DISCUSSION PAPER SERIES

DP15099

SEMI-FLEXIBLE MAJORITY RULES FOR PUBLIC GOOD PROVISION

Hans Gersbach and Oriol Tejada

PUBLIC ECONOMICS

ISSN 0265-8003

SEMI-FLEXIBLE MAJORITY RULES FOR PUBLIC GOOD PROVISION

Hans Gersbach and Oriol Tejada

Discussion Paper DP15099
Published 26 July 2020
Submitted 22 July 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

SEMI-FLEXIBLE MAJORITY RULES FOR PUBLIC GOOD PROVISION

Abstract

We introduce semi-flexible majority rules for public good provision with private valuations. Such rules take the form of a two-stage, multiple-round voting mechanism where the output of the first stage is the default alternative for the second stage and the voting thresholds (a) vary with the proposal on the table and (b) require a qualified majority for final approval in the second stage. We show that the (detail-free) mechanism elicits the information about the valuations and uses it to implement the utilitarian optimal public-good level if valuations can be only high or low. This level is chosen after all potential socially optimal policies have been considered for voting. We explore ways to reduce the number of voting rounds and develop a compound mechanism when there are many types of citizens to approximate the optimal public-good level.

JEL Classification: C72, D70

Keywords: Voting - Utilitarianism - Implementation - Procedural democracy

Hans Gersbach - hgersbach@ethz.ch
ETH Zurich and CEPR

Oriol Tejada - toriol@ethz.ch
ETH Zurich

Acknowledgements
We would like to thank Philip Grech, Matthew Jackson, Vincent Lohmann, and Hervé Moulin, as well as the participants of the 13th SING Conference in Paris, for valuable comments. All errors are our own.

Semi-flexible Majority Rules for Public Good Provision*

Hans Gersbach
CER-ETH – Center of Economic Research
at ETH Zurich and CEPR
Zürichbergstr. 18, 8092 Zurich, Switzerland
hgersbach@ethz.ch

Oriol Tejada
CER-ETH – Center of Economic Research
at ETH Zurich
Zürichbergstr. 18, 8092 Zurich, Switzerland
toriol@ethz.ch

This Version: July 2020

Abstract

We introduce semi-flexible majority rules for public good provision with private valuations. Such rules take the form of a two-stage, multiple-round voting mechanism where the output of the first stage is the default alternative for the second stage and the voting thresholds *(a)* vary with the proposal on the table and *(b)* require a qualified majority for final approval in the second stage. We show that the (detail-free) mechanism elicits the information about the valuations and uses it to implement the utilitarian optimal public-good level if valuations can be only high or low. This level is chosen after all potential socially optimal policies have been considered for voting. We explore ways to reduce the number of voting rounds and develop a compound mechanism when there are many types of citizens to approximate the optimal public-good level.

Keywords: voting; utilitarianism; implementation; procedural democracy

JEL Classification: C72, D70

*We would like to thank Philip Grech, Matthew Jackson, Vincent Lohmann, and Hervé Moulin, as well as the participants of the 13^{th} SING Conference in Paris, for valuable comments. All errors are our own.

1 Introduction

Motivation and goal

Two-stage voting processes in which the output of the first stage is the default alternative for the second stage are common. In direct democracies, citizens can make a proposal—say, by gathering enough signatures—that is later amended, approved, or rejected by parliament. In representative democracies, either the executive or one legislative chamber can make a first proposal, to be ratified or dismissed by a second legislative chamber.[1]

In this paper, we explore whether there exists such a two-stage voting process that yields desirable properties both from a utilitarian welfare perspective (*endstate justice*) as well as from a procedural viewpoint (*procedural justice*) (Moulin, 2008). For this purpose, we introduce semi-flexible majority rules. These rules take the form of a two-stage, multiple-round voting mechanism that takes the output of the first stage as the default alternative for the second stage, in which the thresholds used for the different voting rounds *(i)* vary with the proposal on the table, and *(ii)* require a qualified majority for final approval (in the second stage).

Model

To explore the properties of the above voting process, we consider a society that has to determine the level of some public good, which is financed through uniform taxes. Individuals have private information about their preferred public-good level. The voting process involves the set of alternatives—i.e., the set of possible outcomes—consisting of all public-good levels that are utilitarian optimal for some realization of the citizens' valuations.

The two stages of the voting process are as follows: In each round, individuals vote for one of two options and the only information the mechanism uses is whether or not the ratio between the number of votes for each option reaches a given round-specific threshold. In any first-stage round, individuals either vote to set a given alternative as default and then to jump to the second stage or, alternatively, to proceed to the next round within this first stage—if there is one. In the next round of this first stage, a similar binary voting takes place with a different alternative. In any second-stage round, in turn, individuals vote either to implement a given alternative and stop the mechanism or to proceed to the next round within this same stage. If no second-stage threshold is attained, the default alternative chosen in the first stage is implemented. If no first-

[1] This is the case in many countries such as the US, Germany, and Spain. For Spain, see `http://www.senado.es/web/conocersenado/temasclave/procedimientosparlamentarios/detalle/index.html?id=PROCLEGORD`, retrieved 9 December 2019 (in Spanish).

stage threshold is attained, the lowest public-good level among those that can be utilitarian optimal is taken as default for the second stage.

Main result

We look for equilibria of this voting process (or mechanism). Our equilibrium notion is perfect Bayesian equilibrium in pure strategies. Moreover, to avoid implausible outcomes, agents iteratively eliminate weakly dominated strategies, moving backwards in the game. The main result of the paper—Theorem 1—shows that if the order in which the alternatives are considered across rounds for binary voting and if the thresholds used therein are chosen appropriately, the outcome of the mechanism maximizes (ex-post) utilitarian welfare. An important feature of our mechanism—and thus of Theorem 1—is that the thresholds used across voting rounds do not depend on the prior distribution of individual types, which means that our mechanism is *detail-free*. In other words, for our mechanism to yield the utilitarian optimal solution, no statistical information is needed about how types are (privately) determined.

The voting mechanism that delivers this result determines the following order of alternatives and thresholds: In the first stage, starting from the lowest public-good level as a default alternative for the second stage, moving to higher public-good levels requires the support of a small share of citizens, which is lower than half plus one of the citizens. Moving to even higher levels, however, requires the support of a larger share of citizens, until it reaches the entire society. If every individual agrees, the highest possible public-good level serves as the status quo for the second stage. In the second stage, the approval of any alternative also requires a round-specific qualified majority. In contrast to the first stage, the vote thresholds in any second-stage round are never lower than half of the citizens plus one individual. While the first stage considers policies from low public-good levels to higher ones, the second stage reverses this order. Accordingly, in each stage our mechanism features monotone thresholds.

Our contributions

With this voting process—called *Semi-flexible Majority Rules*—we want to make three contributions to the literature. First, it is well-known that one-stage voting with flexible majority rules can yield utilitarian welfare maximization (Gersbach, 2017; Gershkov et al., 2017). Flexible majority rules allow approval thresholds to depend on the proposal on the table, without imposing that such thresholds must require a majority of votes. Sometimes, however, implementing a mechanism based on flexible majority rules means that the policy eventually adopted *(i)* is explicitly approved only by a minority of the citizenry and *(ii)* is not pitted against all potentially socially optimal alternatives. To avoid these features, we consider two properties, *majority approval* and *inclusiveness*, which semi-flexible majority rules fulfill.

The first property—called *(majority) approval*—is that the final policy decision has to be explicitly approved (in the second, final stage) by at least half of the citizens. Because of this property, our rules are called *semi-flexible majority rules* instead of *flexible majority rules*.[2] Voting procedures based on flexible majority rules that are constrained to require a majority of votes are already applied in practice.[3] The requirement that thresholds must require a majority does not, however, apply to the first stage. This is typically the case in parliaments when a minority can come up with a proposal to be put on the table. The proposal is adopted if it receives support first of the minority and second by the majority of parliament members (see e.g. the Bundestag in Germany). The property that different stages in decision-making face different restrictions with regard to the level of the voting thresholds is also featured by direct democracies such as Switzerland or California—see Section 3. From a normative viewpoint, guaranteeing that proposal-making in the first stage is not constrained by majority requirements is meant to enable small groups to initiate political action, no matter the (initial) support they have in the society. The second property fulfilled by our mechanism is that all potential socially optimal alternatives are really considered in at least one voting round—this property is called *inclusiveness*. There are three rationales for imposing this latter property: procedural fairness, robustness to individual mistakes in voting, and minimal payoff guarantees. These rationales are detailed in Section 3.2.[4]

We impose approval in the design of the mechanism, while we show that inclusiveness obtains in equilibrium. The mechanism we suggest can then be most relevant in (democratic) settings where aspects of the decision-making mechanisms such as those identified by these two properties are important besides the usual requirement that such mechanisms implement the utilitarian optimal solution. In other words, our mechanism can be appealing both in terms of "ends" (endstate justice) and "means" (procedural justice), to follow Moulin (2008). Of course, there exist other mechanisms that implement the utilitarian optimal solution, but none is constrained

[2]Flexible majority rules have been formally introduced in Gersbach (2005) and Gersbach (2009) and surveyed in Gersbach (2017).

[3]We refer to the examples provided by Gersbach (2017) and Gershkov et al. (2017) regarding some voting schemes used in US states such as Nebraska and Florida. In all the examples these authors provide, the thresholds not only vary with the proposal on the table but also require a majority of votes. More generally, the principles that *(i)* larger changes require larger thresholds—this is called *a maiore ad minus*—and that *(ii)* all changes should be approved at least by a majority are embedded in the constitutions of many countries. This is, for instance, the case in Spain (see `http://www.constitutionnet.org/files/constitutional_amendment_procedures.pdf`, retrieved 18 January 2017).

[4]Our mechanism also avoids cycling of policy-making.

by approval and inclusiveness. Because these properties may have a particular appeal for certain democratic decision-making situations, the existing mechanisms may have some disadvantages. Our results also show that two-stage (democratic) voting can be compatible with utilitarian welfare maximization in the provision of public goods.

Our second contribution is to investigate how to reduce the number of voting rounds for our mechanism. This is a relevant issue because the number of voting rounds that occur in equilibrium is equal to the number of citizens (or agents) plus one. Hence, this number can be very large if there are many citizens. The large number of voting rounds is a consequence of the requirements that *(i)* our mechanism must be detail-free and cannot depend on the prior distribution according to which individual types are determined, and that *(ii)* our mechanism must only consider rounds of binary voting, without an (explicit) proposal-making stage. One possibility is to envision our setup as describing a (small) committee, say a few citizens acting as elected members of Parliament representing different parties in a legislature. In such cases, semi-flexible majority rules only entail a few voting rounds. With a large number of citizens, there are nonetheless two ways—which we introduce in Section 4.2—to reduce the number of voting rounds. First, one can adapt the execution of the mechanism by skipping rounds according to some statistical information, in particular, (ex ante) correlation of preferences. Second, one can relax the utilitarian criterion and merely aim at approximating it with some margin of error.

Our third contribution is to explore ways to move beyond two-type societies. This is important since our main theorem is established for environments where individuals are of two types regarding their preferred public-good level.[5] This captures polarized societies and democracies where the power to make proposals is in the hands of two groups (say, two political parties). With two types, the proposed mechanism not only implements the utilitarian optimum, but also elicits the information about how many individuals there are of each type, even if the number of votes cast for each option is not made public for any voting round. With more than two types, we first illustrate with one example that no mechanism, and in particular no voting process, can implement the utilitarian optimal solution for the public-good provision game we consider. This means that utilitarian efficiency cannot be reconciled with strategy-proofness. Yet, one can envision a number of (detail-free) mechanisms that elicit the ex-post type distribution and that can *approximate* utilitarian welfare, of which we explore one instance.

[5]Our paper offers a new mechanism that could be tested in laboratory experiments, following a rich literature on the two-type case, which dates back at least to Palfrey and Rosenthal (1991).

The paper is organized as follows: In Section 2 we review the papers that are most connected to our contribution. In Section 3 we set out the model and introduce our voting mechanism with two citizen types. In Section 4 we present the main result of the paper. In Section 5 we discuss the case where there are three or more types of citizens. Section 6 concludes. The proofs are in the Appendix.

2 Relation to the Literature

Although voting mechanisms (or voting procedures) based on fixed majority thresholds, and on the majority rule in particular, present numerous advantages (Black, 1948; May, 1952; Maskin, 1995; Moulin, 2014b), they have drawbacks (Arrow, 1950; Gibbard, 1973; Satterthwaite, 1975; McKelvey, 1976b).[6] As far as the topic of our paper is concerned, fixed majority rules are not satisfactory because they cannot elicit the intensity of preferences. This generally makes the utilitarian optimum unattainable via such voting procedures (see Green and Laffont, 1977).

A way to reconcile (symmetric) voting and utilitarian welfare maximization in *one* stage is to use flexible majority rules. Gershkov et al. (2017) recently provided a mechanism design foundation of successive voting procedures and showed that every unanimous and anonymous dominant-strategy incentive-compatible mechanism is outcome-equivalent to a successive procedure with decreasing thresholds. Among such mechanisms, Gershkov et al. (2017) further singled out those that yield the (second-best) highest utilitarian welfare, focusing on setups where agents have single-crossing and single-peaked preferences. These include public-good provision setups. In a public-good provision framework with two types similar to ours, Gersbach (2017) showed that a sequential (or successive) procedure based on appropriately designed flexible majority rules implements the welfare-optimal level of the public good. Instead of voting on the final level of the public good immediately, a series of votes on small increments are taken starting from the status quo, and voting goes on until a higher threshold cannot be met.[7] To implement the utilitarian optimal, the above mechanisms must entail voting thresholds that require less than a majority of votes. This means that one-round voting based on flexible majority rules is incompatible with approval. By contrast, our mechanism stipulates thresholds for the final

[6]More recent papers that offer foundations for the majority rule are Aşan and Sanver (2002), Woeginger (2003), and Miroiu (2004). The Median Voter Theorem—an important result associated with the majority rule—is, in turn, the subject of an extensive body of literature (see e.g. Barberà et al., 1993; Sprumont, 1991; Ching, 1997; Chatterji and Sen, 2011).

[7]Incremental voting dates back to Bowen (1943). It is worth noting that sequential procedures with *fixed* majority rules are broadly used in European parliaments (Rasch, 2000).

voting round that require a majority of votes. This allows us to extend the appeal of one-stage voting with flexible-majority rules to two-stage voting while making it compatible with the property of approval (and with the property of inclusiveness).

More generally, we contribute to the literature that tries to find ways to make utilitarian welfare maximization compatible with voting. Some papers have put forward mechanisms that not only determine the public-good level but also the payment (and transfer) scheme to finance the public good. One example can be found in Ledyard and Palfrey (1994) (see also Ledyard and Palfrey, 1999).[8] Without discretionary monetary transfers, Kwiek (2017) recently showed in a Bayesian setting that a weighted majority rule—with type-dependent weights—can be utilitarian efficient if penalties can be used to create incentives for agents to reveal their type.[9] Since penalties are not paid in equilibrium, this mechanism is superior to voting with lump-sum participation fees. Wasteful monetary transfers also typically occur in Vickrey-Clarke-Groves mechanisms, which thus cannot implement the utilitarian social choice function.

