Tailored Bayesian Mechanisms: Experimental Evidence from Two-Stage Voting Games*

Dirk Engelmann[†] Hans Peter Grüner[‡]

February 15, 2015

Abstract

Optimal voting rules have to be tailored to the underlying distribution of preferences. However, in practice a benevolent social planner who has the necessary political authority to perform this task would be difficult to identify. This paper shows that the introduction of a stage at which agents may themselves choose voting rules according to which they decide in a second stage may increase the sum of individuals' payoffs if players are not all completely selfish. We run three closely related treatments (plus two control treatments) to understand how privately informed individuals decide when they choose voting rules and when they vote. Efficiency concerns play an important role on the rule choice stage, whereas selfish behavior seems to dominate at the voting stage. Accordingly, in a setting with an asymmetric distribution of valuations groups that can choose a voting rule do better than those who decide with a given simple majority voting rule.

Keywords: Two-stage voting, Bayesian voting experiments, revelation principle.

JEL-Classification: C91, D70, D82

*We thank Jana Friedrichsen, Sander Renes, Thomas Tröger and Philipp Zahn as well as seminar participants at the University of Zürich, WZB Berlin, DICE Düsseldorf, University of Mannheim, Utrecht University, Tilburg University, Tinbergen Institute/CREED-University of Amsterdam, Erasmus University Rotterdam, University of Erlangen-Nürnberg, University of Konstanz, Royal Holloway, University of Edinburgh, Maastricht University, Max-Planck Institute Bonn, CERGE-EI Prague, University of Mainz, the Schumpeter seminar at HU and FU Berlin, HECER Helsinki, University of British Columbia, and participants at the International Economic Science Association Conference in New York 2012, the Economic Science Association European Conference in Cologne 2012, and the 2013 meeting of the economic theory group of the German Economic Association for their useful comments. We also thank Johannes Bubeck and Christoph Siemroth for excellent research assistance. Financial support from the German research foundation (DFG), SFB 884 is gratefully acknowledged.

†Humboldt-Universität zu Berlin, Department of Economics, Spandauer Str.1. 10099 Berlin, Germany; Centre for Experimental Economics, University of Copenhagen; CESifo, Munich; Economics Institute of the Academcy of Sciences of the Czech Republic. E-mail: dirk.engelmann@hu-berlin.de

‡University of Mannheim, Department of Economics, L7, 3-5. 68131 Mannheim, Germany and CEPR, London. E-mail: gruener@uni-mannheim.de.

1 Introduction

Many collective decisions in civilized societies are governed by rules. Little is known about what motivates people when they select rules that guide decision making. This paper studies the choice of decision making rules with a series of voting experiments.

Our experiments focus on the interim choice of decision rules, i.e., the choice of rules by people whose preferences have already realized. It is a key politico-economic insight that rules governing collective decision making should ideally be determined before individual preferences about outcomes have realized. Consensual and efficient choices are easier to obtain on the constitutional level because individuals cannot be sure about their situation and preferences in the future (Brennan and Buchanan, 1985).[1] While convincing in theory, the concept of a choice of decision making rules, ignorant of own preferences, has limited applicability in practise beyond broadly applicable rules. First, new upcoming collective decision problems may require tailor-made decision rules – such as voting or procedural rules – that cannot be fixed in a constitution beforehand. Second, decision problems often appear simultaneously with private knowledge about individual preferences over possible outcomes. As a result, by the time adjustments to the decision rule have to be made that take specifics of the decision problem at hand into account, these adjustments cannot be made ignorant of own preferences anymore. Therefore, it is important to understand, whether rules can be tailored to major features of the underlying problem at the interim stage, when preferences over outcomes have already realized.

Indeed, interim choices of rules are far more widespread in practice than one would believe based on theoretical reasoning about their usefulness. One important political body that fixes its own rules at the interim stage is the U.S. Congress. According to article I, Section 5 of the U.S. constitution, "Each House may determine the Rules of Its Proceedings". This is done on a case by case basis, which usually implies that preferences of the members of congress have at least to some extent already realized. While these rules choices are mostly procedural (see Oleszek, 2014, for a detailed overview), they may still have a major influence on the chances of a particular outcome. The power of the rules committee has been confirmed by the former House minority leader Robert H. Michel, who argued that "Procedure hasn't simply become more important than substance - it has (...) become the substance of our deliberations. Who rules House procedures rules the House (...)" (quote from Oleszek, 2014, p.12). Also beyond the U.S. Congress it is not uncommon for committees to first decide about procedural rules or voting thresholds

[1] The applicability of rules across numerous decision problems and their validity for longer time horizons is crucial for agreements on efficient rules being easier than agreements on efficient outcomes. Along similarly lines, Rawls (1972) argued that distributional rules should ideally correspond to the ones that an individual uninformed about his own position in society would pick.

before debating and deciding about the issue itself. These rules often take information about the underlying distribution of preferences at least implicitly into account.[2]

There are many different types of collective decision making mechanisms. Our empirical analysis concentrates on the choice among a prominent class - voting mechanisms. Voting mechanisms are widely used for the aggregation of individuals' preferences in binary collective decisions. They are easy to understand and they neither make use of monetary transfers nor do they ask individuals for complex statements about the intensity of their preferences.[3] Moreover, truthful voting is a dominant strategy for selfish voters in any private values setup and hence their behavior can be well predicted. While these are two desirable features, a major problem related to the practical construction of optimal voting mechanisms is that the threshold for a majority vote has to be properly tailored to the underlying distribution of preferences. Several theoretical papers address how optimal majority rules have to be adjusted when the underlying distribution of preferences changes,[4] but very little is known about how this flexibility could be achieved in practice.

---

[2] Another important example of rule choice at the interim stage is the United Nations General Assembly which occasionally chooses to proceed under consensus when the choice is particularly controversial. The Encyclopedia of the Nations notes that "In effect, each member state of the League (of Nations) had the power of the veto, and, except for procedural matters and a few specified topics, a single "nay" killed any resolution. Learning from this mistake, the founders of the UN decided that all its organs and subsidiary bodies should make decisions by some type of majority vote (though, on occasion, committees dealing with a particularly controversial issue have been known to proceed by consensus)." (http://www.nationsencyclopedia.com/United-Nations/Comparison-with-the-League-of-Nations-VOTING.html). Academia also provides examples where at least informally a decision rule is chosen at a time when private preferences have already been realized, such as recruitment committees that may decide to move only based on consensus if they consider a candidate to be potentially controversial.

[3] It is well known that a public good should be provided when the sum of individuals' willingness to pay exceeds the cost of provision. This may be the case when less than a simple majority of citizens has a high willingness to pay but also when a supermajority has a small willingness to pay. More complex mechanisms such as the Vickrey-Clarke-Groves mechanisms do not require any information about the distribution of individuals' willingness to pay at the design stage. However, they have the disadvantage that they require the use of money.

[4] Schmitz and Tröger (2012) study optimal majority rules in a setup with heterogeneous preferences. The simple majority rule is efficient if preferences are stochastically independent and their distribution is not biased in favor of one alternative. In setups with a skewed distribution, one has to adjust the majority threshold in order to avoid that a majority of voters with only a small preference for one outcome makes it impossible to realize substantial welfare gains. For example, Deb, Ghosh, and Seo (2011) show that the optimal threshold quota in referenda for binary public goods depends positively on average intensity of opposition and negatively on average intensity of support. A literature that follows Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1998) studies the effects of various voting thresholds in a Bayesian jury setup with identical preferences and heterogeneous information where

There are several reasons why it is difficult in practice to properly adjust the voting rule to the underlying problem. One theoretical option would be to design a "complete" constitution that lists all possible decision problems. Obviously, this option is far too costly. Not surprisingly, many constitutions very broadly distinguish only a few categories of decision problems and allocate different majority thresholds to these categories (for example, changes of the constitution itself often require a two-thirds majority). Another theoretical option is to let some benevolent planner who knows the underlying distribution of preferences adjust the voting rules to that distribution. This option is unrealistic in practice because perfectly benevolent individuals may be rare and, even if they exist, hard to identify. Furthermore, their political legitimacy for enforcing such a rule may be questioned.

Instead, one may consider to let voters themselves decide on the appropriate majority threshold whenever a new issue comes up. In theory, this can properly solve the problem of selecting an optimal mechanism if this selection takes place at the ex-ante stage where individuals do not yet know their own preferences but know the underlying distribution. Specifically, if all agents' preferences are drawn from the same distribution, they would all suggest a rule that maximizes expected total welfare. However, it is not straightforward that a rule-choice stage at a point when agents are already informed about their preferences (i.e., at the interim stage) increases welfare. As argued above, in many important practical applications, we consider this to be the relevant situation because we expect that individuals form their beliefs about the distribution of other voters' preferences only after a new issue comes up and not before their own preferences are formed.

This paper studies whether the introduction of a stage at which agents may choose the rules according to which they decide may help to solve the problem that voting inefficiently aggregates preferences and whether adding a rule-choice stage thus increases the sum of individuals' payoffs. We conduct an experiment with three closely related treatments to understand how individuals decide when they choose voting rules at the interim stage. More specifically, we consider an environment in which agents can vote on a status quo with predetermined identical payoffs (alternative B) and an alternative with stochastic and often unequal payoffs (alternative A). The difference between both payoffs is hereafter called an agent's valuation of decision A. Agents privately learn about their randomly chosen valuation in the first stage of the game. Then the agents decide on the voting rule and are informed about the implemented rule before voting takes place. In our setup, introducing a procedural choice stage is useless when agents act selfishly because they

strategic voting arises from some voting thresholds.

should then pick the rule which makes their preferred outcome most likely to obtain.[5,6] Consider, for example, the case where the rule itself is determined by a vote under majority rule. All second-stage rules which favor alternative A in the sense that they require a supermajority for alternative B beat all other rules if outcome A makes the majority of agents better off. This is why the two-stage procedure should yield the same results as a simple one-stage vote if all players are selfish. However, our paper shows theoretically that if subjects exhibit efficiency concerns,[7] then the interim rule-choice stage is not redundant. Our experiments study to what extent efficiency-concerned subjects choose welfare-enhancing voting rules, which fit to the underlying distribution of preferences.

The first main finding of our experiments is that agents often do not act selfishly at the rule-choice stage, but instead propose rules which are biased towards the efficient (sum-of-payoffs maximizing) rule. As a result, sequential voting mechanisms may increase social welfare. We consider both symmetric and asymmetric distributions of types. With a symmetric distribution the simple majority rule is efficient and thus at the rule-choice stage efficiency-concerned players deviate from the prediction for selfish players accordingly, that is, if they have a small absolute valuation they choose a rule that deviates from their selfish rule towards majority voting. Because majority voting is the ex-ante efficient rule for a symmetric distribution of valuations, we cannot expect any efficiency gains from adding a rule-choice stage. Our first treatment with a symmetric and fine distribution of valuations therefore serves as an initial test to what degree participants deviate from selfish rules and selfish voting. Our second treatment with a strongly asymmetric distribution of valuations instead permits to investigate the possibility of obtaining efficiency gains through the endogenous determination of voting thresholds. With an asymmetric distribution, the efficient majority threshold changes and efficiency-oriented subjects' deviations from proposing the

[5] More precisely, this holds if all are selfish and this is common knowledge. Otherwise, it is at least conceivable that even if all are selfish, but do not expect all others to be selfish, they could try to use a non-selfish vote at the rule choice stage to signal something about their valuation which might trigger non-selfish voting behavior in others, from which they would in turn benefit. However, at least for the specific rule choice procedures that we study in our first two treatments, this is not a concern, because whenever one's own rule choice becomes relevant and known, one can unilaterally determine the outcome and there is thus no incentive for selfish players to try to signal a different type.

[6] Trebbi, Aghion, and Alesina (2008) provide evidence that the choice of electoral rules is indeed motivated egoistically.

[7] In line with the literature (Charness and Rabin, 2002, Engelmann and Strobel, 2004) we use the term "efficiency concerned" for subjects who care about the - possibly weighted - sum of individual payoffs. Efficiency concerns in this sense should not be equated with a concern for Pareto-optimality. Obviously, in a setup with positive and negative valuations, all collective decisions are Pareto-optimal if one does not permit monetary transfers.

selfish rule adjust accordingly. Indeed, we find that about half of the participants with a small absolute valuation who do not choose the selfish rule, choose the rule that maximizes total welfare (if all participants vote selfishly in the voting stage). This behavior leads to welfare that is slightly higher than in the third treatment that uses the same distribution of preferences but a pre-determined majority voting rule.[8]

As a second main finding, results from our first two treatments suggest that some agents follow different motivations at the rule-choice stage than at the voting stage. Specifically, their behavior at the rule-choice stage seems to be more efficiency driven than at the voting stage. At the rule-choice stage we find substantial evidence for efficiency concerns and no evidence for inequality aversion, whereas at the voting stage, behavior looks mostly selfish and only occasionally in line with inequality aversion or efficiency concerns. Testing for inconsistencies across stages is not straightforward since one has to control for subjects' beliefs regarding other players' motivations, valuations, and strategies. We thus develop two additional treatments that permit to rigorously check whether a subject's behavior is inconsistent across stages. We indeed find that this is the case for a substantial share of efficiency-concerned subjects. Hence adding a rule-choice stage can potentially increase efficiency because efficiency concerns may be more important on the rule-choice stage, apparently because choosing a rule suggests that one should do the right (i.e., efficient) thing.

