Social Mechanism Design: Making Maximally Acceptable Decisions

Ben Abramowitz and Nicholas Mattei

Abstract

Sometimes agents care not only about the outcomes of collective decisions but also about *how* decisions are made. Both the outcome and the procedure affect whether agents see a decision as legitimate or acceptable. We focus on incorporating agents' preferences over decision-making processes into the process itself. Taking whole decisions, including decision rules and outcomes, to be the object of agent preferences rather than only decision outcomes, we (1) identify natural, plausible preference structures and key properties, (2) develop general mechanisms for aggregating these preferences to maximize the acceptability of decisions, and (3) analyze the performance of our acceptance-maximizing mechanisms. We apply our general approach to the setting of dichotomous choice, and compare the worst-case rates of acceptance achievable among populations of agents of different types. We include the special case of rule selection, or amendment, and show that amendment procedures proposed by Abramowitz et al. [2] achieve universal acceptance with certain agent types.

1 Introduction

In the literature on collective decision-making, a *mechanism* is "a specification of how economic decisions are determined as a function of the information that is known by the individuals in the economy" [20]. Social Choice studies the aggregation of preference information in particular. Kenneth Arrow pioneered the axiomatic approach to Social Choice, studying the properties of choice rules in terms of how they map various profiles of agent preferences to singular outcomes [4]. Our motivating observation is that the agents involved in collective decisions might care about the same properties that any mechanism designer or Social Choice theorist might. After all, if agents didn't generally care about such things, there would be little value in studying Social Choice.

Recently, Procaccia [24] proposed that "the central role of axioms should be to help explain the mechanism's outcomes to participants." The main reason given is that "explanations make the solutions more appealing to users, and, therefore, make it more likely that users will **accept** them." Automated procedures have since been developed for providing agents with axiomatic justifications for decisions which consist of an explanation and a set of axioms that serve as a "normative basis" [8].

The generation of axiomatic explanations and justifications faces several issues. The impossibility theorems endemic to Social Choice show that for some sets of axioms no rule can satisfy all of them, and therefore trade-offs must be made. Conflict arises if the agents do not all share a common normative basis of axioms from which acceptable justifications can be generated. Rather than finding one justification that satisfies as many agents as possible, we would like to provide as many agents as possible with a personalized justification they accept, which may differ from those given to other agents. After all, part of what makes collective decision making possible in practice is that different people can accept the same decision for different reasons; otherwise contracts would not exist. Another problem is that the explanations generated in this process may not be proper causal explanations [14]. This is because they may be based on erroneous counterfactuals about what would have happened had the circumstances been different. Therefore, our approach must address how counterfactuals are treated by the agents to ensure they are not being deceived.

Our starting point is that if we know what justifications each agent will accept we can turn the problem on its head and compute decisions justifiable to the maximum number of agents. We model agent preferences over decisions as consisting of (1) preferences over rules and outcomes, and (2) treatment of counterfactuals regarding what would have happened had the preference profile been different. We identify natural, intuitive preference structures built on these properties and show how to aggregate such preferences to make acceptance-maximizing decisions. Whether agents report their preferences in terms of rules or axioms matters in practice, but for our limited purposes they are theoretically equivalent.

We introduce the term *Social Mechanism Design* to unify the literature that seeks to incorporate agents' views on decision-making into the decision-making process itself. Social Mechanism Design recognizes that every choice made in the course of mechanism design is fundamentally a social choice about which reasonable people may disagree.

Contributions Our primary contribution is the identification of several natural, plausible, compact structures for agent preferences over combinations of rules and outcomes. For each preference type we show how to compute acceptance maximizing decisions in the general case and evaluate the worst-case acceptance rate for each type for different classes of reality-aware decision problems. We apply our general framework to asymmetric dichotomous choice problems, in which a proposal to be voted upon is put up against the status quo. We then extend our analysis to the procedures for rule updates or *amendments* proposed by Abramowitz et al. [2], where each amendment is an asymmetric dichotomous choice.

2 General Model

A set of agents N with $|N| = n$ must make a collective decision. The decision consists of selecting a single outcome y from some set of conceivable outcomes $\bar{\mathcal{Y}}$. Each agent $i \in N$ provides some information v_i, and these inputs collectively constitute a *profile* $V = (v_1, \dots, v_n)$. The decision will consist of applying a function, or *rule*, R to the profile to yield an outcome: $R(V) = y$. The set of all conceivable rules is denoted $\bar{\mathcal{R}}$. Rules are resolute in that they return an outcome on every possible profile.[1] We denote a *decision* by the tuple (V, R, y) where $R(V) = y$. If $R(V) \neq y$, then (V, R, y) is not a decision. The notation (V, R, y) is redundant since R and V imply a unique outcome $R(V)$, but the more verbose notation simplifies our presentation.

There is a subset of outcomes $\mathcal{Y} \subseteq \bar{\mathcal{Y}}$ that are said to be *feasible* in each instance. Similarly, only a subset $\mathcal{R} \subseteq \bar{\mathcal{R}}$ of rules are feasible. A decision (V, R, y) is feasible if and only if its rule and outcome are each feasible; $R \in \mathcal{R}$ and $y \in \mathcal{Y}$. The determinants of feasibility will be peculiar to the problem instance. In all instances our agents must collectively make a feasible decision, but we do not assume our agents necessarily know which rules, outcomes, and decisions are feasible. Neither do we assume that they have full knowledge of $\bar{\mathcal{Y}}$ or $\bar{\mathcal{R}}$.[2]

Preferences Over Decisions In our model, agents have preferences over decisions which are not captured by the profile V. We assume that each agent has binary preferences over decisions – they either accept or do not accept any decision. We denote the set of all decisions, including infeasible decisions, by $\mathbb{D}$. For $i \in N$, let $\mathcal{D}_i \subseteq \mathbb{D}$ be the set of all

[1]This can be achieved by defining them to have a default value for all profiles outside a certain set. An example of a default would be to maintain the status quo.

[2]See Appendix A for additional notes on the features of our general model.

decisions that agent i would find acceptable. We refer to $\mathcal{D}_i$ as the *satisfying set* of agent i and let $\mathcal{D}_N = (\mathcal{D}_1, \ldots, \mathcal{D}_n)$ be the collection of agents' satisfying sets.

Problem Instances An instance is characterized by a tuple $I = (V, \mathcal{R}, \mathcal{Y}, \mathcal{D}_N)$.[3] We make a single assumption that precludes indeterminacy: For every instance $I = (V, \mathcal{R}, \mathcal{Y}, \mathcal{D}_N)$, $\exists R \in \mathcal{R}$ such that $R(V) \in \mathcal{Y}$. That is, for all instances we consider, at least one feasible decision exists.

Acceptance Maximization Given any problem instance with profile V, feasible rules $\mathcal{R}$, feasible outcomes $\mathcal{Y}$, and collection of satisfying sets $\mathcal{D}_N$, how should we select a feasible decision (V, R, y)? Without making any further assumptions about the agents, profile, rules, outcomes, or satisfying sets, our answer is to try to select the feasible decision accepted by the greatest number of agents. If there are multiple acceptance-maximizing decisions, one can break ties arbitrarily (e.g., randomly). For any class of instances $\mathcal{I}$, we would like to specify a meta-rule or *mechanism* M that returns an acceptance-maximizing feasible decision for all instances in that class.

Let $\mathcal{F}^I \subseteq \mathbb{D}$ be the set of all feasible decisions in instance I. For each instance I, the acceptance-maximizing mechanism M essentially performs Approval Voting over the set of all feasible decisions $\mathcal{F}^I \subseteq \mathbb{D}$ based on the satisfying sets in $\mathcal{D}_N$, producing a decision $M(I) \in \mathcal{F}^I$.

$$M(I) = \operatorname*{arg\,max}_{(V,R,y)\in\mathcal{F}^I} |\{i \in N : (V, R, y) \in \mathcal{D}_i\}|$$

We are primarily concerned with the best worst-case acceptance rate achievable by an acceptance-maximizing mechanism M for a class of instances $\mathcal{I}$.

$$\alpha_{\mathcal{I}} := \min_{I\in\mathcal{I}} \frac{|\{i \in N : M(I) \in \mathcal{D}_i\}|}{n}$$

2.1 Agent Types

Agent types are characterized by the structure of their satisfying sets. These structures allow for (relatively) compact representation and correspond intuitively to the way we expect agents to express preferences over decisions in many settings.