Our paper also contributes to the literature on the implementation of social choice functions by providing a new mechanism that implements the utilitarian social choice function in the absence of monetary transfers when there are two types of individuals. However, since we consider a mechanism with additional restrictions based on democratic considerations—approval and inclusiveness—, we depart from the standard mechanism design literature—see Bierbrauer and Sahm (2010) or Bierbrauer and Hellwig (2016) for recent papers where such considerations are central. Yet, it is worth mentioning that the mechanism we suggest is incentive-compatible, anonymous, unanimous, and non-manipulable. One way of looking at our results is that they provide a justification for using much simpler direct revelation mechanisms as shortcuts of two-stage, multiple-round voting mechanisms fulfilling certain democratic requirements.

Two features of our analysis connect our contribution to other strands of the literature. First, in a context of local public goods with deadweight costs of redistribution, Gersbach et al. (2019) showed that it is beneficial for society to give the initiative in making proposals to the minority, with the majority having the opportunity to counter-propose and to vote together with the minority on the two proposals. Our main finding complements this result. In contrast to making the first proposal, our setting enables a minority of the citizenry to set the status quo that will be used in the final voting stage.

[8] Other procedures aim to resolve the tension between mean and median voter (Rosar, 2015) and to discern the intensity of preferences (Casella, 2005; Fahrenberger and Gersbach, 2010; Hortalà-Vallvé, 2010).

[9] With two alternatives and independent private values, Azrieli and Kim (2014) recently showed that the utilitarian criterion, subject to incentive compatibility constraints, leads to weighted majority rules.

Second, when citizens are of two types, the problem which decision to adopt can be seen as a bargaining problem between two sets of agents, each corresponding to all citizens of the same type. There is an extensive body of literature on dynamic bargaining models where the outcome of one round is taken as a disagreement point for the next round (see e.g. Fershtman, 1990; John and Raith, 2001; Diskin et al., 2011; Grech and Tejada, 2018). Our paper adds to this literature by studying a mechanism that fulfills the approval and inclusiveness requirements, where both sets of agents jointly determine the status quo (or disagreement point) for the second stage.

3 Model

3.1 Setup

We consider a society with n individuals who decide about the level of a public good. We let $n > 2$ and we assume that n is odd for ease of presentation.[10] Individuals are indexed by i, with $i \in \{1, \ldots, n\}$. Public-good levels are denoted by x or y, with $x, y \in [0, \infty)$. The marginal cost of any unit of investment in the public good is $c > 0$. Costs are distributed equally among individuals. There are two types of individuals, drawn from the type space $\mathcal{T} = \{t^L, t^H\}$, where t^L and t^H denote low and high type, respectively ($0 < t^L < t^H$). The type of individual i is denoted by t_i, with $t_i \in \mathcal{T}$. If an investment x is made, individual i derives utility from the public good equal to

$$v(x, t_i) = t_i \cdot f(x) - \frac{c}{n} \cdot x, \tag{1}$$

with $f(\cdot)$ being a real-valued function on $(0, +\infty)$ that is twice differentiable and satisfies $f(0) = 0$, $f'(x) > 0$, $f''(x) < 0$ for $x > 0$, $\lim_{x \to 0^+} f'(x) = +\infty$, and $\lim_{x \to \infty} f'(x) = 0$. An example is $f(x) = \sqrt{x}$. The first term, $t_i \cdot f(x)$, is the benefit from the public-good level x, while the second term, $\frac{c}{n} \cdot x$, is the per capita cost. Hence, from the implementation of any level of the public good individuals of type t^H benefit more than individuals of type t^L. An immediate consequence of Equation (1) is that the preference of individual i is single-peaked in x, with peak $x_i := x(t_i) > 0$ defined by

$$f'(x(t_i)) = \frac{c}{n \cdot t_i}. \tag{2}$$

[10]The mechanism we introduce can be defined no matter whether n is even or odd. Unlike the case where n is odd, when n is even we need to specify what the outcome of any binary voting is when the threshold of some voting round is exactly $n/2$ and there are $n/2$ individuals voting in favor of one option and $n/2$ in favor of the other. Different rules for such decisions may yield different votes along the equilibrium path, but they all implement the utilitarian optimal solution.

Equation (1) defines a (strict) preference relation for citizen i, denoted by $\succ_i$, over any finite set of public-good levels. This preference relation is single-peaked given the natural order according to which higher public-good levels are labeled with higher indices. Although equilibrium behavior for our mechanism is pinned down by $\{\succ_i\}_{i=1}^n$, utilities $\{v(\cdot,\cdot)_{i=1}^n\}$ are needed for the computation of utilitarian welfare.

Citizen types are privately drawn from a joint (prior) distribution that is common knowledge, with the property that the probability measure that assigns to every number $k \in \{0,\ldots,n\}$ the probability that there are k individuals of the low type has full support. We do not specify the joint distribution as the properties of the mechanism do not depend on it.[11]

Finally, we determine the level of investment that maximizes utilitarian welfare. This investment level is denoted by x^{soc} and can be computed from Equation (1) as follows:

$$f'(x^{soc}) = \frac{c}{\sum_{i=1}^n t_i}. \tag{3}$$

It is convenient to introduce the notation $t^{soc} = \frac{1}{n}\sum_{i=1}^n t_i$. The value t^{soc} can be interpreted as the socially optimal (virtual) type.

3.2 Voting mechanisms: Two crucial properties

To decide which public-good level should be implemented, a variety of procedures (or mechanisms) can be envisioned. One possibility is to focus on *(iterated) voting mechanisms*, which are two-stage mechanisms with multiple voting rounds in each stage, where the outcome of the first stage is taken as default (or as *status quo*) for the second stage. These mechanisms have the following two characteristics. First, in each round, there is a binary voting between moving to the next round within the same stage and either proceeding to the next stage—if there is a further stage—or stopping altogether—if there is no further stage. In the first stage, if the mechanism moves to the next round, the default for the second stage is superseded by a new status quo. In the second stage, a given round-specific outcome is implemented if the mechanism does *not* move to the next round and hence stops. Second, in each voting round, the decision which option to implement depends on whether or not the ratio of individuals voting for either option reaches a certain (round-specific) threshold. If no second-stage threshold is attained, the default

[11] The validity of Theorem 1 does not hinge on the assumption that the prior type distribution has full support, but proceeding with this assumption facilitates the analysis. Moreover, the voting mechanism we consider in Section 3.3 consists of several rounds, some of which may not be needed if the prior type distribution does not have full support. We discuss this issue further in Section 4.

alternative chosen in the first stage is implemented.[12] If no first-stage threshold is attained, the default alternative for the second round is the lowest public-good level among those that can be utilitarian optimal. The mechanisms satisfying the above properties are the focus of this paper. As discussed in the Introduction, two-stage mechanisms where the output of the first stage is considered for further voting in a second stage are common in democracies—more on this below.

Beyond the standard requirements such as efficiency, incentive compatibility, unanimity, and anonymity (of players) that are usually imposed on arbitrary mechanisms, we consider two crucial properties for voting mechanisms. In contrast with the former properties, these are conditions that have to occur along the equilibrium path—and thus are procedural properties that have no direct connection to outcomes. To the best of our knowledge, none of the two properties—which are further justified below—has received scholarly attention in the mechanism design approach to public good provision until now. We formulate them for any arbitrary voting mechanism.

Majority approval: In any round of the *second stage*, for the mechanism to stop with the corresponding round-specific alternative, the number of votes in favor of this option has to be greater than the number of votes against it.[13]

This property, which we call *approval* in short, is best understood when it does *not* hold.[14] Indeed, consider a voting mechanism where approval does not hold and assume that in the last second-stage round played in equilibrium, the option to implement some alternative A has received fewer votes than the option to proceed to the next round. Yet, the mechanism yields A as the final outcome. This can be the case, for instance, in the mechanisms considered by Bowen (1943), Gershkov et al. (2017), and Gersbach (2017). Clearly, an interplay between a minority—who could approve, say, to invest 1 million dollars in road building—and a majority—who could then block further investments—is useful for implementation of the utilitarian optimum. However, giving too much decision power to a minority to implement a policy directly may have a number of disadvantages. First, if the mechanism is repeated, the decision by the minority

[12] We use the term *voting* as coined by Bierbrauer and Hellwig (2016). Our notion of (iterated) voting mechanism can be easily generalized when there are three or more stages and the outcome of each stage is taken as a default for the next stage.

[13] Under the assumption that no abstention occurs, absolute majority is equivalent to relative majority. If some citizens abstain, one would need to differentiate between the two possibilities. In our setup, this distinction is immaterial for the results.

[14] To avoid any confusion, the property we call approval does *not* refer to single-winner electoral systems where each voter may select (or approve) any number of alternatives (see e.g. Brams and Fishburn, 1978).

may be undone by another minority, thereby leading to cycling of policy outcomes. Second, majority support for a final decision is considered essential for democratic decision-making. An approval of a final decision only by a minority may be recognized as a tyranny of this minority and may cast doubts on the legitimacy of the entire mechanism as a democratic decision-making scheme.[15] The (lack of) democratic legitimacy of decisions taken by voting mechanisms has long been the preoccupation of philosophers and political scientists.[16] The approval property is an attempt at capturing some of these concerns formally.

Approval requires a majority of votes for any second-stage round, but not for first-stage voting rounds. Differences across rounds with regard to the minimal support required for making (and voting on) a proposal is common in the case of political action initiated by signature gathering. This can occur in parliaments in representative democracies as discussed in the Introduction or in direct democracies. In Switzerland, for example, 100,000 valid signatures—accounting for less than 3% of the electorate—are needed to propose changes to the Federal Constitution. The decision whether to approve a change is then taken via a popular vote for which the parliament can propose a counter-project instead of the status quo. A double majority (of votes and of cantons) is then required to approve the change.[17] A similar procedure is used in California, where the number of signatures needed to put a proposal to vote even varies depending on the proposal.[18] In Spain, a representative democracy, by collecting 500,000 signatures—accounting for a bit more than 1% of total population—a group of citizens can initiate a vote on certain issues, as can the government, and the regional parliaments.[19] Approval builds on the same premise underlying the above examples: first-stage thresholds are not constrained by the same requirements as the thresholds used in the second, final voting stage. This enables political action to be initiated by a small group.

For the second property we introduce, we let $\mathcal{Y}$ denote a given subset of the set of alternatives.

[15] In the suggested mechanism a minority may be able to trigger a next voting round, but the final decision is only possible if the selected alternative receives a majority of votes.

[16] Our focus is on public good provision, which should not affect any fundamental right. The representativeness of democracies is thoroughly discussed in Bishin (2009)—see also the references therein. Though different in substance, the concerns about the use of thresholds below the majority rule resemble the worries that tend to arise in direct democracies when turnout is very low and hence the share of citizens backing implementation of a certain policy is low as well. Similar concerns arise also in representative democracies. For a discussion about the consequences of low turnout, see Lutz and Marsh (2007) and their references.

[17] See `https://www.bk.admin.ch/bk/de/home/politische-rechte/volksinitiativen.html`, retrieved on 3 December 2019.

[18] See `https://ballotpedia.org/Signature_requirements_for_ballot_measures_in_California`, retrieved on 3 December 2019.

[19] See `http://www.juntaelectoralcentral.es/cs/jec/ley?ambito=1&annoLey=1984&numeroLey=3&p=1379061545835&tipoLey=4` and `https://es.wikipedia.org/wiki/Iniciativa_legislativa`, retrieved on 11 December 2019.

Inclusiveness (w.r.t. $\mathcal{Y}$): In equilibrium, all alternatives from set $\mathcal{Y}$, called *core alternatives*, must be part of at least one voting round.

Core alternatives are deemed relevant for the society. One can conceive of set $\mathcal{Y}$ as being made up of all the alternatives that either gather a significant support among the citizenry—say, in a direct democracy—or represent the position of political parties—say, in a representative democracy. In our (baseline) setup, core alternatives are the public-good levels that are utilitarian optimal for some realizations of citizens' types. Other possibilities are discussed in Section 4.2.

There are three rationales for requiring inclusiveness. First, mechanisms that satisfy this property guarantee that all potentially optimal alternatives are considered (for voting) and, in particular, debated ahead of voting. In real-world settings, citizens may need to incur effort to tell apart different alternatives regarding public good provision, so the possibility to debate about them could be socially desirable. Inclusiveness could then be viewed as a procedural fairness requirement, as it ensures that the public discourse does not exclude potentially optimal alternatives from a social perspective, on which citizens can express their preferences—and even learn them. Second, inclusiveness avoids that mistakes in one voting round have grave consequences. Indeed, with such a property, a voting error of a single individual by voting as if s/he were of a different type leads to a second-best alternative, i.e., the alternative that is chosen yields the second highest level of utilitarian welfare. This is not guaranteed in general for voting mechanisms that do not fulfill inclusiveness. Third, inclusiveness enables the possibility that the worst outcome for an individual—e.g. a public-good level x^* characterized by $f'(x^*) = \frac{c}{n \cdot t^L}$ for an individual of type t^H—will not be selected by the society independently of how other individuals vote. In the voting process we propose, in particular, public-good level x^* can be deterred by a citizen of the high type just through his/her votes. The feature that each individual is guaranteed a minimal payoff independently of what other citizens do has been a prominent theme of a large literature on fair division initiated formally by Steihaus (1948), surveyed by Thomson (2011), Fleurbaey and Maniquet (2011) and Moulin (2004, 2014a, 2019), among others, and further extended recently by Bogomolnaia et al. (2019). With inclusiveness, we therefore introduce a fairness requirement in the voting process.

Core alternatives can be found in real-world political systems. For instance, in the debate about Catalan independence, politicians favoring independence have long argued that they would never accept a voting procedure where the possibility that Catalonia becomes an independent country is not considered. In particular, they haven often refused the possibility of a referendum in which more autonomy from the Spanish central government would be pitted against the status quo, since this would mean that a vote on independence would be off the table.[20,21] Similar concerns have been raised by those who oppose independence but demand more autonomy within Spain. In other words, all wanted their preferred alternative to be effectively taken into consideration for voting. These are the sort of concerns in which inclusiveness is rooted.

Both approval and inclusiveness can be viewed as *procedural* properties rather than *substantive* properties of democracy.[22] That is, if democracy—or mechanism design—is all about outcomes as demanded by *endstate justice*, neither property is required. According to this view, one should select a (voting) mechanism implementing some welfare function only based on how much it is expected to deliver in a given real-world environment. By contrast, if the legitimacy of democracy as a whole—or of one mechanism in particular—also depends on issues such as how much each citizen adheres to its principles, one might want to require properties such as approval or inclusiveness. These are the type of concerns with which *procedural justice* is concerned. We follow Moulin (2008) and adopt the commonplace that "means matter as well as ends", in which case endstate justice and procedural justice complement each other.