Any given voting mechanism can only imperfectly aggregate the available information about individuals' valuation of the relevant alternatives. Therefore, it may occur that many agents with a small positive valuation for some decision trigger a decision which creates much larger losses for a small number of agents. Introducing a rule-choice stage in principle permits subjects to send finer signals about their valuation at that stage if the available set of rule choices is sufficiently rich. In theory, they should not make use of this option and choose extreme voting rules whenever valuations differ from zero. We find that, in practice, a sizable share of subjects suggest voting rules that do react gradually to their valuation. Our experimental design does not fully exploit the possible efficiency gains from gathering finer signals about players' preferences. We implement the rule choice through a random-dictator mechanism. Each player in a group suggests a threshold for the second-stage vote and then one of them is randomly chosen and this rule is implemented in

[8] Interestingly, though, in the treatment where subjects cannot choose the rules, they vote more often non-selfishly in the second stage. There might be several reasons for that. One possibility is that players in the treatment with rule choice vote less often selfishly because if a rule is chosen that is efficient given selfish voting, there is no need for efficiency-concerned players to vote non-selfishly. Given that we do not find much evidence that voting behavior conditions on the rule, so that selfish voting is not more frequent when it is also efficient, this explanation is not that plausible. Another possible explanation is that subjects may feel entitled to vote selfishly if they have proposed a non-selfish rule, even if that rule was not chosen.

the second stage. This allows a player whose rule is chosen to unilaterally implement his preferred outcome. If he prefers A, he picks as the rule that unanimity is required for B and votes for A and vice versa if he prefers B. The random dictator mechanism is obviously not a mechanism that would be suitable for most applications because it gives a lot of power to one randomly chosen person and thus aggregates preferences only poorly. The reason we choose this mechanism in the experiment is that as a first step we wanted to understand participants' preference over rules and this mechanism allows to elicit these in a transparently incentive compatible way. A rule-choice mechanism that makes better use of the information gathered in this stage than the random-dictator mechanism could potentially further improve welfare.

The question how to tailor mechanisms to the underlying distribution of types is of more general relevance and applies beyond voting problems also to other design problems. Often, optimal mechanisms have to be properly adjusted to the underlying distribution of types.[9] While this seems to be a feasible task in theory, it may be difficult to find an appropriately motivated person to perform this task in many practical applications.[10] Our experimental results indicate that efficiency-concerned agents may actually be willing and able to perform this task.

The major part of the mechanism design literature concentrates on the design of normal form games. By providing an example where a theoretically irrelevant procedural-choice stage affects outcomes, our experimental results show that the restriction on the analysis and use of one-stage games may be problematic in practice. In the same context, our results put into question the empirical relevance of the revelation principle. In theory, one could replace any two-stage voting mechanism by a simple one-stage mechanism and produce identical results. According to the revelation principle, this should actually hold for any multi-stage information aggregation procedure. However, if different motivations play a role on different stages, the revelation principle would not apply.[11] In our conclusion we discuss how this effect could be exploited in the design of collective decision mechanisms.

As discussed above, our paper is closely related to a theoretical literature that studies the optimal design of Bayesian voting mechanisms (e.g., Schmitz and Tröger, 2012). It also contributes

[9] Examples range from early work on optimal auction design (Myerson 1981) to recent work on optimal compromising (Börgers and Postl, 2009).

[10] The difficulties that are associated with the task to tailor mechanisms to the distribution of players' types are one reason why a substantial part of the mechanism design literature focuses on different forms of robust implementation (Wilson, 1985, Bergemann and Morris, 2005, 2012, Bierbrauer and Hellwig, 1999).

[11] In a recent paper, Masatlioglu, Taylor und Uler (2012) show experimentally that different one-stage auction games, which in theory produce identical equilibrium social choice functions lead to different bidding behavior in the laboratory.

to a - so far mostly theoretical - literature that studies the acceptance of collective decision making mechanisms at the interim stage. This literature started with the seminal paper on private information in a bilateral trade setup by Myerson and Satterthwaite (1983).[12] In another related theoretical paper, Barbera and Jackson (2004) study the dynamic stability of pairs of voting rules (for constitutions and single issues) when single-issue rules can be changed over time. In our paper we empirically study one single decision problem and we use a random-dictator mechanism to identify voters' preferred single issue rule.

A recent theoretical and experimental literature studies conflict resolution under private information about preferences when several related decisions can be bundled (e.g., Casella, 2005, Casella et al., 2006, Jackson and Sonnenschein, 2007, Hortala-Vallve and Llorente-Saguer, 2010, Engelmann and Grimm, 2012, Kaplan and Ruffle, 2012). Bundling makes it possible to provide individuals with incentives to adjust their messages about preference intensity gradually to their private information. It thus allows for efficient conflict resolution without adding a second decision stage. The present experiment focuses on decision problems when they cannot be bundled.

To the best of our knowledge, there is no experimental research on either the choice of mechanisms or the acceptance of mechanisms at the interim stage. The closest paper in this respect is Balafoutas et al. (2013) who investigate voting about the intensity of redistribution in a subsequent public-good game. Similar to our game, voting happens at the interim stage where participants already know which income group they are in, but in contrast to our experiment subjects do not choose the voting mechanism. Furthermore, our paper contributes to the research about the role of different types of social preferences in economic experiments (e.g., Charness and Rabin, 2002, Engelmann and Strobel, 2004). Inequality-averse individuals should favor the equal payoff option B unless their valuation for option A is high enough to offset the disutility from inequality. Agents who care about the maximization of the sum of individual monetary payoffs should instead be willing to accept outcome A (B) even if their valuation is slightly negative (positive) if this is welfare enhancing. Apart from the observation that we find evidence rather for efficiency concerns than for inequality aversion, one of our main results is that efficiency concerns may play a different role on different stages of a sequential game.

The rest of this paper is organized as follows. In Section 2 we explain our experimental design in detail, including the predictions based on different theories of subjects' (social) preferences. We present the results of treatments with symmetric and asymmetric type distributions in Section 3. Results of two control treatments that test for the consistency of behavior across stages can be found in Section 4. We discuss implications of our results in Section 5. Theoretical predictions

[12] Further contributions on interim participation decisions in the context of other collective choice problems include Güth and Hellwig (1987), Cramton, Gibbons, and Klemperer (1987), Schmitz (2002), Segal and Whinston (2011), and Grüner and Koriyama (2011).

regarding the behavior of efficiency-concerned subjects are derived in the Appendix.

2 Experimental Design

In this section, we first describe the underlying games and derive predictions for different types of preferences for our two main treatments and a control treatment with pre-determined majority voting. Then we report the procedural details for these treatments. The design of two additional treatments, that are inspired by the results from the earlier treatments are reported subsequent to these results.

2.1 Treatment 1: Symmetric Valuation Distribution

Our first treatment is designed to study whether behavior in the two-step procedure differs from rational maximization of one's own payoff and more specifically, to assess the relative importance of different motivations such as selfishness, inequality aversion, and efficiency concerns in our two-step choice procedure.

2.1.1 The Game

There are five players who collectively choose between two options, A and B. If they choose option B, all players receive a payoff of 0. If they choose option A, each player i receives a payoff θ_i (Euro in the experiment). At the beginning of the game, the valuations θ_i are drawn from an i.i.d. distribution on a given set of values. The distribution of types is common knowledge and uniform on the set $\{-7, -3, -1, -0.5, -0.2, -0.1, 0, 0.1, 0.2, 0.5, 1, 3, 7\}$. Each player only learns his own valuation θ_i. Next, players decide between options A and B in a two-stage voting process. In the first of these stages each player proposes a voting rule from the set $\{1, ..., 5\}$. If rule k is chosen, then in the second stage, k votes are required to implement option A. The actual rule is chosen by a random dictator mechanism, that is the rule that will be implemented is chosen randomly from the five voting rule that have been proposed by the five players.[13] In the second

[13] Due to a programming error, our randomization of whose rule was implemented did not work properly and thus the tie-breaking rule too often applied in Treatments 1 and 2. Hence some subjects were far more and others far less often the dictator. However, subjects never learned that their personal choice was implemented, and even for those for whom this was never the case, the implemented rule often coincided with their choice. Hence subjects may have noticed that the chosen rule was surprisingly often or less often than expected the same as their own choice, but only few have apparently detected the systematic error. Three participants in Treatment 1 noted in a post-experimental questionnaire that they were surprised that their choice was (almost) always implemented, but none remarked that their rule choice was surprisingly rarely implemented (which would be more of a concern

stage of the two-stage voting process, players are informed about the chosen rule and then vote about the two options. Either A or B is implemented depending on the rule chosen in the first stage and the votes from the second stage. If rule k is chosen and at least k players vote for A, then A is implemented and each player i receives θ_i. If fewer than k players vote for A, then B is implemented and all players receive payoff 0.

Players are informed about their own θ_i before the first stage but not about the complete vector of valuations. Note in particular that the choice of rules does thus not happen behind a veil of ignorance. Rules are thus to be understood here as ad-hoc rules for individual issues rather than as rules that are applied for a whole set of issues. Players are informed after each period of the game about the realized distributions of valuations in their group. Focusing on subjects' actions (and not counting the two random moves), we will henceforth call the rule proposal stage the first stage and the voting stage the second stage of the game.

2.1.2 Theoretical Predictions

Selfish Subjects Selfish players have weakly dominant strategies. In those strategies, they must choose rule 1 (5) if they have positive (negative) valuations, and, in the second stage, vote for their preferred outcome if they have non-zero valuations. Voters can choose any rule/outcome if their valuation is zero, and hence the dominant strategy is not unique. All other strategies are weakly dominated by those strategies because independent of the strategies of the other players, these strategies maximize the probability for outcome A for $\theta > 0$ and maximize the probability for outcome B if $\theta < 0$.

Inequality-averse Subjects Inequality-averse players in the two-stage game are predicted to behave like selfish players for negative valuations, while their behavior can differ from that of selfish players for positive valuations. We start with a derivation for the case that players assume others vote selfishly in the second stage. We do this to simplify the analysis and also because this is what we primarily observe. We discuss deviations from this assumption below.

Players with valuations $\theta_i \leq 0$ choose rule 5 and vote in favor of alternative B. This is both maximizing their expected payoff and minimizing inequality amongst players. Thus, for negative valuations, they behave just like selfish players (but have an additional reason to prefer B). Players with valuations $\theta_i > 0$, however, face a trade-off. Alternative A yields a higher expected monetary payoff, but at the same time on expectation yields inequality. For small positive θ_i, an inequality-

because this would undermine incentives). No such remarks were made in the post-experimental questionnaire in Treatment 2. We also checked whether behavior differed with the frequency with which the subjects were chosen to be the rule dictator and did not find a significant difference.

averse player will thus still prefer B over A and choose rule 5 and vote for alternative B. For larger θ_i, however, concerns for own income dominate concerns for inequality and the player will prefer alternative A and thus choose rule 1 and vote for alternative A. For example, straightforward calculation shows that for $\alpha_i = 1$ and $\beta_i = 0.3$ in the inequality-aversion model by Fehr and Schmidt (1999),[14] i prefers B for $\theta_i \leq 1$ if he has no further information about the realization of other players' types. According to a classification made by Fehr and Schmidt based on ultimatum game data, about 40% of subjects satisfy $\alpha_i \geq 1$ and $\beta_i \geq 0.3$. According to an estimate of individual parameters by Blanco, Engelmann, and Normann (2011), this condition is satisfied by about 30% in their experiment. Note that $\alpha_i = 1$ implies that a subject would reject offers of less than 1/3 of the total pie in an ultimatum game. Even for substantially weaker inequality aversion, one would still get preferences for B for positive θ_i, e.g., for $\alpha_i = 0.25$ and $\beta_i = 0$ in the Fehr-Schmidt model (implying rejecting offers of less than 1/5 in an ultimatum game), i prefers B for $\theta_i \leq 0.2$.

Also note that subjects who satisfy the assumption $\alpha_i \geq \beta_i$ as in the Fehr-Schmidt model would not choose intermediate rules unless they are strongly inequality averse and θ_i is large. If player i is considering to choose A because θ_i is positive, then if i minds disadvantageous inequality more than advantageous inequality (i.e., $\alpha_i > \beta_i$), she would rather have A if others have smaller valuations than her than if others have larger valuations than her. For θ_i relatively small, this is close to wanting A precisely if others vote for B but not if others vote for A. This implies that A is most attractive precisely if all others vote against it and thus if A results from her own vote alone under rule 1. Every other rule results in A only if several other players have a positive valuation and thus if i likes A less. Therefore, i will switch from rule 5 straight to rule 1. For relatively large θ_i this argument only holds when α_i is substantially larger than β_i, because for large θ_i the expected advantageous inequality is larger than the expected disadvantageous inequality.[15] Note, however, that in the estimation of the distribution of individual (α_i, β_i) pairs by Blanco, Engelmann, and Normann (2011), 38% among the subjects violate the $\alpha_i \geq \beta_i$ constraint.

To summarize, inequality-averse players would show a similar choice pattern as selfish players,

[14] According to the model by Fehr and Schmidt (1999), player i has the utiltiy function $U_i(x) = x_i - \frac{\alpha_i}{n-1}\sum_{j\neq i}\max\{x_j - x_i, 0\} - \frac{\beta_i}{n-1}\sum_{j\neq i}\max\{x_i - x_j, 0\}$ where x_k is the payoff to player k and n is the number of players. α_i and β_i thus measure player i's aversion towards disadvantageous and advantageous inequality, respectively.