Absolutism We begin by looking at absolutist agents. For an absolutist agent, if we know their satisfying set, we do not need to observe the profile to determine whether they accept a decision $(\_\_, R, y)$.

Definition 1 (Absolutism)**.** Agent $i \in N$ is absolutist if for all decisions (V, R, y), $(V', R, y) \in \mathbb{D}$, we have $(V, R, y) \in \mathcal{D}_i \Rightarrow (V', R, y) \in \mathcal{D}_i$.

Consequentialists only care about outcomes. All consequentialists are necessarily absolutist. If we know the satisfying set of a consequentialist agent, we can determine whether they accept a decision $(\_\_, \_\_, y)$ without observing the profile or rule. The complement to consequentialists, who only care about the outcome, are absolute proceduralists, who only care about the rule. For an absolute proceduralist we can determine from their satisfying set whether they accept a decision $(\_\_, R, \_\_)$ without observing the profile or outcome.

Definition 2 (Consequentialism)**.** Agent i is a consequentialist if there exists a set of outcomes Y_i such that for all $(V, R, y) \in \mathbb{D}$, $(V, R, y) \in \mathcal{D}_i$ if and only if $y \in Y_i$.

[3]Our notation excludes $\bar{\mathcal{R}}$ and $\bar{\mathcal{Y}}$ because the only relevant elements of these sets will be those reflected in $\mathcal{R}$, $\mathcal{Y}$, and $\mathcal{D}_N$.

Definition 3 (Absolute Proceduralism)**.** Agent i is an absolute proceduralist if there exists a set of rules R_i such that for all $(V, R, y) \in \mathbb{D}$, $(V, R, y) \in \mathcal{D}_i$ if and only if $R \in R_i$.

We also define two ways in which agents' acceptance of a decision may depend on both the rule and outcome: conjunctivism and disjunctivism.

Definition 4 (Absolute Disjunctivism)**.** Agent i is an absolute disjunctivist if there exists a set of rules R_i and set of outcomes Y_i such that $(V, R, y) \in \mathcal{D}_i$ if and only if $R \in R_i$ **or** $y \in Y_i$

Definition 5 (Absolute Conjunctivism)**.** Agent i is an absolute conjunctivist if there exists a set of rules R_i and set of outcomes Y_i such that $(V, R, y) \in \mathcal{D}_i$ if and only if $R \in R_i$ **and** $y \in Y_i$

Consequentialism and absolute proceduralism are both sub-types of absolute disjunctivism in which $R_i = \emptyset$ or $Y_i = \emptyset$, respectively. If two agents have the same sets R_i and/or Y_i for a given profile, but one is an absolute disjunctivist and the other is an absolute conjunctivist, the disjunctivist is intuitively easier to satisfy; they will accept a greater number of decisions. We neglect the unnatural case in which an agent's acceptance of a decision does not depend on the rule or outcome.

Implementation-Indifference In a decision (V, R, y) we will say that the rule R is *actually implemented* and that all other rules R' such that $R'(V) = y$ are *effectively implemented* by this decision because had R' be used instead the outcome would have been the same. Informally, we refer to an agent as *implementation-indifferent* if we can determine whether they accept a decision $(V, \_\_, y)$ without observing what rule was actually implemented.

Definition 6 (Implementation-Indifference)**.** Agent i is implementation-indifferent (II) if for all decisions $(V, R, y), (V, R', y) \in \mathbb{D}$, we have $(V, R, y) \in \mathcal{D}_i \Rightarrow (V, R', y) \in \mathcal{D}_i$.

Consequentialists are always implementation-indifferent, so there would be no difference between a consequentialist and an "II-Consequentialist."

Definition 7 (II-Proceduralism)**.** Agent i is an II-proceduralist if there exist a set of rules R_i such that for all decisions $(V, R, y) \in \mathbb{D}$, we have $(V, R, y) \in \mathcal{D}_i$ if and only if $\exists R' \in R_i$ such that $R'(V) = y$.

Definition 8 (II-Conjunctivism)**.** Agent i is an implementation-indifferent conjunctivist if there exist a set of rules R_i and set of outcomes Y_i such that for all decisions $(V, R, y) \in \mathbb{D}$, we have $(V, R, y) \in \mathcal{D}_i$ if and only if $y \in Y_i$ **and** $\exists R' \in R_i$ such that $R'(V) = y$.

Definition 9 (II-Disjunctivism)**.** Agent i is an implementation-indifferent disjunctivist if there exist a set of rules R_i and set of outcomes Y_i such that for all decisions $(V, R, y) \in \mathbb{D}$, we have $(V, R, y) \in \mathcal{D}_i$ if and only if $y \in Y_i$ **or** $\exists R' \in R_i$ such that $R'(V) = y$.

All satisfying sets we examine can be represented by a tuple $\mathcal{D}_i = \langle R_i, Y_i, \Psi_i, \Phi_i \rangle$, where Ψ_i is a binary variable that labels them as conjunctivist or disjunctivist, and Φ_i is a binary variable that denotes whether they are implementation-indifferent or absolutist.

Example Satisfying Sets Suppose a group of friends are deciding where to go for dinner between restaurants A, B, and C. They usually use Plurality voting - going to whichever restaurant is preferred by the most people. Here, acceptance means willingness to attend, and if they do not accept they will not join for dinner.

- Consequentialist: I will go to A or B, but not C.

- Proceduralist: I will only go if I get to choose the restaurant this time.
- Conjunctivist: We should use Plurality voting, but I will never go to restaurant B
- Disjunctivist: I will go wherever the most people want to go, but I am always happy going to restaurant A regardless of how we vote.
- II-Proceduralist: I will go wherever the plurality wants to go
- II-Conjunctivist: I will go anywhere except for restaurant A, as long as it's where most people want to go.
- II-Disjunctivist: I will go anywhere except for restaurant A, unless everyone else wants to go to A, then I'll go too.

2.2 Acceptance Maximization

We now construct a general algorithm for maximizing acceptance with absolutist and implementation-indifferent conjunctivist and disjunctivist agents. For brevity we assume here that R_i and Y_i are finite sets for all agents.

Mechanism 1: Acceptance Maximization for Absolute Disjunctivists

Input: $(V, \mathcal{R}, \mathcal{Y}, \mathcal{D}_N)$
  for $a \in \mathcal{Y}$ **do**
    $N^a \leftarrow \{i \in N : a \in Y_i\}$
    $R^a \leftarrow \underset{R \in \mathcal{R}: R(V)=a}{\arg\max} \; |\{i \in N \backslash N^a : R \in R_i\}|$
  end for
  $y = \underset{a \in \mathcal{Y}: R^a \in \mathcal{R}}{\arg\max} \; |N^a \cup \{i \notin N^a : R^a \in R_i\}|$
  return (V, R^y, y)

Theorem 1. Mechanism 1 maximizes acceptance for absolute disjunctivists.

Proof of Theorem 1. For each feasible outcome $a \in \mathcal{Y}$, Mechanism 1 finds the set of agents N^a for whom $a \in Y_i$ and finds the feasible rule R^a that yields this outcome on the given profile for which the maximum number of agents not in N^a have $R^a \in R_i$. For each outcome a it therefore maximizes the number of agents who are either in N^a or have $R^a \in R_i$. The rule and outcome R^y and y are then selected to be the ones that maximize the number of agents who accept the decision based on one of those two reasons. Outcomes for which no feasible rule is found to select them on the given profile are ignored as R^a does not exist. $\square$

Lemma 1. For any fixed profile V^*, for all agent types $\mathcal{D}_i = \langle R_i, Y_i, \Psi_i, \Phi_i \rangle$, there exists an absolute disjunctivist satisfying set $\tilde{\mathcal{D}}_i$ that accepts the exact same set of decisions with this profile: $(V^*, R, y) \in \mathcal{D}_i \Leftrightarrow (V^*, R, y) \in \tilde{\mathcal{D}}_i$.