3.3 A voting mechanism based on semi-flexible majority rules

Next, we introduce a particular (two-stage) voting mechanism, which we call *Semi-Flexible Majority Voting Mechanism (SFM)*. As mentioned earlier, the object of the first stage is to set a status quo $\bar{x}$ for the second stage, where the final outcome is chosen. For all $j \in \{0, \ldots, n\}$, let y^j be the investment level defined by

$$f'(y^j) = \frac{c}{(n-j) \cdot t^L + j \cdot t^H},$$

which yields a unique solution for y^j. That is, y^j is the preferred level of investment for a society with $n-j$ individuals of type t^L and j individuals of type t^H, as well as for a society consisting of n imaginary citizens of identical type $\frac{n-j}{n} \cdot t^L + \frac{j}{n} \cdot t^H$. Since $t^L < t^H$ and $f'(\cdot)$ is decreasing, $y^i < y^j$ if and only if $i < j$. The voting mechanism we consider chooses one *alternative* from the following set:

$$\mathcal{Y} := \{y^0, \ldots, y^n\}.$$

[20] Note that this possibility would allow voters to incorporate their preferences about independence in such referendum by backward induction.

[21] See https://www.ara.cat/politica/Torra-acceptara-referendum-autogovern-Sanchez_0_2091391016.html (in Catalan, retrieved on 20 September 2018). The extent to which these claims are credible is beyond the scope of this paper.

[22] An ample literature deals with the epistemology of democracy (see e.g. Saffon and Urbinati, 2013).

Set $\mathcal{Y}$ consists of all the public-good investment levels that are utilitarian optimal for different combinations of individual types. We note that $y^i > 0$ for all $i \in \{0, \ldots, n\}$. Sometimes we refer to the elements of $\mathcal{Y}$ as *core alternatives*—see Section 3.2 for the rationale behind this set.[23] The (maximum) number of rounds in each of the two stages directly depends on the cardinality of $\mathcal{Y}$ and hence on n. Citizens cannot abstain in any voting round.[24]

We are now in a position to introduce our voting mechanism, SFM. The sequence of events is as follows:

Stage 1

Round 1.1: A vote is held between setting $\bar{x} = y^0$ (and jumping to Stage 2 with such a status-quo policy) and moving to the next round in Stage 1. At least one vote is required to move to Round 1.2.

$\vdots$

Round 1.$(n-1)$: A vote is held between setting $\bar{x} = y^{n-2}$ (and jumping to Stage 2 with such a status-quo policy) and moving to the next round in Stage 1. At least $n-1$ votes are required to move to Round 1.n.

Round 1.n: A vote is held between setting $\bar{x} = y^{n-1}$ and setting $\bar{x} = y^n$. Unanimity is required to set $\bar{x} = y^n$ and move to Stage 2 with such a status-quo policy. Otherwise Stage 2 starts with $\bar{x} = y^{n-1}$.

Stage 2 Let $\bar{x} = y^k$, with $k \in \{0, \ldots, n\}$, be the outcome of Stage 1.

Round 2.1: A vote is held between moving to the next round and choosing y^n as the final outcome. Unanimity is required to choose y^n, in which case the mechanism ends.

$\vdots$

Round 2.$(n-r+1)$ (with $r \geq k$): A vote is held between moving to the next round and choosing y^r as the final outcome. At least $\max\left\{r, \frac{n+1}{2}\right\}$ votes are required to

[23] As already mentioned, in Section 4.2 we discuss different possibilities for this set.

[24] If voting is costless for citizens in any round, abstaining (in any round) can be ruled out by sequential elimination of weakly dominated strategies. We thus assume for simplicity that all citizens vote. If voting were costly and there were no coercion to vote, by contrast, the picture could in principle change substantially, as citizens would like to vote only in those rounds where they are pivotal—if at all. If the cost of voting in each round were small relative to the benefits from public good provision divided by the total number of rounds, however, our results would hold. This can be verified following the logic of the proof of Theorem 1.

choose y^r, in which case the mechanism ends. Otherwise the mechanism moves to Round 2.$(n-r+2)$.[25]

Round 2.$(n-r+1)$ (with $r<k$): A vote is held between moving to the next round and choosing y^r as the final outcome. Unanimity is required to choose y^r, in which case the mechanism ends. Otherwise the mechanism moves to Round 2.$(n-r+2)$.

$\vdots$

Round 2.$(n+2)$: If this round is reached, $\bar{x}$ is chosen as the final outcome.

One can easily verify that all the public-good levels in $\mathcal{Y}$ are the outcome of SFM for certain voting behaviors. Several further remarks are in order. First, we assume that in every voting round all citizens vote simultaneously, and that the precise outcome of the vote is not made public. This means that what citizens see is whether they have moved to the next voting round within the same voting stage, whether they have moved to the next voting stage, or whether they have reached the end of the mechanism.[26] Hence the two-stop voting process that we consider guarantees privacy even if there is only a small group of agents. Concerns about privacy in the context of voting, or, more generally, in the context of social choice, can be connected to concerns about freedom of choice (see e.g. Brandt and Sandholm, 2005; Chevaleyre et al., 2007).

Second, in Stage 1, increasing thresholds have to be met to set higher levels of the public good as the status quo for Stage 2. This conveys the idea that higher levels of the public good as the status quo for subsequent voting necessarily require stronger support from the citizens. This property is illustrated by Figure 1.

Third, in every voting round of Stage 2, a particular qualified majority of votes has to be reached to adopt any alternative from set $\mathcal{Y}$ as the final decision. These thresholds depend on the outcome of Stage 1. More specifically, the required threshold is minimal in the case of the status quo set out in Stage 1 (viz., $\bar{x} = y^k$ for some $k \in \{0, \ldots, n\}$), and becomes higher as the level of the public good deviates from $\bar{x}$. For levels that are lower than the status quo, in particular, unanimity is required. The specific evolution of the voting thresholds along the different rounds is shown in Figure 2.

[25]If we set $h = n+1-r$, the majority threshold considered in Round h, with $h \in \{1, \ldots, n+1-k\}$, is $f^k(h) := \max\{n+1-h, (n+1)/2\}$. It turns out that for Theorem 1 to hold—see Section 4—, it suffices to consider that $\{f^k(\cdot)\}_{k=0}^n$ is a collection of non-increasing, onto functions $f^k : \{1, \ldots, n+1-k\} \to \{\max\{(n+1)/2, k\}, \ldots, n\}$ such that $f^{k+1}(\cdot) \leq f^k(\cdot)$. This is discussed in the Appendix.

[26]Our results hold for other, more detailed disclosure policies. For instance, after each round each individual could obtain a private, noisy about how many votes were cast for each option. Our analysis would also extend to the case where beyond their private type, individuals obtain a private, noisy signal about the distributions of types in the society.

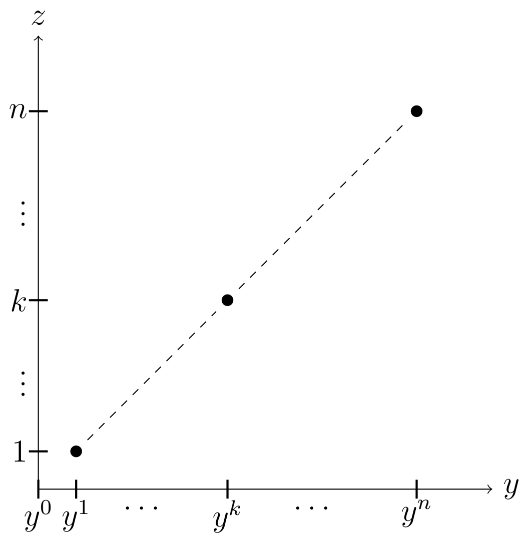
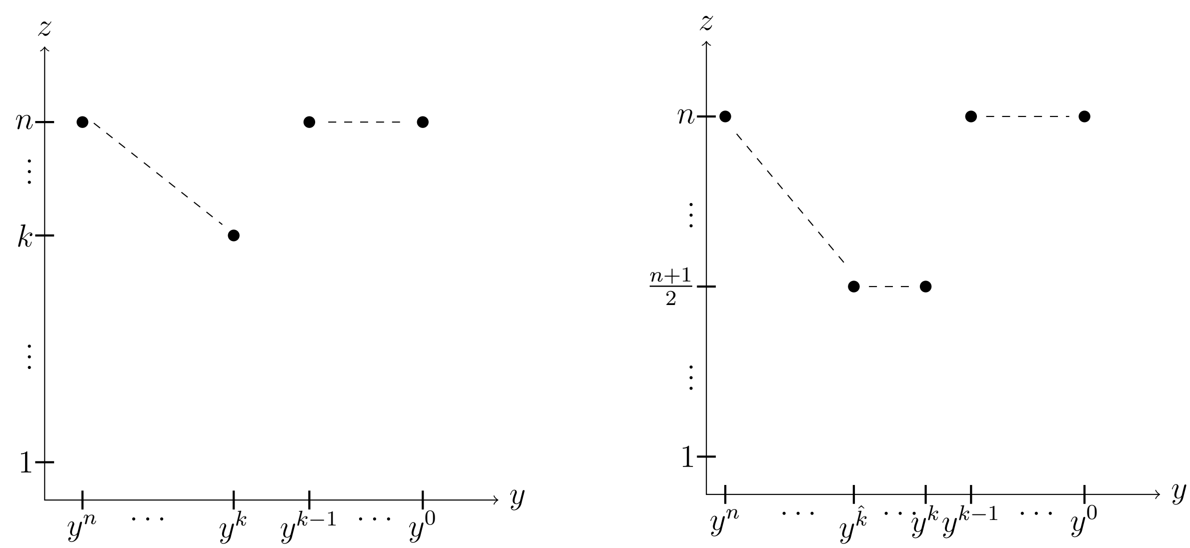


Figure 1: Stage 1—Number of votes (z) required to move from one project level to the next one. The mechanism starts with y^0 as default alternative.



(a) Case $k \geq \frac{n+1}{2}$.

(b) Case $k < \frac{n+1}{2}$ (with $\hat{k} = \frac{n+1}{2}$).

Figure 2: Stage 2—Number of votes (z) required to approve a policy, with $\bar{x} = y^k$ the status quo set out in Stage 1. The mechanism starts with y^n.

Fourth and last, we stress that the status quo chosen in Stage 1 is adopted by default only if no majority threshold is met in all the voting rounds of Stage 2.

4 Equilibrium Results

In this section we present the main result of the paper, which is subsequently discussed in detail.

4.1 Main result

For our analysis, we consider perfect Bayesian equilibria in pure strategies and further assume that citizens eliminate weakly dominated strategies iteratively by moving backwards in the voting mechanism, and that this is common knowledge.[27] We then obtain the following result:

Theorem 1

The outcome of the Semi-Flexible Majority Voting Mechanism (SFM) is x^{soc}.

Proof: See Appendix.

□

According to Theorem 1, SFM implements the utilitarian social choice function (ex post). Although uniqueness of beliefs is not guaranteed, all the equilibria yield the same outcome and thus the outcome is unique.

To see the intuition behind the theorem, take first Stage 2. For high-type citizens, voting for the proposal on the table weakly dominates voting to proceed to the next round in all rounds considering public-good levels that are higher than, or equal to, the status quo level $\bar{x}$ set in Stage 1. In turn, for all rounds considering a lower public-good level than the status quo, it is optimal for any citizen of the high type to vote to proceed to the next voting round. This is because unanimity is required for these rounds, and hence each high-type citizen knows that s/he can ensure that the status quo is eventually chosen if s/he always vote against stopping the mechanism and implementing a lower public-good level. As for low-type citizens, voting for the proposal on the table is weakly dominated by voting to proceed to the next round in *all* rounds considering public-good provision levels that are higher or equal than $\bar{x}$, and it is optimal for the remaining rounds. The most subtle decisions occur in the rounds considering a lower public-good level than $\bar{x}$, in which unanimity is required to stop the mechanism: If a low-type citizen votes in favor of the alternative on the table, her/his vote might help such a policy to be chosen. If, on the contrary, s/he votes for the procedure to continue to the next round, the status quo determined in Stage 1 may be eventually chosen if all remaining alternatives are subsequently rejected. Nevertheless, if it is common knowledge that all citizens play no weakly dominated strategies in subsequent rounds and that high-type citizens hold right beliefs about whether all agents are of the low type, each low-type citizen knows that her/his vote in these rounds is only

[27]Iterative elimination means that, starting from the last voting round and moving backwards across voting rounds, all individuals eliminate iteratively all weakly dominated strategies for *each* voting round.

pivotal when no citizen is of the high type. In such a case, all low-type citizens ensure that the lowest possible public-good level is eventually chosen by always voting to proceed to the next round until the last voting round of Stage 2 is reached.

Consider now Stage 1. If citizens behave in Stage 2 as described above, it turns out that the final outcome cannot be worse (better) for high-type (low-type) citizens as larger values of the status quo chosen in Stage 1 are considered, whatever the type composition of the electorate and the exact beliefs the citizens hold about it. This implies that voting to proceed to the next round in Stage 1 yields higher or equal utility to high-type individuals than choosing to stick with the current status quo policy. For low-type individuals, the optimal voting decisions in Stage 1 are reversed.

We stress that in SFM, the first stage proceeds from low indices to higher indices of the proposal on the table, while the second stage reverses the order. Furthermore, there are as many public-good levels in $\mathcal{Y}$ as there are different voting thresholds. This allows a one-to-one relationship between a given distribution of types and the appropriate majority threshold, which in turn ensures implementation of the utilitarian optimal solution. It also allows the mechanism to function without prior information about how individual types are drawn. In fact, with two types, SFM elicits the information about how many citizens there are of each type.[28] For this result to hold, it is not necessary that the number of votes cast for each alternative in any voting round is made public. It suffices to know the alternative set as status quo for Stage 2 by the outcome of Stage 1.

The next complementary result, which follows directly from the proof of the Theorem 1, identifies further properties of SFM.

Corollary 1

The Semi-Flexible Majority Voting Mechanism (SFM) satisfies majority approval and inclusiveness with respect to $\mathcal{Y}$.

That is, SFM not only implements the utilitarian optimal solution and elicits the ex-post type distribution but it exhibits two properties—approval and inclusiveness—that other known (one-stage) mechanisms based on flexible-majority rules fail to fulfill. The appeal of such properties—and hence of SFM—has been discussed in Section 3.2. Approval can be imposed on a voting

[28] This property is independent of the actual values of t^L and t^H. This means that for Theorem 1 to hold it suffices for low-type citizens to know they are of the low type and for high-type citizens to know they are of the high type. But they need to know neither t^L nor t^H, i.e., they do not need to know the value of their actual type. Of course, the designer needs to know these values to determine the set of admissible alternatives.

mechanism as we do, while inclusiveness depends on equilibrium behavior, and hence on citizens' rationality level. In our public-good provision setup, inclusiveness follows from the order in which the alternatives from set $\mathcal{Y}$ are considered for binary voting: an ascending order regarding the level of public good provision in the first stage, and a descending order in the second stage. The combination of the two opposing orderings strikes a balance between the power of the minority to set the default alternative in the first stage and the power of the majority to choose the final policy in the second stage. This way, not only utilitarian efficiency but also inclusiveness is guaranteed.