[15] One can show that for $\theta_i = 0.5$, i would not choose an intermediate rule if $\alpha_i \geq \frac{3}{2}\beta_i$ and for $\theta_i = 1$, i would never choose an intermediate rule for $\alpha_i \geq 2\beta_i$.If we relax the assumption that i believes that the other players vote selfishly and instead assume that i considers others to be similarly inequality averse, then this result also holds for larger values of θ_i because this decreases the expected conditional advantageous inequality and increased the conditional disadvantageous inequality.

but their switch to rule 1 would not occur at $\theta_i = 0$ or at the smallest positive valuation, but at some higher value. There are a number of alternative assumptions that one could make regarding the subjects' preferences and their expectations about the voting behavior of others. First, subjects could have rational expectations. Given that subjects only observe actual votes of other participants, but not the relation between individual valuations and votes it is rather implausible that they have rational expectations regarding the others' voting strategies. Furthermore, actual voting behavior is very close to selfish voting, so for the actual behavior that we observe replacing the assumption that subjects expect selfish voting with the assumption that they have rational expectations regarding the other participants' voting behavior makes very little difference.

As a second alternative assumption, players could be subject to a so-called "false" consensus effect (Ross, Greene, and House, 1977, Engelmann and Strobel, 2000, Engelmann and Strobel, 2012) and have beliefs that are biased towards their own type, in the extreme case believing that others behave exactly as they do. If an inequality-averse player i expects other players to be equally inequality averse, as a strong consensus effect would predict, then i would expect others with positive valuation to occasionally vote for B. This increases i's estimate of the average valuation conditional on i being pivotal, which in turn would reduce the expected advantageous inequality but increase the expected disadvantageous inequality. Depending on i's inequality aversion parameters, this may make A more or less attractive and hence change i's voting behavior for $\theta_i > 0$. More importantly, however, it does not affect i's rule choice and voting behavior for $\theta_i \leq 0$. Outcome A then can only yield negative payoffs and inequality and is thus never preferred over B. The same holds also for rational expectations or indeed all others beliefs that i might have. So also for alternative assumptions regarding players' beliefs, the prediction for inequality-averse players remains that they behave like selfish players for negative valuations, but may behave differently for positive valuations.

Efficiency-concerned Subjects According to Schmitz and Tröger (2012), with a symmetric distribution of types, the simple majority rule maximizes the expected sum of individual payoffs. Suppose for a moment that all voters can be expected to vote "selfishly" in the second stage (we address below the case that others do not vote selfishly in the second stage). In this case, players who are fully efficiency concerned, i.e., those who consider the maximization of the sum of payoffs as their objective, should propose rule 3 independently of their own valuation. In the second stage these voters must also reply by voting in favor of the alternative that maximizes their own payoff if indeed rule 3 is chosen. Therefore, suggesting rule 3 and voting selfishly if rule 3 is chosen is an equilibrium of the two-stage voting game amongst such players.

The choice of less than fully efficiency-concerned subjects is more complex. Such agents care more about themselves than about others. In the appendix we assume that subjects maximize a convex combination of all agents' monetary payoffs (with a larger weight on their own monetary payoff) and show that when they expect selfish play in stage 2, they may propose rules 1, 2, 3, 4, or 5, depending on the value of θ_i. The proposed rule varies monotonically with the valuation. We note in particular that for the symmetric distribution of valuations in Treatment 1, the deviation from selfish rule choices is symmetric around 0 and that the suggested rule will always be between the selfish rule and majority rule. That is, if i has efficiency concerns and $\theta_i < 0$, i may choose rule 3, 4, or 5, but not rules 1 or 2 and if $\theta_i > 0$, i may choose rule 1, 2, or 3, but not rules 4 or 5.

We can again consider alternative assumptions regarding the beliefs of subjects concerning the other participants' behavior. As argued above, rational expectations are neither a very plausible assumption, nor would that assumption make much of a difference as compared to the assumption that all vote selfishly. Assuming alternatively that others are equally efficiency concerned, as suggested by a strong consensus effect, leads to complications if efficiency concerns are relatively strong. In this case, it could happen that a player proposes a different rule than under the assumption that all vote selfishly and then in the second stage votes against her preferences if she takes into account that choosing a different rule will change the voting behavior of others. As a result, the proposed rules may not change monotonically in the valuation. However, in the experiment we observe essentially no evidence of strategic voting, i.e., conditioning of the vote on the chosen rule, and therefore it does not appear to be a realistic assumption that players expect others to vote strategically.

Preferences for "Democracy" and Selfishness with Error We note that in Treatment 1 one cannot perfectly distinguish efficiency-concerned subjects from those that have a preference for majority voting per se (for example because they equate that with democracy and have a preference for democracy). A relevant question is whether a participant with preferences for majority voting in this sense would also consider rules 2 and 4 a reasonable compromise between selfishness and this concern for majority voting. In this case, such a participant would be indistinguishable from one who is efficiency concerned. It might be more plausible that a subject with preferences for majority voting would find rules 2 and 4 poor compromises and thus only choose between rules 1, 3 and 5, depending on θ_i and the intensity of her preferences for majority voting. This implies that at least subjects who choose all five rules in line with efficiency concerns cannot be explained by concerns for majority voting.

Furthermore, we note that selfish behavior with error where less costly errors are more likely to be made could produce patterns of behavior that look similar to that of efficiency-concerned subjects in the sense that for large absolute valuations, rule choices are more selfish and non-selfish rule choices are most often for the majority rule or the rule in between the selfish and the majority rule. However, efficiency concerns, but not selfishness with error would be consistent with the majority rule being chosen more often among these two rules. Furthermore, in the asymmetric Treatment 2, we can distinguish between efficiency-concerned behavior and both alternatives, preferences for majority voting per se and selfishness with error.

2.2 Treatment 2: Asymmetric Valuation Distribution

The main purpose of our second and third treatment is to establish whether the two-step voting procedure can increase efficiency over majority voting. We thus change the distribution of valuations such that majority voting is not ex-ante efficient. We then compare Treatment 2 which features a two-step voting procedure as in Treatment 1, with Treatment 3 where no rule choice takes place and majority voting is always implemented. The game we use in these two treatments is intentionally designed to give the two-step procedure a chance to provide efficiency gains compared to the simple majority rule. However, note also that the random dictator mechanism, which we chose primarily to make the design not too complex for the participants, is a rather inefficient way of aggregating preferences. Therefore, choosing a two-step procedure that more efficiently aggregates the preferences of participants regarding the voting rule may be more likely to produce efficiency gains.

2.2.1 The Game

In the second treatment, we have replaced the symmetric probability distribution from Treatment 1 by an asymmetric one with positive expected payoffs (of 2 Euro per player). The distribution of types is

$$\theta_i = \begin{cases} 14 & \text{with probability } 0.2 \\ -1 & \text{with probability } 0.8 \end{cases}$$

Apart from this change in the distribution of valuations, the design is identical to Treatment 1. This game has the same type of Bayesian Nash equilibrium for selfish players as the game that underlies Treatment 1. The efficient rule here is rule 1. Indeed the outcome that maximizes the sum of payoffs will always be reached if rule 1 is implemented and all players then vote according to their material preferences.[16] Expected payoffs given rule 1 and voting according to material preferences are

$$0.2^5 \cdot 14 + 5 \cdot 0.8^1 \cdot 0.2^4 \cdot \frac{4 \cdot 14 - 1}{5} + 10 \cdot 0.8^2 \cdot 0.2^3 \cdot \frac{3 \cdot 14 - 2}{5} + 10 \cdot 0.8^3 \cdot 0.2^2 \cdot \frac{2 \cdot 14 - 3}{5} + 5 \cdot 0.8^4 \cdot 0.2^1 \cdot \frac{14 - 4}{5} = 2.327.$$

[16] This is a difference to Treatment 1 where the sum of payoffs is not necessarily maximized even if the ex-ante efficient simple majority rule has been chosen because a majority with small absolute valuations could outvote a minority with large absolute valuations. In Treatment 2, given the efficient rule 1 and selfish voting in the second stage, A will be implemented if and only if at least one player has $\theta = 14$, which makes the total payoff positive.

2.2.2 Theoretical Predictions

A *purely selfish* subject should always vote in favor of his preferred alternative in stage 2. Moreover, he should propose rule 1 if $\theta_i = 14$ and rule 5 if $\theta_i = -1$.

An *inequality-averse* player i acts like a selfish subject if $\theta_i = -1$ and also if $\theta_i = 14$ unless i's aversion to advantageous inequality is extreme (that is for $\beta_i < \frac{14}{15}$ in the Fehr-Schmidt model, i would choose rule 1 if $\theta_i = 14$ and i believes others vote in line with payoff maximization in second stage).

A *sufficiently efficiency-concerned* subject (i.e., one who puts at least a third of the weight on the average payoffs of the others as on her own) would choose for $\theta_i = -1$ rule 1 and vote for B. Note in particular that efficiency concerns here impact on the rule choice, not on the voting behavior if indeed the efficient rule 1 is chosen. However, if the rule is rule 2, 3, 4, or 5, an efficiency-concerned player with $\theta_i = -1$ would in equilibrium vote with positive probability for A. An efficiency-concerned player with $\theta_i = 14$ would choose rule 1 and vote for A unless we make the extreme assumption that he actually cares substantially more for each of the other players' earnings than his own, in which case he would vote for rule 2 and vote for A.

Considering alternative beliefs about others' voting behavior is not of much interest. Inequality-averse players should always vote in line with their monetary preferences (they will always do so for $\theta_i = -1$ and for $\theta_i = 14$ they will do so for any rule and any beliefs about the others' behavior as long as they are not excessively inequality averse, $\beta_i < \frac{14}{15}$), as should efficiency-concerned players whenever the efficient rule 1 is chosen. So if a player assumes that others are inequality averse or assumes that others are efficiency concerned and rule 1 is chosen, he would still expect the others to vote selfishly. If efficiency-concerned players expect others to be efficiency concerned and thus vote for A with positive probability if they have $\theta = -1$ and rule 2, 3, 4, or 5 is chosen, this does still not make choosing these rules preferable to choosing rule 1 and voting for B (assuming all others to do so when they have $\theta = -1$ and rule 1 is chosen). The reason is that only choosing rule 1 and then voting for B ensures that A is chosen if and only if at least one player has $\theta = 14$.

2.3 Treatment 3: Asymmetric Valuation Distribution with Pre-determined Majority Voting

The purpose of our third treatment is to serve as a benchmark for Treatment 2 to assess whether the two-step procedure affects welfare compared to simple majority voting, the commonly used unbiased alternative for decision making in the given context.

2.3.1 The Game

Treatment 3 works with the same probability distribution as Treatment 2. However, it starts directly with the voting stage and the rule is always the simple majority rule. Hence players are informed about their valuations and then vote directly.

2.3.2 Theoretical Predictions

Purely selfish subjects should vote for the alternative that maximizes their payoff. Expected equilibrium payoffs are then given by

$$0.2^5 \cdot 14 + 5 \cdot 0.8^1 \cdot 0.2^4 \cdot \frac{4 \cdot 14 - 1}{5} + 10 \cdot 0.8^2 \cdot 0.2^3 \cdot \frac{3 \cdot 14 - 2}{5} = 0.484.$$

Efficiency-concerned subjects with a positive valuation should always vote for alternative A. If all agents are sufficiently efficiency concerned, there is no sincere voting equilibrium. Instead, in a symmetric equilibrium voters with a negative valuation either play a mixed strategy (if their efficiency concerns are such that they would pay at most 1 to increase all other players' payoffs by slightly less than 2) or (if their efficiency concerns are stronger) always vote for A independent of their valuation.[17] Inequality-averse subjects with a negative valuation vote for alternative B, and with a positive valuation they vote for A (if they are pivotal, then two players have $\theta = -1$ so the advantageous inequality term in the Fehr-Schmidt model is $\frac{1}{4}(15 + 15) = \frac{15}{2}$, so for $\beta < 1$, they would not be willing to give up 14 in order to eliminate this inequality.) Hence when all agents are either selfish or inequality averse, there is an equilibrium in which all agents vote sincerely.

To summarize, sincere voting can be motivated by selfishness and inequality aversion. Insincere voting of subjects with a negative valuation is compatible with efficiency concerns but not with selfishness or inequality aversion. If subjects are sufficiently efficiency concerned to always vote

[17] There is an equilibrium where altruistic players always vote for A if all players are altruistic, even though this also implies that A is chosen for sure when all have $\theta = -1$. If all others always vote for A, one is actually never pivotal and hence indifferent. Moreover, for sufficiently altruistic players, this equilibrium is trembling-hand perfect and there is also no equilibrium in mixed strategies. Assume that other altruistic players vote for A with a probability smaller than 1. Then considering the cases when one is pivotal, the expected gains in total payoffs from cases where one or two other players have $\theta = 14$ exceed the expected losses when all have $\theta = -1$, so that a sufficiently efficiency concerned player would then want to vote for A for sure. Hence for sufficiently strongly efficiency concerned players, voting for A is weakly dominant and thus it is the only symmetric equilibrium.

for A, then the expected payoff is[18]

$$0.2^5 \cdot 14 + 5 \cdot 0.8^1 \cdot 0.2^4 \cdot \frac{4 \cdot 14 - 1}{5} + 10 \cdot 0.8^2 \cdot 0.2^3 \cdot \frac{3 \cdot 14 - 2}{5} + 10 \cdot 0.8^3 \cdot 0.2^2 \cdot \frac{2 \cdot 14 - 3}{5}$$

$$+5 \cdot 0.8^4 \cdot 0.2^1 \cdot \frac{14 - 4}{5} + 0.8^5 \cdot (-1) = 2.$$

The efficiency gains compared to the equilibrium with majority voting and selfish players that can be reached by efficiency concerned players with the two-step procedure thus exceed those that can be achieved through always voting for A given majority voting by $\frac{(2.327-0.484)-(2-0.484)}{2-0.484} = 21.57\%$.

