Proposal Assessment Voting

Volker Britz
CER-ETH – Center of Economic
Research at ETH Zurich
Zürichbergstrasse 18
8092 Zurich, Switzerland
vbritz@ethz.ch

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

This Version: January 2020

Abstract

We introduce a voting procedure with proposal assessment called *Proposal Assessment Voting.* The procedure works as follows: A selfish agenda-setter chooses a proposal meant to replace a given status quo. In a first stage, only a random sample of the population votes on the proposal. The result of the first stage is made public, and may therefore reveal information about the underlying distribution of preferences in the electorate. Depending on the outcome of the first stage, a third alternative (next to the proposal and the status quo) is added: This alternative is either closer to or more distant from the status quo. Then, a second stage takes place: The entire electorate expresses pairwise preferences over the status quo, the initial proposal, and the newly added third alternative. We investigate the manipulability of this voting procedure, and we establish conditions under which information is truthfully revealed and under which Proposal Assessment Voting yields strict welfare gains.

Keywords: Democracy, Manipulation, Information Sharing, Referendum
JEL Codes: C72, D70, D72

1 Introduction

Good democratic procedures should accomplish two goals: First, they should facilitate accurate revelation and aggregation of information about the consequences of policy proposals. Second, they should adopt policies that concur with the true underlying preferences of the population. For instance, if a Condorcet winner exists, it should be adopted and prevail against any other feasible alternative, including the status quo.

Achieving these objectives in a voting procedure is difficult for several reasons: First, there is a large number of policy proposals that can be proposed for a single issue. Second, there is uncertainty about the underlying distribution of preferences in the electorate and thus ambiguity. Third, once a proposal has been adopted, it is often quite difficult to reverse the policy. The irreversibility is obvious if the policy involves physical or human capital investments such as infrastructure investments in highways, bridges, public buildings or the use of environmental resources. However, the irreversibility extends to many other policies as well. A particular pertinent example in current events is Brexit: The delays and procedural complications in implementing Brexit have illustrated how difficult it is to join and, in particular, leave a political union or even an interconnected set of multilateral treaties. Such decisions are reversible only at a high cost. Many other examples can be found for collective decisions that show a high degree of irreversibility.

How should we design good voting systems in such environments? In particular, how can we induce information revelation before proposals are made or how can proposals be adjusted in the light of new information before final irreversible collective decisions are taken?

In this paper, we introduce a new approach called *Proposal Assessment Voting* and examine to which extent it can resolve these issues. We aim for a democratic procedure that is robust against manipulation of information revelation by citizens, but also against exploitation of information revelation by a selfish agenda-setter. In previous work (Britz and Gersbach (2019)), we have explored how manipulation and exploitation in democratic procedures can be addressed by a combination of tax incentives and transfers, which set the right incentives to reveal and share information accurately and accomplish the implementation of the Condorcet winner. In this paper, we use only a voting procedure to address the problem.[1]

In this paper, we study a voting procedure called *Proposal Assessment Voting (PAV).* The procedure works as follows: A selfish agenda-setter chooses a proposal meant to replace a given status quo. In a first stage, only a random sample of the population votes on the proposal. The result of the first stage is made public, and may therefore reveal information about the underlying distribution of preferences in the electorate. Depending on the outcome of the first stage, a third alternative (next to the proposal and the status quo) is added: This alternative is either closer or more distant to the status quo. Then, a second stage takes place: The entire electorate expresses pairwise preferences over the status quo, the initial proposal, and the newly added third alternative. We investigate the manipulability of this voting procedure, and we establish conditions under which information is truthfully revealed and under which Proposal Assessment Voting yields strict welfare gains.

The present paper makes three main contributions:

First, the concept of Proposal Assessment Voting allows us to further explore one of the fundamental questions in democracy research: How can information revelation be accomplished in democratic procedures? This question has received extensive attention in the literature, see for instance and Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1997). More recent contributions are Bierbrauer and Hellwig (2016) and Britz and Gersbach (2019). When information about an underlying state of nature is dispersed among agents, there are complex incentives which may or may not lead agents to reveal, share, and aggregate their private information through a democratic procedure. For instance, if there are turnout costs, an individual agent may free-ride on the information provided by other agents. If a group of agents can use some coordination device, they may

[1] If waiting can reveal the information, then waiting with the adoption of an irreversible alternative may be desirable, see Gersbach (1993) for an assessment of whether a majority benefits from a "wait and see" choice.

be interested in misrepresenting their information and strategically mislead other voters. If some agents have information of better quality than others, they may also strategically decide to keep it to themselves. On the other hand, it is also conceivable that citizens share information through voting because they like to vote for the winner. In the present paper, we study how voting processes can be organized when there is not only private information about individual preference but also about the distribution of these preferences. Ideally, a multi-stage voting process entails that initial proposals are made upon voting reveals information about the underlying distribution of preferences and –in light of this information– new proposals are made that are "better" for the society in terms of welfare.

