A Little Fairness may Induce a Lot of Redistribution in Democracy

Jean-Robert Tyran and Rupert Sausgruber *

revised, August 2004

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.

*Tyran: University of Copenhagen, Institute of Economics, Studiestraede 6, DK-1455 Copenhagen. E-mail: Jean-Robert.Tyran@econ.ku.dk. Sausgruber: University of Innsbruck, Department of Public Economics, Universitätsstr. 15, A-6020 Innsbruck. E-mail: Rupert.Sausgruber@uibk.ac.at. We thank Jonathan Baron, Simon Gächter, Randolph Sloof, and seminar participants at CREED (Amsterdam) for helpful comments. We are grateful for financial support by the Austrian National Bank (Jubiläumsfonds) under Project no. 9134.

1 Introduction

In his authoritative textbook on public choice, Dennis Mueller (1989: 456) concludes 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 (e.g., Roberts 1977, Persson and Tabellini 2000). 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 redistribution 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 2003: Ch. 2 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 heterogeneous fairness preferences (e.g., Rabin 1993, Levine 1998, Bolton and Ockenfels 2000).

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 (non-voting) experimental studies having measured social preferences to predict voting on redistribution. These predictions are applied to a particular redistribution proposal that we implement in a laboratory experiment to test the adapted FS-model against the standard model.

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.

We proceed as follows. Section 2 explains how the FS-framework can be adapted to analyze voting on redistribution. Section 3 identifies conditions under which fairness preferences can explain democratic redistribution. Section 4 reports experimental results and section 5 provides some concluding remarks.

2 Fairness and voting on redistribution

One possible interpretation of fairness is inequity-aversion (see Konow 2003 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 Amiel and Cowell (1992), Clark and Oswald (1996), Solnick and Hemenway (1998), and Johansson-Stenman et al. (2002).

We now provide intuitions for why 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 exclusively motivated by material self-interest they would reject the redistribution policy by a narrow margin. The first reason 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 reason for why fairness preferences can have a large effect on voting outcomes is the presence of voters who are materially unaffected by redistribution. These voters may vote for redistribution 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.[1]

[1]An important advantage of the FS-approach is that it allows to model the effect of fairness pref-

A third reason is provided by strategic low-cost considerations. According to this explanation voters consider the fact that they are not pivotal most of the time (Mulligan and Hunter 2003). If an individual voter does not affect the outcome, he or she may just as well cast a fair vote to feel good (Brennan and Hamlin 2000). 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. However, we do test the expectations-based low-cost theory in our experiment (see section 4.3).

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 the difference between i's income and that of other voters j.

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

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 ($\alpha_i \geq \beta_i$). 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.

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

erences 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.

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 $w_{kh} \equiv \frac{n_h}{N - n_k}$ the weight a voter in income class k attaches on inequality with respect to class h, where $\sum_{k \neq h} w_{kh} = 1$.

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

$$
\begin{aligned}
U_i(\mathbf{x}_0 \,|\, r) &= x_r - \beta_i \left[w_{rm}(x_r - x_m) + w_{rp}(x_r - x_p) \right], && (2) \\
U_i(\mathbf{x}_0 \,|\, m) &= x_m - \alpha_i w_{mr}(x_r - x_m) - \beta_i w_{mp}(x_m - x_p), && (3) \\
U_i(\mathbf{x}_0 \,|\, p) &= x_p - \alpha_i \left[w_{pr}(x_r - x_p) + w_{pm}(x_m - x_p) \right]. && (4)
\end{aligned}
$$

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. 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, we assume that $x_r - t_r > x_m > x_p + b_p$. The vector $\mathbf{x_1}$ denotes the post-redistribution income distribution. In our example, $\mathbf{x_1} = (x_r - t_r,\ x_m,\ x_p + b_p)$. The utility of a voter i in income class k in case of acceptance of R_{rp} is

$$
\begin{aligned}
U_i(\mathbf{x}_1 \,|\, r) &= x_r - t_r - \beta_i \left[w_{rm}(x_r - t_r - x_m) + w_{rp}(x_r - t_r - x_p - b_p) \right], && (5) \\
U_i(\mathbf{x}_1 \,|\, m) &= x_m - \alpha_i w_{mr}(x_r - t_r - x_m) - \beta_i w_{mp} \left(x_m - x_p - b_p \right), && (6) \\
U_i(\mathbf{x}_1 \,|\, p) &= x_p + b_p - \alpha_i \left[w_{pr}(x_r - t_r - x_p - b_p) + w_{pm}(x_m - x_p - b_p) \right]. && (7)
\end{aligned}
$$

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, 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, if $U_i(\mathbf{x}_1 \,|\, k) > U_i(\mathbf{x}_0 \,|\, k)$, i votes for the proposal with probability $\pi = 1$, if $U_i(\mathbf{x}_1 \,|\, k) = U_i(\mathbf{x}_0 \,|\, k)$, i randomizes between voting yes and no with probability $\pi^{ind.} = 0.5$, and if $U_i(\mathbf{x}_1 \,|\, k) < U_i(\mathbf{x}_0 \,|\, k)$, i votes 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 votes for policy R_{rp} if $U_i(\mathbf{x}_1 \,|\, r) > U_i(\mathbf{x}_0 \,|\, r)$, i.e., if

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

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 $\overline{\beta}(R_{rp})$.