2.4 Procedures

The computerized experiments were run at the experimental laboratory mLab at the University of Mannheim with software programmed in z-Tree (Fischbacher 2007). Subjects were recruited using ORSEE (Greiner, 2004). In each session, 15 periods were played with random matching among subjects, and between 10 and 20 subjects took part. We ran three sessions each for Treatments 1, 2, and 3 for a total of 45 subjects in Treatment 1 and 55 in both Treatment 2 and 3. To avoid any income effects we did not pay all periods, but at the end of each session, one period was chosen randomly, and that period determined the payoffs of all players. If in that period alternative A was chosen in subject i's group, then i's payoff was given by $\pi_i = \theta_i$, otherwise $\pi_i = 0$. Each subject i earned a show-up fee of 9 Euro plus the experimental payoff π_i, hence earnings could be between 2 and 16 Euro in Treatment 1 and either 8, 9 or 23 Euro in Treatments 2 and 3. In all treatments, at the end of each period, the participants learned the number of votes in favor of alternative A, the resulting outcome of the vote as well as all five valuations from their group without being able to identify the other players.

3 Results

3.1 Treatment 1

3.1.1 Second-stage Voting Behavior

We begin by discussing the behavior in the voting stage. Voting in the second stage is overwhelmingly in line with the maximization of the subjects' own payoff. Overall, only 4.3 percent of the votes cast are in favor of the alternative which does not maximize the subject's monetary payoff

[18] Compared to choosing the efficient rule in Treatment 2, efficiency losses occur because A is also implemented when all players have $\theta = -1$.

(i.e., are either for A when $\theta_i < 0$ or for B when $\theta_i > 0$). Table 1 shows the share of deviations for different valuations. We aggregate large negative, small negative, small positive and large positive valuations, respectively, and drop $\theta_i = 0$ because either vote is in line with payoff maximization in this case. As we can see in Table 1, deviations from payoff maximization occur most frequently for small positive valuations $0.1 \leq \theta_i \leq 0.5$. This is where inequality aversion would predict votes against own payoff maximization, so this suggests that inequality aversion may play a role for some subjects. We note, however, that the rate is still very small with 10.4%. Furthermore, voting for A when $\theta_i < 0$ is difficult to reconcile with any preference model, so seems likely to result from errors. (Alternatively, this might result from a misguided attempt to signal non-selfishness to the experimenter or themselves.) Thus taking the 3.3 percent of votes against own payoff maximization for negative valuations as an estimate of the error rate, only about 7 % of the observations are in line with inequality aversion, where this is expected to matter most. Remember that only a very small degrees of inequality aversion is needed in order to imply a preference for B when $0.1 \leq \theta_i \leq 0.5$.

An important question is whether participants show evidence of strategic voting, i.e., taking into account what being pivotal implies about the others' valuations. For example, if the rule implemented is rule 4 and one has a valuation of 0.1, then one is much more likely to have a positive effect on efficiency by voting for A than if rule 2 has been chosen (at least as long as the other participants are not all efficiency concerned and strategic).[19] In this sense, we observed very little evidence of strategic voting. Votes are against own monetary preferences least often for rule 5 (2%) and most often for rule 4 (8%), which is only chosen 50 times altogether, so that this amounts to only 4 observations. Specifically, the rule that is in place does not have much impact on the pattern presented in Table 1, with the interesting exception that of the 16 votes most indicative of inequality aversion (for B if $0.1 \leq \theta_i \leq 0.5$) 11 occur when rule 1 is implemented and thus when only one vote is needed to implement A. Note that in this case a player is actually pivotal only if all others vote for B (including the rule chooser who thus must have made an error in one of the stages), and are thus likely to have negative valuations, and hence B is also on expectation maximizing efficiency. Hence more than two thirds of the votes that appear to support inequality aversion are also in line with efficiency concerns.

[19] Exact predictions in how voting behavior should depend on the implemented rule require assumptions regarding participants' beliefs about other participants' voting behavior. Since we do not have information about these beliefs, we cannot derive such predictions. Nevertheless, except for very specific beliefs, strategic voting should lead to differences in voting behavior for different rules. Such differences would thus be evidence of strategic voting.

$-7 \leq \theta_i \leq -1$	$-0.5 \leq \theta_i \leq -0.1$	$0.1 \leq \theta_i \leq 0.5$	$1 \leq \theta_i \leq 7$
3.5%	3.0%	10.4%	1.8%

Table 1: Deviations from selfish voting in the second stage

$-7 \leq \theta_i \leq -1$	$-0.5 \leq \theta_i \leq -0.1$	$0.1 \leq \theta_i \leq 0.5$	$1 \leq \theta_i \leq 7$
14.6%	39.3%	32.5%	18.8%

Table 2: Deviations from selfish voting in the first stage

3.1.2 First-stage Behavior

In contrast to the voting behavior in the second stage, subjects' proposals of rules in the first stage for $\theta_i \neq 0$ frequently deviate from the payoff-maximizing rule. Overall, 25.4 percent of the proposals made do not coincide with a payoff maximizing voting rule. According to Table 2, which shows the rate of deviations from the selfish rule and again groups valuations into large negative, small negative, small positive and large positive valuations, there is a clear pattern. First, deviations from the payoff-maximizing rule are more frequent for small absolute valuations than for large absolute valuations, which is in line with any model that combines selfishness with some other motive. Second, deviations are about equally likely for negative valuations than for positive valuations, a pattern that is in line with efficiency concerns (or selfishness with error or preferences for majority voting). In the appendix of this paper we show that for our model of efficiency concerns, the proposed rule is monotonous in the agent's payoff and that the deviation from a selfish rule proposal is stronger, the smaller the absolute value of the valuation.

One obvious reason why there may be more deviations from selfishness in the first stage than in the second stage is that finding the selfish rule is more complex. Furthermore, for the choice of rules, there is only one way to be "right" (i.e., payoff maximizing) but four ways to be "wrong". Both these arguments would suggest that pure confusion may contribute to the higher share of deviations from selfishness in the first stage. However, the smaller rate of deviations for larger absolute valuations suggests that the non-selfish rules are not chosen plainly out of confusion. While errors being rarer for larger absolute valuations is consistent with standard models of choice with error, the pattern of rules that are chosen for small absolute nonzero valuations suggests that the deviations from the selfish rule are not only the result of confusion. Specifically, some rules make sense given our model of efficiency concerns, namely those that range from the selfish rule to majority voting whereas others do not, namely those that are biased against the maximization of subjects' own payoff, i.e., the chosen rule favors alternative A (rules 1 and 2) while the subject has a negative valuation or it favors B while the subject has a positive valuation (rules 4 and 5).

θ_i	$\sum$	1	2	3	4	5	average rule
-7	59	4	0	4	1	50	4.57
-3	51	2	2	3	0	44	4.61
-1	61	1	3	3	2	52	4.66
-0.5	42	1	2	3	8	28	4.43
-0.2	46	1	3	5	11	26	4.26
-0.1	47	2	0	11	6	28	4.23
0	45	7	1	22	2	13	3.29
0.1	49	32	3	7	3	4	1.86
0.2	46	29	9	4	0	4	1.72
0.5	59	43	5	10	0	1	1.49
1	59	47	2	5	3	2	1.49
3	57	46	3	7	1	0	1.35
7	54	45	1	3	4	1	1.43

Table 3: Distribution and average of proposed rules for the different valuations in Treatment 1.

We find about 7 percent of the latter type of rule choices in each of the four groups of valuations. Interestingly, for $0.1 \leq \theta_i \leq 0.5$, where mild degrees of inequality aversion are consistent with the choice of rule 5 (and possibly 4), rules that are biased against the maximization of subjects' own payoff are not chosen more frequently than in any of the other groups of valuations. This suggests that inequality aversion does not play an important role in the rule-choice stage. Given that 7 percent of rule proposals are for the two rules that make little sense given θ_i, a fair estimate of the error rate might be $2 * 7 = 14$ percent, given that there are four non-selfish rules. This would also explain most of the non-selfish choices for $|\theta_i| \geq 1$ as errors but would leave us with about 22 percent of non-selfish rule choices for small (≤ 0.5) absolute valuations.

A key prediction of our efficiency-concerns model and key distinction to the predictions by inequality aversion are that deviations from the selfish rule are not only decreasing in absolute valuation but are also symmetric around 0. In order to assess this prediction, we show the distribution of the chosen rules for each θ_i in Table 3. Looking at the disaggregated distribution, we see that for $|\theta_i| \geq 1$, the vast majority of choices (79.7% or more) are for the selfish rule, while the remaining choices are spread fairly evenly across the other rules, which would be in line with noise. In contrast, for $|\theta_i| < 1$, while still more than half of the rule choices are for the selfish rule, a substantial share is in each case in favor of the efficient rule 3 or the rule in between the selfish and the efficient rule (between 20.4% and 36.2%, overall 28.7%). As we can see, the deviations from the selfish rule are quite symmetric. This pattern is in line with efficiency concerns. To test for the symmetry of the distribution, we compare for each valuation θ_i the distribution of chosen rules with the inverted distribution for $-\theta_i$. According to Fisher exact tests, the distributions are not significantly different for any $\theta_i \in \{0.1, ..., 7\}$ ($p > 10\%$ in each case).[20]

Rule choices that are biased against own monetary preferences are not chosen more often for $0 < \theta_i \leq 0.5$, where they would be in line with inequality aversion than for $\theta_i \geq 1$ or $\theta_i < 0$. We also see that for $\theta_i = 0$ the most frequently chosen rule is 3, as predicted by efficiency concerns, and not 5, which would be predicted by inequality aversion. Hence, Table 3 provides very little evidence in favor of inequality aversion playing a role for subjects' rule choices. We also note that the median rule is always the selfish rule, so even for very small own monetary gains or losses, an absolute majority always goes for the selfish rule.

For a more thorough assessment of the rule choice, we run ordered probit regressions of the rank number of the chosen rule on the valuation.[21] We also run probit regressions for a dummy whether subjects are choosing a non-selfish rule because the key distinction appears to be the one between selfish and non-selfish rules. We run both the ordered probit and probit regressions separately for negative and positive valuations, because this allows us to discriminate between the predictions of different preference types. Efficiency concerns would predict a roughly symmetric impact of the valuation both for negative and positive valuations, because the efficiency model presented in the appendix is symmetric and the closer the valuation is to 0, the more one should deviate from the selfish rule.[22] In contrast, inequality aversion predicts that the valuation has no impact for $\theta_i < 0$, because then the selfish rule should always be chosen, whereas for $\theta_i > 0$, subjects should deviate more the smaller their valuation, as for the efficiency-concerns model. The

[20] The table also shows the average chosen rule. We note that one should not take these average rules too seriously as it is fair to argue that the rules are on an ordinal scale and not a cardinal scale. Thus, the averages serve primarily for illustration. For any given level of θ_i, comparing the absolute difference between the average rule and the selfish rule (1 or 5), with this difference for $-\theta_i$, we see that they never differ by more than 0.15. Note also that the Fisher test does not correct for possible dependence of observations due to repeated observations for the same person and within the same group. Given that we demonstrate here the absence of a significant difference, this is arguably not problematic.

[21] Ordered probit is more appropriate than OLS because the rule number cannot really be considered as cardinal, given that one rule is the optimal one for a selfish player and the others vary rather gradually.

[22] In theory, the effect should be exactly symmetric. However, since valuations are randomly assigned, valuations are typically not obtained by all subjects the same number of times and because subjects are likely to be heterogenous this randomness will lead to some noise in the estimates, even if all subjects followed the efficiency concerns model consistently.

		(1)	(2)	(3)	(4)
$\theta_i < 0$	θ_i	$-.0729412^{**}$	$-.0729412^{*}$	$.1728693^{***}$	$.1100306^{**}$
		(.0316187)	(.0429341)	(.0574938)	(.0454685)
	const.			$-.9027392^{***}$	$-.4480241^{**}$
				(.2980232)	(.1837908)
	Pseudo R^2	0.0102	0.0102		
$\theta_i > 0$	θ_i	$-.0659364^{**}$	$-.0659364$	$-.2054207^{***}$	$-.077422^{*}$
		(.0318391)	(.0440765)	(.0570684)	(.0432327)
	const.			-1.045793^{***}	$-.5213664^{***}$
				(.3661205)	(.1773873)
	Pseudo R^2	0.0078	0.0078		

Table 4: Regressions for chosen rule in Treatment 1. Columns (1) and (2) present order probit regressions for the rule chosen in stage 1 without and with clustering of standard errors at the individual level, respectively. Columns (3) and (4) present probit regressions of a dummy for not haven chosen the selfish rule in stage 1 with individual-level random effects and with clustering of standard errors at the individual level. The top part shows the results for valuations $\theta < 0$, the bottom part for valuations $\theta > 0$. Standard errors in parentheses. $^{*}: p < 0.1$, $^{**}: p < 0.05$, $^{***}: p < 0.01$.

regressions again support the efficiency-concerns model. As can be seen in Table 4, the coefficient on the valuation is very similar for negative valuations and positive valuations in the ordered probit both without (column 1) and with (column 2) clustering of standard errors on the individual level as well as in the probit, both with individual-level random effects (column 3) and with clustering of standard errors on the individual level (column 4). Moreover, if there are differences in the significance levels, then in the direction of stronger significance for negative valuations, in contrast to the inequality aversion prediction.