Specifically, we show in the proof of Theorem 1 that due to off-equilibrium threats, the alternative that maximizes utilitarian welfare, say $y^k \in \mathcal{Y}$, is already chosen in the first stage of SFM as a default alternative for the second stage, possibly with the explicit approval of a minority only—when a minority of citizens are of the low type. This property of our two-stage mechanism can then be used as a theoretical justification for employing (much simpler and shorter) revelation mechanisms in democracies. Back to SFM, alternatives $y^1, \dots, y^k \in \mathcal{Y}$ are considered for a round of binary voting in the first stage.[29] Then, the second stage serves the purpose of ratifying the choice of the first stage, but with an effective threshold of at least half plus one of the citizens. When only a minority of citizens are of the low type (i.e., when $k \geq (n+1)/2$), alternatives $y^n, \dots, y^k \in \mathcal{Y}$ are considered for a round of binary voting in Stage 2. When a majority of citizens are of the low type (i.e., when $k < (n+1)/2$), at least alternatives $y^n, \dots, y^k \in \mathcal{Y}$ are considered for a round of binary voting in Stage 2. Figures 1 and 2 are useful for illustrating equilibrium behavior.

For our analysis we have assumed that the same individuals participate in both rounds of the two-stage mechanism. However, one can also ask what the outcome would be if two different groups (with different aggregate preferences) participated in each stage of the mechanism. As mentioned in the Introduction, this is the case in direct democracies, as well as in representative democracies. It turns out that, as we show in the proof of Theorem 1, equilibrium behavior with two different groups can be described as in the case where the two groups are the same. That is, the first-stage group would choose the alternative maximizing the first group's utilitarian welfare as status quo for the second stage. In the second stage, however, the second group might end up choosing a different public-good level. Depending on the aggregate preferences of the two groups, this level could maximize the first group's utilitarian welfare, the second group's

[29] We stress that y^0 is the default alternative for the first stage. One could always add a dummy stage where y^0 is considered for binary voting and the threshold needed for approval is zero.

utilitarian welfare or neither. In particular, it is not true that our mechanism systematically favors one group over the other. Regardless of the outcome, our mechanism would elicit the aggregate preferences of *both* groups, since all individuals would vote sincerely according to such preferences.

Finally, one can infer from the proof of Theorem 1 that inclusiveness guarantees that individual mistakes have no grave consequences: If a high-type (low-type) citizen votes according to the opposite type, a second-best alternative is selected. It also follows from the proof of Theorem 1 that each individual can ensure through their vote that the worst possible outcome for him/her is never implemented, no matter what the other citizens do.

4.2 Further discussions

Several further discussions of Theorem 1 help to fully grasp its relevance and implications.

Less demanding majority thresholds

SFM requires qualified majorities in Stage 2 for changes to the outcome of Stage 1, and unanimity, in particular, for public levels that are lower than the one specified in the status quo chosen therein. For a range of rounds, however, a whole family of qualified majority thresholds ensures utilitarian optimality, which includes the thresholds set out in Section 3. This is shown in the Appendix—see also Footnote 25. More specifically, given that y^k is the outcome of Stage 1, it suffices for the majority thresholds of Round 2.1 to Round 2.$(n+1-k)$ to be non-increasing, ranging from unanimity to a certain qualified majority (never lower than half plus one of the votes). This guarantees that high-type citizens cannot impose a public-good level that is higher than the socially optimal level. By contrast, the unanimity rule required in any voting round after Round 2.$(n+1-k)$ grants any individual the veto power to impose the status quo as the final outcome, which is essential for optimality in the general case.

The voting thresholds of Stage 2 can be lowered further when the possible number of high- or low-type individuals is bounded, and in particular when individual types are not drawn independently from each other. For instance, suppose that the number of high-type individuals has support $\{0, \underline{n}, \underline{n}+1, \ldots, n\}$ for some $\underline{n}$ with $1 < \underline{n} < n$. Then the unanimity thresholds that apply to public-good levels lower than the status quo can be lowered to $\max\{n - \underline{n} + 1, \frac{n+1}{2}\}$. The reason is that situations with only a few high-type individuals (below $\underline{n}$) cannot occur. In these circumstances, each high-type individual enjoys a *de facto* veto power enabling her/him to block the approval of any alternative that requires more than the support of $n - \underline{n}$ citizens of the low type. Knowing this, low-type citizens also vote to proceed to the next round, since their vote only makes a difference if there are no citizens of the high type at all.

Another example is a situation where the number of low-type individuals has support $\{\overline{n}-1, \overline{n}, \ldots, n\}$ for some $\overline{n}$ with $1 < \overline{n} < n$. In this case, the second-stage thresholds for the approval of public-good levels that are higher than the one prescribed by the outcome of Stage 1 can be lowered to $\max\{n-\overline{n}+1, \frac{n+1}{2}\}$. The reason is that there are never more than $n-\overline{n}+1$ individuals of the high type, and voting in favor of any alternative that proposes a public level y^r, with $r \geq n-\overline{n}$, is weakly dominated for low-type individuals by voting to proceed to the next round. When either $\underline{n} = 1$ or $\overline{n} = 1$, the same logic serves the purpose of further explaining the main mechanisms behind Theorem 1.

Fewer rounds

To achieve utilitarian optimality through SFM, a number of proposals equal to the number of voters plus one would have to be considered in general. This is because our mechanism involves binary voting rounds and works without any information about the distribution according to which individual types are determined, but is not always necessary if some statistical information can be used. On the one hand, if the joint distribution of types has rather small support—e.g. because the preferences of different individuals are highly correlated—, all those rounds can be skipped in which the alternative considered for voting cannot be a socially optimal alternative for some preference profile. On the other hand, an approximately socially optimal solution may suffice when the number of citizens is considerable. If n is large and the citizens' types are i.i.d., in particular, the Central Limit Theorem guarantees that the socially optimal type is distributed approximately normally, with a mean μ and a variance σ that could be estimated. Then one could add a criterion to our procedure to exclude the tails of the distribution and hence the most extreme policies. For instance, for a given $\alpha \geq 0$, the set

$$\mathcal{X} = \left\{ x : f'(x) = \frac{c}{n \cdot t}, t = \frac{n-j}{n} \cdot t^L + \frac{j}{n} \cdot t^H, \mu - \alpha \cdot \sigma \leq \frac{n-j}{n} \cdot t^L + \frac{j}{n} \cdot t^H \leq \mu + \alpha \cdot \sigma \right\}$$

could be considered instead of the entire set $\mathcal{Y}$ to run the mechanism. In such case, the maximal total number of rounds of the modified mechanism would be twice the number of alternatives contained in $\mathcal{X}$. The parameter α determines the (expected) loss of efficiency that such a mechanism would induce. The larger α, the lower the loss.

More generally, one could consider an arbitrary finite subset $\mathcal{Z} = \{z_1, z_2, \ldots\} \subseteq \mathcal{Y}$ consisting of a (small) number of provision levels of the public good, with $z_k < z_l$ if and only if $k < l$. One possibility is that $\mathcal{Z}$ consists of public-good levels proposed by political parties, by regions within the same country, or by a popular initiative (in direct democracies). As before, one could then easily adapt SFM to run only over the rounds corresponding to elements of $\mathcal{Z}$, with the corresponding voting thresholds. That is, alternatives $z_1, z_2, \ldots$ would be considered in ascending order in the first stage, and in descending order in the second stage. While such variation of SFM would not implement the utilitarian optimal solution in general, it can be verified that it would yield the element in $\mathcal{Z}$ that is closest to the socially optimal level. With significant costs of performing voting rounds, say because there is a delay in implementation, this modified mechanism would be attractive in real-world environments. This modified mechanism would satisfy approval and inclusiveness with respect to $\mathcal{Z}$.

Yet another possibility is to consider a set $\mathcal{Z}$ containing all public-good levels that maximize a given weighted utilitarian social function. Following the logic of the proof of Theorem 1, one can verify that the entire Pareto frontier could be traced out by running SFM for all such social choice functions over the corresponding set $\mathcal{Z}$. This means that variations of SFM can produce outcomes in which some types have a particular weight in the collective decision. For instance, only some types may matter when citizens form parties or factions, and a representative individual, say the median member of the coalition, casts the votes for this group.

Conditioning on types

By construction, SFM depends on the citizens' utility function, since the elements of $\mathcal{Y}$ are calculated with the benefit function f. However, SFM could be easily adapted to hinge on types rather than on policies if types are constant—though private—and the function f can take several forms. This would expand SFM's applicability to a wider range of problems.[30] This follows from the fact that a one-to-one correspondence generally exists between the set of potential socially optimal types and the set of feasible alternatives. Of course, for such a procedure to be implementable, it should be possible, i.e. legal, to base a democratic procedure for public-good provision on the aggregate type distribution, e.g. on the aggregate income distribution, rather than on the policies themselves.

Different default alternative

In our analysis of SFM, we have proceeded on the assumption that the default alternative is y^0, i.e., the lowest level of public good that *all* societies accept, which could be arbitrarily close to zero. This could correspond to one-shot decisions on the provision of a certain public good such as building a high-speed train network of a given size from scratch, as the status quo is

[30]Recall Footnote 28.

zero by definition. In other cases, however, the decisions on public goods are recurrent, and the outcome of one decision is the status quo for the next decision. This is typically the case for expenditures on education—say, on public schools—or on defense. SFM could also be used under such circumstances. On the one hand, it would still deliver the utilitarian solution while satisfying approval and inclusiveness. On the other hand, suppose that y^k was chosen in some period and that the same society wants to choose a public-good level in the subsequent period using SFM. Then one can interpret the first k voting rounds of Stage 1, in which a vote is held between setting $\bar{x} = y^l$ ($l \in \{0, \ldots, k-1\}$) as status quo for Stage 2, as a (preliminary) vote on whether Stage 1 of SFM should start from Round 1.(k+1)—in which y^k is voted upon—or it should start from Round 1.(l+1)—in which y^l is voted upon.[31] Finally, the fact that SFM implements the utilitarian optimal solution regardless of the policy chosen in the past ensures that no cycling in policy-making occurs, provided that aggregate preferences remain unaltered.

Alternative rationality assumptions

The equilibrium notion we have used in our analysis requires that it is common knowledge that all citizens iteratively eliminate weakly dominated strategies by moving backwards in the game. Yet, Theorem 1 could also be obtained if citizens used cut-off strategies in Stage 2, according to which an individual can only change her/his vote (*proceed* or *stop*) at most once along the different voting rounds of *a* stage. Iterative elimination of weakly dominated strategies by moving backwards in the game can thus be seen as a foundation for such cut-off (behavioral) rules, in the case of two-stage, multiple-round voting mechanisms.[32]

5 Multiple Types

In our analysis thus far, we have assumed that individuals are of two types, low and high. In this section we address the general case, in which individuals can be of any finite number of types. First, we prove an impossibility result for three or more types. Second, we explore *one* way to reconcile utilitarian welfare maximization with approval and inclusiveness to some extent when there are at least three types of individuals.

[31] This interpretation entails duplicating the round of Stage 1 of SFM that is voted upon to be the first.

[32] Monotone (or cut-off) strategies have been used by Gershkov et al. (2017). They have also been justified by an iterative process of elimination of weakly dominated strategies. The subtleties, however, are different here. One reason is that the last step of the procedure considered by Gershkov et al. (2017) consists of a vote between the last two exogenously-given alternatives, while our procedure consists of a vote between the status quo determined endogenously in Stage 1, namely $\bar{x} = y^k$, and the ex-ante status quo, y^0.

5.1 A counter-example

With two types, SFM implements the utilitarian optimal solution and elicits the information about how many individuals are of each type. This occurs even if individuals receive noisy (private and/or public) signals in the beginning of the game about the ex-ante type distribution. Two additional noteworthy properties hold when there are only two types of agents. First, there are as many different thresholds as there are different ratios of citizens of the two types. This makes it possible for the agents to express in SFM the distribution of types. Second, with two types, individuals have single-peaked preferences with peaks at one of the two extremes of the set of alternatives, viz. y^0 or y^n.

The picture changes dramatically when individuals are of three or more types. To see this, we consider the following example: $f(x) = \sqrt{x}$, $c = 1$, $n = 3$ and $\mathcal{T} = \{t^L, t^M, t^H\}$, with $t^L = 1$, $t^M = 2$, and $t^H = 4$. The individual optimal public-good levels for each type are

$$x(t^L) = \frac{9}{4}, \quad x(t^M) = 9, \quad \text{and} \quad x(t^H) = 36.$$

Using Equation (3), the socially optimal level is equal to

$$x^{soc} = \frac{(t_1 + t_2 + t_3)^2}{4},$$

where $t_i \in \mathcal{T}$ is the type of individual $i \in \{1, 2, 3\}$. Assume now that beyond their private type, individuals receive signals that allow them to know that types are drawn i.i.d. from a probability distribution such that for any individual i,

$$Prob[t_i = t^L] = 1 - 2\varepsilon \quad \text{and} \quad Prob[t_i = t^M] = Prob[t_i = t^H] = \varepsilon,$$

where ε is strictly positive but arbitrarily small. Then consider the *direct* mechanism in which every individual $i \in \{1, 2, 3\}$ reports $\tilde{t}_i \in \mathcal{T}$ and the following public-good level is implemented:

$$x^D = \frac{(\tilde{t}_1 + \tilde{t}_2 + \tilde{t}_3)^2}{4}. \tag{4}$$

This mechanism would be the *only* direct mechanism that implements the utilitarian optimum, provided that individuals report their types truthfully. Clearly, for individuals of types t^L and t^H, it is weakly dominant to report their type truthfully. What about individuals of type t^M? By reporting their type truthfully and anticipating that the other two individuals are of type t^L with very high probability, they expect a payoff v^{truth}, with

$$v^{truth} \approx v\left(\frac{(2t^L + t^M)^2}{4}, t^M\right) = 2 \cdot \frac{1+1+2}{2} - \frac{1}{3} \cdot \frac{(1+1+2)^2}{4} = \frac{8}{3}.$$

If individuals of type t^M report t^H instead, they expect a payoff $v^{no\ truth}$, with

$$v^{no\ truth} \approx v\left(\frac{(2t^L + t^H)^2}{4}, t^M\right) = 2 \cdot \frac{1+1+4}{2} - \frac{1}{3} \cdot \frac{(1+1+4)^2}{4} = 3 > v^{truth}.$$

Hence, the direct mechanism is not Bayesian incentive compatible, and by invoking the revelation principle there is no mechanism that is Bayesian incentive compatible and can implement the utilitarian optimal solution x^{soc} in general. This means, in turn, that no strategy-proof mechanism, no matter whether it is a voting mechanism or not, can implement x^{soc} using our equilibrium concept. In particular, neither suitable extensions of SFM, of the mechanisms in Gersbach (2017), nor of the mechanism in Gershkov et al. (2017) can achieve utilitarian efficiency through truth-telling.