Proof of Lemma 1. For II-disjunctivists, for every rule R in R_i we add the outcome $R(V)$ to their set Y_i. We can then set R_i to the empty set. When they are treated as an absolute disjunctivist the set of decisions they accept is the same because any outcome in Y_i was either in their original set Y_i or would have been selected by some rule in their original R_i on the given profile. For II-conjunctivists, we remove outcomes from Y_i if no rule $R \in R_i$ selects that outcome on the given profile. We can then set R_i to the empty set. When they are treated as an absolute disjunctivist the set of decisions they accept is the same because

Mechanism 2: Acceptance Maximization with All 7 Agent Types

```
Input: (V, 𝓡, 𝓨, 𝓓_N)
  for i ∈ N do
    Ỹ_i ← ∅
    R̃_i ← ∅
    if i is an II-disjunctivist then
      Ỹ_i ← Y_i
      for R ∈ R_i do
        Ỹ_i ← Ỹ_i ∪ R(V)
      end for
    else if i is II-conjunctivist then
      for R ∈ R_i do
        if R(V) ∈ Y_i then
          Ỹ_i ← Ỹ_i ∪ R(V)
        end if
      end for
    else if i is an absolute conjunctivist then
      for R ∈ R_i do
        if R(V) ∈ Y_i then
          R̃_i ← R̃_i ∪ R
        end if
      end for
    end if
    𝓓_i ← ⟨R̃_i, Ỹ_i, Ψ_i, Φ_i⟩
  end for
  return Mechanism 1(V, 𝓡, 𝓨, 𝓓_N)
```

the outcome is in the newly constructed set Y_i if and only if it was in the original set Y_i and a rule in the original R_i selects it on the given profile. Lastly, for absolute conjunctivists, we remove any rule R from R_i if it selects an outcome on the given profile $R(V)$ that is not in Y_i. We then set Y_i to be the empty set. When they are treated as an absolute disjunctivist the set of decisions they accept is the same because with Y_i empty they only accept a decision if the rule is in R_i, and the rule implemented is in the new set R_i if and only if it was in their original set R_i and the outcome it selects on V was in their original set Y_i. □

Theorem 2. Mechanism 2 maximizes acceptance with all agent types.

Proof of Theorem 2. Mechanism 2 takes the collection of satisfying sets, and for all agents $i \in N$ finds an absolute disjunctivist satisfying set that accepts the exact same set of decisions on the given profile V, in accordance with the proof of Lemma 1. The satisfying sets for all agents who are originally absolute disjunctivists remains the same. The mechanism then calls Mechanism 1 on the new collection of satisfying sets, which maximize acceptance among absolute disjunctivist agents. □

Theorem 3. If all agents are implementation-indifferent, and for each agent, for every profile, there exists at least one feasible decision they accept, then $\alpha_{\mathcal{I}} \geq \frac{1}{|\mathcal{Y}|}$.

Proof. When all agents are implementation-indifferent, we know whether they accept a decision $(V, \_, y)$ without observing the rule that was implemented. Since V is fixed, the number of possible unique tuples (V, y) is $|\mathcal{Y}|$. With n agents and $|\mathcal{Y}|$ feasible outcomes, for at least one outcome there must be at least $\frac{n}{|\mathcal{Y}|}$ agents who will accept any decision with that outcome by the pigeonhole principle. This is a fraction $\frac{1}{|\mathcal{Y}|}$ of the total number of agents. $\square$

We note that the type of an agent should properly be defined with respect to some class of instances. We drop this dependence in our notation for brevity. Next we will examine asymmetric dichotomous choice problems. Here we will make concrete assumptions about how the agents' satisfying sets are structured within this class and how their satisfying sets may correspond to the profile.

3 Asymmetric Dichotomous Choice

Model For all instances of asymmetric dichotomous choice problems, the feasible outcomes are $\mathcal{Y} = \{r, p\}$ where r is the status quo and p is a competing proposal. Each agent must express a vote $v_i \in \{r, p\}$, so the set of possible profiles is $\{r, p\}^n$ with n agents. We also stipulate that the set of feasible rules $\mathcal{R}$ is the set of supermajority rules for all instances. We define a supermajority rule, denoted R^δ where $\frac{1}{2} \leq \delta < 1$, as a rule that chooses p as the winning outcome if and only if strictly greater than δn agents vote for it; otherwise selecting r. For brevity, our definition of (reality-aware) supermajority rules includes majority rule. Here, $R^{\frac{1}{2}}$ is the majority rule which breaks ties in favor of the status quo, and $R^{1-\frac{1}{n}}$ is unanimity rule. There are exactly $|\mathcal{R}| = \lfloor \frac{n+1}{2} \rfloor$ distinct supermajority rules with n agents. To simplify notation we refer interchangeably to a rule R^δ and its threshold δ. We now look at acceptance in instances of asymmetric dichotomous choice problems with homogeneous agents of each type. The proofs of each result in Table 1 follow from counting arguments and can be found in Appendix B.1.

Agent Type	Assumptions $\forall i \in N$	$\alpha_{\mathcal{I}}$
Any	None	0
Absolute Conjunctivists	$v_i \in Y_i$ and $\lvert R_i \cap \mathcal{R}\rvert \geq 1$	0
Absolute Conjunctivists	$\lvert Y_i \cap \mathcal{Y}\rvert \geq 1$ and $\exists R \in R_i \cap \mathcal{R} : R(V) \in Y_i \cap \mathcal{Y}$	$2/n$
Absolute Disjunctivists	$\lvert R_i \cap \mathcal{R}\rvert \geq 1$	$2/n$
Absolute Disjunctivists	$\lvert R_i \cap \mathcal{R}\rvert \geq k$	$\frac{1}{n} \cdot \left\lceil \frac{nk}{\lfloor \frac{n+1}{2} \rfloor} \right\rceil$
Absolute Disjunctivists	$\lvert Y_i \cap \mathcal{Y}\rvert \geq 1$	$1/2$
II-Conjunctivists	$\lvert Y_i \cap \mathcal{Y}\rvert \geq 1$ and $\exists R \in R_i : R(V) \in Y_i \cap \mathcal{Y}$	$1/2$
II-Disjunctivists	$\exists R \in R_i : R(V) \in \mathcal{Y}$	$1/2$
II-Disjunctivists	$\lvert Y_i \cap \mathcal{Y}\rvert \geq 1$	$1/2$
II-Disjunctivists	$\lvert Y_i \cap \mathcal{Y}\rvert = 1$ and $\exists R \in R_i : R(V) \in \mathcal{Y} \setminus Y_i$	1

Table 1: Worst-case acceptance rates for asymmetric dichotomous choice problems when all agents are of the same type, given additional assumptions about satisfying sets.

Below we show the acceptance maximization algorithms for homogeneous sets of consequentialist, absolute disjunctivist, and II-disjunctivist agents for asymmetric dichotomous choice problems, and refer the reader to the appendix for the remaining cases and proofs. The mechanism that maximizes acceptance for the class of all asymmetric dichotomous choice instances with only consequentialists is given by Mechanism 3. We can already see that sometimes the outcome will go against the voter majority in V if this maximizes acceptance.

Mechanism 3: Acceptance Maximization for Asymmetric Dichotomous Choice with Consequentialists

Input: $(V, \mathcal{R}, \mathcal{Y}, \mathcal{D}_N)$
  $V_r \leftarrow |\{i \in N : v_i = r\}|$
  $V_p \leftarrow |\{i \in N : v_i = p\}|$
  $N_r \leftarrow |\{i \in N : r \in Y_i\}|$
  $N_p \leftarrow |\{i \in N : p \in Y_i\}|$
  if $V_p = |N|$ **then**
    return $(V, R^{1-\frac{1}{n}}, p)$
  else if $N_r \geq N_p$ **or** $V_r \geq V_p$ **then**
    return $(V, R^{1-\frac{1}{n}}, r)$
  else
    return $(V, R^{\frac{1}{2}}, p)$
  end if

Proposition 1. Mechanism 3 maximizes acceptance for all asymmetric dichotomous choice instances with consequentialist agents.

Proof. If the status quo r receives at least as many votes as the proposal p, then no supermajority rule will select p as the outcome so the only feasible decisions are those that select r, and the decision can be made according to any supermajority rule. If $N_r \geq N_p$, a feasible decision that maintains the status quo as the outcome will satisfy at least as many agents as a feasible decision that selects the proposal as the outcome. Unanimity rule will always select the status quo as the outcome unless the agents unanimously voted for the proposal. If $N_r \geq N_p$ but the agents unanimously voted for the proposal, then no feasible decision maintains the status quo. The outcome must be p, and the decision can be made by any supermajority rule. Lastly, if the proposal p receives more votes than the status quo and $N_p > N_r$, we maximize acceptance with a feasible decision that has p as the outcome. Majority rule selects p, so the decision $(V, R^{\frac{1}{2}}, p)$ is feasible and maximizes acceptance. $\square$

Consequentialist satisfying sets pay no attention to the rule, making the choice of rule in Mechanism 3 entirely *ad hoc.* Since the outcome is determined by the mechanism M with no regard for the rule R other than to ensure feasibility, the choice of rule feels superfluous. With absolute disjunctivists, the choice of rule matters both for the outcome it selects and for how many of the agents will accept any decision using that rule.