Deviations from selfish behavior do not decrease substantially over time. Figure 1 shows the trend over time of votes and rule choices that are not selfish. Except for a drop of non-selfish rule choices from the first to the second period, there is no clear discernible time trend. Hence, deviations from the selfish prediction are unlikely exclusively errors because we would then expect a decreasing time trend.[23]

[23]If we look at the sessions of Treatment 1 separately, the pattern is very similar in all three of them, with

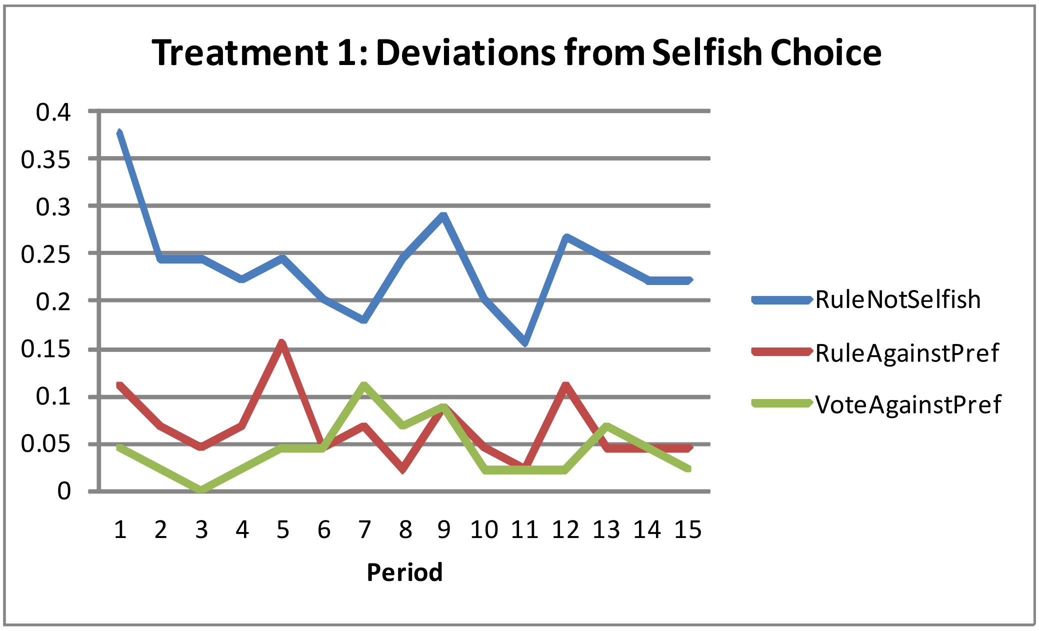

Figure 1: Share of choices that differ from the selfish prediction across periods. VoteAgainstPref corresponds to votes for A given $\theta < 0$ or votes for B given $\theta > 0$. RuleNotSelfish summarizes all rule choices that are not for rule 5 when $\theta < 0$ or not for rule 1 when $\theta > 0$. RuleAgainstPref is the subset where the rule is biased against the subject's own monetary preferences, i.e., for rule 1 or 2 when $\theta < 0$ or for rule 4 or 5 when $\theta > 0$.

The analysis so far has only taken aggregate data into account and has found support that deviations from selfishness are largely in line with the predictions of our model of efficiency concerns. An important question is whether individual subjects behave according to a specific preference model across the periods where they draw different valuations. In order to do this, we classified the 45 subjects according to the following four criteria, which are key predictions of our efficiency model.

1. The subject always proposes rule 3 when $\theta_i = 0$.
2. The subject never proposes a rule which is biased against selfish preferences.
3. The subject proposes at least once a rule that is not implementing selfish preferences.
4. The subject has a rule profile that is weakly monotone in the sense that the average rank number of the proposed rules is weakly decreasing in θ_i.

We call a subject consistently efficiency oriented if it satisfies all four criteria, with one deviation against (2) or (4) permitted. Out of 45 subjects, 11 are in this sense consistently efficiency oriented. We can also classify subjects as consistently selfish if the following conditions are satisfied.

1. The subject proposes rule 5 whenever $\theta_i < 0$.
2. The subject proposes rule 1 whenever $\theta_i > 0$.
3. The subject votes for alternative B whenever $\theta_i < 0$.
4. The subject votes for alternative A whenever $\theta_i > 0$.

We find 17 subjects who are consistently selfish. This number increases to 23 if we allow for one deviation on each stage. Finally, we classify subjects as inequality averse if the following conditions are satisfied.

1. The subject proposes rule 5 whenever $\theta_i \leq 0$.
2. The subject proposes at least once rule 4 or 5 for $\theta_i > 0$.
3. The subject votes for alternative B whenever $\theta_i \leq 0$.
4. The subject votes for alternative B at least once when $\theta_i > 0$.

rule choices and votes against preferences fluctuating between 0% and 10%, and non-selfish rule choices fluctuating around a 20-25% range. The fluctuations are more extreme, however, due to the smaller number of observations. Most importantly, neither of the sessions shows a clear upward or downward trend in either of the three measures.

We find no subject that satisfies these criteria perfectly, and three who satisfy it on the rule choice stage only, where two of them only once choose a non-selfish rule.

Remember that for the symmetric distribution of valuations, the simple majority rule is efficient and given an exogenously chosen majority rule, selfish voting is efficient. Thus, because the random dictator rule that we employ cannot result in a rule that is ex-ante more efficient than the majority rule,[24] but often results in less efficient rules, it is not surprising that the total payoffs in Treatment 1 are about 35 percent lower than what would be obtained with pre-determined majority voting and selfish voting behavior given the draws of valuations we observe. However, given the rules of our experiment, the deviation from selfish behavior leads to total payoffs being about 8 percent higher than the payoff that selfish behavior on both stages would yield given the draws of valuations realized in the experiment.

3.2 Treatments 2 and 3

We again start by considering voting behavior in the second stage. In the asymmetric Treatment 2 with rule-choice stage, the observed voting behavior of subjects with positive valuations ($\theta_i = 14$) is consistent with selfishness as well as with efficiency concerns (as well as non-extreme form of inequality aversion). Only in 3 percent of cases, subjects with $\theta_i = 14$ voted for alternative B. This would correspond to the same error rate as observed in Treatment 1. Subjects with a negative valuation ($\theta_i = -1$) vote for alternative A in 13.6 percent of the cases. Their behavior is inconsistent with selfishness and inequality aversion, but is consistent with efficiency concerns if the rule is not rule 1 (because in that case, A would only maximize total payoffs if at least one player has $\theta_i = 14$, when one could expect that this player votes for A herself). Note, again, that the rate of non-selfish votes is relatively low, given that quite some efficiency gains are possible, so even moderately efficiency-concerned players could be expected to vote for A if the rule is not rule 1.

Interestingly, subjects' voting behavior is statistically independent of the implemented voting rule.[25] This is inconsistent with efficiency concerns because these imply for $\theta_i = -1$ to vote for A

[24] Though at an interim-stage if the valuation of the dictator is $\theta = 7$ or $\theta = -7$, it might result in the more efficient rules 2 and 4, respectively.

[25] Specifically, for Rules 1, 2, 3, and 5, the share of votes for A by subjects with $\theta_i = -1$ is between 12.6% and 14.8%. Only for Rule 4 the share is lower with just 4.7%. However, there are only 21 observations for subjects with $\theta_i = -1$ when Rule 4 is implemented. According to Fisher's exact test, the distribution of votes does not differ between Rule 4 and any of the other rules, or between Rule 4 and all other rules pooled or between any pair of rules ($p > 0.3$ for all two-sided tests). Note that this test does not correct for dependence due to multiple choices of the same individual and interaction within sessions. Negelecting the dependence would rather lead to identifying

θ_i	1	2	3	4	5
-1	123	72	58	38	365
14	149	5	10	4	1

Table 5: Distribution of proposed rules for both valuations in Treatment 2.

only for rules unequal to rule 1. Not very surprisingly, though, voting behavior is correlated with subjects' own rule choice. In only 3 percent of the cases where subjects choose the selfish rule do they vote for an outcome that does not maximize their own payoff (so they are consistently selfish), but in between 19 and 33 percent of the cases where subjects choose a non-selfish rule do they vote against their monetary preference in the second stage.

Regarding the rule-choice stage, subjects with a valuation of $\theta_i = 14$ choose the selfish rule (rule 1) in 88 percent of cases (see Table 5). Among the 12 percent remaining proposals 6 percent are for the simple majority rule. The latter observation indicates that these subjects may have a preference for democracy – interpreted as a preference for majority voting. Subjects with a valuation of $\theta_i = -1$ choose the selfish rule (rule 5) in only 56 percent of cases and the efficient rule (rule 1) in 19 percent of cases. The behavior of these 19 percent is inconsistent with selfishness and inequality aversion, but in line with efficiency concerns. Furthermore, 9 percent choose the simple majority rule, which could suggest a preference for majority voting per se or for mild efficiency concerns. While we thus observe that substantially more subjects deviate from the selfish rule if they have $\theta_i = -1$ than if they have $\theta_i = 14$, still a majority chooses the selfish rule 5 and in 97 percent of the cases then also votes for B and thus enforces B even if all other subjects have $\theta_i = 14$, which amounts to quite substantial selfishness.

The actual average payoff in Treatment 2 per period was $W_a = 1.30$. This value exceeds the hypothetical average payoff with majority rule, selfish voting behavior, and the observed valuation distribution of $W_s = 0.44$. Thus the combination of letting subjects choose the rule and their deviation from selfishness yields a substantial efficiency gain. Part of this gain stems from the fact that the efficient rule is sometimes chosen, part of it results from non-selfish voting when the rule is not rule 1. We thus also compare the average payoff to the hypothetical average payoff that would result under exogenously chosen majority rule given the average voting behavior observed if the majority rule was chosen in the experiment and the observed valuation draws, which is $W_o = 0.90$. Since $W_o > W_s$, but $W_o < W_a$, the observed efficiency gains are derived indeed both from the choice of the efficient rule and from non-selfish voting. We also note that the efficient

significance where there is none, so since the point here is to show that there is no dependence of voting on the rule, this does not appear problematic.

rule is sometimes chosen because a player with $\theta_i = 14$ is chosen as the dictator. Hence the choice of the efficient rule is not exclusively based on altruism, but sometimes also simply on allowing somebody with strong interests to have his will.

The comparison of Treatments 2 and 3 allows us to assess whether our two-stage decision process leads to efficiency gains compared to actual behavior given an exogenously given majority rule. Using an OLS regression with robust standard errors (controlling for the number of positive valuations in the group), we find that with rule-choice stage, the average individual payoff is about 0.47 higher per period than under exogenously given majority voting. This difference is far from significant ($p > 0.5$). It is also smaller than what one would expect from looking at the behavior in Treatment 2. The reason is that for exogenously given majority rule, voting behavior is less selfish than observed with the two-stage procedure. Specifically, the probability to vote for A for $\theta_i = -1$ is by 11 percentage points lower in Treatment 2 with rule choice than in Treatment 3. According to a linear probability model, this difference is highly statistically significant ($p < 0.01$, also if standard errors are clustered at group level). Hence in Treatment 3 with pre-determined majority voting, efficiency gains are rather obtained by non-selfish voting, in Treatment 2 with rule-choice stage by choosing the efficient rule. This suggests that subjects consider non-selfish behavior at the two stages partly as substitutes. They may engage in moral licensing, voting selfishly in the second stage if they have chosen a non-selfish rule, even if that rule was not chosen.

3.3 Interpretation of the Results from Treatments 1-3

Overall, we observe non-trivial shares of rule choices as well as of votes that are in line with some degree of efficiency concerns. There is a non-trivial share of non-selfish choices on the rule-choice stage of Treatment 1 that is consistent with efficiency concerns. Similarly, in the asymmetric Treatment 2, 19 percent of the subjects' rule proposals given a negative valuation can be explained by efficiency concerns, and another 25 percent would be consistent with at least some degree of efficiency concerns. In the second stage, however, in only 14 percent of cases, subjects with negative valuation vote for alternative B in Treatment 2. This share is, however, higher in Treatment 3, with exogenously given majority rule. Since somewhat, but not very strongly, efficiency-concerned subjects should actually consider mixed strategies in Treatment 3, this may constitute a lower bound on the share of efficiency-concerned individuals in Treatment 3.

Interestingly, in both Treatment 1 and 2, the share of own-payoff maximizing votes on the second stage is higher than the share of own-payoff maximizing choices on the rule-choice stage. While this may reflect that efficiency concerns have little effect in the voting stage as argued above, it might also be that subjects approach the two stages differently and do not behave consistently with the same preference model across both stages. Partly, this may reflect moral licensing, as subjects who have chosen a non-selfish rule feel justified to cast a selfish vote, even if a different rule has been chosen. This would also be in line with the higher share of non-selfish votes in Treatment 3 than in Treatment 2 (if the majority rule is chosen in the latter) because in Treatment 3 such moral licensing does not apply as no rule is chosen. Alternatively, subjects may perceive the rule-choice stage as a more principled decision, where they should do the "right thing", whereas voting selfishly in the second stage is perhaps considered more acceptable. This would, however, not explain the higher share of non-selfish votes in Treatment 3. We address this apparent inconsistency in Treatments 4 and 5, where we eliminate uncertainty about what it means to be pivotal and can thus make clear predictions about consistent behavior across the two stages, independent of assumptions about beliefs.