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 the percentage of rich voters voting for R_{rp} by $\lambda \in [0, 1]$. This percentage is a decreasing function of $\overline{\beta}(R_{rp})$:

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

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 $\beta_i > 0$. The restriction $\alpha_i \geq \beta_i \geq 0$ implies that we only have to consider the voters with strictly positve α_i. Define $\mu = const.$ as the percentage of voters with $\alpha_i > 0$. Hence, the number of voters with $\alpha_i = 0$ in income class n_m is $(1 - \mu)\, n_m$. These voters are indifferent about whether redistribution is implemented. Since we assume that indifferent voters randomize between voting yes and no with probability $\pi^{ind.} = 0.5$, 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. From (8) to (11), it follows that the total number of yes-votes $y(R_{rp})$ for R_{rp} is

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

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. Hence, R_{rp} is accepted if $y(R_{rp}) > qN$.

3 When fairness induces redistribution

We now discuss two reasons why fairness preferences may induce redistribution in democratic referenda. First, when all voters have material interests at stake, *few* sufficiently inequality averse voters deciding against material self-interest may tip the balance. Second, voters who are *only weakly* inequality averse may vote for redistribution if it reduces inequality but is costless to them.

To analytically 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

Result 1 summarizes the main conclusion of this section.

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. The frontier 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.

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 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).

This discussion has important implications for the robustness of the predictions of the standard theory for referendum outcomes. Remarkably, the standard theory is not necessarily unreliable if some voters are fairness-minded. Rather, standard theory's robustness depends on the proportion of rich and poor voters.

3.2 Low-cost decisions of monetarily unaffected voters

While middle-class voters are not affected in monetary terms by policy R_{rp} its adoption affects their relative income position. Hence, minimally inequality-averse middle-class voters vote for R_{rp}.

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 $\mu > 2q - 1$.*

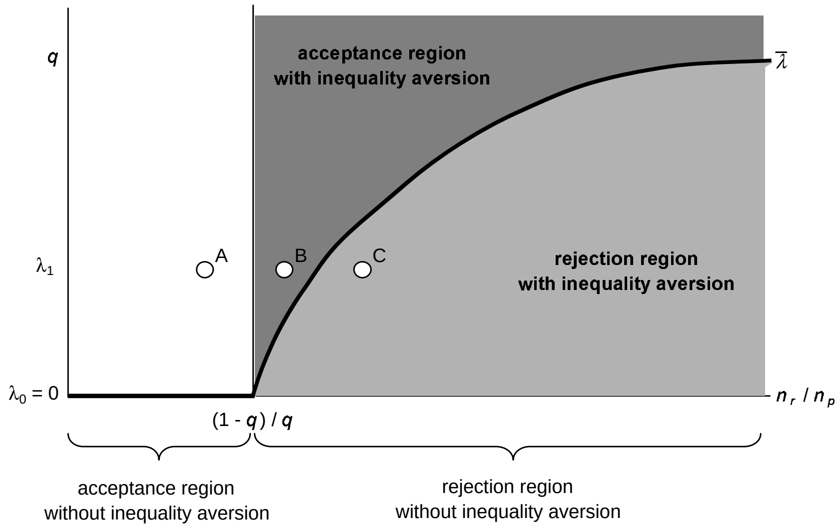

Figure 1: Acceptance frontier $\overline{\lambda}$ for R_{rp} if $n_m = 0$.

In the presence of middle-class voters ($n_m > 0$), 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 $A = q + \frac{1}{n_r} \left[\left(q - \frac{1}{2} \right) n_m + (q-1) n_p \right]$ is the intercept and $B = -n_m/2n_r$ is the slope in the λ-μ-space for $\overline{\lambda} > 0$ (see figure 2). In constructing the acceptance region in the λ-μ-space, we use the restriction on preference parameters suggested by FS 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 $\mu \geqslant \lambda$. The acceptance region of R_{rp} is bounded from below by the frontier $\overline{\lambda}$, and bounded from above by the restriction $\mu \geqslant \lambda$ (see figure 2). The negative slope B means that the higher the 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. This effect is more pronounced if the percentage of fairness-minded middle-class voters μ is large. To see this, note that $\partial\overline{\lambda}/\partial n_m|_{\overline{\lambda}>0} = [2q - 1 - \mu]/2n_r \leq 0$ at $q = 0.5$. If the middle class is at least weakly inequality averse

