situations involving efficiency gains or losses from redistribution. Concerns for efficiency are not explicitly modelled in the FS-approach but may importantly affect voting decisions (Beckman et al. 2002).

Finally, the model of Fehr and Schmidt (1999) is a rational choice model. Therefore, framing cannot play a role. However, framing is quite likely to play an important role in political choice (Quattrone and Tversky 1988), and in how people vote in referenda (see Sausgruber and Tyran 2002 for an experiment involving taxation).

5   Concluding remarks

We have adapted a model by Fehr and Schmidt (1999) to analyze voting on redistribution with three income classes. We have opted for this approach because it is parsimonious, tractable, and generates clear-cut predictions that differ from those emanating from the standard model, assuming rationality and selfishness. The first main virtue of the adapted FS-framework is its parsimony. However, while a parsimonious modelling strategy facilitates mathematical tractability, it runs into the risk of ignoring factors which may importantly affect redistribution decisions. We believe that there is a plethora of such factors. Their relative importance, however, appears to be difficult to assess and seems to be strongly context-dependent (see section 4.3 for a discussion). That the adapted FS-approach indeed remains tractable is apparent from the simplicity of our formal reasoning. This is quite a remarkable advantage for a model that allows to discuss fairness issues in voting. However, the adapted FS-approach also has its limitations. For example, it does not allow to incorporate expectations and, for example, strategic voting. We therefore believe that while the adapted FS-approach can be used to investigate some interesting issues in voting on redistribution in simple settings, it may be inadequate to simply extrapolate its predictions to rich, context-laden environments. The third virtue of the adapted FS-approach is that it proved to clearly outperform the standard model of voting, in the sense that it provided much more accurate predictions of the outcome of the experimental referendum. We take this as a warning that conclusions derived from the standard model may be grossly misleading in voting on redistribution, and that incorporating fairness may provide more realistic predictions.

References

Akerlof, G.A. and Yellen, J.L. (1985): A Near-Rational Model of the Business Cycle, with Wage and Price Inertia. *Quarterly Journal of Economics* 100(5): 823-38.

Alesina, A. and Angeletos, G.-M. (2002): Fairness and Redistribution: US versus Europe. *Harvard Institute of Economics Working Paper* 1983, October 2002.

Alesina, A. and LaFerrara, E. (2001): Preferences for Redistribution in the Land of Opportunities. *NBER working paper* 8267, May 2001.

Amiel, Y. and Cowell, F.A. (2001): Attitudes to Risk and Inequality: A New Twist on the Transfer Principle. *Distributional Analysis Discussion Paper*, No. 56, STICERD, London School of Economics.

Amiel, Y. and Cowell, F.A. (1992): Measurement of Income Inequality - Experimental Test by Questionnaire. *Journal of Public Economics* 47(1): 3-26.

Baron, J. (2002): The Rational Voter: Moralistic Values in Judgments about Political Behavior. *Mimeo*, Department of Psychology, University of Pennsylvania.

Beckman, S.R.; Formby, J.P.; Smith, W.J. and Zheng, B. (2002): Envy, Malice and Pareto Efficiency: An Experimental Examination. *Social Choice and Welfare* 19(2): 349-67.

Boadway R. and Keen, M. (2000): Redistribution. In: A. Atkinson and F. Bourguignon (eds.): *Handbook of Income Distribution.* Amsterdam: North-Holland: Vol. 1, Ch. 12.

Boeri, T.; Börsch-Supan, A. and Tabellini, G. (2001): Would You Like to Shrink the Welfare State? A Survey of European Citizens. *Economic Policy* 16(32): 9-50.

Bolton, G. and Ockenfels, A. (2000): A Theory of Equity, Reciprocity, and Competition. *American Economic Review* 100(1): 166-93.

Brandts, J. and Charness, G. (2000): Hot vs. Cold, Sequential Responses and Preference Stability in Experimental Games. *Experimental Economics* 2(3): 227-38.

Brennan, G. and Lomasky, L. (1993): *Democracy and Decision.* Cambridge: Cambridge University Press.

Bütler, M. (2000): The Political Feasibility of Pension Reform Options: The Case of Switzerland. *Journal of Public Economics* 75(3): 389-416.

Butler, D. and Ranney, A. (1994, eds.): *Referendums around the World. The Growing Use of Direct Democracy.* Washington, D.C.: AEI Press.

Camerer, C.F. (2001): *Behavioral Game Theory:* Experiments on Strategic Interaction. Princeton: Princeton Univ Press. Forthcoming (April 2003).

Carlsson, F.; Daruvala, D. and Johansson-Stenman, O. (2001): Are People Inequality Averse or just Risk Averse? *Working Paper in Economics* No. 43, Department of Economics, Göteborg University.