Another interesting result in Treatments 1 and 2 is that some participants appear to have a preference for the majority rule (rule 3). It is chosen more often than both rule 2 and rule 4 in Treatment 1 and in Treatment 2 by those subjects with $\theta = 14$.

From the comparison of Treatments 1 and 2 we see that subjects very clearly react to the distribution of valuations. In particular, there are more choices for the efficient rule in Treatment 2 where efficiency gains are higher. This change in behavior is in line with subjects being concerned with the maximization of total payoffs, but not with a preference for majority voting per se or with a selfishness with noise model.

4 Control Treatments: Identifying Inconsistent Behavior across Stages

4.1 Motivation and Experimental Design

The results of Treatments 1 and 2 indicate that several participants behave differently across both stages. In particular, some participants consistently vote selfishly in the second stage but propose efficient rules in the first stage of the game. This raises the question whether subjects' behavior is consistent with utility maximization across stages. Related to that is the question whether each stage appeals to a different behavioral motive in the sense that subjects act more altruistically in the rule-choice stage than in the voting stage. Our last two treatments are designed to rigorously test for inconsistencies across stages. In particular, Treatment 4 implements the following game where we do not need to worry about beliefs of subjects about the other subjects' rule choices and hence whether the chosen rule reveals something about what being pivotal implies for the other subjects' valuations. Treatment 5 is a simplified variant that will be explained below.

In this game we consider three players. First, nature determines players' valuations, where $\theta = -1$ or $\theta = 14$, both with probability $\frac{1}{2}$ (because the game is rather complicated, we decided to simplify the assignment of valuations and both outcomes being equally likely is easier to explain and understand than one occurring with probability $\frac{4}{5}$). Then, as in the other treatments a two-stage voting procedure follows. In the first stage a rule is suggested by each player. Each player's choice is limited to rules 1 and 2, i.e., rules that require either one or two votes to pass A.[26] Then either one player's rule choice is randomly selected to be implemented (with probability $\frac{1}{4}$ for each player's rule choice), or rule 2 is randomly chosen to be implemented as the default, which also happens with probability $\frac{1}{4}$. In the second stage of the voting procedure, all players cast their intended votes in favor of alternative A or B. However, only one randomly chosen subject can freely decide how to vote. The others are "forced" to vote in line with their monetary preferences, i.e., to vote for B if $\theta < 0$ and to vote for A if $\theta > 0$. To gather more data, all vote, but all but one of the votes will be replaced with the money-maximizing vote.

We represent a strategy of a player who has a given valuation θ by a vector (R, v_1, v_2, v_c) , where $R \in \{1, 2\}$ denotes the rule proposed, and $v_1, v_2, v_c \in \{A, B\}$ denote the vote under rule 1, rule 2 proposed by a player, or rule 2 selected by the computer, respectively. Our test for consistency is based on the assumptions that (i) all subjects expect players with a positive valuation to always choose rule 1 and to vote in favor of alternative A for any implemented rule and (ii) all players with a negative valuation vote in favor of alternative B if rule 1 is implemented, and all players expect the others to do so.

We first show that a player i who proposes rule 1 even though $\theta_i < 0$ weakly prefers alternative A in case that exactly one other player has a positive valuation. A player i with $\theta_i < 0$ who proposes rule 1 chooses, under the above assumption (ii), strategy $(1, B, B, B)$, or $(1, B, A, A)$, or $(1, B, B, A)$, or $(1, B, A, B)$ (under the standard assumption that by the time they propose a rule the players actually have a complete plan how to vote contingent on all possible implemented rules, an assumption that may well be violated for experimental participants).

Consider first strategy $(1, B, B, B)$. Alternative A is chosen if (i) rule 1 is selected and exactly one other subject has a valuation $\theta > 0$ or (ii) two other subjects have a valuation $\theta > 0$ or (iii) exactly one other subject has $\theta > 0$ and one other subject with $\theta < 0$ votes for A and her vote is chosen to count. Compare this to the strategy $(2, B, B, B)$. Here, alternative A is chosen under the same conditions as above. Since rule 1 obtains with a lower probability under the second strategy and if exactly one other player has $\theta > 0$, A is more likely to result under rule 1 than under rule 2,[27] i must be weakly in favor of alternative A, conditional on exactly one player having a positive valuation, if i chooses $(1, B, B, B)$ rather than $(2, B, B, B)$. The same reasoning holds

[26] A similar argument could be made by restricting the rule set to rules 2 and 3.

[27] Note that given that i votes for B under rule 2, and her own vote may be chosen to count, A will not result for sure if exactly one other player has $\theta > 0$, even if i assumes the remaining player would vote for A under rule 2 for sure.

for comparing strategy $(1, B, B, A)$ to $(2, B, B, A)$ because this only possibly leads to a different outcome if rule 2 was chosen by default, which is beyond the control of the players. For strategies $(1, B, A, A)$ and $(1, B, A, B)$ the same argument applies unless player i believes that both other players vote for A for sure if rule 2 is chosen by a player. For such a belief, the outcome would actually always be the same for $(1, B, A, v_c)$ and $(2, B, A, v_c)$ because in that case both when rule 1 is implemented and when rule 2, chosen by a player, is implemented, A will result whenever at least for one player $\theta > 0$. Thus we make the mild additional assumption that i does not assume that with certainty both other players are sufficiently altruistic such that they will vote for A even if they have $\theta < 0$ when rule 2 is implemented.

Now consider i in the voting stage when the rule has been determined to be rule 2, not chosen by any subject, so i has not learned anything about the other subjects' valuations. If i's vote is relevant, then the other two players are forced to vote according to their monetary preferences. Thus i is pivotal and voting for A implements A if exactly one of the other players has $\theta > 0$. Thus if i prefers A when exactly one of the others has $\theta > 0$, then i should vote for A. Hence if i chooses rule 1, but votes for B if the rule is exogenously chosen to be rule 2, then this is inconsistent with our model, but more generally with any model that is consistent with choosing rule 1 in the first place (which must include altruism whereas social preferences such as inequality aversion do not imply voting against monetary preferences for anybody with $\theta < 0$).

Note that even if player i thinks the others are altruists who would want to vote for A given rule 2 to help one player with a positive payoff, i does not have an incentive to strategically choose rule 1 for $\theta_i < 0$. Given that only one player can freely choose in the voting stage, it cannot happen given rule 2 that A gets implemented even though both other players have a negative valuation but are altruistic. Thus given rule 2, A can only result if at least one of the others has a positive valuation. But if that is the case, A also results if i chooses rule 1 (unless the player with a positive valuation is chosen to be the one who can vote and votes for B, which is unlikely), and A will also result if rule 1 is chosen and a confused altruist's vote is chosen to count. Thus if $\theta_i < 0$ and i wants B, then i should choose rule 2.

Because the participants exhibited substantial problems in understanding the instructions in Treatment 4, which implemented this game, Treatment 5 was based on a simplified variant. Specifically, the option that the computer automatically implements rule 2 with probability $\frac{1}{4}$ was eliminated. This simplifies the description of the first stage substantially. In the first stage now simply all three players choose a rule and one of them is then randomly selected and her chosen rule is implemented. The second stage remains unchanged, because for the test of inconsistency, it is essential that players know that if their vote counts, the other two players are voting in line with their monetary preferences. Simplifying the first stage now means that if a player learns that rule 2 has been implemented, he knows that this has been chosen by another player. Since it is highly implausible that a player with $\theta = 14$ chooses rule 2, the player should thus infer that another player has $\theta = -1$. Nevertheless, if rule 2 is implemented, a player can still infer that he is only pivotal (in the sense that his vote is chosen to count and changes the outcome) if exactly one other player has $\theta = 14$. Hence if he prefers outcome A in that case, he should vote for A. Regarding the choice in the first stage, the same argument as above holds, that is a player with $\theta = -1$ should choose rule 1 only if he prefers A if exactly one player has $\theta = 14$. Hence a player with $\theta = -1$ choosing rule 1 but voting for B if rule 2 is implemented behaves inconsistently in the sense that his first-stage rule choice suggests a higher degree of altruism than his second-stage vote. The only difference between Treatments 4 and 5 is that in the latter, a player might (implausibly) assume that a player with $\theta = 14$ will choose rule 2 and hence conclude from the implementation of rule 2 that at least one other player has $\theta = 14$. However, this does not change anything about the conclusions from being pivotal. A wrong conditioning on being pivotal might possibly make such a player be more inclined to vote for A if rule 2 is implemented (because he now "knows" that one other player has $\theta = 14$ rather than inferring that from being pivotal). This, however, only strengthens the argument for voting A and thus still a player who chooses rule 1 but votes for B when rule 2 is implemented is behaving inconsistently.

Treatments 4 and 5 were implemented with 45 and 63 subjects, respectively, with random matching into groups of three in sessions with 12 to 24 participants. Otherwise, the same procedures were followed as for the other treatments.

4.2 Results

In Treatment 4, we observe 62 instances where a subject had $\theta = -1$ and votes for rule 1 (18 percent of the cases where $\theta = -1$). In only 10 of these cases the rule is then chosen by default. In three cases the participants then make the inconsistent vote B. Thus, the number of observations where an unambiguous inconsistency across stages could in principle be detected is lower than expected and it is clearly too low to draw any clear inferences.[28] There are two further reasons why the results from Treatment 4 were not as informative as we expected. First, it turned out that the participants exhibited substantial problems understanding the complicated instructions, which is the reason we then implemented the simpler Treatment 5. Second, due to a programming error, while the default rule choice was announced, it was not actually implemented. As a consequence, while at the voting stage subjects were informed in case of a default rule choice that the rule would be rule 2, the actually implemented rule was sometimes instead rule 1, which subjects then learned at the end of the period. As a result, their confusion has likely grown (though surprisingly,

[28] In particular, the 10 observations where an inconsistency could potentially be observed are concentrated on only six distinct participants, three of which then behave inconsistently exactly once.

nobody raised this issue during the experiment or in the post-experimental questionnaire). In any case, while Treatment 4 shows some evidence of a behavioral inconsistency across stages, its results should not be over-interpreted because the number of relevant observations was lower than expected and because many subjects were likely too confused.[29]

Treatment 5 provides clearer support to the view that there is an inconsistency across stages. In this Treatment, in 66 instances a subject with $\theta = -1$ suggests rule 1 (only about 14 percent of the cases where $\theta = -1$). In 18 of these cases, rule 2 is chosen instead. In only 4 of these cases, the subject then votes for A. Thus in 14 out of 18 cases where inconsistent behavior could occur, this is indeed observed. The 14 inconsistent choices are made by 10 distinct subjects, the four consistent choices by three distinct subjects, where one subject decides once consistently and once inconsistently. According to a binomial test, 14 out of 18 is significantly different from an equal distribution ($p < 0.05$). Taking into account repeat observations and that only 12 distinct subjects were ever in a situation where they could behave inconsistently, the 10 out of 12 who decide at least once inconsistently are also significantly different from an equal distribution ($p < 0.05$), whereas the nine out of 12 who are never consistent are not ($p > 0.1$).

Apparently, confusion also matters to a large degree in Treatment 5, because in 18 of the 48 cases where rule 1 is implemented after a subjects chose rule 1, the subject votes for A, which reduces both own and total payoff if the vote becomes relevant. Still, the share of 10 out of 12 participants choosing inconsistently would require an even larger share of confused subjects.[30] This indicates that indeed inconsistencies in the sense of subjects being more concerned about efficiency in the rule-choice stage than in the voting stage could be a relevant phenomenon. Given the low absolute number of these instances observed even in Treatment 5, this phenomenon should be subject to further research.

[29] Due to the fact that given the relatively low share of choices for rule 1 by construction only few instances of possible inconsistency could occur and because of the high degree of confusion, we decided not to re-run Treatment 4 with corrected software because this would very likely not have generated clearly interpretable data either. In post-experimental questionnaires, subjects in Treatment 4 mentioned more often than in the other treatments to have decided randomly. Some explain correctly how to decide selfishly in Treatment 4, but those that try to explain that they maximized expected total payoffs actually got it wrong because they explained that it would be good to always try to get outcome A. Hence there was indeed substantial confusion in Treatment 4.

[30] Just for illustration, ignoring the dependence between observations, 14 out of 18 and 18 out of 48 differ significantly according to a Fisher exact test ($p < 0.01$). Also in the post-experimental questionnaire of Treatment 5, some confusion is present in the sense that those who explain that they tried to maximize total payoffs do not show that they understand what it means to be pivotal for the different rules.

5 Discussion and Concluding Remarks

Our paper experimentally addresses a fundamental problem in political economics. Voting rules should adjust to the problem at hand while frequently, they are fixed in practice. Our results indicate that many voters are willing to efficiently adjust the voting rule even when they are already privately informed about their preferences for the voting outcome. Choosing in a two-step procedure, i.e. choosing a rule first and then voting according to this rule, can yield higher aggregate payoffs than simple majority voting. On the one hand, voters with a weak preference can express their low stakes by supporting a rule that does not maximize the probability that the preferred outcome according to their selfish preferences is implemented and then voting in line with their preference. On the other hand, people with strong preferences can choose rules that increase the probability that their preferred outcome is implemented, so their preference carries more weight. If strong preferences are substantially stronger than weak ones and voters are sufficiently efficiency concerned, the two-step procedure can yield more efficient outcomes than simple majority voting.