5.2 A generalized voting mechanism

In this section, we show that with an arbitrary, finite number of types, we can still design a (democratic) mechanism based on the principles behind approval and inclusiveness that *approximates* implementation of the utilitarian optimal solution and elicits the information about how many citizens there are of each type.

Accordingly, suppose that there are $T \geq 2$ possible types and let the type of individual i be denoted by t_i, where $t_i \in \mathcal{T} = \{t^1, \ldots, t^T\}$ and $0 < t^1 < \ldots < t^T$. The utilitarian optimum outcome is still given by Equation (3). This means that any (voting) mechanism intending to implement such an outcome should involve the set of alternatives $\mathcal{Y}$, where now

$$\mathcal{Y} := \left\{ y : f'(y) = \frac{c}{t}, t = \sum_{l=1}^{T} n_l \cdot t^l, \sum_{l=1}^{T} n_l = n, n_l \in \mathbb{N} \text{ for all } l \in \{1, \ldots, T\} \right\}$$

and n_l is the number of individuals of type t_l. A first important observation is that, unlike for the two-type case, we cannot establish in general a one-to-one correspondence between elements of $\mathcal{Y}$ and voting thresholds in a binary vote, as the latter are fewer than the former when there are at least three types. This means that citizens cannot express in general how many of them there are of each type if they use *one* (two-stage) voting mechanism. It is also immediate to observe that implementing the utilitarian optimum through (arbitrarily many rounds of) binary voting may require a very large number of rounds if there are multiple types, even if the population of individuals is small.

In the following, we consider a mechanism that is based on SFM and takes the above observations into account. For this purpose, define the level of public good y_l^j for all $l \in \{1, \ldots, T-1\}$ and

$j \in \{0, \ldots, n\}$, by

$$f'(y_l^j) = \frac{c}{(n-j) \cdot t^l + j \cdot t^{l+1}},$$

as well as the set

$$\mathcal{Y}_l = \{y_l^0, \ldots, y_l^n\}. \tag{5}$$

We note that y_l^j is the optimal public-good level when $n - j$ individuals are of type t^l and j individuals are of type t^{l+1}. By construction,

$$\bigcup_{l=1}^{T-1} \mathcal{Y}_l \subsetneq \mathcal{Y}.$$

While sets $\mathcal{Y}_1, \ldots, \mathcal{Y}_{T-1}$ do no cover set $\mathcal{Y}$, they span it.[33] Then recall from Equation (1) that all individual types have single-peaked preferences with respect to the order that labels higher public-good levels with higher indices, with peak given by (2). One can easily verify that the optimal policy (i.e. the peak) for type t^1 is

$$y_1^0 \in \mathcal{Y}_1, \tag{6}$$

for type t^T is

$$y_{T-1}^n \in \mathcal{Y}_{T-1}, \tag{7}$$

while for type t^l, with $l \in \{2, \ldots, T-2\}$, the optimal policy is

$$y_{l-1}^n = y_l^0 \in \mathcal{Y}_{l-1} \cap \mathcal{Y}_l. \tag{8}$$

To construct a mechanism that generalizes SFM to account for multiple types, we introduce an initial communication stage, denoted by Stage 0, before SFM is applied. In this communication stage, citizens can express their preference regarding which element of $\mathcal{Z}$, and thus which set of alternatives, should be chosen for running SFM. To describe this communication stage, we introduce further notation.

Given citizen i's preference relation $\succ_i$ over the set of alternatives $\mathcal{Y}$, define the preference relation $\succ_i^T$ over the set (of sets)

$$\mathcal{Z} := \{\mathcal{Y}_1, \ldots, \mathcal{Y}_{T-1}\}$$

as follows (for $j, j' \in \{1, \ldots, T-1\}$):

$$\mathcal{Y}_j \succ_i^T \mathcal{Y}_{j'} \iff \neg(y' \succ_i y) \text{ for all } y \in \mathcal{Y}_j, y' \in \mathcal{Y}_{j'}, \text{ and } y \succ_i y' \text{ for some } y \in \mathcal{Y}_j, y \in \mathcal{Y}_{j'}.$$

[33] That is, every element of $\mathcal{Y}$ can be represented by a linear combination of elements in $\mathcal{Y}_1, \ldots, \mathcal{Y}_{T-1}$.

One can see that $\succ_i^T$ orders the elements of $\mathcal{Z}$ completely. Moreover, from (6)–(8) it follows that $\succ_i^T$ is single-peaked (given the order $(\mathcal{Y}_1, \ldots, \mathcal{Y}_{T-1})$) with peak $\mathcal{Y}_1$ if citizen i's type is t^1, peak $\mathcal{Y}_{T-1}$ if citizen i's type is t^T, and peak(s) in $\{\mathcal{Y}_{l-1}, \mathcal{Y}_l\}$ if citizen i's type is t^l ($l \in \{2, \ldots, T-2\}$).[34] Since we are assuming that citizens play no weakly dominated strategies, it is clear that to decide which two elements of $\mathcal{Z}$ should be chosen for SFM to be run, citizen i chooses according to $\succ_i^T$,

For each citizen i, we let $\tau_i \in \mathcal{Z}$ denote his/her peak (or type) according to $\succ_i^T$. Then consider

$$G : \mathcal{Z}^n \to \mathcal{Z}$$
$$(z_1, \ldots, z_n) \to G(z_1, \ldots, z_n).$$

to be any (direct) mechanism, possibly non-deterministic, that assigns reported types $(\tau_i)_{i=1}^n$ (i.e., a vector of elements of $\mathcal{Z}$) to an element of $\mathcal{Z}$. We assume that G is Pareto efficient, anonymous, and dominant strategy incentive compatible. By Moulin (1980), we know that such (deterministic) mechanisms exist, e.g. variants of the Condorcet procedure.

Taking G as given, we are now in a position to introduce a mechanism for the case of multiple types, which we call *G-generalized SFM*. The (non-deterministic) mechanism specifies the following course of events:

Step 0: Apply G to $\mathcal{Z}$. We let $p_l = p_l(\tau_1', \ldots, \tau_n')$ denote the probability that $\mathcal{Y}_l$, with $l \in \{1, \ldots, T\}$, is the outcome of G when citizens send messages $(\tau_i')_{i=1}^n$. The numbers $p_1, \ldots, p_{T-1} \geq 0$ satisfy $\sum_{j=1}^{T-1} p_j = 1$ and are not made public.

Step 1: Apply SFM to $\mathcal{Y}_1$, which yields as outcome an element $y_1^{i_1} \in \mathcal{Y}_1$. Then define $N^1 \in \{0, \ldots, n\}$ by $N^1 = n - i_1$.

$\vdots$

Step $T-1$: Apply SFM to $\mathcal{Y}_{T-1}$, which yields as outcome an element $y_{T-1}^{i_{T-1}} \in \mathcal{Y}_{T-1}$. Then define $N^{T-1} \in \{0, \ldots, n\}$ by $N^{T-1} = n - i_{T-1}$.

Step T: The final outcome is the allocation $y \in \mathcal{Y}$ chosen in accordance with the following rule:

$$y = \begin{cases} y_1^{n-N^1} & \text{with probability } p_1, \\ \ldots & \\ y_{T-1}^{n-N^{T-1}} & \text{with probability } p_{T-1}. \end{cases}$$

[34] For type t^l, with $l \in \{2, \ldots, T-2\}$, the peak depends on the values of $t^1, \ldots, t^T$ as well as on the beliefs held by the citizens.

In other words, the above mechanism operates as follows. The (final) outcome is chosen randomly from one of the (potential) outcomes of $T-1$ applications of SFM, each of which considers a different set of alternatives. The weights according to which the non-deterministic mechanism chooses one particular application of SFM are determined by the individuals themselves prior to running all SFM applications, although they are not disclosed.

With most of the formal analysis in the Appendix, some remarks about the G-generalized SFM are in order here.

Principles behind the G-generalized SFM

The G-generalized SFM revolves around the principles behind inclusiveness and approval. As for inclusiveness, note that all elements of $\cup_{l=1}^{T-1}\mathcal{Y}_l$ are voted upon in at least one round of the corresponding application of SFM. Recall that $\cup_{l=1}^{T-1}\mathcal{Y}_l$ spans set $\mathcal{Y}$, with the latter set containing all public-good levels that can be utilitarian optimal for at least one preference profile in the population. Because with multiple types this set can be prohibitively large, focusing on $\cup_{l=1}^{T-1}\mathcal{Y}_l$ might suffice to comply with inclusiveness. As for approval, the second-stage outcome of the selected SFM must have been ratified by a majority of individuals, as in the two-type case. This means that the concerns that approval tries to address are also dealt with by the G-generalized SFM in the case of multiple types. Finally, any fairness and efficiency property of G carries over to the G-generalized SFM, since mechanism G determines the set(s) where SFM is run over. For instance, if G chooses the median generalized type (see Moulin, 1980), the G-generalized SFM combines majoritarian concerns in a first step, and then utilitarian concerns in a second step.

Elicitation of types: Sampling in large elections

For every application of SFM, individuals report their preferences *sincerely*. This follows from (6)–(8) because (within each SFM) individuals iteratively eliminate strategies that are weakly dominated, and, in particular, they never play strategies that are weakly dominated. Take $\mathcal{Y}_1$, for example. Then individuals of type t^1 prefer the lowest possible public-good level (within $\mathcal{Y}_1$). For their part, individuals of type $t^2, \ldots, t^T$ prefer the highest possible public-good level (within $\mathcal{Y}_1$). Because the outcome of SFM running over $\mathcal{Y}_1$ does not depend on any other applications of SFM, every individual who is not of type t^1 votes as if s/he were of type t^2. Analogous comments hold for the application of SFM to sets $\mathcal{Y}_2, \ldots, \mathcal{Y}_{T-1}$. This means that if we let $N^T = n$ and solve

$$\begin{cases} n_1 & = N^1, \\ \dots & \\ n_1 + \ldots + n_{T-1} & = N^{T-1}, \\ n_1 + \ldots + n_{T-1} + n_T & = N^T, \end{cases} \tag{9}$$

then n_l is equal to the number of individuals of type t^l, with $l \in \{1, \ldots, T\}$. Hence, the G-generalized SFM elicits the information about how many individuals there are of each type.[35]

Outcome of the G-generalized SFM

In general, x^{soc} is not implemented by the G-generalized SFM (see the Appendix for a detailed account of the equilibria of the underlying game). Yet, the outcome of the G-generalized SMF is an element of $\cup_{l=1}^{T-1} \mathcal{Y}_l$ and is not very far away from the utilitarian optimum. Moreover, the G-generalized SFM satisfies a number of other noteworthy properties. First, if all agents are of the same type and they receive sufficiently informative, yet noisy, public signals about the ex-ante type distribution, their preferred public-good level is always chosen.

Second, let us consider the case where agents are only of two types. We also assume that citizens receive sufficiently informative, yet noisy, public signals about the type distribution. This can be a good representation of societies that are polarized in some issue, with the power to make proposals being in the hands of two organized groups. Let t^L, t^R ($L, R \in \{1, \ldots, T-1\}$) be the two types that are present in the population, with $L \leq R - 1$. For simplicity, we focus on the case where G selects the median (with probability one).

On the one hand, consider the case where $L = R - 1$. Then, the utilitarian optimal solution belongs to $\mathcal{Y}_L$. This outcome would be chosen by the G-generalized SFM if there are more individuals of type t^L than of type t^R. If there are more individuals of type t^R than of type t^L, then $y_R^0 \in \mathcal{Y}_R$ would be chosen, which is itself close to the utilitarian optimal. On the other hand, consider the general case, in which $L \leq R - 1$. Then the utilitarian optimal solution belongs to $\cup_{l=L}^{R-1} \mathcal{Y}_l$. For its part, the outcome of the G-generalized SFM would be $y_L^j \in \mathcal{Y}_L$ ($y_R^n \in \mathcal{Y}_{R-1}$) if $j > n - j$ ($j < n - j$), where j denotes the number of individuals of type t^L.[36]

Third, note that Equation (1) implies that all citizens have concave utility in the public-good level, and hence that they are risk-averse. This means that if the probability that either type is the majority is close to 1/2, it would be in the best interest of both groups of individuals to

[35] One could also envision that this information is used by a social planner. It would suffice that citizens cannot foresee how this information is going to be used, so that they cannot inform their decisions in the G-generalized SFM in order to manipulate the information collected by the social planner in their favor.

[36] Recall that n is an odd integer, so $j = n - j$ cannot occur.

run SFM with only two types—t^L and t^R—instead of running the G-generalized SFM over the entire set $\mathcal{Z}$, which corresponds to types $t^1, \ldots, t^T$. The former mechanism would implement the utilitarian optimal solution.

Fourth and last, consider the case of large populations. Then a social planner could randomly sample a (representative) subpopulation of individuals, who would then participate in the generalized voting mechanism. The social planner would observe the actions taken by the individuals from the subpopulation when participating in the mechanism. S/he would also commit to transferring an amount of money to the participants that matches the utility these individuals would have gotten had the public-good level been implemented with the G-generalized SFM, plus a sufficiently large fixed amount to ensure individual rationality. Then, with the information yielded by the mechanism, the social planner could implement a public-good level that approximates the utilitarian optimum for the entire population with arbitrarily high precision (in the case of large populations). It would suffice to sample sufficiently many individuals. With a very large population, the relative size of the sample would be negligible, and so would be the required transfers. These transfers could be financed by uniform taxation by the rest of the population.

6 Conclusions

We have presented a new (voting) mechanism that spans over two stages, is detail-free, and implements the utilitarian optimum in a standard problem of public-good provision when there are two types of citizens. Unlike other mechanisms described in the literature which use multiple-round voting with varying thresholds, we have imposed the property that such thresholds require more than half of the votes for the policy finally approved in the second voting stage. This is a restriction often encountered in real settings that reflects the majoritarian logic of democracy; we have called this property *approval*. Our mechanism also displays the property that in equilibrium, all potential socially optimal proposals are considered for voting at some point in time. This property, which we have called *inclusiveness*, may also be desirable in real voting applications. Voting schemes based on our mechanism could therefore be introduced on an experimental basis, since they may be appealing both from the perspective of endstate justice as well as from the perspective of procedural justice.