Absolute Disjunctivists Absolute disjunctivist satisfying sets generalize both the consequentialist and absolute proceduralist satisfying sets. Therefore, $\alpha_{\mathcal{I}}$ cannot be greater than the worst-case acceptance rate with only consequentialists or only absolute proceduralists, but it also cannot be worse than the minimum of these two. With absolute disjunctivists agents whose satisfying sets are non-empty ($|R_i| + |Y_i| > 0$), we have $\alpha_{\mathcal{I}} = \frac{2}{n}$.

Recall that agents may have infeasible decisions in their satisfying sets. For absolutists, only the feasible rules in their sets R_i were relevant for maximizing acceptance, but this is not the case for implementation-indifferent agents.

II-Disjunctivists To calculate an acceptance maximizing decision efficiently, we can take advantage of the fact that if $R^\delta(V) = p$ then $R^{\delta'}(V) = p$ for all $\delta' > \delta$ and if $R^\delta(V) = r$ then $R^{\delta'}(V) = r$ for all $\delta' < \delta$. In Mechanism 5 we use $\max(R_i)$ to denote the δ of the rule R^δ in the set R_i with the largest δ (i.e. the highest threshold to overturn the status quo), and similarly $\min(R_i)$ for the smallest δ.

Mechanism 4: Acceptance Maximization for Asymmetric Dichotomous Choice with Absolute Disjunctivists

Input: $(V, \mathcal{R}, \mathcal{Y}, \mathcal{D}_N)$
  $\delta_r = \frac{|\{i \in N : v_i = p\}|}{|N|}$
  $N_r \leftarrow \{i \in N : r \in Y_i\}$
  $N_p \leftarrow \{i \in N : p \in Y_i\}$
  $R_r^* \leftarrow \underset{R^\delta \in \mathcal{R}: \delta \geq \delta_r}{\arg\max} \; |\{i \in N \backslash N_r : R^\delta \in R_i\}|$
  $R_p^* \leftarrow \underset{R^\delta \in \mathcal{R}: \delta < \delta_r}{\arg\max} \; |\{i \in N \backslash N_p : R^\delta \in R_i\}|$
  $N_r \leftarrow N_r \cup \{i \in N : R_r^* \in R_i\}$
  $N_p \leftarrow N_p \cup \{i \in N : R_p^* \in R_i\}$
  if $|N_r| \geq |N_p|$ **then**
    return (V, R_r^*, r)
  else
    return (V, R_p^*, p)
  end if

Mechanism 5: Acceptance Maximization Among II-Disjunctivists for Asymmetric Dichotomous Choices

Input: $(V, \mathcal{R}, \mathcal{Y}, \mathcal{D}_N)$
  $\delta_r = \frac{|\{i \in N : v_i = p\}|}{|N|}$
  $\delta_p = \frac{|\{i \in N : v_i = p\}| - 1}{|N|}$
  $N_r \leftarrow \{i \in N : r \in Y_i \;\texttt{or}\; \max(R_i) \geq \delta_r\}$
  $N_p \leftarrow \{i \in N : p \in Y_i \;\texttt{or}\; \min(R_i) \leq \delta_p\}$
  if $|N_r| \geq |N_p|$ **then**
    return (V, R^{δ_r}, r)
  else
    return (V, R^{δ_p}, p)
  end if

4 Supermajority Rule Selection (Amendment)

Model Let $\mathcal{R}$ be the set of supermajority rules, including majority, with n agents. To simplify notation we refer interchangeably to a rule R^δ and its threshold δ. For each amendment problem, there is a supermajority rule $r \in \mathcal{R}$ currently in use. The agents are to decide whether or not to change the rule to be $p \in \mathcal{R}$, so $\mathcal{Y} = \{r, p\}$. Each agent casts a vote $v_i \in \{r, p\}$ forming collective profile by V_{rp}. An amendment is a special case of an asymmetric dichotomous choice where the potential outcomes are also rules which could be used to decide on the amendment itself. Since the votes v_i are rules, we assume a correspondence between the profile and the satisfying sets of the agents. Each agent $i \in N$ has a strict preference ordering $\succ_i$ over $\mathcal{R}$. We denote the most preferred rule at the top of i's ordering by δ_i, and assume that agent preferences are single-peaked over the numerical order of $\mathcal{R}$ with the peak at δ_i. In other words, for any two supermajority rules $\delta_1, \delta_2 \in \mathcal{R}$, if $\delta_1 < \delta_2 < \delta_i$ then $\delta_2 \succ_i \delta_1$ and if $\delta_i < \delta_1 < \delta_2$ then $\delta_1 \succ_i \delta_2$. Between any two rules, agents naturally vote for the one they prefer in V_{rp}. With any status quo r and proposed rule p, if $r \succ_i p$ then we assume $v_i = r$, and similarly, if $p \succ_i r$ then $v_i = p$. Following [1], our focus is on II-disjunctivists voting on amendments. We make two natural assumptions about how the profile corresponds to their satisfying sets: for all $i \in N$, $v_i \in Y_i$ and $\delta_i \in R_i$. In other words, an agent accepts the outcome of an amendment whenever their preferred outcome wins and whenever their ideal rule would have selected the same outcome on the profile V_{rp}.

Universal Acceptance One would hope that if the decision made for choosing a rule reaches universal acceptance then all future decisions using the chosen rule will be unanimously accepted on procedural grounds [13]. Since we have assumed our agents are II-disjunctivists, we can use Mechanism 5 to compute an asymmetric binary decision that maximizes acceptance. Unfortunately, for some instances there is no decision that achieves universal acceptance. Given status quo r, we want to know for what proposal p and amendment rule $R \in \mathcal{R}$ does the decision $(V_{rp}, R, R(V_{rp}))$ maximize acceptance. Amazingly, given any status quo $r < 1 - \frac{1}{n}$, there exists at least one proposal $p \in \mathcal{R}$, $p \neq r$, and amendment rule $R \in \mathcal{R}$, such that $(V_{rp}, R, R(V_{rp}))$ is guaranteed to achieve universal acceptance for any profile V_{rp} induced by agents' underlying single-peaked preferences over supermajority rules.

Lemma 2. The amendment decision $(V_{r,r+\frac{1}{n}}, R^r, R^r(V_{r,r+\frac{1}{n}}))$ is universally accepted by II-disjunctivists with single-peaked preferences over supermajority rules ordered on the real line such that $v_i \in \{r, p\}$, $v_i \in Y_i$, and $\delta_i \in R_i$ for all agents.

Proof. Let $p = r + \frac{1}{n}$. Suppose $R^r(V_{rp}) = r$. Then for all $i \in N$ such that $\delta_i \geq r$, $(V_{rp}, R^r, r) \in \mathcal{D}_i$ because R^{δ_i} is effectively implemented. For all $i \in N$ such that $\delta_i < r$, $r \in Y_i$. Now suppose $R^r(V_{rp}) = p$. Then for all $i \in N$ such that $\delta_i \leq r$, $(V_{rp}, R^r, p) \in \mathcal{D}_i$ because R^{δ_i} is effectively implemented. For all $i \in N$ such that $\delta_i > r$, $p \in Y_i$. □

Based on Lemma 2, we can build an iterative algorithm, Mechanism 6, for considering sequences of proposals to change the supermajority rule in incrementally increasing order such that each decision along the way is universally accepted. Mechanism 6 does not consider all possible proposals. It only considers those greater than r. When $r = \frac{1}{2}$ all other supermajority rules have the potential to be proposed. There are additional benefits to ensuring the original status quo be $\frac{1}{2}$, or at least that it not be too large. The following special supermajority rule is labeled h due to its resemblance to the h-index in bibliometrics:

$$h = \arg\max_{p \in \mathcal{R}} |\{i \in N : \delta_i \geq p\}| \geq np$$

Mechanism 6: Supermajority Rule Amendment

Input: $(V, \mathcal{R}, \mathcal{Y}, \mathcal{D}_N)$

```
p ← r + 1/n
while p < 1 do
    if R^r(V_rp) = p then
        r ← p
        p ← p + 1/n
    else
        return (V, R^r, R^r(V))
    end if
end while
```

Theorem 4. If $r \leq h$, then Mechanism 6 returns (V, R^h, h), otherwise it returns (V, R^r, r).