Our results suggest that it may be useful to let voters fix or alter voting thresholds in referenda. However, drawing concrete welfare or policy conclusions from our analysis requires additional research. In our analysis we have compared the aggregate payoffs that arise under different mechanisms, not welfare. This is not a problem if everybody is efficiency concerned or selfish, because then whatever maximizes payoffs also maximizes welfare. However, if some voters are indeed inequality averse, our criterion would be paternalistic since it only takes into account the payoffs from the project itself but not any disutility that might arise from inequality. Put differently, if voters are inequality averse, our analysis reveals little about the optimal (i.e., welfare-maximizing) design of voting institutions.

Another finding of our paper is that when choosing rules, different types of social preferences may matter than when choosing outcomes. Thus when predicting how experimental participants choose rules, one should be careful in taking experimental results from games with no procedural choice stage as a guidance of the prevalence of preferences. Efficiency appears to be a more important behavioral motive when subjects decide on procedures than if they decide on outcomes according to a given procedure. This effect could in principle be exploited to increase expected payoffs in democratic decisions or to end political gridlock in reform processes. Adding more pre-voting stages might result in even more altruistic behavior. However, it is difficult to draw conclusions regarding the optimal number of voting stages when preferences are not stable from one stage to another. Drawing conclusions regarding institutional design requires a theoretical framework that takes the instability of preferences across voting stages into account. A potential complication of trying to raise efficiency gains through a multi-stage procedure arises from strategic voting given that the rule choices reveal something about other players' valuations. This in turn can change strategic rule choices. Such effects can both harm or enhance efficiency gains, depending on whether the dominating effect is altruists taking the information they gain into account in their attempts to maximize total welfare or selfish players strategically exploiting the altruists' reactions.

In general, choosing rules (possibly at multiple levels) allows citizens to condition their voting behavior on information at various stages. In applications, people may know more or less about the others' preferences than in our experiment. For example, it is clearly not generally true that people know the underlying distribution of preferences. They may, however, often know crucial aspects of the distribution, such as that preferences against a reform project are much stronger than those in favor of it. Such information could often be enough to know that asymmetric voting rules are more efficient than majority voting in these cases.

How can we interpret that apparently subjects are more likely to deviate from selfish maximization of monetary payoffs in the rule-choice stage than in the voting stage? A first possible explanation is that at the rule-choice stage, the final outcome is still somewhat remote and this may reduce the temptation to act selfishly and hence make subjects choose somewhat more altruistically. A second possibility is that at the rule-choice stage, at least some participants believe they should choose what is the right thing to do, and choosing the efficient rule is likely perceived as the right thing to do. In spite of the fact that rules are chosen at an interim stage, choosing a rule may have a flavor of choosing what is best at an ex-ante stage, or put differently, even though rule choice does not happen behind a veil of ignorance, participants may think that it should.[31] In the voting stage, however, voting in line with one's own monetary preferences is likely to be perceived as legitimate.[32] Voting selfishly is quite generally accepted, whereas tilting rules in one's favor is not. Therefore, a simple outcome-oriented preference model is unable to adequately capture behavior across both stages in our two-stage voting game.[33] We also note that while we find that

[31] Another possibility is that subjects tend to interpret the choice of a rule as the choice of a general rule, i.e., they might misinterpret the instructions that their choice actually does happen behind a veil of ignorance in the sense that it also binds them for future periods, where they do not yet know their valuation. However, they should quickly learn that the rules may actually change from period to period. It turns out that subjects' behavior is fairly stable across periods and hence misunderstanding the temporary character of rules can at best explain the higher deviation from selfish rule choices in the first period.

[32] An alternative explanation why participants rather choose non-selfishly in the rule-choice stage is that because in this stage they act as a dicator (if their choice matters), they feel more responsible for the outcome and hence obliged to make the efficient choice.

[33] A further potential alternative explanation would be self-image concerns if participants believe that their choice in the first stage is less likely to be relevant than in the second stage and hence a non-selfish choice is less likely to

rule choices are less selfish than voting behavior, overall behavior in our experiment is relatively selfish compared to results in many other experimental games. For example, in Treatment 1, even with $|\theta| = 0.1$, the majority of players chose the selfish rule. A possible explanation is that our experiment is complex, which might remove the focus from concerns for other players and gains from cooperation to the task of trying to find out what works for oneself in the first place. This is in line with the observation that behavior in the more complicated Treatments 4 and 5 is more selfish than in Treatment 2.[34]

Voting mechanisms where players first decide about the majority threshold are also related to the concept of liberal paternalism (Thaler and Sunstein, 2008). A social planner might know that given the distribution of valuations, the optimal voting threshold may deviate from the simple majority rule. However, he may lack the political legitimacy to implement this rule. Instead, he may thus leave the decision regarding the threshold to the participants, who will, if they are sufficiently efficiency concerned, choose the optimal threshold. Naturally, this will not work perfectly because participants in this mechanism will not all be sufficiently efficiency concerned. However, it allows for some gains compared to exogenously assigned majority voting. These gains can be increased, if the rule choice optimally exploits the efficiency concerns of those who have any.

We note that the specific two-stage procedure we have implemented is not well suited at capturing efficiency gains because the random-dictator mechanism at the rule-choice stage is very inefficient since it does not aggregate information about preferences and because it allows a randomly chosen person to implement her preferred choice independent of the preferences and behavior of the other participants. This leads in many individual decisions problems to worse outcomes than majority voting. We have chosen this procedure to obtain evidence regarding subjects' individually preferred voting rules that is not affected by first-stage strategic considerations. A more subtle way of aggregating preferences for rules makes use of more than just one subject's proposal. In our asymmetric treatment we found that in about 44 percent of the cases subjects suggested a rule that was not completely biased in favor of alternative B. This indicates that an appropriately designed majority rule for the choice of the voting rule may produce better results than our simple

be costly in monetary terms. Tonin and Vlassopoulos (2013) make a corresponding observation in repeated dictator games that are not all paid out. This assumption is, however, not very plausible in our experiment. With the unanimity rules, participants are rarely pivotal and can expect not to be so, so that non-selfish votes are very likely not costly, yet we see few of such votes. Furthermore, the image gain from altruistic acts could go down with each act, but then we should see also a decrease of non-selfish behavior across periods and not only across stages, which we do not.

[34]Interestingly, in the post-experimental questionnaire both selfish participants as well as efficiency maximizers often call those choosing differently confused or stupid.

random dictator mechanism. Testing and comparing such more complex multi-stage mechanisms is worth additional research.

Overall, we still know rather little about how people behave in two-stage voting procedures. Our experiment provides only a few early steps. More research is required to understand whether, when and how one would want to implement such procedures in practice. As explained in the introduction, informal rule-choice procedures can be observed in various social groups. A challenging task is to develop and experimentally test more formal real-world decision procedures. One first step could be to experiment with hybrid voting mechanisms in local referenda. Such hybrid mechanisms would ask voters to simultaneously approve both a decision rule and, contingent on the realized rule, a decision outcome. More generally, other mechanisms such as matching and double auction mechanisms might also benefit from the introduction of a pre-play stage that enables efficiency-concerned participants to express their beliefs about the underlying distribution of types as well as their own intensity of preferences and, related to that, their preferred mechanisms.

References

[1] Austen-Smith, D. and J.S. Banks (1996). "Information Aggregation, Rationality, and the Condorcet Jury Theorem." *American Political Science Review*, 90, 34–45.

[2] Balafoutas, L., M.G. Kocher, L. Putterman, and M. Sutter (2013). "Equality, Equity and Incentives: An Experiment." *European Economic Review* 60, 32–51.

[3] Barbera, S. and M.O. Jackson (2004). "Choosing How to Choose: Self-Stable Majority Rules and Constitutions." *Quarterly Journal of Economics* 119, 1011–1048.

[4] Bergemann, D. and S. Morris (2005). "Robust Mechanism Design." *Econometrica* 73, 1771–1813.

[5] Bergemann, D. and S. Morris (2012). *Robust Mechanism Design: The Role of Private Information and Higher Order Beliefs.* World Scientific, Singapore.

[6] Bierbrauer F. and M. Hellwig (2011). "Mechanism Design and Voting for Public-Good Provision." Max Planck Institute for Research on Collective Goods Working Paper 2011/31.

[7] Blanco, M., D. Engelmann, and H.-T. Normann (2011). "A Within-Subject Analysis of Other-Regarding Preferences." *Games and Economics Behavior* 72, 321–338.

[8] Börgers, T. and P. Postl (2009). "Efficient Compromising." *Journal of Economic Theory* 144, 2057–2076.

[9] Brennan, G. and J.M. Buchanan (1985). "The Reason of Rules: Constitutional Political Economy." Cambridge: Cambridge University Press.

[10] Casella, A. (2005). "Storable Votes." *Games and Economic Behavior* 51, 391–419.

[11] Casella, A., A. Gelman, and T.R. Palfrey (2006). "An Experimental Study on Storable Votes." *Games and Economic Behavior*, 57, 123–154.

[12] Charness G. and M. Rabin (2002). "Understanding Social Preferences with Simple Tests. " *Quarterly Journal of Economics* 117, 817–869.

[13] Cramton, P., R. Gibbons, and P. Klemperer (1987). "Dissolving a Partnership Efficiently." *Econometrica* 55, 615–32.

[14] Deb, R., I.K. Ghosh, and T.K. Seo (2011). "Referenda and the Provision of Public Goods." Mimeo, Southern Methodist University.

[15] Engelmann, D. and V. Grimm (2012). "Mechanisms for Efficient Voting with Private Information about Preferences." *Economic Journal* 122, 1010–1041.

[16] Engelmann, D. and M. Strobel (2000). "The False Consensus Effect Disappears if Representative Information and Monetary Incentives are Given." *Experimental Economics* 3, 241–260.

[17] Engelmann, D. and M. Strobel (2004). "Inequality Aversion, Efficiency and Maximin Preferences in Simple Distribution Experiments." *American Economic Review* 94, 857–869.

[18] Engelmann, D. and M. Strobel (2012). "Deconstruction and Reconstruction of an Anomaly." *Games and Economics Behavior* 76, 678–689.

[19] Feddersen, T.J. and W. Pesendorfer (1998). "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting." *American Political Science Review* 92, 23–35.

[20] Fehr, E. and K.M. Schmidt (1999). "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114, 817–868.

[21] Fischbacher, U. (2007). "z-Tree: Zurich Toolbox for Ready-made Economic experiments." *Experimental Economics* 10, 171–178.

[22] Güth, W. and M.F. Hellwig (1986). "The Private Supply of a Public Good." *Journal of Economics* 5, 121–159.

[23] Greiner, B. (2004). "An Online Recruitment System for Economic Experiments." In: K. Kremer and V. Macho (Eds.) *Forschung und wissenschaftliches Rechnen* 2003. GWDG Bericht 63, 79–93. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

[24] Grüner, H.P. and Y. Koriyama (2012). "Public Goods, Participation Constraints, and Democracy: A Possibility Theorem." *Games and Economic Behavior* 75, 152–167.

[25] Hortala-Vallve, R. and A. Llorente-Saguer (2010). "A Simple Mechanism for Resolving Conflict." *Games and Economic Behavior*, 70, 375–391.

[26] Jackson, M.O. and H.F. Sonnenschein (2007). "Overcoming Incentive Constraints by Linking Decisions." *Econometrica* 75, 241–258.

[27] Kaplan, T.R. and B.J. Ruffle (2012). "Which Way to Cooperate." *Economic Journal* 122, 1042–1068.

[28] Masatlioglu, Y., S. Taylor and N. Uler (2012). "Behavioral Mechanism Design: Evidence from the Modified First-Price Auction", *Review of Economic Design* 6, 159–173.

[29] Myerson, R. (1981). "Optimal Auction Design." *Mathematics of Operations Research* 6, 58–73.

[30] Myerson, R. and M. Satterthwaite (1983). "Efficient Mechanisms for Bilateral Trading." *Journal of Economic Theory* 29, 265–81.

[31] Oleszek, W.J. (2014). "Congressional Procedures and the Policy Process." 9^{th} Edition. Washington DC: CQ Press.

[32] Rawls. J (1972). "A Theory of Justice." Oxford: Clarendon Press.

[33] Ross, L., D. Greene, and P. House, (1977). "The 'False Consensus Effect': An Egocentric Bias in Social Perception and Attribution Processes." *Journal of Experimental Social Psychology* 13, 279–301.

[34] Segal, I. and M. Whinston (2011). "A Simple Status Quo that Ensures Participation (with Application to Efficient Bargaining)." *Theoretical Economics* 6, 109–125 .

[35] Schmitz, P. W. (2002). "Simple Contracts, Renegotiation under Asymmetric Information, and the Hold-up Problem." *European Economic Review* 46, 169–188.

[36] Schmitz, P.W. and T. Tröger (2012). "The (Sub-)Optimality of the Majority Rule." *Games and Economic Behavior* 74, 651–665.

[37] Thaler., R.H. and C.R. Sunstein (2008). "Nudge: Improving Decisions about Health, Wealth, and Happiness." Yale University Press.