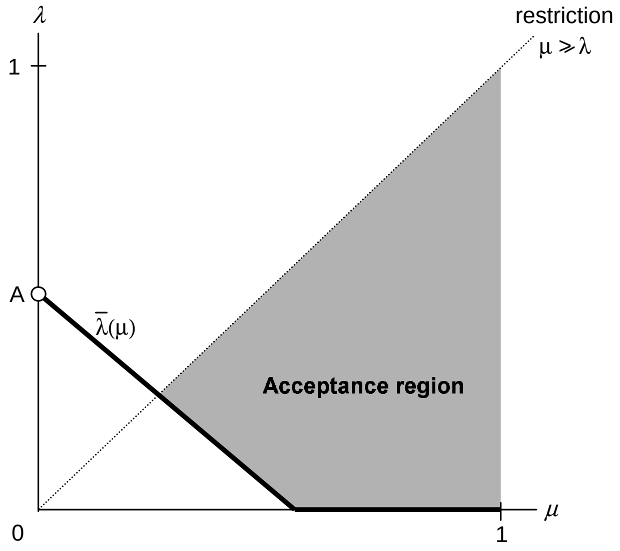

Figure 2: Combinations of λ and μ implying the acceptance of R_{rp} for $n_m > 0$.

$(\mu > 0)$, an increase in n_m reduces the critical $\overline{\lambda}$ and it increases the actual λ for any given distribution of preferences β_i.

Supermajority requirements $(q > 0.5)$ reduce the effect of fairness preferences on two accounts. First, a higher quorum 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 ambiguous $(\partial\overline{\lambda}/\partial n_m|_{\overline{\lambda}>0} \gtreqless 0)$ since 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 middle class has a strong effect on redistribution if it involves classes above and below the own income class (e.g., R_{rp}). In contrast, the effects of materially unaffected voters are less pronounced for an 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. Costless redistribution (in the sense that all tax revenues are redistributed) among voters who are both richer or poorer than voter i does not affect voter i's utility in our framework.

4 An experimental test

We now report results from a simple experiment to test the adapted FS-framework against 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 differ. 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 $t_r = 50$ on each 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 voters approved.

The experiment was explained to subjects in written instructions[2]. We used the neutral labels A, B, and C for rich, middle, and poor voters, respectively. When subjects made 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 made their decisions. Subjects knew that they had to make their decisions only once and that they were paid according to their decisions in the randomly assigned roles and committees.

Predictions. The standard theory (with $\alpha_i = \beta_i = 0$ for all i) predicts that 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

[2] Instructions are available from the authors on request.

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 the following discrete distribution of fairness preferences: 30% of the voters have $\alpha_i = 0$, 30% have $\alpha_i = 0.5$, and 40% have $\alpha_i \geq 1$. Furthermore, 30% of voters have $\beta_i = 0$, 30% have $\beta_i = 0.25$, and 40% have $\beta_i = 0.6$. 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 $\beta_i > t_r/[w_{rm}(t_r - t_m) + w_{rp}(t_r + b_p)] \equiv \overline{\beta}(R_{rp})$. 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 $\alpha_i w_{mr}(t_r - t_m) + \beta_i w_{mp}(b_p + t_m) > t_m$ (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% of the rich ($\lambda \leq 0.4$), 70% of the middle-class ($\mu = 0.7$), and 100% of the poor to vote for redistribution. On average, the proposal R_{rp} is predicted by FS to be accepted by a margin of $[\lambda n_r + \mu n_m + n_p]/N = 64\%$ (see table 1). In contrast, the standard theory assuming strictly self-interested voters predicts that 0% of the rich, 0% of the middle-class, and 100% of the poor voters to approve. Hence, the standard theory predicts rejection with an average approval rate of $n_p/N = 20\%$.

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.[3] The main result of our experiment is summarized as follows.

[3]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.

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.*

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
total	20	64	61.4

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% yes-votes. This observation is much closer to the prediction of the FS-framework of an average approval rate of 64% than of the standard theory. 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% (= 27/80), and for the middle-class it is 70.0% (= 56/80) (see table 1). These observations are far away from the predictions of the standard theory (0% in both cases), but strikingly close to the predictions of the adapted FS-framework of $\lambda \leq 40\%$, and of $\mu = 70\%$.