Proof. Assume $r < p \leq h$ where $p = r + \frac{1}{n}$. If $R^r(V_{rp}) = r$, this means that there are fewer than pn with peaks $\delta_i \geq p$, but this violates the definition of h, since there must be at least $hn > pn$ agents with peaks $\delta_i \geq h \geq p$. Thus, all amendments to increment the status quo will be successful if $r < h$. If $p > r \geq h$, there are at most rn agents with $\delta_i > r$ by the definition of h and all other agents must prefer $r \succ_i p$, so no proposal $p > r$ can possibly succeed as an amendment. □

One of the constitutions proposed by Abramowitz et al. [2], is derived from axioms that uniquely imply the initial supermajority rule when founding the constitution should be $r = \frac{1}{2}$, and then Mechanism 6 should be applied. Our analysis shows a different perspective; when agents are II-disjunctivists with votes and satisfying sets based on δ_i values, every amendment decision can be universally accepted under such a constitution.

Implementing Mechanism 6 with its iterative, incremental changes to the status quo is tedious. Do the agents need to observe every update to the status quo? Suppose we know all agents' ideal rules δ_i, and can therefore infer a partial ordering consistent with their preference ordering over $\mathcal{R}$ based on single-peakedness. Let this collection of partial orderings be the profile V. If we can compute h directly from V, and $r < h$, we could directly implement the decision (V, R, h) with the rule that always selects h. This can be achieved by implementing the amendment decision (V_{rh}, R^r, h). We know that any further proposals to amend h will fail, and the decision to maintain h against any proposal will be universally accepted. This more efficient algorithm corresponds to the second amendment procedure proposed by Abramowitz et al. [2], derived from a similar set of axioms to the first. What happens to the acceptance rate for the amendment decision (V_{rh}, R, h)?

Theorem 5. If the status quo is $r < h$, the amendment (V_{rh}, R^r, h) is universally accepted by II-disjunctivists with $v_i \in Y_i$ and $\delta_i \in R_i$.

Proof. Assume $r < h$. Any agent who prefers $h \succ_i r$ will accept the decision (V_{rh}, R^r, h) because $h \in Y_i$. For all $i \in N$ such that $r \succ_i h$, we know that $\delta_i < h$ due to single-peakedness. We also know there must be at least hn agents prefer $h \succ_i r$ by the definition of h. Thus, for any agent for whom $r \succ_i h$, they accept $R^{\delta_i}(V_{rh}) = h$ because R^{δ_i} is effectively implemented. □

5 Related Work

In the introduction to *The Calculus of Consent*, Buchanan and Tullock state, "The selection of a decision-making rule is itself a group choice, and it is not possible to discuss positively the basic choice-making of a social group except under carefully specified assumptions about rules. We confront a problem of infinite regression here." In Chapter 2, they state the implication explicitly, "...in discussing decision rules, we get into the familiar infinite regress if we adopt particular rules for adopting rules. To avoid this, we turn to the unanimity rule..." [11]. In practice, such infinite regress does not prevent collective decision making. While disagreements can lead to an impasse, the inexorable passage of time guarantees that *something* will happen. Ultimately, there will be an outcome, and whatever leads to that outcome was a feasible decision all along. This is consistent with the view of Reality-Aware Social Choice, which recognizes that there is always a status quo, even if implicit [27].

A major source of inspiration for us is the burgeoning research on generating axiomatic justifications for collective decisions [7, 8, 9, 10, 18, 21, 23, 24, 29] and arguing about voting rules [12, 15]. Rather than constructing justifications, we consider the dual problem of making decisions that are justifiable to the maximum number of agents, given that different agents may accept different justifications. Personalized justifications may be beneficial not only for collective decisions, but for many applications of "explainable AI." The personalization of justifications raises important questions about what it means for multiple justifications to be compatible with one another. Our model formalizes one aspect of how agents might view the relevant counterfactuals with our notion of *implementation-indifference*, which is related to the difference between intra-profile and inter-profile axioms [26].

The literature on constitutional amendments and "voting on the voting rule", where agents have preferences over rules, is typically focused on types of stability or idempotency [3, 5, 16] and separated into consequentialist and non-consequentialist approaches [22].

We present a unifying framework capturing both approaches because whether people accept decisions can depend on both outcomes and procedures [17, 19]. Our framework for acceptance can be seen as a broad generalization of the nascent idea of "complaint-freeness" proposed by [1]. From the literature on founding and amending constitutions, Abramowitz et al. [2] put forward two constitutional mechanisms that yield an idempotent rule, along with their axiomatic characterizations. In Section 4 we consider the problem of amendments with supermajority rules in which agents either accept or reject an amendment decision. We demonstrate that under mild assumptions, which do not require all agents to accept a common set of axioms, both constitutional mechanisms proposed by Abramowitz et al. [2] can achieve a form of universal acceptance. Here we also mention Bhattacharya [6] which looks at voting over voting rules in single-peaked domains.

Lastly, the works perhaps closest to ours in spirit are Dietrich [13] and Schmidtlein [25]. Dietrich argues for maximizing acceptance where all agents are treated as II-proceduralists with $|R_i| = 1$ based on a principle of "Procedural Autonomy." Schmidtlein demonstrates how decisions can be made by applying different sets of axioms depending on the profile, which implicitly define a rule.

6 Discussion

Through an approach we term *Social Mechanism Design*, we demonstrate how one can incorporate agents' views about collective choice into the identification, design, selection, and implementation of collective choice mechanisms. We have identified plausible preference structures expressed as combinations of rules and outcomes and shown how structured preferences can be used to maximize acceptance. We examined worst-case acceptance in two settings; one in which the agents must make a generic single asymmetric binary choice, and the other in which agents must make decisions about amending a supermajority rule. In the case of amendments with implementation-indifferent disjunctivists, we showed that universal acceptance is possible and achieved by mechanisms in the existing literature.

Implicitly, the structure of agents' preferences over decisions determines what justifications they will accept. When agents are implementation-indifferent they can be given different justifications, separate from causal explanations of how the decision was actually made, without issue. However, offering agents multiple justifications based on conflicting counterfactuals leaves room for deception. Such deception has not been addressed in the literature on generating axiomatic explanations and justifications for collective decisions.

There are several promising directions for future work. Once a profile is fixed, an acceptance-maximizing mechanism M effectively performs approval voting over the feasible decisions using the satisfying sets as approval ballots, so agents cannot benefit by misreporting whether they accept any decision. However, agents may be strategic in reporting information in the profile. It is an open question how to develop mechanisms that incentivize agents to report only truthful information in both the profile and satisfying sets. Moreover, the general mechanisms we have provided are not particularly efficient for all instance classes. Computing the outcomes of many different rules on a single profile can be computationally expensive. Developing more efficient application-specific mechanisms is an important open challenge.

Lastly, agents may have preferences over decisions reflecting the fact that rules are algorithms, potentially implemented as programs, and not just functions. Agents might care about the computational complexity of rules, whether rules are easy to understand, whether the rules preserve privacy, the probability distribution over outcomes with randomization, etc.

Acknowledgements Nicholas Mattei was supported by NSF Awards IIS-RI-2007955, IIS-III-2107505, and IIS-RI-2134857, as well as an IBM Faculty Award and a Google Research Scholar Award. Ben Abramowitz was supported by the NSF under Grant #2127309 to the Computing Research Association for the CIFellows Project.

References

[1] Ben Abramowitz, Ehud Shapiro, and Nimrod Talmon. How to amend a constitution? model, axioms, and supermajority rules. *arXiv preprint arXiv:2011.03111v2*, 2020.

[2] Ben Abramowitz, Ehud Shapiro, and Nimrod Talmon. In the beginning there were n agents: Founding and amending a constitution. In *International Conference on Algorithmic Decision Theory*, pages 119–131. Springer, 2021.

[3] Daron Acemoglu, Georgy Egorov, and Konstantin Sonin. Dynamics and stability of constitutions, coalitions, and clubs. *American Economic Review*, 102(4):1446–76, 2012.

[4] Kenneth J Arrow. *Social choice and individual values.* Yale university press, 2012.