Second, our analysis of PAV extends the work of Gersbach et al. (2017) on Assessment Voting with only two fixed alternatives. Assessment Voting was originally suggested by Gersbach (2015), and has been studied from a game-theoretic point of view in a recent paper by Gersbach et al. (2017). The idea of assessment voting is as follows: When an initiative comes up for a vote, a small random sample of the population is drawn (the so-called *Assessment Group*). Now, only this group votes on the proposal, and the result is publicly observed. The group initiating the proposal can then decide to withdraw it or to have the rest of the population vote. This voting procedure does not compromise the standard rules of a liberal democracy: it is still true that all citizens' votes count equally in the end. The only pitfall is that a partial result of the referendum can be observed before most citizens cast their votes. Assessment voting implies that each citizen has to express himself on fewer "hopeless" proposals, since those proposals are eliminated at the assessment stage. Voter frustration with too many extreme and hopeless proposals is thus avoided. In addition, Assessment Voting makes it harder for a small but extreme minority to impose their will on society as a whole: If an extreme proposal were to proceed to the second voting stage, this would be a strong wakeup call for the silent and moderate majority that they have to mobilize. Assessment Voting as studied by Gersbach et al. (2017) has one crucial limitation, however: Between the assessment stage and the final voting stage, it only allows for two courses of action: A proposal can be either maintained or withdrawn. It is not possible to make an amendment to the proposal. The proposers cannot re-optimize their proposal based on the information they learn from the result of voting in the assessment group. This is the gap that the present work aims to fill. We allow for "proposal assessment." That is, upon observing the result of a vote at the assessment stage, one can decide to change the proposal. This is best understood with the following example: Suppose that someone makes an extreme policy proposal. The assessment stage reveals that this extreme proposal is supported by only 30 per cent of voters. Now, it might be tempting for the proposers to amend their proposal in such a way that it becomes a little bit less extreme. They might hope that some moderation pushes support for their motion just above a 50 per cent threshold. By first selecting a random sample of the population and letting them vote, the remaining electorate may decide to abstain in the second voting round if the result from the first round favors clearly one alternative. In a costly voting setting, Gersbach et al. (2017) have shown that Assessment Voting can save costs and increase the likelihood that the alternative preferred by a majority is selected. In this paper, we develop Proposal Assessment Voting to increase the probabiltiy that proposals are made that are superior in terms of welfare compared to one or two round voting with fixed proposals.

Third, PAV allows us to suggest potential remedies for some of the problems faced by modern democracies. In the Swiss system of direct democracy, for instance, any citizen can in principle propose a law. If the proposal is backed by a certain number of citizens in a signature collection process, then it must be put to a popular vote.[2] This rule has existed for more than a century, but one important recent development has been that the rise of social media and internet campaigns as well as population growth have made it much easier to collect the required number of signatures. At first sight, this seems to be a positive development: It can increase citizen participation. However, there are also significant drawbacks. First, an increasing number of initiatives and thus popular votes may lead to frustration among the citizenry and undermine citizens' willingness to participate in the process. Second, it becomes easier to propose extreme or unrealistic initiatives.

[2]For a more detailed discussion of the direct democratic system in Switzerland, see Rühli and Adler (2015).

Sometimes, such initiatives may be used by a party or interest group to strengthen their visibility and attractiveness for subgroups in society. For instance, a party might advocate an unrealistic change to be perceived as defender of the poor. Third, both of the above effects combined increase the likelihood that small but highly mobilized minorities with extreme views can implement their agenda, while a moderate but aloof majority stands idly by. Proposal Assessment Voting might help to address these concerns by introducing the possibility that an extreme proposal may be moderated in the process.

The aftermath of the Brexit referendum is a case in point for how PAV could improve democratic decision-making: While the referendum allowed for only two options ("Leave the European Union" vs. "Remain in the European Union"), it has since become apparent that there are many more options in reality, such as various forms of a "negotiated" Brexit, a free trade zone, or a "no-deal" Brexit. As of January 2020, it appears clear that Brexit is about to occur, however, negotiations about future relations with the EU are scheduled to take another year. Brexiteers and Remainers cite various opinion polls which deliver contradictory information about the British public's preferences over these options. Applying Proposal Assessment Voting could have led to a timely discovery of the underlying preferences, to a more precise and meaningful formulation of the referendum question, and ultimately to greater certainty that the outcome corresponds to the underlying preferences of the population.

Introducing the possibility of proposal assessment into the model, however, leads to important information and manipulation problems, which are not present in Gersbach et al. (2017)'s study of Assessment Voting, and which do resemble the problems of exploitation and manipulation as discussed in Britz and Gersbach (2019): In an ideal world, voters would reveal their preferences during the assessment stage. The proposer could then adjust his proposal to fit these preferences, and the popular vote would then sanction the outcome. In reality, however, voters may not express their true preferences in the assessment stage. For instance, voters may want to feign support for an extreme policy because they hope that this will lead to a continuation of the status quo. This is similar to the problem of holding "open primaries." Supporters of the Democratic Party may want to vote for a very extreme Republican candidate, hoping that the extreme candidate wins the nomination, but then goes on to lose the election. With proposal assessment voting, we can eliminate or at least limit these attempts and can ensure that the scheme is welfare-improving.[3]

2 Model

A society collectively decides to chooses a policy $\theta \in [0, 1]$. It is convenient to define the notation:

$$\begin{aligned} \alpha^+(\theta) &= \min\{\theta + \mu,\ 1\}, \\ \alpha_-(\theta) &= \max\{\theta - \mu,\ 0\}. \end{aligned}$$

for some $\mu > 0$. We will sometimes refer to $\alpha_-(\theta)$ and $\alpha^+(\theta)$ as the *predecessor* and the *successor* of θ, respectively.

The status quo is zero. The society consists of a unit mass of citizens, each of them privately informed about their type, which is some point $z \in [0, 1]$. It is convenient to refer to a citizen of type z as *citizen z.*

Citizen z's utility from policy θ is $u(z, \theta) = -(z - \theta)^2$.

For our results, it is only important that each citizen's preferences are single-peaked and symmetric in the sense that equal deviations from the peak on either side lead to the same utility loss.

There is uncertainty at both the individual and aggregate levels, which we model in the same way as in Britz and Gersbach (2019): That is, we assume that there are finitely many states of nature. We denote the state space by $N = \{1, \ldots, n\}$, and use k to index the elements of N. There is a family of probability distributions on Z associated with the states

[3]Our paper is also part of the broadening literature on learning in dynamic collective decisions. Strulovici (2010) examines how long a committee invests in learning until a majority takes a final decision. The duration of learning in committees with heterogeneous members is characterized in Chan et al. (2018).

of nature. We use f_k and F_k to denote the probability density function and cumulative distribution function, respectively, of the probability distribution associated to state k. In each state of nature, citizens' types are independent draws from the relevant probability distribution. Regardless of their type, citizens share a common prior belief p about the state of nature, where we assume that $p_k > 0$ for every $k \in N$. Due to Bayesian updating, citizen z has a posterior belief that assigns to state k the following probability:

$$\gamma_k(z) = \frac{f_k(z)p_k}{\sum_{j=1}^{n} f_j(z)p_j} > 0.$$

We make the following assumptions:

Assumption 1

1. *For any $z \in int(Z)$, we have $F_1(z) > \ldots > F_n(z)$.*
2. *For every $k \in N$ and every $z \in Z$, it holds that $\gamma_k(z) > 0$ and Bayesian updating is monotone.*[4]

3 The Proposal Assessment Voting Procedure

In this section, we give the formal description of the *Proposal Assessment Voting (PAV) procedure.*

An agenda–setter makes a proposal $\theta \in [0, 1]$. We will allow for two different cases: The agenda–setter may be a benevolent social planner who seeks to implement the Condorcet winner. Alternatively, he may also have his own interests, and be of a particular type denoted by θ_{AS}. The voting procedure consists of two stages:

1. Once the agenda–setter has made a proposal, say $\bar{\theta}$, a random sample of size λ of the population is drawn. Each member of the sample group may vote in favor of $\bar{\theta}$, or in favor of the status quo. The share of sample group members who vote in favor of $\bar{\theta}$

[4]Monotone Bayesian updating means that for any $z_1, z_2 \in Z$ with $z_1 < z_2$, the posterior probability distribution $\{\gamma_k(z_2)\}_{k=1}^{n}$ stochastically dominates $\{\gamma_k(z_1)\}_{k=1}^{n}$.

is denoted by δ. We define

$$\beta(\bar{\theta}, \delta) = \begin{cases} \alpha^+(\bar{\theta}) & \text{if } \delta \geq 1/2, \\ \alpha_-(\bar{\theta}) & \text{if } \delta < 1/2. \end{cases}$$

2. In the second stage, the entire population votes. Each voter is asked to submit his pairwise preferences over the three alternatives $\{0, \bar{\theta}, \beta(\bar{\theta}, \delta)\}$. The outcome is then determined as follows: If the majority pairwise prefers 0 to both $\bar{\theta}$ and $\beta(\bar{\theta}, \delta)$, then the outcome is the status quo 0. If the majority pairwise prefers $\bar{\theta}$ to 0, but 0 to $\beta(\bar{\theta}, \delta)$, then the outcome is $\bar{\theta}$. If the majority pairwise prefers $\beta(\bar{\theta}, \delta)$ to 0, but 0 to $\bar{\theta}$, then the outcome is $\beta(\bar{\theta}, \delta)$. If the majority pairwise prefers both $\bar{\theta}$ and $\beta(\bar{\theta}, \delta)$ to 0, then the outcome is either $\bar{\theta}$ or $\beta(\bar{\theta}, \delta)$, whichever is pairwise preferred by a majority to the other.

We assume that citizens with the same preference ranking over the alternatives $0, \bar{\theta}$, and $\beta(\bar{\theta}, \delta)$ coordinate their votes.

4 Sincere Voting by the Population

In this section, we establish the following claim: At the second stage of the PAV procedure, all citizens find it optimal to vote sincerely, that is, in accordance with their true preferences.

In order to show this claim, the following reformulation is convenient:

For any choices of $\bar{\theta} \in \Theta$ and $\delta \in [0, 1]$, we can always restate the problem of choosing from $\{0, \bar{\theta}, \beta(\bar{\theta}, \delta)\}$ as the problem of choosing from three alternatives x_0, x_1, and x_2 such that $0 = x_0 < x_1 < x_2$. Citizens cast three pairwise votes:

$$\begin{aligned} x_0 &\leftrightarrow x_1, \\ x_0 &\leftrightarrow x_2, \\ x_1 &\leftrightarrow x_2. \end{aligned}$$

The rules as described in the previous section can be restated as follows:

If any alternative wins two of the three pairwise votes, then it becomes the outcome of the voting procedure. If each of the three votes is won by a different alternative, then the alternative that has defeated x_0 becomes the outcome of the voting procedure.

In principle, there are six ways to rank–order the alternatives $\{x_0, x_1, x_2\}$ by a preference order $\succsim$. They are as follows:

$$x_2 \succsim x_1 \succsim x_0,$$
$$x_1 \succsim x_2 \succsim x_0,$$
$$x_1 \succsim x_0 \succsim x_2,$$
$$x_0 \succsim x_1 \succsim x_2,$$
$$x_0 \succsim x_2 \succsim x_1,$$
$$x_2 \succsim x_0 \succsim x_1.$$

Note that the last two preference orders above are inconsistent with our assumption on single–peaked preferences. Hence, we can restrict attention to the following four preference orders:

$$x_2 \succsim x_1 \succsim x_0,$$
$$x_1 \succsim x_2 \succsim x_0,$$
$$x_1 \succsim x_0 \succsim x_2,$$
$$x_0 \succsim x_1 \succsim x_2.$$

We assume that all citizens which share one of these four preference orders can coordinate their votes. Thus, we have to check for profitable deviations by each of the four groups, assuming that the remaining three groups vote sincerely.

As a first step, we show that citizens express their preference sincerely over the alternatives that differ from the status quo. This is the claim of the following proposition:

Proposition 1 *Citizens vote sincerely between $x_1 \leftrightarrow x_2$.*

Proof. **Case 1.** Suppose that x_0 wins against both x_1 and x_2. Then, the outcome of the voting procedure is x_0, regardless of the vote between $x_1 \leftrightarrow x_2$.

Case 2. Suppose that x_1 wins against x_0, but x_0 wins against x_2. Then, the outcome of the voting procedure is x_1, regardless of the vote between $x_1 \leftrightarrow x_2$.

Case 3. Suppose that x_2 wins against x_0, but x_0 wins against x_1. Then, the outcome of the voting procedure is x_2, regardless of the vote between $x_1 \leftrightarrow x_2$.

Case 4. Suppose that x_0 wins against neither x_1 nor x_2. Then, the outcome of the vote between $x_1 \leftrightarrow x_2$ selects the outcome of the whole voting procedure. In that case, it is optimal to vote sincerely.