Our model and results are also relevant from a purely positive perspective. It is known that reference points as default policies may have an impact on the outcome of voting procedures through varies channels (see e.g. McKelvey, 1976a; Kahneman and Tversky, 2013). We have shown that the utilitarian optimal outcome can be attained when the reference point—i.e., the status quo—and the vote requirements for changing this reference point are chosen appropriately. This is particularly relevant for the (optimal) design of thresholds for political action initiated by citizens through signature gathering or through other means by which minorities can be granted the right to propose policies. Our results indicate, in particular, that utilitarian welfare maximization can be reconciled with proposal-making procedures that span over two periods and that satisfy approval. It suffices for the first-round thresholds, say the thresholds determining the minimum number of signatures to be collected, to vary with the proposal on the table and *not* to be constrained by approval. The latter not only is a property that enables (almost) any group of individuals to initiate political action, it also helps to reach the utilitarian efficient outcome, at least in the case of two types.

References

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy*, 58(4):328–346.

Aşan, G. and Sanver, M. R. (2002). Another characterization of the majority rule. *Economics Letters*, 75(3):409–413.

Azrieli, Y. and Kim, S. (2014). Pareto efficiency and weighted majority rules. *International Economic Review*, 55(4):1067–1088.

Barberà, S., Gul, F., and Stacchetti, E. (1993). Generalized median voter schemes and committees. *Journal of Economic Theory*, 61(2):262–289.

Bierbrauer, F. and Sahm, M. (2010). Optimal democratic mechanisms for taxation and public good provision. *Journal of Public Economics*, 94(7-8):453–466.

Bierbrauer, F. J. and Hellwig, M. F. (2016). Robustly coalition-proof incentive mechanisms for public good provision are voting mechanisms and vice versa. *Review of Economic Studies*, 83(4):1440–1464.

Bishin, B. (2009). *Tyranny of the Minority: The Subconstituency Politics Theory of Representation*. Temple University Press.

Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy*, 56(1):23–34.

Bogomolnaia, A., Moulin, H., and Stong, R. (2019). Guarantees in fair division: beyond divide & choose and moving knifes. Technical report, arXiv. org.

Bowen, H. R. (1943). The interpretation of voting in the allocation of economic resources. *Quarterly Journal of Economics*, 58(1):27–48.

Brams, S. J. and Fishburn, P. C. (1978). Approval voting. *American Political Science Review*, 72(3):831–847.

Brandt, F. and Sandholm, T. (2005). Unconditional privacy in social choice. In *TARK*, pages 207–218.

Casella, A. (2005). Storable votes. *Games and Economic Behavior*, 51(2):391–419.

Chatterji, S. and Sen, A. (2011). Tops-only domains. *Economic Theory*, 46(2):255–282.

Chevaleyre, Y., Endriss, U., Lang, J., and Maudet, N. (2007). A short introduction to computational social choice. In *International Conference on Current Trends in Theory and Practice of Computer Science*, pages 51–69. Springer.

Ching, S. (1997). Strategy-proofness and median voters. *International Journal of Game Theory*, 26(4):473–490.

Diskin, A., Koppel, M., and Samet, D. (2011). Generalized Raiffa solutions. *Games and Economic Behavior*, 73(2):452–458.

Fahrenberger, T. and Gersbach, H. (2010). Minority voting and long-term decisions. *Games and Economic Behavior*, 69(2):329–345.

Fershtman, C. (1990). The importance of the agenda in bargaining. *Games and Economic Behavior*, 2(3):224–238.

Fleurbaey, M. and Maniquet, F. (2011). *A theory of fairness and social welfare*, volume 48. Cambridge University Press.

Gersbach, H. (2005). *Desinging Democracy: Ideas for Better Rules*. Springer, Heidelberg.

Gersbach, H. (2009). Democratic Mechanisms. *Journal of the European Economic Association*, 7(6):1436–1469.

Gersbach, H. (2017). Flexible majority rules in Democracyville: A guided tour. *Mathematical Social Sciences*, 85:37–43.

Gersbach, H., Imhof, S., and Tejada, O. (2019). Channeling the final say in politics: A simple mechanism. *forthcoming in Economic Theory*.

Gershkov, A., Moldovanu, B., and Shi, X. (2017). Optimal voting rules. *Review of Economic Studies*, 84(2):688–717.

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601.

Grech, P. and Tejada, O. (2018). Divide the dollar and conquer more: sequential bargaining and risk aversion. *International Journal of Game Theory*, 47(4):1261–1286.

Green, J. and Laffont, J.-J. (1977). Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica*, 45(2):427–438.

Hortalà-Vallvé, R. (2010). Qualitative voting. *Cuadernos de economía*, 33(92):5–43.

John, R. and Raith, M. G. (2001). Optimizing multi-stage negotiations. *Journal of Economic Behavior & Organization*, 45(2):155–173.

Kahneman, D. and Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the Fundamentals of Financial Decision Making: Part I*, pages 99–127. World Scientific.

Kwiek, M. (2017). Efficient voting with penalties. *Games and Economic Behavior*, 104:468–485.

Ledyard, J. O. and Palfrey, T. R. (1994). Voting and lottery drafts as efficient public goods mechanisms. *Review of Economic Studies*, 61(2):327–355.

Ledyard, J. O. and Palfrey, T. R. (1999). A characterization of interim efficiency with public goods. *Econometrica*, 67(2):435–448.

Lutz, G. and Marsh, M. (2007). Introduction: Consequences of low turnout. *Electoral Studies*, 26(3):539–547.

Maskin, E. (1995). *Majority Rule, Social Welfare Functions, and Games Forms (essays in honor of Amartya Sen)*, pages 100–109. Oxford University Press, Oxford, UK.

May, K. O. (1952). A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica*, 20(4):680–684.

McKelvey, R. (1976a). D.(1976),'intransitivities in multidimensional voting models and some implications of agenda control'. *Journal of Economic Theory*, 12(3):472–473.

McKelvey, R. D. (1976b). Intransitivities in multidimensional voting models and some implications for agenda control. *Journal of Economic Theory*, 12(3):472–482.

Miroiu, A. (2004). Characterizing majority rule: From profiles to societies. *Economics Letters*, 85(3):359–363.

Moulin, H. (1980). On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455.

Moulin, H. (2004). *Fair division and collective welfare*. MIT press.

Moulin, H. (2008). Social choice. In Weingast, B. R. and Wittman, D., editors, *The Oxford Handbook of Political Economy*, volume 4. Oxford University Press.

Moulin, H. (2014a). *Cooperative microeconomics: A game-theoretic introduction*, volume 313. Princeton University Press, Princeton, NJ.

Moulin, H. (2014b). *The Strategy of Social Choice*, volume 18. North Holland, Amsterdam.

Moulin, H. (2019). Fair division in the internet age. *Annual Review of Economics*, 11:407–441.

Palfrey, T. R. and Rosenthal, H. (1991). Testing for effects of cheap talk in a public goods game with private information. *Games and Economic Behavior*, 3(2):183–220.

Rasch, B. E. (2000). Parliamentary floor voting procedures and agenda setting in Europe. *Legislative Studies Quarterly*, 25(1):3–23.

Rosar, F. (2015). Continuous decisions by a committee: Median versus average mechanisms. *Journal of Economic Theory*, 159:15–65.

Saffon, M. P. and Urbinati, N. (2013). Procedural democracy, the bulwark of equal liberty. *Political Theory*, 41(3):441–481.

Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217.

Sprumont, Y. (1991). The division problem with single-peaked preferences: A characterization of the uniform allocation rule. *Econometrica*, 59(2):509–519.

Steihaus, H. (1948). The problem of fair division. *Econometrica*, 16:101–104.

Thomson, W. (2011). Fair allocation rules. In *Handbook of Social Choice and Welfare*, volume 2, pages 393–506. Elsevier.

Woeginger, G. J. (2003). A new characterization of the majority rule. *Economics Letters*, 81(1):89–94.

Appendix

In this appendix, we do three things. First, in Section A.1, we introduce a family of voting mechanisms that includes SFM, the voting mechanism analyzed in the main text of the paper. All our results apply to each element of this family of mechanisms. Second, in Section A.2, we provide the proof of Theorem 1, which is itself structured along several items, following the logic of backward induction. Third and last, in Section A.3, we provide the details needed in addition to the discussion of Section 5 for the generalized SFM.

A.1 A family of voting mechanisms based on semi-flexible majority rules

As discussed in Section 3—see Footnote 25—, the result of Theorem 1 not only holds for the voting mechanism described in this section, but also for a broader family of voting mechanisms (or voting procedures) containing the former as an instance. Each voting mechanism of this family is contingent on the choice of certain non-increasing sequences of majority thresholds to be applied to the first rounds of Stage 2, more specifically, to the rounds involving policies y^n, y^{n-1},..., and $y^k = \bar{x}$ (the status quo chosen in Stage 1) as defined in Section 3.3. Formally, for each $k \in \{0, \ldots, n\}$ let

$$A^k := \{1, 2, \ldots, n+1-k\}$$

and

$$B^k := \left\{ \max\left\{\frac{n+1}{2}, k\right\}, \max\left\{\frac{n+1}{2}, k\right\} + 1, \ldots, n \right\}.$$

While we interpret A^k as round indices (which, in turn, are associated with voting over particular alternatives), the elements of B^k are majority thresholds to be used for voting—see below. Note in particular that $B^k = \{k, \ldots, n\}$ if $k \geq \frac{n+1}{2}$. Then for each $k \in \{0, \ldots, n\}$, with $\bar{x} = y^k$ being the status quo chosen in Stage 1, let

$$\begin{aligned} f^k : A^k &\longrightarrow B^k \\ h &\longrightarrow f^k(h) \end{aligned}$$

be a function with the properties described below. The only difference with regard to the description of the two-stage voting mechanism defined in Section 3 is that now, for Round 2.h, with $h \in \{1, \ldots, n+1-k\}$, the majority threshold required to adopt the proposal on the table y^{n-h+1} is $f^k(h)$. Hence, each choice of $\left\{f^k(\cdot)\right\}_{k=0}^{n}$ defines a voting mechanism. We assume that

$f^k(\cdot)$ is non-increasing and onto.[37] It follows, in particular, that $f^k(1) = n$ and

$$f^k(n+1-k) = \max\left\{\frac{n+1}{2}, k\right\}. \tag{10}$$

Moreover, we assume that

$$f^k(h) \geq f^{k+1}(h), \text{ for all } h \in A^{k+1} = \{1, \ldots, n-k\}. \tag{11}$$

For each $k \in \{0, \ldots, n\}$, we let $\mathcal{F}^k$ be the set of all such functions. It follows directly that if $k \geq \frac{n+1}{2}$, there is only one possible element of $\mathcal{F}^k$, namely

$$f^k(h) = n + 1 - h.$$

This is the threshold function considered in the main body of the paper. Finally, consider that $k < \frac{n+1}{2}$, in which case

$$B^k = \left\{\frac{n+1}{2}, \ldots, n\right\}.$$

Then let

$$\underline{f}^k(h) := \max\left\{n + 1 - h, \frac{n+1}{2}\right\},$$

and

$$\overline{f}^k(h) = \min\left\{n, \frac{3(n+1)}{2} - k - h\right\}.$$

Note that

$$n + 1 - h \geq \frac{n+1}{2} \Longleftrightarrow h \leq \frac{n+1}{2}\Big(< n + 1 - k\Big)$$

and

$$n \geq \frac{3(n+1)}{2} - k - h \Longleftrightarrow h \geq \frac{n+1}{2} + 1 - k\Big(> 1\Big).$$

Hence,

$$\underline{f}^k(1) = n \quad \text{ and } \quad \underline{f}^k(n+1-k) = \frac{n+1}{2}$$

and

$$\overline{f}^k(1) = n \quad \text{ and } \quad \overline{f}^k(n+1-k) = \frac{n+1}{2}.$$

Because $\underline{f}^k(h)$ and $\overline{f}^k(h)$ are clearly non-decreasing in h, both functions belong to $\mathcal{F}^k$. Note, in particular, that $\underline{f}^k(\cdot)$ is the threshold function considered in the main body of the paper. It is also straightforward to verify that for each $f^k(\cdot) \in \mathcal{F}^k$ and $h \in A^k$,

$$\underline{f}^k(h) \leq f^k(h) \leq \overline{f}^k(h). \tag{12}$$

[37] A mapping $f : A \to B$ is *onto* if for all $y \in B$ there is $x \in A$ such that $y = f(x)$.

The shape of the possible functions $f^k(\cdot) \in \mathcal{F}^k$ is illustrated in Figure 3 for the case where $k < \frac{n+1}{2}$. It shows that $\mathcal{F}^k$ is the convex hull of $\underline{f}^k(\cdot)$ and $\overline{f}^k(\cdot)$, which are the supremum and the infimum of the set of functions, respectively.[38]

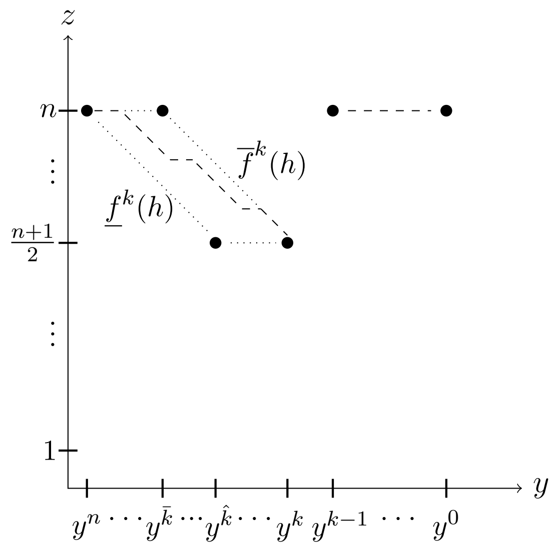


Figure 3: Generalized Stage 2—Number of votes (z) required to approve a policy, with $\bar{x} = y^k$ the status quo set out in Stage 1. The mechanism starts with y^n. Case $k < \frac{n+1}{2}$ (with $\hat{k} = \frac{n+1}{2}$ and $\bar{k} = \frac{n-1}{2} + k$). Functions $\underline{f}^k(\cdot)$ and $\overline{f}^k(\cdot)$ are depicted by dotted lines.

For each status-quo proposal $\bar{x} = y^k$ chosen in Stage 1, $\mathcal{F}^k$ has been defined as the set of possible majority thresholds for voting rounds in Stage 2 where the proposal on the table specifies a public-good level higher than, or equal to, the one specified by the status quo chosen in Stage 1. This set is reminiscent of the class of decreasing thresholds considered by Gershkov et al. (2017). Indeed, when $\bar{x}$ coincides with the ex-ante status-quo, namely y^0, our class of thresholds coincides with the subclass of thresholds considered by Gershkov et al. (2017) satisfying the additional property that no threshold can be below half plus one of the number of citizens. Such a class of thresholds can be rationalized from a mechanism-design viewpoint when the majority requirement is imposed beyond standard properties such as anonymity, unanimity, and incentive compatibility.

A.2 The proof of the main result

In this section we solve the game induced by any of the voting mechanisms described above, and, in particular, by the voting mechanism considered in the main body of the paper, viz.