[5] Salvador Barbera and Matthew O Jackson. Choosing how to choose: Self-stable majority rules and constitutions. *The Quarterly Journal of Economics*, 119(3):1011–1048, 2004.

[6] Mihir Bhattacharya. Constitutionally consistent voting rules over single-peaked domains. *Social Choice and Welfare*, 52(2):225–246, 2019. doi: 10.1007/s00355-018-1145-1. URL `https://doi.org/10.1007/s00355-018-1145-1`.

[7] Arthur Boixel and Ronald de Haan. On the complexity of finding justifications for collective decisions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 5194–5201, 2021.

[8] Arthur Boixel and Ulle Endriss. Automated justification of collective decisions via constraint solving. 2020.

[9] Arthur Boixel, Ulle Endriss, and Ronald de Haan. A calculus for computing structured justifications for election outcomes. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI-2022). AAAI Press*, 2022.

[10] Arthur Boixel, Ulle Endriss, and Oliviero Nardi. Displaying justifications for collective decisions. *IJCAI Demo Paper*, 2022.

[11] James M Buchanan and Gordon Tullock. *The calculus of consent: Logical foundations of constitutional democracy*, volume 100. University of Michigan press, 1965.

[12] Olivier Cailloux and Ulle Endriss. Arguing about voting rules. 2016.

[13] Franz Dietrich. How to reach legitimate decisions when the procedure is controversial. *Social Choice and Welfare*, 24(2):363–393, 2005.

[14] Joseph Y Halpern. *Actual causality.* MiT Press, 2016.

[15] Michael Kirsten and Olivier Cailloux. Towards automatic argumentation about voting rules. In *4ème conférence sur les Applications Pratiques de l'Intelligence Artificielle APIA2018*, 2018.

[16] Semih Koray and Arkadii Slinko. Self-selective social choice functions. *Social Choice and Welfare*, 31(1):129–149, 2008.

[17] E Allan Lind and Tom R Tyler. *The social psychology of procedural justice.* Springer Science & Business Media, 1988.

[18] Yves Meinard and Olivier Cailloux. On justifying the norms underlying decision support. *European Journal of Operational Research*, 285(3):1002–1010, 2020.

[19] Vanessa Mertins. Procedural satisfaction matters: Procedural fairness does not; an experiment studying the effects of procedural judgments on outcome acceptance. Technical report, IAAEG Discussion Paper Series, 2008.

[20] Roger B Myerson. Mechanism design. In *Allocation, Information and Markets*, pages 191–206. Springer, 1989.

[21] Oliviero Nardi, Arthur Boixel, and Ulle Endriss. A graph-based algorithm for the automated justification of collective decisions. In *AAMAS*, pages 935–943, 2022.

[22] Houy Nicolas. Dynamics of stable sets of constitutions. 2007.

[23] Dominik Peters, Ariel D Procaccia, Alexandros Psomas, and Zixin Zhou. Explainable voting. *Advances in Neural Information Processing Systems*, 33:1525–1534, 2020.

[24] Ariel D Procaccia. Axioms should explain solutions. *The Future of Economic Design: The Continuing Development of a Field as Envisioned by Its Researchers*, pages 195–199, 2019.

[25] Marie Christin Schmidtlein. Voting by axioms. 2022.

[26] Amartya Sen. *Collective Choice and Social Welfare.* Harvard University Press, 2018.

[27] Ehud Shapiro and Nimrod Talmon. Reality-aware social choice. *arXiv preprint arXiv:1710.10117*, 2017.

[28] Ehud Shapiro and Nimrod Talmon. Incorporating reality into social choice. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1188–1192, 2018.

[29] Sharadhi Alape Suryanarayana, David Sarne, and Sarit Kraus. Justifying social-choice mechanism outcome for improving participant satisfaction.

Ben Abramowitz
Tulane University
New Orleans, LA
Email: `babramow@tulane.edu`

Nicholas Mattei
Tulane University
New Orleans, LA
Email: `nsmattei@tulane.edu`

A Details of the General Model

A.1 Acceptance

Our notion of acceptance is entirely generic. It is fundamentally a binary classification of possible decisions by agents. We do not address what it means for an agent to accept a decision or the consequences if one or more agents do not accept a decision. We relate acceptance to the existence of an axiomatic justification for a decision that would satisfy an agent. This could mean that they see the decision as legitimate, just, fair, proper, or morally permissible. Acceptance could also mean, for example, that an agent will not dispute or appeal a decision.

A.2 Feasibility

Feasibility is a restriction on what rules and outcomes might be part of the decision output by a mechanism. Feasibility is not a constraint on what agents can report in the profile or in their satisfying sets. Similarly, whether an agent accepts a decision cannot depend on knowledge of what rules and outcomes are feasible. We do not assume that agents know what rules and outcomes are feasible. We do not need to assume the agents know $\mathcal{R}$ and $\mathcal{Y}$, much less the broader of all possible rules and outcomes $\bar{\mathcal{R}}$ and $\bar{\mathcal{Y}}$ from which the elements of their sets R_i and Y_i are drawn.

The feasibility of rules and outcomes are not assumed to be dependent. A feasible rule may not yield a feasible outcome when applied to the given profile; $R(V) \notin \mathcal{Y}$ for some $R \in \mathcal{R}$. Likewise, a feasible outcome may not be selected as the winning outcome by any feasible rule. A decision (V, R, y) is feasible if and only if its rule and outcome are both feasible.

One might wonder why, once a profile V is fixed in some instance, we do not narrow the set of feasible rules to only those that select feasible outcomes and narrow the set of feasible outcomes to only those selected by feasible rules on that profile. In practice, this computation may be inefficient, unnecessary, and impractical when there are many agents, many possible rules, and many possible outcomes. We want to be able to make decisions without this step, though sometimes it is the most straightforward approach.

Our model is agnostic to the determinants of feasibility, and there are many possibilities that differ from application to application. For example, feasibility might be constrained by budget limits and resource availability, willingness of participants, time constraints, privacy and security requirements, or what rules are currently implemented programmatically in a computerized voting system being used.

In future work it will be important to consider what information agents have about feasibility, as this can impact the profile, the structure of their satisfying sets, and their use of counterfactual reasoning.

A.3 Rules and Outcomes

It is to be expected that the structure of agent preferences depends on the object of their preference. Much of the literature on voting and social choice considers abstract alternatives. Here, preferences over rule are preferences over functions, while preferences over outcomes are preferences over abstract alternatives. Rules are functions they are defined with respect to how they map a domain of inputs (profiles) to a range of possible outputs (outcomes). Since only one profile is observed in any instance, the very definition of a rule depends on counterfactuals about what outcome it would have yielded had the profile been different. In our problems every agent observes the same outcome, but the rule cannot generally be observed, leading to the distinction between what singular rule is *actually implemented* and all the other rules that are *effectively implemented* in a given instance.

We have assumed that the rules are "resolute" in the sense that for all possible profiles V, every rule returns some outcome y. This is easily justified by assuming that all rules have some subset of profiles in their relevant domain and return a default outcome for all profiles outside of their domain. This default could be to maintain the status quo or some designated null value.

We also note that we have not assumed the rules in our model to be anonymous. Therefore our model is sufficiently general to include, say, delegative voting rules and satisfying sets where acceptance of outcomes depends on which agents prefer that outcome.

A.4 Implementation

In our framework, the mechanism M selects a single rule R to be actually implemented to produce a decision $R(V) = y$. One can view the feasible rules in $\mathcal{R}$ as subroutines of the algorithm M and the rule R in the decision corresponds to which subroutine gets called. Recall that M selects the rule and outcome (R, y) based on the profile V and the collection of satisfying sets $\mathcal{D}_N$. If the mechanism M selects the rule R based on the agents' satisfying sets independent of the profile V observed, then the mechanism is only capable of reflecting the proceduralist aspects of the agents' satisfying sets. One might argue that it is M that is actually implemented, and any rule R is only effectively implemented, because had the profile V been different, M may have selected a decision with some other rule R' actually implemented. This gets at the heart of how the mechanism operates and how counterfactuals are treated.