[38] Tonin, M. and M. Vlassopoulos (2013). "Experimental Evidence of Self-Image Concerns as Motivation for Giving." *Journal of Economic Behavior and Organization* 90, 19–27.

[39] Trebbi, F., P. Aghion, and A. Alesina (2008). "Electoral Rules and Minority Representation in U.S. Cities". *Quarterly Journal of Economics* 123, 325–357.

[40] Wilson, R. (1985). "Incentive Efficiency in Double Auctions". *Econometrica* 53, 1101-1115.

A Appendix

A.1 Modelling Efficiency Concerns

In this section we study the optimal rule choice of an agent who cares about a weighted sum of all $2n+1$ players' monetary payoffs in the game underlying Treatment 1. Assume that this player has altruistic preferences captured by a utility function $u_i(\theta) = \gamma_i \theta_i + (1-\gamma_i)\overline{\theta}_{-i}$, where $\overline{\theta}_{-i} = \frac{1}{2n}\sum_{j \neq i} \theta_j$. We assume that agent i expects that, for any rule, all players (including player i himself) vote symmetrically and strictly selfishly in the second stage.[35] That is, player j votes for A if $\theta_j > 0$, votes for B if $\theta_j < 0$ and flips a fair coin if $\theta_j = 0$. Given that the rule that player i chooses only affects outcomes if that rule is chosen, we can treat the rule choice as if player i could decide the rule for sure.

It is useful to first introduce some additional notation. Denote with p_k^+ the probability that alternative A is chosen if the rule is k and player i votes for A and p_k^- the probability that A is chosen if the rule is k and i votes for B.

Let E_k^+ be the expected average payoff of the $2n$ other players given rule k and player i voting for A and E_k^- the expected average payoff of the other players given rule k and player i voting for B. Let ω_j be the expected average of the valuations of the other players conditional on at least j among these players voting for A. Then $E_k^+ = p_k^+ \omega_{k-1}$ (because if at least $k-1$ of the other players vote for A and player i votes for A, then at least k vote for A, so A is chosen) and $E_k^- = p_k^- \omega_k$ (because if at least k of the other players vote for A and player i votes for B, then at least k vote for A, so A is chosen). Moreover $E_k^+ = p_k^+ \omega_{k-1} = p_{k-1}^- \omega_{k-1} = E_{k-1}^-$ for all $k \in \{2, .., n\}$

Finally, denote by

$$u_i^+(\theta_i, k) := \gamma p_k^+ \theta_i + (1-\gamma)\, E_k^+$$

the expected utility of player i under rule k conditional on all other players voting selfishly in the above defined sense and player i voting for alternative A. We define $u_i^-(\theta_i, k)$ and $u_i^0(\theta_i, k)$, the expected utility when player i votes for B or flips a fair coin, respectively, in an analogous way.

Note that for all $k = 1, .., n$ the functions $u_i^+(\theta_i, k)$ and $u_i^-(\theta_i, k)$ are linear in the valuation θ_i. Moreover $u_i^+(\theta_i, 1) = \gamma \theta_i$ because if player i votes for A if rule 1 is chosen, A is implemented for sure and hence nothing can be inferred about the valuations of the other players and hence their conditional expected valuation is 0 given the symmetry of the underlying distribution of valuations, $u_i^-(\theta_i, 2n+1) = 0$ (because in that case, B is chosen for sure since unanimity would be required for A but is prevented by i), $u_i^+(0, k) = (1-\gamma)\, E_k^+ \geq 0$ because for rule 1, nothing is learned about the other players' valuations from the implementation of A, whereas for the other rules it can be inferred that at least one of the other players has a positive valuation, and hence

[35] The optimality of the assumed second stage voting behavior will be discussed below.

in that case the inequality is actually strict, and $u_i^-(0,k) = (1-\gamma)\, E_k^- > 0$ because for any rule it can be inferred that at least one of the other players has a positive valuation.

It is a direct consequence of Proposition 2 in Schmitz and Tröger (2012) that the simple majority rule is optimal for agent i when his valuation is zero. This, in combination with the strict monotonicity of the slopes of $u_i^+(\theta_i, k)$ and $u_i^-(\theta_i, k)$ (p_k^+ and p_k^-) in k implies that rules $k > n+1$ ($k < n+1$) are suboptimal for agents with strictly positive (negative) valuations. Therefore, for agents with strictly positive (negative) valuations, it remains to compare all rules k with $k \leq n+1$ ($k \geq n+1$). For $k \geq n+1$, $u_i^-(0,k)$ is strictly monotonously decreasing in k and for $k \leq n+1$, $u_i^+(0,k)$ is strictly monotonously increasing in k.[36]

This and the strict monotonicity of the slopes of $u_i^+(\theta_i, k)$ and $u_i^-(\theta_i, k)$ in k implies that there are unique valuations $\theta_{kk'} < 0$ at which players are indifferent between any two rules $k \neq k'$ with $k, k' \geq n+1$. For the same reason, there are unique valuations $\theta_{kk'} > 0$ at which players are indifferent between any two rules $k \neq k'$ with $k, k' \leq n+1$. The following symmetry property is useful to prove our main result.

Lemma 1: For all $n \geq 1$, for all $j \in \{-n, ..., n\}$ and for all $\theta_i > 0$:

$$u_i^+(\theta_i, n+1+j) - \frac{\gamma}{2}\theta_i = u_i^-(-\theta_i, n+1-j) - \frac{\gamma}{2}(-\theta_i).$$

Proof

$$u_i^+(\theta_i, n+1+j) = p_{n+1+j}^+ \left(\gamma\theta_i + (1-\gamma)\omega_{n+j}\right)$$

$$u_i^-(-\theta_i, n+1-j) = p_{n+1-j}^- \left(-\gamma\theta_i + (1-\gamma)\omega_{n+1-j}\right)$$

Hence the claim is that

$$p_{n+1+j}^+ \left(\gamma\theta_i + (1-\gamma)\omega_{n+j}\right) = p_{n+1-j}^- \left(-\gamma\theta_i + (1-\gamma)\omega_{n+1-j}\right) + \gamma\theta_i.$$

We know that $p_{n+1+j}^+ = 1 - p_{n+1-j}^-$ (if the rule is $n+1+j$ and i votes for A, then A is implemented if at least $n+j$ others vote for A and thus if they have a positive valuation, while if the rule is $n+1-j$ and i votes for B, then B is implemented if at least $2n-(n+1-j)+1 = n+j$ others vote for B and thus if they have a negative valuation, and at least $n+j$ of the others having

[36] When $k \leq n+1$, E_k^+ is increasing in k, because then cases where fewer than half of the other players have a positive valuation are dropped in the calculation of E_k^+. Similarly, when $k \geq n+1$, E_k^- is decreasing in k, because then cases where more than half of the other players have a positive valuation are dropped in the calculation of E_k^-.

a positive valuation is equally likely as at least $n+j$ of them having a negative valuation). Hence the previous equality follows from

$$p^{+}_{n+1+j}\omega_{n+j} = p^{-}_{n+1-j}\omega_{n+1-j}.$$

Call $p_{m,2n}$ the probability that m out of $2n$ agents vote for A. We have, with $\bar{\theta}^{+}$ denoting the expected valuation conditional on voting for A:

$$\begin{aligned}
p^{+}_{n+1+j}\omega_{n+j} &= \sum_{m=n+j}^{2n} p_{m,2n}\left(m-(2n-m)\right)\bar{\theta}^{+} \\
&= 2\bar{\theta}^{+}\sum_{m=n+j}^{2n} p_{m,2n}\left(m-n\right),
\end{aligned}$$

and

$$\begin{aligned}
p^{-}_{n+1-j}\omega_{n+1-j} &= \sum_{m=n+1-j}^{2n} p_{m,2n}\left(m-(2n-m)\right)\bar{\theta}^{+} \\
&= 2\bar{\theta}^{+}\sum_{m=n+1-j}^{2n} p_{m,2n}\left(m-n\right).
\end{aligned}$$

Therefore, our claim can be rewritten as:

$$\begin{aligned}
2\bar{\theta}^{+}\sum_{m=n+j}^{2n} p_{m,2n}\left(m-n\right) &= 2\bar{\theta}^{+}\sum_{m=n+1-j}^{2n} p_{m,2n}\left(m-n\right) \\
\sum_{m=n+j}^{2n} p_{m,2n}\left(m-n\right) &= \sum_{m=n+1-j}^{2n} p_{m,2n}\left(m-n\right) \\
\sum_{m=n+1-j}^{n+j-1} p_{m,2n}\left(m-n\right) &= 0 \\
\sum_{m=n+1-j}^{n-1} p_{m,2n}\left(m-n\right) + \sum_{m=n+1}^{n+j-1} p_{m,2n}\left(m-n\right) &= 0 \\
\sum_{m=n+1}^{n+j-1} p_{m,2n}\left(m-n\right) &= \sum_{m=n+1-j}^{n-1} p_{m,2n}\left(n-m\right).
\end{aligned}$$

The last inequality holds because $p_{k,2n} = p_{2n-k,2n}$.

Q.E.D.

Based on these preliminaries, we can prove the main result for the case of five players, i.e. $n = 2$.

Proposition 1: For $n = 2$, there exist $\underline{\theta}_i < \overline{\theta}_i < 0$ such that player i's preferred rule is rule 5 for $\theta_i < \underline{\theta}_i$, rule 4 for $\underline{\theta}_i < \theta_i < \overline{\theta}_i$, rule 3 (majority rule) for $\overline{\theta}_i < \theta_i < -\overline{\theta}_i$, rule 2 for $-\overline{\theta}_i < \theta_i < -\underline{\theta}_i$, and rule 1 for $\theta_i > -\underline{\theta}_i$.

Proof Consider first $\theta_i < 0$. If player i chooses rule 5, then her utility is 0 for sure (because she will vote for B in the second stage). If she chooses rule 3 (majority rule) and her choice is implemented, then her expected utility is $p_3^-(\gamma\theta_i + (1-\gamma)\,\omega_3)$. So she is indifferent between rule 5, which implements B for sure and the majority rule, if $p_3^-(\gamma\theta_i + (1-\gamma)\,\omega_3) = 0$ or $\theta_i = -\frac{(1-\gamma)}{\gamma}\omega_3$.

Now consider the intermediate rule 4. Player i's expected utility for $\theta_i = -\frac{(1-\gamma)}{\gamma}\omega_3$ and rule 4 is

$$p_4^-(\gamma\theta_i + (1-\gamma)\,\omega_4) = p_4^-(-(1-\gamma)\,\omega_3 + (1-\gamma)\,\omega_4) > 0,$$

because $\omega_4 > \omega_3$, since the expected average payoff ω_k increases strictly in k. Thus at the point where i is indifferent between rule 5 and the majority rule, i strictly prefers the intermediate rule.

The first part of the proposition follows from this and from (i) the linearity of all conditional utility functions (ii) the strict monotonicity of their slopes in k and (iii) the fact that rule 3 strictly maximizes welfare conditional on $\theta_i = 0$. The second part of the proposition follows from the symmetry property shown in Lemma 1, since, for all $\theta_i > 0$ and $j \in \{0, 2\}$,

$$\begin{aligned}
& u_i^+(\theta_i, 2) = u_i^+(\theta_i, 3-j) \\
\Rightarrow\quad & u_i^+(\theta_i, 2) - \frac{\gamma}{2}\theta_i = u_i^+(\theta_i, 3-j) - \frac{\gamma}{2}\theta_i \\
\Rightarrow\quad & u_i^-(-\theta_i, 4) - \frac{\gamma}{2}(-\theta_i) = u_i^-(-\theta_i, 3+j) - \frac{\gamma}{2}(-\theta_i) \\
\Rightarrow\quad & u_i^-(-\theta_i, 4) = u_i^-(-\theta_i, 3+j).
\end{aligned}$$

Hence, indifference between rules 2 and 3 (2 and 1) for players with some type $\theta_i > 0$ implies indifference of a player with type $-\theta_i$ between rules 4 and 3 (4 and 5). Q.E.D.

Hence, assuming selfish voting behavior, the optimal rule choice of efficiency-concerned subjects is monotonous and symmetric and, when the type space is large enough, it covers the entire set of available rules.

Moreover, one can easily show that, when player i has no additional knowledge about other players' types, one can rule out that player i prefers to vote non-selfishly under any voting rule that he has proposed himself. This follows because $u_i^+(\theta_i, k) = u_i^-(\theta_i, k-1)$ for all $k = 2, .., n$. Hence, if $\theta_i > 0$ and $u_i^+(\theta_i, k) \geq u_i^+(\theta_i, k')$ for all $k' = 1, .., 5$ then $u_i^+(\theta_i, k) \geq u_i^-(\theta_i, k'-1)$ for all $k' = 2, .., 5$. Furthermore, choosing rule 5 and voting for B is worse than choosing rule k and voting for A, i.e., if $\theta_i > 0$, $u_i^+(\theta_i, k) \geq u_i^-(\theta_i, 5) = 0$. The same type of argument can be made to rule out that players with negative valuations are better off by voting insincerely.

Selfish voting may instead not be optimal if another player's rule proposal has been successful. In such cases, player i can infer something about other players' types from being pivotal in the voting stage. Moreover, player i may learn something about other player's expected valuations from the rule chosen in stage 1. This is why non-selfish voting may be optimal for small positive or negative values of i's willingness to pay. Similarly, if an efficiency-concerned player expects other players to vote non-selfishly if certain rules are chosen, the rule choice may deviate from the one derived in Proposition 1.