□

Now we are ready to show that citizens vote sincerely.

Proposition 2 *In the second stage of the PAV procedure, there is an equilibrium in which all citizens vote sincerely.*

Proof.

1. Consider the group with preference order $x_2 \succsim x_1 \succsim x_0$. If all citizens vote sincerely, x_2 is the outcome of the voting procedure if and only if a majority of the population belongs to the group at hand. If the group with preference order $x_2 \succsim x_1 \succsim x_0$ is indeed a majority, they get their most preferred alternative by sincere voting. Hence, a deviation from sincere voting can only be profitable in states of nature where the group at hand is a minority. In that case, the alternative x_2 will never be implemented – this follows from the premise that the three other groups vote sincerely. Now we see that the group with preference order $x_2 \succsim x_1 \succsim x_0$ can only benefit from a deviation if (i) sincere voting would lead to the outcome x_0, and (ii) some strategic voting by the group at hand leads to the outcome x_1 instead. It follows from point (i) that, under sincere voting, x_0 wins the pairwise votes against both x_1 and x_2. But only the group with preference ranking $x_0 \succsim x_1 \succsim x_2$ sincerely prefers x_0 over

x_1. It follows that a majority of the population belongs to the group with preference ranking $x_0 \succsim x_1 \succsim x_2$. Due to the premise that this group votes sincerely, the outcome of the voting procedure is x_0, regardless of any votes by the group with preference order $x_2 \succsim x_1 \succsim x_0$. Hence, point (ii) is certainly violated. We conclude that the group with preference order $x_2 \succsim x_1 \succsim x_0$ cannot have a profitable deviation from sincere voting.

2. Consider the group with preference order $x_1 \succsim x_2 \succsim x_0$. Suppose first that a majority of citizens belong to the group with preference ranking $x_2 \succsim x_1 \succsim x_0$. Since that group votes sincerely, the outcome is going to be x_2, no matter what the group with preference order $x_1 \succsim x_2 \succsim x_0$ does. Indeed, the group under consideration can only profit from a deviation if a majority of the population prefers x_1 over x_2. Hence, the vote between $x_1 \leftrightarrow x_2$ is always won by x_1 and so the outcome of the vote $x_0 \leftrightarrow x_2$ is irrelevant for the outcome of the voting procedure. Indeed, x_1 is the outcome of the voting procedure if and only if it wins in the vote between $x_0 \leftrightarrow x_1$. If any strategic voting is beneficial for the group under consideration, then it must be because the vote $x_0 \leftrightarrow x_1$ is won by x_1 under strategic voting, but would be won by x_0 under sincere voting. But with sincere voting, x_1 wins against x_0. We have now shown that the group with preference order $x_1 \succsim x_2 \succsim x_0$ cannot gain by deviating from sincere voting.

3. The argument above also applies to the group with preference order $x_1 \succsim x_0 \succsim x_2$.

4. Finally, consider the group with preference order $x_0 \succsim x_1 \succsim x_2$. In a state of nature where the majority prefers x_2 to the other two alternatives, the votes of the group at hand are inconsequential. In states of nature where the group at hand is in the majority, they get their most preferred outcome from voting sincerely. Suppose that the group at hand is not in the majority, and a majority is also not in favor of x_2. Then, the outcome under sincere voting is x_1. The only way a deviation could benefit the group at hand is if they could change the outcome from x_1 to x_0. But this is impossible, because all other citizens vote sincerely for x_1 in the vote $x_1 \leftrightarrow x_0$.

□

This result is reminiscent of the well-known *Median Voter Theorem.* It is, however, not readily implied by the Median Voter Theorem since the agenda-setter is unaware of the location of the median voter. This is a consequence of the individual and aggregate uncertainties inherent in our model.

5 Benchmark: Proposal against Status Quo

The key implication of the above result is the following: Suppose that the alternative preferred by the median voter over all alternatives qualifies for the second stage of the PAV procedure. Then, it is certain that this alternative will also be the outcome of the entire procedure. Hence, the question is: How can we ensure that the alternatives present in the second round are as close as possible to the alternative preferred by the median voter?

The purpose of the PAV procedure is to perform better under this criterion than a benchmark scenario in which the agenda-setter uses only his prior belief to choose a proposal that is then voted upon.

We now briefly consider that benchmark scenario.

Proposition 3 *In any state k with median voter $\widehat{z}_k$, if the agenda-setter chooses a proposal $\theta < 2\widehat{z}_k$, then the proposal is accepted by the majority. Otherwise, the status quo prevails.*

The intuition is as follows: Due to our assumption of single-peaked preferences, there is for each citizen $z > 0$ some proposal $\theta > 0$ such that citizen z is indifferent between θ and the status quo. He strictly prefers any $\theta' \in (0, \theta)$ to either zero or θ, and he strictly prefers θ to any $\theta' > \theta$. Due to the quadratic functional form we have chosen for the utility function, this critical proposal θ equals $2z$ for each citizen z. Indeed $-(z - \theta)^2$ returns the same utility for $\theta = 0$ as for $\theta = 2z$.

6 Information Revelation through PAV

We take a proposal $\bar{\theta}$ as given, and verify under what conditions the PAV procedure reliably implements that choice from $\{0, \alpha_-(\bar{\theta}), \bar{\theta}, \alpha^+(\bar{\theta})\}$ which is closest to the median voter. This means that we assess social welfare based on the median voter's preference, which concurs with the approach in Britz and Gersbach (2019). In our earlier work, we argue that this makes sense when discussing democratic procedures: Indeed, a democratic procedure should satisfy a requirement such as "stability to majority voting," which is similar to assessing social welfare based on the median voter's preferences.