If agents base their acceptance on the single rule R output by the mechanism M rather than the mechanism M itself, this is a form of higher-level implementation-indifference. For example, when decisions are made by a Parliament or Congress according to some voting rule R determined by a constitution, the constitution provides the mechanism M, but whenever it is impractical, infeasible, or unlikely for the rule to be changed for a particular decision, whether agents' accept the decision might be based on whether they accept R. Their acceptance of a decision may not need to appeal to the broader legitimacy of the constitutional mechanism and history of how that rule was selected. For all decisions, when we construct a model of what actually caused the decision to be what it was, we must decide what variables are exogenous and endogenous to the causal model [14]. Accepting that there are necessarily some exogenous variables is a form of implementation-indifference at some level. The willingness to accept exogenous variables as part of the causal model is a fundamental component of preventing the problem of infinite regress, which we reflect in the concept of implementation-indifference. In practice, agents whose preferences are based on coherent normative principles are necessarily implementation-indifferent at some level.

A.5 Elicitation

We have left aside for now how agents report their information v_i and satisfying sets $\mathcal{D}_i$, and assume that they report sufficient information for our mechanism. Our mechanism only requires enough information from the agents about their satisfying sets to determine whether they accept various decisions for a specific profile V. Since each instance only contains one profile V, only decisions in their satisfying sets with that profile $(V, \_, \_)$ will be relevant for that instance. In practice, agents may reveal their satisfying sets with reference to the specific profile or independent of it. An agent might express that the rule used should belong to a certain class like scoring rules, or list axioms such that they will accept a decision if and only if it obeys those axioms.

Notice that in principle, the sets R_i and Y_i for each agent and feasible sets $\mathcal{R}$ and $\mathcal{Y}$ can be finite or infinite, continuous or discrete. For example, an agent might accept any outcome that lies within a particular interval on the real line. The set of scoring rules is infinite, although the set of unique scoring rules is finite in any instance with a finite number of alternatives. The mechanisms we have provided assume finite sets of discrete elements, and may need to be varied to suit different applications.

Sometimes information about the full preferences and satisfying sets of agents must be inferred. When preferences have structure it enables them to be elicited more efficiently. For example, with supermajority rule amendments, we assumed we can infer an agent's vote v_i and entire satisfying set just by knowing their ideal rule, or peak, δ_i and the fact that they are II-disjunctivists. When there are many possible decisions that can be made, efficient elicitation becomes increasingly important, and it may be necessary to implement mechanisms with incomplete and imperfect information about agent satisfying sets.

Since the mechanisms we have given effectively implement Approval Voting over decisions using agent satisfying sets, a form of strategyproofness is necessarily inherited: Once the profile is fixed, no agent can lead the mechanism to yield a decision they accept by misreporting their satisfying set $\mathcal{D}_i$ when reporting their true satisfying set would have rendered a decision they do not accept. The same holds with group strategyproofness for sets of agents. There are many important open questions about how to design strategyproof mechanisms when agents can misreport their information v_i.

A.6 Reality-Awareness

In Reality-Aware Social Choice, 'Reality' is, "an ever-present, always-relevant, and evolving social state, distinguished from hypothetical social states" [27, 28]. Before a decision is made there is some status quo or state of the world constituting Reality. We take this view to it's logical conclusion – in the future, there will also be a Reality. That future reality, regardless of whether it is explicitly identifiable in advance, is always a feasible outcome. Whatever process leads from the current reality to that future reality constitutes a decision of some kind, as long as any choices were made or actions taken or avoided along the way. Hence, the existence at least one feasible decision in every problem instance is necessarily implied by reality-awareness, even if it cannot be identified because the set of feasible outcomes is not fully known.

B Acceptance Maximization for Homogeneous Agents

We provide here simpler mechanisms for settings in which all agents are consequentialists or absolute proceduralists which can be implemented more efficiently than our more general Mechanism 1. Maximizing acceptance with consequentialists is straightforward (Mechanism 7). We choose the feasible outcome that appears in the most satisfying sets such that there exists a feasible rule which selects this outcome on the given profile. Maximizing acceptance with proceduralists is similar (Mechanism 8). We choose the feasible rule that appears in the most satisfying sets such that it selects a feasible outcome on the given profile.

Proposition 2. Mechanism 7 maximizes acceptance when all agents are consequentialists.

Proof. Mechanism 7 begins by taking the set of feasible outcomes and removing from it any outcomes that do not appear in the set Y_i of any agent because those will be accepted by no one. If the resulting set is empty, the mechanism returns any feasible decision, though it will not be accepted by any agent. If the set is non-empty, then the algorithm counts for each remaining feasible outcome how many agents have it in their set Y_i. The outcome of every feasible rule on the given profile is then computed, and the outcome among them that appeared in the most sets Y_i is returned. □

Mechanism 7: Acceptance Maximization with Consequentialists

Input: $(V, \mathcal{R}, \mathcal{Y}, \mathcal{D}_N)$
  $\mathcal{Y} \leftarrow \{\bigcup_{i \in N} Y_i\} \bigcap \mathcal{Y}$
  if $\mathcal{Y} \neq \emptyset$ **then**
    for $a \in \mathcal{Y}$ **do**
      $N^a \leftarrow |\{i \in N : a \in Y_i\}|$
      $R^a = \emptyset$
    end for
  end if
  for $R \in \mathcal{R}$ **do**
    $a \leftarrow R(V)$
    if $\mathcal{Y} = \emptyset$ **then**
      return (V, R, a)
    else if $a \in \mathcal{Y}$ and $R^a = \emptyset$ **then**
      $R^a \leftarrow R$
    end if
  end for
  $\mathcal{Y} \leftarrow \{a \in \mathcal{Y} : R^a \neq \emptyset\}$
  $y \leftarrow \arg\max_{a \in \mathcal{Y}} N^a$
  return (V, R^y, y)

Mechanism 8: Acceptance Maximization with Proceduralists

Input: $(V, \mathcal{R}, \mathcal{Y}, \mathcal{D}_N)$
  $\mathcal{R} \leftarrow \{R \in \mathcal{R} : R(V) \in \mathcal{Y}\}$
  $\mathcal{R} \leftarrow \{\bigcup_{i \in N} R_i\} \bigcap \mathcal{R}$
  $R \leftarrow \arg\max_{R' \in \mathcal{R}} |\{i \in N : R' \in R_i\}|$
  return $(V, R, R(V))$

Proposition 3. Mechanism 8 maximizes acceptance when all agents are absolute proceduralists.

Proof. The optimality of Mechanism 8 follows from its use of brute-force search to find the feasible rule that yields a feasible outcome on the given profile and appears in the most sets R_i. It includes a small optimization to only consider feasible rules that appear in at least one set R_i. □

We can also implement a more direct mechanism when all agents are absolute conjunctivists.

Proposition 4. Mechanism 9 maximizes acceptance when all agents are absolute conjunctivists.

Proof. Mechanism 9 begins by removing all rules from R_i that do not select an outcome in Y_i on the given profile. Thus, a rule R is in R_i for agent i if and only if i would accept the decision $(V, R, R(V))$. The mechanism then finds the rule that is in the set R_i of the largest number of agents. □

Mechanism 9: Acceptance Maximization for Absolute Conjunctivists

Input: $(V, \mathcal{R}, \mathcal{Y}, \mathcal{D}_N)$
  $R_i \leftarrow \{R \in R_i : R(V) \in Y_i\}$
  $R = \arg\max_{R' \in \mathcal{R}} |\{i \in N : R' \in R_i\}|$
  return $(V, R, R(V))$

B.1 Asymmetric Dichotomous Choice

Proposition 5. For any set of agents, if we make no assumptions about Y_i or R_i, then $\alpha_{\mathcal{I}} = 0$ for the class of asymmetric dichotomous choice problems.

Proof. Proposition 5 follows simply because if $|R_i| = |Y_i| = 0$ for all agents i, then we have a degenerate case where no agent accepts any decision. So in the worst case $\alpha_I = 0$, and of course it cannot be less than 0. □

Proposition 6. For any set of absolute conjunctivist agents, if $v_i \in Y_i$ and $|R_i \cap \mathcal{R}| \geq 1$ for all $i \in N$, then $\alpha_{\mathcal{I}} = 0$ for the class of asymmetric dichotomous choice problems.

Proof. Even if we assume that for all absolute consequentialist agents there is at least one feasible rule in R_i, and that their set Y_i contains their vote v_i, it is still possible that no agent accepts any feasible decision on a given profile because there is no feasible rule $R \in R_i$ such that $R(V) \in Y_i$. □

Proposition 7 says that if all agents are absolute conjunctivists, then even if every agent accepts at least one feasible decision, there still might be no feasible decision that more than 2 agents accept.