[38]Note that $\underline{f}^k(h)$ and $\overline{f}^k(h)$ are functions that assign integers (and their corresponding alternatives) to integers, so the convex hull refers only to functions of the same type.

SFM. We analyze the game backwards. We thus start with the analysis of Stage 2.

Proof of Theorem 1

First, we analyze Stage 2, then we analyze Stage 1. Finally we discuss the equilibria of the game.

Analysis of Stage 2

Let $\bar{x} = y^k$ be the outcome of Stage 1, with $k \in \{0, \ldots, n\}$. We start by considering Round 2.h of Stage 2, with $h \in \{n+1, \ldots, 1\}$. We next prove the following claim:

Claim 1

In any round 2.h, with $h \in \{n+1, \ldots, 1\}$,

(i) Any high-type citizen votes to implement the proposal on the table (if $1 \leq h \leq n-k+1$, i.e., in the initial rounds of Stage 2) or to proceed to the next round (if $n-k+1 < h \leq n+1$, i.e., in the last rounds of Stage 2).

(ii) Any low-type citizen votes to proceed to the next round in all rounds (except the final round of Stage 2, i.e. in Round 2.(n+1)).

To prove the above claim, we distinguish three cases.

Case I: $h = n+1$

The decision in Round 2.$(n+1)$ consists in choosing either y^0 or $\bar{x} = y^k$ as the final outcome. If $k = 0$, there is no real choice, and $\bar{x} = y^0$ is adopted regardless of the citizens' vote. Hence, let $k > 0$ and thus $y^0 < \bar{x}$. In this case, unanimity is required for approval of y^0. Then for any observed history of previous play, it is in any citizen's best interest to vote for the alternative that yields higher utility, either y^0 or $\bar{x}$, because doing so makes a positive difference when the vote is pivotal—i.e., when all other citizens vote for y^0—and makes no difference otherwise. In other words, voting for the proposal on the table is weakly dominated for high-type citizens by voting to proceed to Round 2.$(n+2)$, while the weak-domination relation is reversed for low-type citizens. Hence, if citizens play no strategies that are weakly dominated, low-type citizens vote for y^0 and high-type citizens vote to proceed to the last round, where $\bar{x}$ is automatically adopted.

Case II: $n-k+1 < h < n+1$

The decision in Round 2.h, with $n-k+1 < h < n+1$, consists in choosing between y^{n-h+1} as the final outcome and proceeding to Round 2.$(h+1)$. Unanimity is required for y^{n-h+1} to be adopted. This implies that any individual has the (veto) power to ensure that the procedure continues to the next round and ultimately to Round 2.$(n+2)$, a round in which $\bar{x}$ is adopted—if in all these rounds such a citizen votes in favor of this possibility.

Let us first consider an individual i of type t^H. Such an individual votes to proceed to the next round until Round 2.$(n+2)$ is eventually reached. The reason is that s/he prefers $\bar{x} = y^k$ over y^{n-h+1} (recall that $n-k+1 < h$) and s/he anticipates that weakly undominated strategies (in particular, her/his own strategies) will be eliminated in future rounds. Indeed, consider Round 2.n as a base case for an induction argument. Then citizen i knows that in Round 2.$(n+1)$ s/he will vote to proceed to Round 2.$(n+2)$, as doing otherwise is weakly dominated for her/him. This has been proven in Case I above. Assume now that citizen i's vote is pivotal in Round 2.n, i.e., all other citizens vote for the proposal on the table y^1 (in all other cases citizen i's vote does not matter for the outcome). If citizen i votes for the proposal on the table, y^1 is implemented. If, by contrast, citizen i votes to proceed to the next round, namely to Round 2.$(n+1)$, then $\bar{x}$ will be implemented. Since individual i prefers alternative $\bar{x}$ to alternative y^1, s/he votes in Round 2.n to proceed to the next round.

Besides the fact that it is common knowledge that agents eliminate recursively strategies that are weakly dominated by moving backwards, and in particular that it is common knowledge that agents do not play weakly dominated strategies, in the above reasoning we have used the fact that citizen i knows that s/he himself/herself will take part in the binary voting of Round 2.$(n+1)$. This means, in particular, that s/he can form *correct* beliefs about whether all citizens are of the low type (s/he knows they are not). Then, iterating the above argument backwards from Round 2.n to Round 2.$(n-k+2)$, we obtain that for any observed history of previous play, any high-type citizen votes in favor of proceeding to the next round in all Rounds 2.h, with $h \in \{n-k+2, \ldots, n\}$. This results in adoption of $\bar{x}$ regardless of the vote of the remaining citizens whenever there is at least one individual of the high type.

We now consider an individual of type t^L. In Round 2.h, s/he faces a more subtle choice than a citizen of type t^H. If s/he votes in favor of the alternative on the table, y^{n-h+1}, her/his vote may theoretically help to guarantee that such a policy is chosen. If, by contrast, s/he votes for the procedure to continue to the next round, the risk s/he takes is that the status quo $\bar{x}$ will eventually be chosen if all alternatives $y^{n-h}, \ldots, y^0$ are subsequently rejected. Because s/he prefers alternative y^{n-h+1} to $\bar{x} = y^k$ (recall that $n-k+1 < h$), the choice is not obvious. Nevertheless, we next show that for an individual i of type t^L, it is always optimal to vote to proceed to the next voting round.

We develop a further induction argument and consider Round 2.n as a base case. We focus on the situation where citizen i's vote is pivotal, as her/his vote does not matter for outcomes otherwise. If s/he votes for the proposal on the table, alternative y^1 is implemented. Instead, if s/he votes to proceed to the next round, namely to Round 2.$(n+1)$, then either y^0 or $\bar{x}$ will be implemented. Citizen i knows that the latter will occur if and only if there is at least one citizen who is of the high type. Moreover, since it is common knowledge that (high-type) citizens eliminate iteratively weakly dominated strategies, a high-type citizen votes in Round 2.n to proceed to the next round and hence citizen i's vote is not be pivotal in this round. That is, a necessary condition for i's vote to be pivotal in Round 2.n is that all agents are of the low type. In this case, our previous analysis has shown that y^0 is implemented in Round 2.$(n+1)$. This means that if, in Round 2.n, citizen i votes for the proposal on the table when all agents are of the low type, y^1 is implemented. In contrast, if in this same case citizen i votes to proceed to the next round, namely to Round 2.n, alternative y^0 is implemented. Since individual i prefers alternative y^0 to alternative y^1, s/he votes in Round 2.n to proceed to the next round.

In the above reasoning we have used that citizen i of type t^L knows that citizens of the high type can form correct beliefs about whether all agents are of the low type. Then, iterating the above argument backwards from Round 2.n to Round 2.$(n-k+2)$, we obtain that for any observed history of previous play leading to correct beliefs about whether all citizens are of the low type or not, any low-type citizen votes to proceed to the next round in all Rounds 2.h, with $h \in \{n-k+2, \ldots, n\}$. This results in adoption of y^0 if and only if all individuals are of the low type. If not, $\bar{x}$ is implemented.

Case III: $1 \leq h \leq n-k+1$

Again, the decision in Round 2.h, now with $1 \leq h \leq n-k+1$, consists in choosing y^{n-h+1} as the final outcome or proceeding to the next stage. Note that $r = n+1-h$ is the index used in Section 3 to describe policy y^r. This time, however, a majority of $f^k(h)$ of votes is required for y^{n-h+1} to be adopted.[39] If such a majority does not materialize, the procedure continues, yielding some outcome $y \in \mathcal{Y}$ satisfying the property that $y \leq y^{n-h+1}$, with $y^{n-h+1} \geq \bar{x} = y^k$. Recall that $n-k+1 \geq h$. This means that if the procedure does not stop at Round 2.h the outcome will leave high-type individuals worse off and low-type individuals better off (if $h < n-k+1$), or will leave high-type individuals *weakly* worse off and low-type individuals

[39]We stress that in the main body of the paper, we have considered $f^k(h) = \max\left\{\frac{n+1}{2}, n+1-h\right\}$.

weakly better off (if $h = n - k + 1$).

We distinguish two cases. First, assume that $h < n - k + 1$. Then, high-type individuals vote in favor of y^{n-h+1} and low-type individuals vote to proceed to the next round. The reason is that for any high-type citizen i, voting in favor of the proposal on the table weakly dominates voting to proceed to the next round. For this result to hold, no assumption about future voting behavior by other agents is needed. Similarly, for any low-type citizen, voting to proceed to the next round weakly dominates voting in favor of the proposal on the table. For this result to hold, the fact that agents will iteratively eliminate weakly dominated strategies in all future rounds and can form correct beliefs about whether all agents are of the high type does not matter. These two properties hold independently of the majority required in the voting round, viz. $f^k(h)$.

Second, let us now examine the case $h = n - k + 1$. Consider first a citizen i of type t^H. If s/he votes for the proposal on the table, i.e. for y^k, s/he may contribute to the adoption of such a proposal if her/his vote is pivotal (in any other case, her/his vote makes no difference for the outcome). If, on the other hand, citizen i's vote is pivotal for the procedure to continue to the next round, s/he knows that s/he can ensure that $\bar{x} = y^k$ will also be adopted in subsequent voting rounds thanks to her/his veto power. The latter requires that citizen i knows that s/he will vote in these subsequent rounds, and in particular that s/he can form correct beliefs about whether or not all citizens are of the low type. Yet, it is clear that, in Round 2.$(n - k + 1)$, voting to proceed to the next round is weakly dominated for citizen i by voting for the proposal on the table, since doing so would produce an outcome $y \leq y^k = \bar{x}$. This holds independently of the fact that s/he can use her/his veto power in all subsequent rounds. Consider now a citizen of type t^L. Reversely, voting to proceed to the next round produces an outcome $y \leq y^k = \bar{x}$, which in the case where $y < y^k$ (an outcome that is attainable for certain voting behavior) is strictly preferred over y^k. That is, for a low-type citizen, voting to proceed to the next round weakly dominates voting for the proposal on the table, y^k. This holds independently of the majority required in the voting round, viz. $f^k(n + 1 - k) = \max\{\frac{n+1}{2}, k\}$. This completes the proof of Claim 1.

Having established the citizens' behavior in all voting rounds of Stage 2, we now let k (with $\bar{x} = y^k$) vary from 0 to n and obtain the outcome of Stage 2 for different distributions of citizen types. More specifically, we next prove the following claim:

Claim 2

The outcome of Stage 2 for different distributions of citizen types as function of $\bar{x}$ is given by the following table:

$\bar{x} \backslash t^{soc}$	t^L	$\frac{n-1}{n}t^L + \frac{1}{n}t^H$	...	$\frac{n+1}{2n}t^L + \frac{n-1}{2n}t^H$	$\frac{n-1}{2n}t^L + \frac{n+1}{2n}t^H$	...	$\frac{1}{n}t^L + \frac{n-1}{n}t^H$	t^H
y^n	y^0	y^n	...	y^n	y^n	...	y^n	y^n
y^{n-1}	y^0	y^{n-1}	...	y^{n-1}	y^{n-1}	...	y^{n-1}	y^n
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
$y^{\frac{n+1}{2}}$	y^0	$y^{\frac{n+1}{2}}$	...	$y^{\frac{n+1}{2}}$	$y^{\frac{n+1}{2}}$	...	y^{n-1}	y^n
$y^{\frac{n-1}{2}}$	y^0	$y^{\frac{n-1}{2}}$	...	$y^{\frac{n-1}{2}}$	$y^{r^*((n-1)/2,(n+1)/2)}$	...	$y^{r^*((n-1)/2,n-1)}$	y^n
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
y^1	y^0	y^1	...	y^1	$y^{r^*(1,(n+1)/2)}$	...	$y^{r^*(1,n-1)}$	y^n
y^0	y^0	y^0	...	y^0	$y^{r^*(0,(n+1)/2)}$	...	$y^{r^*(0,n-1)}$	y^n

Table 1: Outcome of Stage 2 as a function of the status quo $\bar{x} = y^k$ (rows) and the optimal utilitarian type $t^{soc} = \frac{n-j}{n} \cdot t^L + \frac{j}{n} \cdot t^H$ (columns).

The numbers $r^*(k, j)$ are defined in (14) below. To prove the above claim, we build on the insights provided in Cases *I–III* in the proof of Claim 1. Recall that the socially optimal level of public good is denoted by $t^{soc} = \frac{n-j}{n} \cdot t^L + \frac{j}{n} \cdot t^H$, with $j \in \{0, \ldots, n\}$. That is, t^{soc} denotes the utilitarian solution when the society is made up of j citizens of type t^H and $n - j$ citizens of type t^L.

First, if $j = 0$, all citizens—who are of the low type—always vote to proceed to the next voting round, until Round 2.$(n+1)$ is eventually reached, where all of them vote in favor of the proposal on the table, y^0. Second, if $0 < j \leq \frac{n-1}{2}$, all low-type citizens, who constitute a majority of the electorate, will block the approval of any proposal of Rounds 2.1 to 2.$(n+1-k)$. The reason is that a (qualified) majority is needed in all these rounds for the approval of the proposal on the table. However, since there is always at least one citizen of the high type, any such individual can guarantee that the status quo $\bar{x}$ is eventually chosen. Third, if $\frac{n+1}{2} \leq j \leq n$, a majority of the electorate is made up of citizens of the high type. If $j = n$, in particular, all citizens, who are of type t^H, vote for y^n, and the procedure ends just after Round 2.1.

Accordingly, we are left with the constellation

$$\frac{n+1}{2} \leq j < n. \tag{13}$$

Note that because there is always one citizen of type t^H, the procedure yields some proposal y

with the property that $y \geq y^k = \bar{x}$. We distinguish two cases, depending on the status quo chosen in Stage 1, viz. $\bar{x}$.

Case A: $\frac{n+1}{2} \leq k \leq n$

In this case, for $h \in \{1, \ldots, n+1-k\}$, we have $f^k(h) = n+1-h$ as the majority threshold required for the approval of the proposal on the table of Round 2.h, namely y^{n+1-h}. The reason is that for $k \geq (n+1)/2$, the only non-increasing, onto function belonging to $\mathcal{F}^k$ is $f^k(h) = n+1-h$, since A^k and B^k have the same cardinality. Because in each of these rounds all individuals of the high type vote in favor of the proposal on the table and all individuals of the low type vote to proceed to the next voting round, the outcome of the procedure is $\max\{y^j, \bar{x}\}$, where $\bar{x} = y^k$ is the status quo and j is the number of citizens of type t^H. In particular, if $\bar{x} = y^j$, the outcome is the status quo proposal chosen in Stage 1.