At the first stage of the voting procedure, we take as given the proposal $\bar{\theta}$ and define the following groups:

- $Z_3(\bar{\theta})$ are citizens with preference order $\alpha^+(\theta) \succsim \theta \succsim \alpha_-(\theta) \succsim 0$.
- $Z_2(\bar{\theta})$ are citizens who prefer $\theta \succsim \alpha_-(\theta) \succsim 0$ but also $\theta \succsim \alpha^+(\theta)$.
- $Z_1(\bar{\theta})$ are citizens who prefer $\alpha_-(\theta) \succsim \theta \succsim \alpha^+(\theta)$, but also $\alpha_-(\theta) \succsim 0$.
- $Z_0(\bar{\theta})$ are citizens with preference order $0 \succsim \alpha_-(\theta) \succsim \theta \succsim \alpha^+(\theta)$.

In what follows, we omit the argument $\bar{\theta}$. Note that the four groups are mutually exclusive and exhaustive.

We consider the following equilibrium candidate: At the first stage of the voting procedure, $Z_0 \cup Z_1$ vote No and $Z_2 \cup Z_3$ vote Yes.

We will be interested in the issue of *manipulation.* Our notion of manipulation is that one of the four groups as defined above has an incentive to deviate from the aforementioned equilibrium candidate. Indeed, we are going to claim that Z_1, Z_2, and Z_3 have no incentive to make such a deviation, and we examine the conditions under which Z_0 may have an incentive to deviate.

It is an important feature of our result that we do not need to assume anything about the beliefs of the group members about the underlying state of nature.

We do assume that voting behavior within each of the four groups is coordinated. Note that this is a conservative assumption that biases our results in favor of manipulation, and therefore against the benefits of proposal assessment voting. Of course, obtaining equilibria without manipulation would be trivial if they would have to be robust only against individual citizens' deviations, and it would at least be easier if we allowed only smaller groups of individuals to coordinate their votes. Allowing groups of citizens to coordinate their deviation is an approach which also concurs with the insights of a recent paper by Bierbrauer and Hellwig (2016).

The argument now runs as follows: We consider the four groups defined above in turn, in descending order. We need to check for various triples of a state of nature, a proposal, and a group whether manipulations are possible. In order to appreciate the gist of the argument, note that for many such triples, the absence of manipulation is trivial: For instance, a group can never gain from manipulation if it constitutes a majority by itself. In that case, it can simply be sincere and enforce its preferred alternative by virtue of its majority. Moreover, a group can also not manipulate the process if some other group constitutes a majority – they simply do not have the power to do so. Repeating these considerations will allow us to show that the scope for any manipulation is quite restricted. Only members of Z_0 can manipulate, and they can only do so in states in which several conditions on the relative size of the various groups are simultaneously satisfied. This will give us a set of necessary conditions for manipulation. Conversely, we will contain a set of conditions each of which is sufficient for non-manipulation. This result is formally stated in Theorem 1 below.

Theorem 1 *If the PAV procedure is manipulated, then its outcome is either the status quo or the proposal θ, whichever is preferred by a majority.*

Proof. Consider the strategy profile where groups Z_0 and Z_1 vote No and groups Z_2 and Z_3 vote Yes. We check whether any one group has a profitable deviation from that strategy profile. We show first that groups Z_3, Z_2, and Z_1 have no such deviation. Then, we show that group Z_0 may have an incentive to deviate, but then the outcome is either the status quo or the proposal, whichever the majority prefers.

1. Consider Z_3's voting decision in the first round. In states where Z_3 itself is a majority, Z_3 can have their most preferred option $\alpha^+(\theta)$ if and only if it qualifies for the second round. So voting Yes at the first stage is optimal for Z_3. We will show that Z_3's choice is inconsequential in any state of nature where Z_3 is not a majority. It is straightforward that Z_3's choice is inconsequential in those states where $Z_0 \cup Z_1$ have a majority. Now consider states where neither Z_3 nor $Z_0 \cup Z_1$ are a majority. Then, $\alpha^+(\theta)$ could never win against θ in the second round. But since $Z_2 \cup Z_3$ is a majority, neither zero nor $\alpha_-(\theta)$ could win against θ in the second round either. So θ is the outcome, regardless of what happens in the first voting round.

2. Consider Z_2's voting decision in the first round. In states where Z_2 has a majority, they can obtain their most preferred outcome θ in the second round, regardless of the outcome of the first round. Hence, their decision at the first stage is inconsequential. In states where $Z_0 \cup Z_1$ are the majority, Z_2's choice in the first round is again inconsequential. The same is true in states where Z_3 alone has a majority. Now it remains to consider those states where neither $Z_0 \cup Z_1$, nor Z_2, nor Z_3 are a majority. In such a state, the second stage of voting cannot be won by $\alpha^+(\theta)$ (because Z_3 is a minority), cannot be won by $\alpha_-(\theta)$ (because $Z_0 \cup Z_1$ is a minority), and cannot be won by zero because the majority $Z_2 \cup Z_3$ which prefers θ to α_- also prefers θ to zero. Hence, the outcome is θ regardless of the votes cast in the first stage. Whatever the state, Z_2 is indifferent between voting Yes or No in the first stage. (Intuition: In those states where Z_2 is powerful enough to change the outcome of the first stage, they are also powerful enough to ensure that the outcome of the second stage is θ, which is their preferred option, and which is in the race anyway!)

3. Consider Z_1's voting decision in the first round. It is inconsequential in all states in which either $Z_2 \cup Z_3$ or Z_0 is a majority. Indeed, consider the remaining states. Suppose that Z_1 votes No. Then $\alpha_-(\theta)$ qualifies for the second round. Since $Z_2 \cup Z_3$ is a minority, θ cannot win the second round. Since Z_0 is also a minority, zero cannot win either. So the outcome is $\alpha_-(\theta)$. Now suppose Z_1 switches from No to Yes. Then

$\alpha^+(\theta)$ qualifies for the second round. Again because $Z_2 \cup Z_3$ is a minority, $\alpha^+(\theta)$ cannot win in the second round, thus the outcome is either zero or θ. But Z_1 likes $\alpha_-(\theta)$ better than zero and better than θ. Hence, whatever the state, Z_1 cannot gain from the deviation.