Proposition 7. For any set of absolute conjunctivist agents, if $|Y_i \cap \mathcal{Y}| \geq 1$ and $\exists R \in R_i \cap \mathcal{R} : R(V) \in Y_i \cap \mathcal{Y}$ for all $i \in N$, then $\alpha_{\mathcal{I}} = \frac{2}{n}$ for the class of asymmetric dichotomous choice problems.

Proof. Let $\tilde{R}_i = \{R \in R_i \cap \mathcal{R} : R(V) \in Y_i \cap \mathcal{Y}\}$ be the subset of feasible rules in R_i that select a feasible outcome on the given profile V that is within Y_i. Thus, the absolute conjunctivist agent i will accept a decision $(V, R, R(V))$ if and only if $R \in \tilde{R}_i$. Are assumption states that $\tilde{R}_i$ is non-empty for all agents. With $|\mathcal{R}| = \left\lfloor \frac{n+1}{2} \right\rfloor$ feasible rules and n agents, our bound of $\alpha_{\mathcal{I}} \geq \frac{2}{n}$ follows by the pigeonhole principle (with n pigeons and $\left\lfloor \frac{n+1}{2} \right\rfloor$ holes). So there must be at least two agents who accept a decision that uses the same rule. □

Proposition 8. For any set of absolute disjunctivist agents, if $|R_i \cap \mathcal{R}| \geq 1$ for all $i \in N$, then $\alpha_{\mathcal{I}} = \frac{2}{n}$ for the class of asymmetric dichotomous choice problems.

Proof. With n absolute disjunctivist agents, all of whom have at least one feasible rule in their set R_i, and $|\mathcal{R}| = \lfloor \frac{n+1}{2} \rfloor$ feasible rules, there must be at least 2 agents $i, j \in N$ such that $|R_i \cap R_j| \geq 1$ by the pigeonhole principle (with n pigeons and $\lfloor \frac{n+1}{2} \rfloor$ holes). Therefore $\alpha_{\mathcal{I}} \geq \frac{2}{n}$, because we can always find some decision $(V, R, R(V))$ accepted by at least two agents. However, there may be no rule accepted by 3 or more agents, and so $\alpha_{\mathcal{I}} < \frac{3}{n}$. □

Proposition 9. For any set of absolute disjunctivist agents, if $|R_i \cap \mathcal{R}| \geq k$ for all $i \in N$, then $\alpha_{\mathcal{I}} = \frac{1}{n} \cdot \left\lceil \frac{nk}{\lfloor \frac{n+1}{2} \rfloor} \right\rceil$ for the class of asymmetric dichotomous choice problems.

Proof. With n absolute disjunctivist agents all with at least k feasible rules out of the $\lfloor \frac{n+1}{2} \rfloor$ possible feasible rules in their sets R_i, the bound follows from the pigeonhole principle (with nk pigeons and $|\mathcal{R}| = \lfloor \frac{n+1}{2} \rfloor$ holes) which sets a tight lower bound on the minimum number of agents that must have the same rule in their sets R_i. □

Corollary 1. For any set of absolute disjunctivist agents, if $|Y_i \cap \mathcal{Y}| \geq 1$ for all $i \in N$, then $\alpha_{\mathcal{I}} = \frac{1}{2}$ for the class of asymmetric dichotomous choice problems.

Proof. The proof follows directly from the proof of Proposition **??**, because consequentialists are a sub-type of absolute disjunctivists. If the agents' rule sets R_i are non-empty, this could only increase $\alpha_{\mathcal{I}}$. □

The following four corollaries follow directly from Theorem 3, which applies to all sets of II agents, because $\mathcal{Y} = 2$ in asymmetric dichotomous choice problems.

Corollary 2. For any set of II-conjunctivist agents, if $|Y_i \cap \mathcal{Y}| \geq 1$ $\exists R \in R_i : R(V) \in \mathcal{Y}$ for all $i \in N$, then $\alpha_{\mathcal{I}} = \frac{1}{2}$ for the class of asymmetric dichotomous choice problems.

Corollary 3. For any set of II-disjunctivist agents, if $\exists R \in R_i : R(V) \in \mathcal{Y}$ for all $i \in N$, then $\alpha_{\mathcal{I}} = \frac{1}{2}$ for the class of asymmetric dichotomous choice problems.

Corollary 4. For any set of II-disjunctivist agents, if $|Y_i \cap \mathcal{Y}| \geq 1$ for all $i \in N$, then $\alpha_{\mathcal{I}} = \frac{1}{2}$ for the class of asymmetric dichotomous choice problems.

Corollary 5. For any set of II-disjunctivist agents, if $|Y_i \cap \mathcal{Y}| \geq 1$ and $\exists R \in R_i : R(V) \in \mathcal{Y} \backslash Y_i$ for all $i \in N$, then $\alpha_{\mathcal{I}} = \frac{1}{2}$ for the class of asymmetric dichotomous choice problems.

Consequentialists Suppose all agents are consequentialists and their satisfying sets are *consistent* with their votes $v_i \in \{r, p\}$ such that $v_i \in Y_i$ and $\mathcal{D}_i = \{(V, R, y) \in \mathbb{D} : y \in Y_i\}$ for all $i \in N$. Universal acceptance with consequentialists is only possible if there is an outcome everyone accepts: $\bigcap_{i \in N} Y_i \neq \emptyset$. The worst-case acceptance rate is determined by the minimum size of the majority, which is half the agents.

Proposition 10. For asymmetric dichotomous choice with all consequentialist agents such that $v_i \in Y_i$, the worst-case acceptance rate is $\alpha_{\mathcal{I}} = \frac{1}{2}$, which is achieved by the simple mechanism M that applies majority rule $R^{\frac{1}{2}}$ to the profile V in every instance.

Proof. The proof of the lower bound of $\frac{1}{2}$ on $\alpha_{\mathcal{I}}$ in Proposition 10 follows from the pigeonhole principle. One of the two potential outcomes, r or p, must be accepted by at least half of the agents because we forbid abstention and empty satisfying sets. This follows from the requirement that $v_i \in Y_i$. With n pigeons and 2 holes there must be at least $\lceil \frac{n}{2} \rceil$ pigeons in one of the holes. Therefore, for any instance, we can always achieve an acceptance rate of at least $\frac{1}{2}$ by making the decision according to majority rule: $(V, R^{\frac{1}{2}}, R^{\frac{1}{2}}(V))$.

In the worst case, $Y_i = \{v_i\}$ for every agent $i \in N$, the number of agents n is even, and exactly half of the agents vote for each alternative. Any feasible decision satisfies exactly half of the agents, so we know $\alpha_{\mathcal{I}} \leq \frac{1}{2}$ because no mechanism can achieve a higher acceptance rate on this instance. This gives us our upper bound of $\frac{1}{2}$. $\square$

Notice that the mechanism M in Proposition 10 does not maximize acceptance for every instance. A decision with the alternative that receives fewer votes as its outcome might still be acceptable to a greater fraction of the agents. However, the proof above shows that the mechanism M that maximizes acceptance must also have $\alpha_{\mathcal{I}} = \frac{1}{2}$. The same value of $\alpha_{\mathcal{I}} = \frac{1}{2}$ is obtained for the smaller class of instances with the stronger requirement that $Y_i = \{v_i\}$ for all agents. The proof is essentially the same as the one above. If agents can be inconsistent, where $v_i \notin Y_i$, then $\alpha_{\mathcal{I}} = 0$ because of the degenerate instances where $v_i = r$ and $Y_i = \{p\}$ for all $i \in N$.

Absolute Proceduralists Suppose every agent has a vote $v_i \in \{r, p\}$ and a set of one or more supermajority rules $R_i \subseteq \mathcal{R}$ such that $(V, R, y) \in \mathcal{D}_i$ if and only if $R \in R_i$. The optimal mechanism counts for each supermajority rule R the number of sets R_i that contain R and then picks among these rules the one with the highest count. The outcome is uniquely determined by the choice of rule and the profile. For there to be universal acceptance, it would have to be that $\bigcap_{i \in N} R_i \neq \emptyset$.

II-conjunctivists An acceptance maximizing mechanism for II-conjunctivists can be created by changing the ‘or’ to ‘and’ in lines 3 and 4 of Mechanism 5 defining N_r and N_p.

Ben Abramowitz
Tulane University
New Orleans, LA
Email: babramow@tulane.edu

Nicholas Mattei
Tulane University
New Orleans, LA
Email: nsmattei@tulane.edu