Table 1 also reports the finding that 96.3% of poor voters voted for redistribution. This result is interesting despite the fact both theories predict that 100% of poor voters to approve 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 make 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% of the subjects take decisions that are consistent with strict self-interest, 37.5% are consistent with weak inequality aversion, and 30% with strong inequality aversion. The remaining 7.5% of subjects were consistent with neither type. 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 confidence intervals 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 s yes-votes in our sample? Call ρ the unknown population proportion of voters in class k that would vote in favor of redistribution. The interval is constructed such that it contains the true parameter with probability $(1-\alpha)$. 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 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 is far from the prediction $\mu = 0.0$ of the standard theory. Finally, for the poor voters, the upper bound of the confidence limit is close to 100%.[4]

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 mentioned in section 2, a strategic low-cost argument provides a third potential explanation. According to this explanation, 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 at the same time they took the voting decision. 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% (= 9/21) of these voters voted

[4] Note that the confidence interval cannot contain the value 1 by definition of the binomial distribution. However, it can come arbitrarily close to this value.

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% (= 18/59), but not significantly so ($p = 0.304$ according to a χ^2 test). The same is true of the middle class. Pivotal middle-class voters tend to approve more (23/29) than non-pivotal voters (33/51), but not significantly so ($p = 0.171$). 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 expect the strategic low-cost hypothesis to have more predicting power in larger electorates. However, Tyran (2004) finds in experimental referenda with 30 voters that the low-cost hypothesis fails to predict individual voting behavior.

Strategy method vs. veil of ignorance. In our experiment, we use 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, middle-class or a poor voter. When making 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). However, in this situation each voter has to take one *un*conditional voting decision without knowing his or her future income position, and this voting decision is binding for 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 1992). In contrast, the strategy method applied in our experiment does not provide incentives to vote in a fair manner in theory. Whether the use of the strategy method elicits different behavioral responses than when subjects take their decisions knowing their income position (i.e., in a 'hot' state) is an open issue (e.g., Brandts and Charness 2000, Bolton and Ockenfels 2002, Weber et al. 2004), and seems to depend on context.

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. 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 and that these beliefs are to some extent self-serving (Alesina and LaFerrara 2001, Fong 2001, Johansson-Stenman and Martinsson 2003). There is also some experimental evidence supporting this view (e.g., Hoffman et al. 1994).

Some limitations of the adapted FS-approach. 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. 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 comparison of alternative models of inequality aversion in simple redistribution games see, for example, Charness and Rabin (2002).

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). Whether distributive equity has more or less behavioral attraction power than efficiency seems to be a debated issue (compare Engelmann and Strobel 2002 and Bolton and Ockenfels 2002).

Finally, the FS-model 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. 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 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 in an 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

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. (1992): Measurement of Income Inequality - Experimental Test by Questionnaire. *Journal of Public Economics* 47(1): 3-26.

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.

Bolton, G. and Ockenfels, A. (2002): The Behavioral Tradeoff between Efficiency and Equity when a Majority Rules. Working Paper, Max-Planck-Institute for Research into Economic Systems, Jena.

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 Hamlin, A. (2000): *Democratic Devices and Desires.* New York: Cambridge University Press.

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. (2003): *Behavioral Game Theory: Experiments on Strategic Interaction.* Princeton: Princeton University Press.

Charness, G. and Rabin, M. (2002): Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics* 117(3): 817-69.

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

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

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.

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

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.

Johansson-Stenman, O., Carlsson, F. and Durvala, D. (2002): Measuring Future Grandparents' Prefrences for Equality and Relative Standing. *Economic Journal* 112(479): 362-83.

Johansson-Stenman, O. and Martinsson, P. (2003): Anyone for Higher Speed Limits? Self-Interested and Adaptive Political Preferences. Forthcoming in *Public Choice*.

Konow, J. (2003): Which is the Fairest One of All? A Positive Analysis of Justice Theories. *Journal of Economic Literature* 40(4): 1188-239.

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.

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

Mulligan C.B. and Hunter C.G. (2003): The Empirical Frequency of a Pivotal Vote. *Public Choice* 116(1): 31-54

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

Quattrone G.A. and Tversky, A. (1988): Contrasting Rational and Psychological Analyses of Political Choice. *American Political Science Review* 82(3): 720-36.

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

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

Roberts, K.W.S. (1977): Voting over Income Tax Schedules. *Journal of Public Economics* 8: 329-40.

Sausgruber R. and Tyran, J.-R. (2002): Testing the Mill Hypothesis of Fiscal Illusion. Forthcoming in *Public Choice*.

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. (2004): Voting when Money and Morals Conflict. An Experimental Analysis of Expressive Voting. *Journal of Public Economics* 88(7): 1645-64.

Weber, R.A., Knez, M. and Camerer, C.F. (2004): Timing and Virtual Observability in Ultimatum Bargaining and 'Weak Link' Coordination Games. *Experimental Economics* 7(1): 25-48.