4. Consider Z_0's voting decision at the first stage. It is inconsequential in all states in which $Z_2 \cup Z_3$ is a majority, or in which Z_1 is a majority. In states where Z_0 by itself is a majority, the outcome of the second round is always zero, no matter what happens in the first round – again, Z_0's decision in the first round is inconsequential. Now consider the remaining states, in which neither $Z_2 \cup Z_3$, nor Z_0, nor Z_1 are a majority. Suppose first that Z_0 votes No. Then, because $Z_2 \cup Z_3$ is a minority, $\alpha_-(\theta)$ qualifies for the second round. Again because $Z_2 \cup Z_3$ is a minority, θ cannot win the second round. Since Z_0 is also a minority, zero cannot win either. So the outcome is $\alpha_-(\theta)$. Now suppose Z_0 switches from No to Yes. Then $\alpha^+(\theta)$ qualifies for the second round. Again because $Z_2 \cup Z_3$ is a minority, $\alpha^+(\theta)$ cannot win in the second round, so the outcome must be either zero or θ, whichever is preferred by a majority.

Recall from the previous section that we are comparing the PAV procedure to a benchmark procedure in which citizens simply choose between the proposal and the status quo without any preliminary proposal assessment. The key implication of Theorem 1 is that, whenever the PAV procedure is manipulated, its outcome coincides with the outcome of the benchmark procedure, so that welfare does not change. If the PAV procedure is not manipulated, it may still be the case that the outcome of PAV coincides with that of the benchmark procedure. Again, welfare does not change. Finally, there is a case where the PAV procedure is not manipulated, and yet leads to a different outcome than the benchmark procedure. In that case, the outcome under PAV is a welfare improvement in the sense that this outcome is preferred by a majority to that of the benchmark procedure. Hence, Theorem 1 implies the following corollary.

Corollary 1 *Moving from the benchmark procedure to the PAV procedure, while holding the proposal constant, never leads to a welfare loss, regardless of the state.*

It is important to note that Theorem 1 and its corollary hold regardless of the prior or posterior beliefs held by any of the citizens.

Now let us consider the beliefs of group Z_0. The proof of Theorem 1 implies that a deviation by group Z_0 can only be profitable if the state of nature is such that all of the following conditions are simultaneously satisfied:

1. $Z_0 \cup Z_1$ constitutes a majority
2. Z_0 alone does not constitute a majority.
3. Z_1 alone does not constitute a majority.
4. A majority prefers the status quo to θ.

We note that the first condition above is implied by the fourth condition, hence we are left with three necessary conditions for manipulation.

Consider the following three inequalities:

$$F_k\left(\theta - \frac{1}{2}\mu\right) - F_k\left(\frac{1}{2}\theta - \frac{1}{2}\mu\right) \geq \frac{1}{2} \quad (1)$$

$$\theta \leq 2\widehat{z}_k, \quad (2)$$

$$\theta \geq 2\widehat{z}_k + \mu. \quad (3)$$

If the first inequality is satisfied, then Z_1 has a majority, in which case no manipulation is possible. If the second inequality is satisfied, then a majority prefers θ to the status quo – again, no manipulation is possible. If the , inequality is satisfied, then the majority belongs to Z_0, in which case manipulation is impossible. Hence, we have the following theorem.

Corollary 2 *Manipulation of the PAV procedure does not occur if group Z_0 believes that at least one of the inequalities (1)–(3) is satified.*

Due to our assumptions on the probability distribution functions, every state is believed to occur with strictly positive probability.

Let us now consider the case where $\theta + \mu \leq 2\widehat{z}_n$. This means that each of the three proposals α_-, θ, and $\alpha^+(\theta)$ are preferred by a majority to the status quo, and this is true in *each* state. Therefore, group Z_0 never finds it optimal to vote strategically, and hence, no manipulation is possible.

Corollary 3 *Suppose that $\theta + \mu \leq 2\widehat{z}_n$. Furthermore, suppose that there is some state $k \in N$ so that $\widehat{z}_k > \theta + \mu/2$. Then, moving from the benchmark procedure to the PAV procedure, while holding the proposal θ constant, leads to a welfare gain with strictly positive probability.*

The two necessary conditions in the corollary above boil down to a requirement that the proposal θ, which we are holding fixed here, should not be "too high." Later in the paper, we will argue that an agenda-setter, regardless of his motivation, would never have an incentive to make an excessively high proposal in the first place. Hence, the interpretation of the corollary is that switching from the benchmark procedure to the PAV procedure can indeed be expected to lead to a welfare gain.

7 Limit Results on Welfare Gains

The results in the previous section reveal that no manipulation can occur if any one of the four groups alone has a majority. It is important to note that the model introduced here generalizes some of the frameworks that exist in the literature as limit cases.

7.1 Homogeneous preferences

Suppose we modify citizen's types as follows: In state k, each citizen's type is determined by drawing some $z \in [0, 1]$ from the distribution f_k, and the type is then equal to

$$z' = pz + (1 - p)\widehat{z}_k,$$

where $\widehat{z}_k$ is defined as the median voter in state k, that is $F_k(\widehat{z}_k) = 1/2$. For any $p \in [0, 1]$, the type $\widehat{z}_k$ of the median voter is the same. However, the degree to which citizens' preferences differ from each other is scaled by p. If the parameter p is close to zero, citizens have nearly homogeneous preferences. In that case, for a generic choice of the probability distributions and the concomitant median voter types, a majority of citizens will have the same ranking over the proposals and the status quo at the second stage of the PAV procedure. This is intuitive: For small values of p, there is little conflict of interest among citizens. Hence, the problem reduces to an election in which all citizens agree to implement the best alternative but have noisy private information about which alternative is indeed "best." (References Voting Literature) One attractive feature of the PAV model is that the case of homogeneous preferences can be obtained as a special case. It helps embed the idea of PAV in the voting literature.