Charness, G. and Rabin, M. (2000): Social Preferences: Some Simple Tests an a New Model. Department of Economics and Business, University Pompeu Fabra.

Clark, A.E. and Oswald, A.J. (1996): Satisfaction and Comparison Income. *Journal of Public Economics* 61(3): 359-81.

Corneo, G. and Grüner, H.P. (2002): Individual Preferences for Political Redistribution. *Journal of Public Economics* 83(1); 83-107.

Downs, A. (1957): *An Economic Theory of Democracy,* New York: Harper & Row.

Dufwenberg, M. and Kirchsteiger, G. (2001): A Theory of Sequential Reciprocity. *Working paper* Stockholm University, July 2001.

Engelmann, D. and Strobel, S. (2002): Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. *Working paper* Maastricht University, March 2002.

Falk, A. and Fischbacher, U. (2000): A Theory of Reciprocity. Institute for Empirical Research in Economics, *Working Paper* no. 6, University of Zurich.

Fehr, E. and Schmidt, K.M. (2001): Theories of Fairness and Reciprocity - Evidence and Economic Applications. Institute for Empirical Research in Economics, University of Zurich Working paper no. 75. Forthcoming in: *Advances in Economic Theory*: Eighth World Congress.

Fehr, E. and Schmidt, K.M. (1999): A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics* 114 (3): 817-68.

Fong, C. (2001): Social Preferences, Self-Interest, and the Demand for Redistribution. *Journal of Public Economics* 82(2): 225-46.

Frohlich, N. and Oppenheimer, J.A. (1992): *Choosing Justice:* An Experimental Approach to Ethical Theory. Berkeley: University of California Press.

Frohlich, N. and Oppenheimer, J.A. (1990): Choosing Justice in Experimental Democracies with Production. *American Political Science Review* 84(2): 461-77.

Gächter, S. and Riedl, A. (2002): Moral Property Rights in Bargaining. *Working paper* CREED University of Amsterdam.

Haltiwanger, J. and Waldman, M. (1989): Limited Rationality and Strategic Complements: The Implications for Macroeconomics. *Quarterly Journal of Economics* 104(3): 463-84.

Harsanyi, J.C. (1955): Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility. *Journal of Political Economy* 63: 309-21.

Hochman, H.M. and Rodgers, J.D. (1969): Pareto Optimal Distribution. *American Economic Review* 59(4): 542-57.

Hoffman, E.; McCabe, K.; Shachat, K. and Smith, V.L. (1994): Preferences, Property Rights and Anonymity in Bargaining Games. *Games and Economic Behavior* 7(3): 346-80.

Kirchgässner, G. (1992): Towards a Theory of Low-Cost Decisions. *European Journal of Political Economy* 8: 305-20.

Kirchgässner, G. (2000): *Homo oeconomicus.* 2nd ed. Tübingen: Mohr Siebeck.

Konow, J. (2002): Which is the Fairest One of All?: A Positive Analysis of Justice Theories. *Working Paper* Loyola Marymount University, June 2002.

Levine, D.K. (1998): Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics* 1(3): 593-622.

Meltzer, A.H. and Richard, S.F. (1981): A Rational Theory of the Size of Government. *Journal of Political Economy* 89(5): 914-27.

28

Mueller, D.C. (1989): *Public Choice II.* Cambridge: Cambridge University Press.

Mulligan, C.B. and Hunter, C.G. (2001): The Empirical Frequency of a Pivotal Vote. *NBER Working Paper* No.w8590. November 2001.

Persson, T. and Tabellini, G. (2000): *Political Economics.* MIT Press.

Quattrone G.A. and Tversky, A. (1988): Contrasting Rational and Psychological Analyses of Political Choice. *American Political Science Review* 82: 720-36.

Rabin, M. (1993): Incorporating Fariness into Game Theory and Economics. *American Economic Review* 83(5): 1281-302.

Rawls, J.A. (1971): *A Theory of Justice.* Cambridge: Belknap Press.

Rubinstein, A. (1998): *Modelling Bounded Rationality.* Cambridge, Mass.: MIT Press.

Rutström, E.E. and Williams, M.B. (2000): Entitlements and Fairness: An Experimental Study of Distributive Preferences. *Journal of Economic Behavior and Organization* 43(1): 75-89.

Sausgruber R. and Tyran, J.-R. (2002): Testing the Mill Hypothesis of Fiscal Illusion, *Mimeo*, University of St.Gallen.

Simon, H.A. (1997): *Models of Bounded Rationality.* Volume 3: Empirically Grounded Economic Reason. Cambridge, Mass.: MIT Press.