Case B: $0 \leq k \leq \frac{n-1}{2}$

For $h \in \{1, \ldots, n+1-k\}$, we again have $f^k(h)$ as the majority threshold required for the approval of the proposal on the table of Round 2.h, namely y^{n+1-h}. As in Case A, in each of these rounds all individuals of the high type vote in favor of the proposal on the table and all individuals of the low type vote to proceed to the next voting round. Accordingly, the final outcome of the procedure is $\max\{y^{r^*(k,j)}, \bar{x}\}$, where $\bar{x} = y^k$ is the status quo, j is the number of citizens of type t^H, and[40]

$$r^*(k,j) = \max\{r : f^k(n+1-r) \leq j\}, \tag{14}$$

That is, $y^{r^*(k,j)}$ is the highest level of the public good that the j citizens of the high type can guarantee as an outcome throughout Rounds 2.1 to 2.$(n+1-k)$, given the status quo $\bar{x} = y^k$ chosen in Stage 1. Using Equations (10) and (13), we obtain

$$j \geq \frac{n+1}{2} = \max\left\{\frac{n+1}{2}, k\right\} = f^k(n+1-k).$$

It then follows that $r^*(k,j)$ is well-defined and that $r^*(k,j) \geq k$. What is more, $r^*(k,l)$ is non-decreasing in k, due to Equation (11), and non-decreasing in j, by construction. We further note that also due to Equations (10) and (13), we have

$$f^{(n-1)/2}\left(n+1-\frac{n-1}{2}\right) = \max\left\{\frac{n+1}{2}, \frac{n-1}{2}\right\} = \frac{n+1}{2} \leq j.$$

This implies

$$r^* := r^*\left((n-1)/2, j\right) \geq \frac{n-1}{2}. \tag{15}$$

[40] For notational convenience, we have suppressed the dependence of $r^*(k,j)$ on the function $f^k(\cdot)$.

Trivially, Equation (13) also implies that

$$j \geq \frac{n-1}{2}. \tag{16}$$

We next claim that

$$r^* \leq j. \tag{17}$$

Since $f^{(n-1)/2}(\cdot)$ is non-increasing and since Equations (15) and (16) guarantee that $n+1-r^*$ and $n+1-j$ belong to the domain of function $f^{(n-1)/2}(\cdot)$, Inequality (17) implies

$$f^{(n-1)/2}(n+1-j) \geq f^{(n-1)/2}(n+1-r^*).$$

Finally, it remains to note that

$$f^{(n-1)/2}(n+1-j) \geq \underline{f}^{(n-1)/2}(n+1-j) = \max\left\{\frac{n+1}{2}, j\right\} = j \geq f^{(n-1)/2}(n+1-r^*).$$

The above chain of inequalities can be explained as follows: The first inequality is due to Equation (12). Note that Equation (16) guarantees that $n+1-j$ belongs to the domain of $f^{(n-1)/2}(\cdot)$ and $\underline{f}^{(n-1)/2}(\cdot)$. The first equality is a direct consequence of the definition of $\underline{f}(\cdot)$. The second equality holds because $j \geq \frac{n+1}{2}$. Finally, the second inequality holds by definition of $r^* = r^*\left((n-1)/2, j\right)$.

To sum up, building on our entire analysis, we can arrange the outcome of Stage 2 for different status-quo choices in Stage 1 as was done in Table 1. This completes the proof of Claim 2. That is, as we are considering a larger value of $\bar{x}$, the final outcome cannot be worse (or better) for high-type (low-type) citizens, whatever the type composition of the electorate. Note that together with the fact that $f^k(\cdot)$ is non-increasing and onto, Equation (17) is crucial for Table 1. Moreover, the outcome is strictly better (worse) in some cases. This completes the analysis of Stage 2.

Analysis of Stage 1

Next we consider the problem which actions are taken by the citizens in each Round $1.h$ of Stage 1, with $h \in \{n, \ldots, 1\}$. Because agents eliminate weakly dominated strategies, we can focus our analysis on the case where a citizen's vote is pivotal for the outcome of voting round. Let us focus on high-type citizens first. From Table 1, it immediately follows that in Round $1.n$, voting for y^n as status quo yields higher or equal utility than voting for y^{n-1}, no matter the composition of the electorate, and hence regardless of the beliefs held by the citizens about such composition. As a matter of fact, unless a citizen of type t^H believes that all citizens are of the same type (t^L or t^H), voting to set $\bar{x} = y^n$ as status quo yields higher utility than voting to set $\bar{x} = y^{n-1}$. This means that in Round $1.n$, for any citizen i of the high type, *(a)* it is always optimal to set $\bar{x} = y^n$ and move to Stage 2, no matter the previous play of game, and *(b)* it is never optimal to set $\bar{x} = y^{n-1}$ and move to Stage 2, unless citizen i believes that all agents are of the high type. Because i herself/himself is of the high type, s/he cannot believe that all agents are of the low type.[41] However, if the prior distribution of types has full support, i.e., the probability measure that assigns to every number $k \in \{0, \ldots, n\}$ the probability that there are k individuals of the low type at the beginning of the first voting round has full support, then at the beginning of Round $1.(n-1)$, where only $n-2$ binary voting rounds have taken place, no individual can be certain that all individuals are of the high type after having observed the play and updated their beliefs using the Bayes' rule.

At any prior round of Stage 1, in turn, voting to proceed to the next round yields higher or equal utility than voting to stick with the current status quo policy. This property holds no matter the composition of the electorate, and hence regardless of the beliefs held by the citizens about such composition. This means that for any citizen of the high type, voting to proceed to the next voting round in Stage 1 is optimal in any Round $1.h$ (with $1 \leq h < n$). To see whether there are other optimal choices for some given beliefs, we distinguish two cases depending on what the proposal on the table, y^{h-1}, is. Recall that we are focusing on the case where the vote of a citizen i of the high type is pivotal for the outcome of the voting round. If $\frac{n+1}{2} \leq h-1 < n-1$, then voting to proceed to the next round yields higher utility than voting to stick with the current status quo policy unless citizen i believes that all citizens are of the high type. As before, if the prior distribution of types has full support, then at the beginning of Stage $1.h$ no individual can be certain that all individuals are of the high type, after having observed the play and updated their beliefs using the Bayes' rule. If $0 \leq h-1 < \frac{n+1}{2}$, voting to proceed to the next round yields higher utility than voting to stick with the current status quo policy unless (possibly) citizen i believes that the majority of the citizens are of the high type. However, if the prior distribution of types has full support, then at the beginning of Round $1.h$, where only $h-1$ binary voting rounds have taken place, no individual can be certain that a majority of the individuals are of the high type after having observed the play and updated their beliefs using the Bayes' rule.

For low-type individuals, the optimal decisions in Rounds $1.h$ (with $1 \leq h \leq n$) are reversed. It is then important to point out that for the voting decision to hold as best responses for

[41] This would contradict Bayesian update given i's private signal about her/his own type

both citizen types, it is necessary for it to be common knowledge that in Stage 2 agents will iteratively eliminate weakly dominated strategies by moving backwards and that they will be able to form correct beliefs about whether all agents are of type t^L.

It then follows that if citizens vote as described above, the elements of the off-diagonal of Table 1 are chosen in Stage 1. This yields the following table (recall that at least k votes are needed to set $\bar{x} = y^k$):

t^{soc}	t^L	$\frac{n-1}{n}t^L + \frac{1}{n}t^H$	$\dots$	$\frac{n+1}{2n}t^L + \frac{n-1}{2n}t^H$	$\frac{n-1}{2n}t^L + \frac{n+1}{2n}t^H$	$\dots$	$\frac{1}{n}t^L + \frac{n-1}{n}t^H$	t^H
y	y^0	y^1	$\dots$	$y^{\frac{n-1}{2}}$	$y^{\frac{n+1}{2}}$	$\dots$	y^{n-1}	y^n

Table 2: Outcome y of Stage 2 as a function of the optimal utilitarian type t^{soc}.

In short, combining the first-stage result with the second-stage result produces the outcome y^j, where j is the number of high-type citizens.

Analysis of Equilibria

We start by noting that if citizens vote in Stage 1 as described above, then at the end of this stage they reveal exactly how many citizens are of each type. This property obtains from any initial type distribution (with full support) through the Bayes' rule because high-type individuals always vote to proceed to the next voting round, while low-type citizens always vote for the proposal on the table (with the exception of Round $1.n$ where high-type citizens vote for y^n and low-type citizens vote for y^{n-1}). Accordingly, $\bar{x} = y^k$ if and only if k individuals are of the high type and $n - k$ individuals are of the low type. This means that all citizens can form correct beliefs at the beginning of Stage 2 about how many citizens there are of each type, and in particular about whether all citizens are of type t^L (and that this is common knowledge). To sum up, any pair of strategies and beliefs satisfying that

(i) in any but the last voting round of Stage 1, all citizens of type t^L (type t^H) vote to (not to) set the proposal on the table for status quo for Stage 2,

(ii) in the last voting round of Stage 1, all citizens of type t^L (type t^H) vote to set y^{n-1} (y^n) as status quo for Stage 2,

(iii) in any voting round of Stage 2, all citizens of type t^L vote to proceed to the next round (except for the last round, namely for Round $2.(n+1)$, where they vote for y^0),

(iv) In any voting round of Stage 2, all citizens of type t^H vote in favor of the proposal of the table if and only it prescribes a higher or equal public-good level than in the status quo $\bar{x}$ determined in Stage 1,

(v) The beliefs of all citizens at the end of Stage 1 are such that all citizens know the distribution of the types, and in particular whether all citizens are of the low type, and that this is common knowledge.

is a perfect Bayesian equilibrium. First, the actions taken at any voting round of Stage 2 are best responses given the beliefs and the strategies for the current and subsequent voting rounds of Stage 2 (if any). Second, the decisions taken at any voting round of Stage 1 are best responses (regardless of the exact beliefs of each individual at these nodes), given the strategies for the remaining voting rounds of Stage 1 and for all voting rounds of Stage 2. Third, as mentioned above, the beliefs at the end of Stage 1 can be computed through the Bayes' rule from the prior beliefs. Moreover, the resulting beliefs are consistent with the play along the rest of the equilibrium path. Accordingly, we have proved that the utilitarian optimal solution x^{soc} is an equilibrium outcome. Moreover, for any pair of strategy and beliefs satisfying the above conditions, citizens choose pure strategies that survive iterative elimination of weakly dominated strategies by moving backwards in the game, as is required by our equilibrium notion.

Are there other equilibria in the game underlying SFM? As we have shown above, if *(v)* holds, then in any perfect Bayesian equilibrium in which players eliminate iteratively weakly dominated strategies, *(ii)*, *(iii)*, and *(iv)* must, in turn, hold. Being independent of the beliefs in Stage 1, also *(i)* must hold. Therefore, if there are other equilibria, it must be the case that *(v)* does not hold. However, no citizen of the high type can hold wrong or non-conclusive beliefs about whether all citizens are of the low type at any node of the game, as this would contradict Bayesian updating.

To sum up, we obtain that the utilitarian optimal outcome x^{soc} is implemented uniquely by the mechanism. This completes the proof of Theorem 1.

□

A.3 The case of multiple types

In this section, we complete the analysis of Section 5.2.

We start by noting that because we assume that individuals do not play strategies that are weakly dominated and eliminate such strategies iteratively, to find out what individuals do in Steps 1 to $T-1$ in any equilibrium, we can apply the logic of the proof of Theorem 1 to each instance of SFM. This is for three reasons: *(i)* once a particular instance of SFM is chosen (randomly) through mechanism G in Step 0, the outcome is entirely determined by the actions chosen by the individuals in this application of SFM; *(ii)* when participating in any instance of SFM, individuals act as if it was chosen, since if it is not chosen their actions are irrelevant for payoffs; *(iii)* as discussed in Section 5.2, for each set $\mathcal{Y}_l$, agents of types $t^1, \ldots, t^l$ behave as if they were of type t^l, while all other individuals behave as if they were of type t^{l+1}.

Analogously to Claim 1, we therefore obtain the following claim about Stage 2 for any Step τ of the G-generalized SFM:

Claim 3

In any round 2.h of Step τ of the G-generalized SFM, with $\tau \in \{1, \ldots, T-1\}$ and $h \in \{n+1, \ldots, 1\}$,

(i) Any citizen of type t^l, with $l \in \{\tau+1, \ldots, T\}$, votes to implement the proposal on the table (if $1 \leq h \leq n-k+1$, i.e., in the initial rounds of Stage 2) or to proceed to the next round (if $n-k+1 < h \leq n+1$, i.e., in the latest rounds of Stage 2).

(ii) Any citizen of type t^l, with with $l \in \{1, \ldots, \tau\}$, votes to proceed to the next round in all rounds (except the final voting round of Stage 2), i.e. in Round 2.(n+1).

As for the first stage of each of the instances of SFM, one can proceed in a similar way and apply the counterpart of Claim 2 in the proof of Proposition 1 to each step, starting from the last step, and then using induction across all of them. As mentioned in Section 5, the latter combined with Claim 3 elicits the information about how many individuals are of each type through the system of linear equations in (9).

Finally, we focus on Step 0. One can easily verify that for an agent of type t^1, it is weakly dominated to report any $\mathcal{Y}_l \neq \mathcal{Y}_1$ due to the properties of $\succ_i^T$ and mechanism G. A symmetric reasoning shows that individuals of type t^T vote for $\mathcal{Y}_{T-1}$. Take now $l \in \{2, \ldots, T-1\}$. A reasoning along the same lines shows that an individual of type t^l votes either for $\mathcal{Y}_{l-1}$ or $\mathcal{Y}_l$ but never for another set. The exact vote depends on the exact values of types t^{l-1}, t^l, and t^{l+1}, and the beliefs the citizens have about the aggregate type distribution.

To sum up, recall that for any type $l \in \{1, \ldots, T\}$, we use n_l to denote the individuals who (report they) are of type t^l. Then also recall from (9) that letting $N^T = n$,

$$\begin{cases} n_1 & = N^1, \\ \dots & \\ n_1 + \ldots + n_{T-1} & = N^{T-1}, \\ n_1 + \ldots + n_{T-1} + n_T & = N^T. \end{cases}$$

That is, for $l \in \{1, \ldots, T\}$, N^l denotes the number of individuals of types $t^1, \ldots, t^l$. Then we have shown the following:

1. In Step 0,
   - any individual i of type t^1 reports $\tau_i' = \mathcal{Y}_1$,
   - any individual i of type t^l (with $l \in \{2, \ldots, T-1\}$) reports either $\tau_i' = \mathcal{Y}_{l-1}$ or $\tau_i' = \mathcal{Y}_l$,
   - any individual i of type t^T reports $\tau_i' = \mathcal{Y}_{T-1}$.
2. The outcome of Step 1 is
   - $y_1^{n-N^1}$ with probability $p_1(\tau_1', \ldots, \tau_n')$,
   - $y_l^{n-N^l}$ (with $l \in \{2, \ldots, T-1\}$) with probability $p_l(\tau_1', \ldots, \tau_n')$,
   - $y_{T-1}^{n-N^{T-1}}$ with probability $p_{T-1}(\tau_1', \ldots, \tau_n')$.
3. The numbers $n_1, \ldots, n_T$ reflect the actual shares in the population in terms of types.

□