7.2 Small Steps Between Proposals

Another special case that helps the intuitive understanding of the model is to consider a sufficiently small value of the parameter μ. Verbally, this means that the three alternatives that may go against the status quo are close to each other. When this is the case, then in each state, either the group Z_0 or the group Z_3 constitutes a majority. Therefore, no manipulation can occur. At the same time, in the limit as $\mu \to 0$, the outcome of the PAV procedure must lie in an ever smaller neighborhood around the outcome of the benchmark procedure. The interpretation is as follows: The PAV procedure always allows for some gains in social welfare compared to the benchmark scenario. How large these gains are depends on the choice of the model parameters, and, in particular, the probability distribution functions. From a practical point of view, this is a question of institutional design: Depending on the distribution of the underlying preferences, one can choose the parameter μ in such a way that welfare gains are realized in expectation. One alternative that could also be explored is that the agenda-setter himself chooses μ.

8 Agenda-setter's Motivation

8.1 Welfare gain with same proposal

The analysis so far has led to some insights about the benefits of the PAV procedure relative to the benchmark procedure, under the premise that the proposal made under both procedures is the same. In a nutshell, we have found the following: In expectation, social welfare increases with the PAV procedure compared to the benchmark. The size of the welfare gain depends on μ and goes to zero as μ goes to zero.

After the analysis so far, a natural follow-up question is this: Suppose that the PAV procedure is used, and the agenda-setter anticipates the conditions for its manipulability. Would the agenda-setter then want to make the same proposal as in the benchmark procedure? If he does so, we have shown that a welfare gain can be realized. But what happens if the agenda-setter re-optimizes even his original proposal in anticipation of proposal assessment?

8.2 Welfare change with a neighboring proposal

Suppose that an agenda-setter, whatever his preferences, would make the proposal θ_0 under the benchmark procedure. If the proposal he would make under the PAV procedure belongs to the set $\{\theta_0 - \mu,\ \theta_0,\ \theta_0 + \mu\}$, then social welfare under the PAV procedure is not less than under the benchmark procedure.

8.3 Benevolent Agenda-Setter

Theorem 2 *With a benevolent agenda–setter, social welfare is higher in expectation with PAV than without it.*

The agenda-setter has the option of sticking to the same proposal as under the benchmark procedure, which already implies the above statement. In addition, an agenda-setter could use re-optimization of the proposal in order to improve expected welfare gains even further.

Which proposal is optimal, however, always depends on the underlying model parameters, and in particular on the probability distributions from which the types are drawn.

8.4 Selfish Agenda-Setter

Another important case is where the agenda-setter is selfish. One natural case to think about is the one where the agenda-setter wants to move as far as possible away from the status quo. This case can be interpreted as follows:

Consider once more the example in the introduction where a small but well-organized minority proposes a referendum on an extreme policy plan, which it may want to moderate in order to gain popular support. In the policy space $\theta \in [0, 1]$, we can think of 0 as "no policy change" and think of 1 as the extreme policy most preferred by the proposers. Then, the interval $(0, 1)$ is the space of all the possible compromises or moderate versions of that extreme policy. In view of this interpretation, it is natural to think of a selfish agenda-setter as having a preference for the extreme point of the policy space.

More formally, in our analysis of the *selfish agenda-setter,* we assume that this agenda-setter's utility from any policy increases in that policy's proximity to an ideal point, which itself is greater than the median voter's type in any state of nature.

Proposition 4 *A selfish agenda–setter is better off in expectation under PAV than without it.*

Note: Manipulation can only occur if the agenda–setter makes a proposal such that the majority would rather remain at zero than accept the proposal. But if that is true, nothing higher than the proposal will ever go through. Hence, if the agenda–setter is selfish, and sticks to the same proposal as in the benchmark scenario, he can never become any worse off.

The question is whether moving from the benchmark procedure to the PAV procedure also leads to social welfare gains even if the agenda-setter is selfish. In order to show that this is true, we have to demonstrate that, under the PAV procedure, the selfish agenda-setter wants to choose a proposal which is either the same proposal as in the benchmark procedure, or a proposal that is at a distance of μ from the benchmark proposal. This is the claim of the following proposition.

Proposition 5 *Under the PAV procedure, the selfish agenda-setter makes a proposal from the set* $\{\theta_0 - \mu, \theta_0, \theta_0 + \mu\}$.

In order to see why this is true, the following considerations are needed: Under the benchmark procedure, the agenda-setter choses proposal θ_0 as follows: For each $\theta \in [0, 1]$, he has a belief $\pi(\theta)$ indicating the probability with which a majority prefers θ over the status quo. Then, the agenda-setter weighs the benefits of a policy θ against the probability that it will be the outcome of the procedure. That is, he chooses

$$\theta_0 \in \arg\max_{\theta\in[0,1]} \; -\pi(\theta)(1-\theta)^2 + (1-\pi(\theta)).$$

This optimization problem can be simplified to

$$\theta_0 \in \arg\min_{\theta\in[0,1]} \; \pi(\theta)(\theta^2 - 2\theta + 2).$$

Define $\lambda(\theta) = -\pi(\theta)(1-\theta)^2 + (1-\pi(\theta))$.

Now let us suppose that, under the PAV procedure, the agenda-setter makes a proposal $\widetilde{\theta}$. Then, the maximum utility that he can obtain is $\max\{\lambda(\widetilde{\theta} - \mu),\ \lambda(\widetilde{\theta}),\ \lambda(\widetilde{\theta} + \mu)\}$. Now suppose that $\arg\max_{\theta\in[0,1]} \; -\pi(\theta)(1-\theta)^2 + (1-\pi(\theta))$ is singleton, and $\theta_0 \notin \{\lambda(\widetilde{\theta} - \mu),\ \lambda(\widetilde{\theta}),\ \lambda(\widetilde{\theta} + \mu)\}$. Then, the agenda-setter would be better off proposing θ_0.

So far, we have considered the case where the optimal proposal for the selfish agenda-setter under the benchmark procedure is unique. In principle, however, it is possible that $\arg\max_{\theta\in[0,1]} \; -\pi(\theta)(1-\theta)^2 + (1-\pi(\theta))$ is not singleton. Then, the above argument can be generalized to show the following: The selfish agenda-setter's proposal under the PAV procedure is either a proposal that he could optimally make under the benchmark procedure, or it is a proposal that is at a distance of μ from a proposal that he could optimally make under the benchmark procedure.