Solnick, S.J. and Hemenway, D. (1998): Is More always Better?: A Survey on Positional Concerns. *Journal of Economic Behavior and Organization* 37(3): 373-83.

Tyran, J.-R. (2002): Voting when Money and Morals Conflict. An Experimental Analysis of Expressive Voting. *Working Paper* 2002-07, University of St. Gallen, April 2002.

van der Heijden , E.C.M., Nelissen, J.H.M. and Verbon, H.A.A. (1997): Altruism and Fairness in a Public Pension System. *Journal of Economic Behavior and Organization* 32(4): 505-18.

van Winden, F. (2002): Interest Group Behavior and Influence. Forthcoming in: C.K. Rowley and F. Schneider (eds.): *Encyclopedia of Public Choice.* Boston: Kluwer Academic Publishers.

Appendix A: Robustness with respect to bounded rationality

This appendix shows that for majority voting the predictions of the FS-approach are robust with respect to a particular form of bounded rationality: random and across income classes uniformly distributed decision errors. With supermajority requirements, however, bounded rationality reduces the effects of inequality aversion.

We now assume that voters for whom $U_i(\mathbf{x}_1) > U_i(\mathbf{x}_0)$ vote for redistribution with probability $1 - \varepsilon$, voters for whom $U_i(\mathbf{x}_1) < U_i(\mathbf{x}_0)$ vote against redistribution with probability $1 - \varepsilon$, and voters with $U_i(\mathbf{x}_1) = U_i(\mathbf{x}_0)$ vote for redistribution with probability $\pi^{ind.} = 0.5$. In addition, we assume that the number of voters of each type is large such that $\varepsilon \in [0, 0.5]$ is the percentage of voters who are boundedly rational. Finally, we assume that the incidence of bounded rationality is exogenous and in particular independent of income.

To analyze how bounded rationality interacts with fairness preferences, we first derive the number of yes-voters under the assumption of bounded rationality as defined above. With perfectly rational voters, the expected approval for policy R_{rp} among rich voters is given by λn_r (see (12)). In the presence of boundedly rational voters this modifies to $\varepsilon(1 - \lambda)n_r + (1 - \varepsilon)\lambda n_r$. Therefore, the additional number of rich yes-voters due to bounded rationality is $\varepsilon n_r(1 - 2\lambda)$. Similarly, for the middle class the expected approval changes from $(1 + \mu)n_m/2$ to $\varepsilon(1 + \mu)n_m/2 + (1 - \varepsilon)(1 + \mu)n_m/2$, resulting in a decrease in the number of approving middle-class of $-\varepsilon \mu n_m$. Finally, fully rational poor voters all vote for R_{rp}, and because of bounded rationality $-\varepsilon n_p$ is the reduction in the number of poor voters who vote for redistribution. Hence, the number of yes-voters changes by $\varepsilon[(1 - 2\lambda)n_r - \mu n_m - n_p]$, and, therefore, the policy is accepted if

$$y(R_{rp}, \varepsilon) = \lambda n_r + \frac{(1 + \mu)}{2}n_m + n_p + \varepsilon[(1 - 2\lambda)n_r - \mu n_m - n_p] > qN. \quad (17)$$

Consider first the case in which all voters are completely egoistic ($\alpha_i = \beta_i = 0$ for all i). In the absence of inequality aversion, (17) reduces to $\varepsilon[n_r - n_p] >$

$n_r q + n_m(q - 0.5) + n_p(q - 1)$. For $q = 0.5$, (17) further reduces to $\varepsilon[n_r - n_p] > 0.5[n_r - n_p]$. Since $\varepsilon \in [0, 0.5]$ this inequality is never satisfied for $n_r - n_p > 0$. Therefore, bounded rationality does not induce acceptance when the standard theory predicts rejection of R_{rp}. We conclude that the prediction of the standard theory is robust to this particular form of bounded rationality.

We now show that the prediction of our model incorporating inequality aversion is also robust to this type of bounded rationality. Solving (17) for λ, we obtain the acceptance frontier $\overline{\lambda}$, i.e., the critical percentage of rich voters necessary to tip the balance in favor of redistribution

$$\overline{\lambda}_\varepsilon = max \left[0, \frac{1}{1 - 2\varepsilon} \left\{ (q - \varepsilon) + \left(q - \frac{1 + \mu}{2} + \varepsilon\mu \right) \frac{n_m}{n_r} + (q - 1 + \varepsilon)\frac{n_p}{n_r} \right\} \right]. \quad (18)$$

The effect of bounded rationality on the acceptance frontier is $\partial\overline{\lambda}_\varepsilon / \partial\varepsilon|_{\overline{\lambda}_\varepsilon > 0} = [N(2q - 1)] / [n_r(1 - 2\varepsilon)^2]$. For majority voting ($q = 0.5$), $\partial\overline{\lambda}_\varepsilon / \partial\varepsilon|_{\overline{\lambda}_\varepsilon > 0} = 0$. Therefore, the presence of bounded rationality does not dilute the effect of inequality aversion discussed in sections 3.1 and 3.2. For supermajority requirements, however, $\partial\overline{\lambda}_\varepsilon / \partial\varepsilon|_{\overline{\lambda}_\varepsilon > 0} > 0$, and bounded rationality does reduce the effect of inequality aversion.

31

Appendix B: Instructions

Original instructions were in German.

General instructions You are participating in an economics experiment which is financed by the Austrian National Bank. The purpose of this experiment is to analyze voting behavior. You have the opportunity to earn money in this experiment. All the money earned during the experiment will immediately be paid to you at the end of the experiment in cash. You are not allowed to communicate with other participants during the experiment. If you have any questions, please ask us. We will gladly answer your questions in private. It is very important that you follow this rule. Otherwise the results of this experiment will be worthless from a scientific perspective. Here is a short description of the experiment. Detailed instructions follow below.

In this experiment, participants are assigned to roles A, B, and C. These roles differ with respect to their initial endowment of money. All participants vote on a monetary transfer from persons in role A to persons in role C. Whether a transfer takes place depends of your own decision as well as on the decisions of the other participants. The following pages explain the details of the experiment.

Detailed Instructions You and 4 more persons are in a group. Decisions in all other groups are irrelevant for your group. No participant will learn about the decisions of the others. That is, your decision is anonymous.

You and 4 other persons in your group are assigned to the following roles:

- 2 persons are in role A.
- 2 persons are in role B.
- 1 person is in role C.

The assignment of roles determines your initial **endowment of money**:

- The 2 persons in role A get an initial endowment of **ATS 250**.
- The 2 persons in role B get an initial endowment of **ATS 185**.
- The 1 person in role C gets an initial endowment of **ATS 60**.

Participants vote on the following proposal:

Proposal The 2 persons in role A pay ATS 50 to finance a transfer ATS 100 to the person in role C. The 2 persons in role B receive a deduction ATS 5 from their endowment.

Voting rule The proposal is accepted if the majority votes for it (that is, if 3, 4, or 5 persons vote YES). Otherwise it is rejected. Everybody has to decide in *each* role whether to vote YES or NO. That is everybody decides once in role A, once in role B, and once in role C. After everyone has voted it will be *randomly determined* which role participants actually are in. The probabilities of role A and B are two fifth. The probability of role C is one fifth. *Relevant for your payment* are only those decisions that you and the other participants have taken in the assigned roles.

Example: Suppose you are assigned role A. In that case only your decision in role A is relevant (see 1st part of the voting form); in that case the decisions you have taken in the role of B and C are irrelevant for the outcome of the voting and your payoff.

Consequences of voting

In role **A** you either pay ATS 0 or ATS 50 to finance a transfer to the person in role C. If the proposal is accepted you are paid ATS 200 (=250 *minus* 50) at the end of the experiment. If the proposal is rejected you are paid ATS 250.

In role **B** you are paid ATS 180 (185 *minus* 5) if the proposal is accepted. You are paid ATS 185 if the proposal is rejected.

In role **C** you either get a transfer of ATS 0 or ATS 100. If the proposal is accepted you are paid ATS 160 (=60 *plus* 2 x 50) at the end of the experiment. If the proposal is rejected you are paid ATS 60.

What you have to do

Decide in each of the three roles on the separate voting form whether you vote YES or NO. In addition, indicate your expectation on how many of the other 4 voters in your group vote for the proposal. If your expectation is correct, i.e., if it is equal to the actual number of YES-votes in the relevant case you get paid an additional ATS 50 (If you predict correctly you are paid ATS 50 independent of the outcome of the voting and independent of your decision on the proposal).

Voting Form

Please vote in each role A, B, and C on the proposal and indicate your expectation on the number of YES-votes in your group.

Given that you are in role **A**:

• Do you vote for or against the proposal (please tick)?

◯ YES ◯ NO

• How many of the other 4 voters in your group do you expect to vote for the proposal? Please tick a number between 0 and 4.

0	1	2	3	4

Given that you are in role **B**:

• Do you vote for or against the proposal (please tick)?

◯ YES ◯ NO

• How many of the other 4 voters in your group do you expect to vote for the proposal? Please tick a number between 0 and 4.

0	1	2	3	4

Given that you are in role **C**:

• Do you vote for or against the proposal (please tick)?

◯ YES ◯ NO

• How many of the other 4 voters in your group do you expect to vote for the proposal? Please tick a number between 0 and 4.

0	1	2	3	4