Hence, for any equilibrium outcome under the benchmark procedure, there is an equilibrium outcome of the PAV procedure which generates at least as much social welfare.

Alternatively, this potential multiplicity of outcomes could be excluded by assuming that (i) the selfish agenda-setter chooses a proposal based on his prior beliefs and (ii) these prior beliefs are such that the optimal proposal is uniquely determined.

Either way, the key insight of our analysis is that the PAV procedure leads to welfare gains in expectation even when the agenda-setter is selfish.

9 Illustration with an example

The simplest way to illustrate the main point of the paper is with the following stylized example.

Suppose that there are two states of nature, which we call *"high"* and *"low."*

Suppose that in the high state, a majority would prefer any proposal $\theta \in (0, 1]$ to the status quo. This implies that $y^*_{high} = 0.5$ equals is the Condorcet winner in the high state because it solves the equality $(1 - y)^2 = y^2$. Similarly, in the low state, a majority prefers a proposal θ to the status quo if and only if $\theta \in (0, 0.5]$. This implies that $y^*_{low} = 0.25$ is the Condorcet winner in the low state because it solves the equality $(0.5 - y)^2 = y^2$.

Consider first the benchmark scenario: The agenda-setter makes a proposal, and this proposal is then pitted against the status quo. A benevolent agenda-setter would choose a proposal θ to maximize expected utility

$$-0.6(\theta - 0.25)^2 - 0.4(\theta - 0.5)^2.$$

This yields $\theta_0 = 0.35$. This proposal would be accepted regardless of the state. Let us suppose that $\mu = 0.05$. Then, the benevolent agenda-setter could stick to the proposal θ_0. In the high state, the outcome would be $\theta_0 + \mu = 0.4$, while in the low state it would be $\theta_0 - \mu = 0.3$: Clearly, social welfare improves.

While this welfare improvement is a good outcome, it hinges on the benevolence of the agenda-setter. Hence, we want to make sure that PAV does not lead to a decrease in social welfare if the agenda-setter is selfish. Indeed, let us suppose for the sake of this example that the agenda-setter is selfish and his bliss point is 1. Due to the monotonicity of the agenda-setter's preferences over $\theta \in [0, 1]$, it can only be optimal for him to propose either 1 or 0.5. Suppose that his prior belief is that the low state obtains with probability $p = 0.6$ and the high state obtains with complementary probability $1 - p = 0.4$.

If the agenda-setter proposes 1, then with probability 0.4, he obtains his bliss point and thus a payoff of zero, while with probability 0.6, the status quo persists, which means a payoff of -1 for the agenda-setter: Indeed, his expected payoff is -0.6. On the other hand, if he proposes 0.5, he receives the payoff of -0.25 for sure. Clearly, the agenda-setter chooses to propose 0.5. Since the Condorcet winner would be 0.25 with probability 0.6 and 0.5 with probability 0.4 we can take as a measure for social welfare: $-0.6 \times (0.5 - 0.25)^2 = -0.6 \times 0.25^2 = -0.0375$.

Suppose first that the agenda-setter decides to make some proposal $\theta' > 0.5 + \mu$. In that case, the status quo prevails if the state is low. Hence, expected payoff is bounded above by -0.6. Now suppose the agenda-setter proposes $0.5 + \mu$. In that case, 0.5 prevails whatever the state is. This gives the agenda-setter the sure payoff of -0.25. Thus, proposing $0.5 + \mu$ dominates any higher proposal. The agenda-setter has no possibility to exploit PAV to achieve a higher outcome than the one that prevails without PAV.

10 Conclusion

We have provided a first analysis of voting procedures involving proposal assessment. On a qualitative level, we find that a voting procedure with proposal assessment leads to social welfare gains in expectation. A particularly attractive aspect of this result is that it holds regardless of what one assumes about the agenda-setter's self-interest. Moreover, the results are independent of any assumptions about the posterior beliefs of citizens at any point during the democratic process. On a quantitative level, the size of the expected welfare gains from proposal assessment depend on the parameters of the model, such as the underlying distribution of preferences.

The paper links with the voting literature in general, and also with an emerging strand of literature on new forms of democracy.

The insights in this paper further enhance our understanding of the democratic process and, in particular, of the challenges associated with direct democracy. For instance, the present paper could help provide a theoretical foundation for the concept of a "counter-proposal" within the Swiss system of direct democracy. When an extreme policy proposal to change the constitution is put to a popular vote, the Swiss parliament has the right to design a "counter-proposal" which becomes may end up as the change of the constitution if it wins against the status quo and the original proposal. These counter-proposals are typically used to offer the public an opportunity for moderate changes.

References

Austen-Smith, D. and J. Banks (1996), Information Aggregation, Rationality, and the Condorcet Jury Theorem, *American Political Science Review, 90* pp.34-45

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

Britz, V. and H. Gersbach (2019), Information Sharing in Democratic Mechanisms, *International Journal of Game Theory,* (forthcoming).

Feddersen, T. and W. Pesendorfer (1997), Voting Behavior and Information Aggregation in Elections with Private Information, *Econometrica, 64,* 1029-1058.

Gersbach, H. (1993), Environmental Preservation and Majority Decisions, *Land Economics, 69,* 147-155.

Gersbach, H., Tejada, O. and A. Mamageishvili (2017), Assessment Voting in Large Electorates, ETH Working Paper 17/284, https://doi.org/10.3929/ethz-b-000221605

Callander, S. (2008), Majority Rule when Voters Like to Win, *Games and Economic Behavior, 64,* 393-420.

Chan, J., A. Lizzeri, W. Suen, and L.Yariv (2018), Deliberating Collective Decisions, *Revue of Economic Studies, 85,* 929-963.

Rüli, L. and T. Adler (2015), Die Volksinitiative, Avenir Suisse Diskussionspapier.

Strulovici, N. (2010), Learning While Voting: Determinants of Collective Experimentation, *Econometrica, 78,* 933-971.