NBER WORKING PAPER SERIES

INTERPRETING THE WILL OF THE PEOPLE:
A POSITIVE ANALYSIS OF ORDINAL PREFERENCE AGGREGATION

Sandro Ambuehl
B. Douglas Bernheim

Working Paper 29389
http://www.nber.org/papers/w29389

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2021

We are grateful to Matias Cersosimo for research assistance. This project was funded by the Department of Economics at Stanford University and by the Department of Economics at the University of Zurich. All experiments were approved by the Stanford IRB and by the Ethics Review Board of the Department of Economics at the University of Zurich. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Interpreting the Will of the People: A Positive Analysis of Ordinal Preference Aggregation
Sandro Ambuehl and B. Douglas Bernheim
NBER Working Paper No. 29389
October 2021
JEL No. C91,D7

ABSTRACT

Collective decision making requires preference aggregation even if no ideal aggregation method exists (Arrow, 1950). We investigate how individuals think groups should aggregate members' ordinal preferences—that is, how they interpret "the will of the people." Our experiment elicits revealed attitudes toward ordinal preference aggregation and classifies subjects according to the rules they implicitly deploy. Majoritarianism is rare while rules that promote compromise are common. People evaluate relative sacrifice by inferring cardinal utility from ordinal ranks. Cluster analysis reveals that our classification encompasses all important aggregation rules. Aggregation methods exhibit stability across domains and across countries with divergent traditions.

Sandro Ambuehl
Department of Economics
University of Zurich
Bluemlisalpstrasse 10
Zurich, ZH 8006
Switzerland
sandro.ambuehl@econ.uzh.ch

B. Douglas Bernheim
Department of Economics
Stanford University
Stanford, CA 94305-6072
and NBER
bernheim@stanford.edu

An online appendix and code replication is available at
http://www.nber.org/data-appendix/w29389

1 Introduction

Envision a benevolent party (the planner) who must make a decision that impacts the members of some group, and who possesses accurate information about the group members' preferences. The foundational question of *social welfare* asks, how should the planner determine what is best for the group? Potential answers to this question have been the subject of enduring debate since at least the late 18th century, when Marie Jean Antoine Nicholas Caritat, better known as Marquis de Condorcet, and Jean-Charles de Borda proposed the competing aggregation criteria that still bear their names (Borda, 1781; Condorcet, 1785). Condorcet advocated selecting an option that majority-defeats all alternatives. Borda insisted that the best rule would instead make use of all the information contained in each individual's ranking of the options. His proposal amounts to selecting the alternative for which the sum of the assigned ranks is smallest.

Potential answers to the social welfare question necessarily depend on the types of information concerning preferences the planner deems meaningful. In particular, most of modern economic theory treats utility as identified only up to an individual-specific monotonic transformation, which precludes cardinal interpersonal comparisons.[1] Arrow's canonical formulation of the social welfare problem (Arrow, 1950) therefore assumes that the planner can only use ordinal information concerning preferences.[2] Invoking the aggregation axiom known as Independence of Irrelevant Alternatives (IIA) also effectively precludes the use of information on ranks.[3] The logic of IIA is traceable to Condorcet (Condorcet, 1788), who criticized the Borda rule on the grounds that it "relies on irrelevant factors to form its judgments," in that it "confuses votes comparing Peter and Paul with those comparing either Peter or Paul to Jack and uses them to judge the relative merits of Peter and Paul." While the Borda rule respects the modern proscription on using cardinal measures of well-being, it slips cardinality back into the social calculus through these types of third-option comparisons. Indeed, Borda' justification for the rule explicitly references cardinal inferences.[4]

Arrow's celebrated Impossibility Theorem (Arrow, 1950) precludes the existence of any rule for aggregating ordinal preferences satisfying IIA and certain other reasonable axioms. Yet groups cannot avoid preference aggregation merely because they might run afoul of a theoretical axiom. One way or another, they routinely make implicit or explicit judgments about tradeoffs between different members' objectives. Our overall goal in this paper is to understand those judgments.[5] In other words, we

[1]There is, however, a branch of the literature that attempts to justify such comparisons; see Roberts (2005) for a review.

[2]Arrow's assumption is "that interpersonal comparison of utilities has no meaning and ... that there is no meaning relevant to welfare comparisons in the measurability of individual utility" (Arrow, 1951).

[3]According to IIA, if a change in group members' preference rankings leaves the relative rankings of options A and B unchanged, then it also leaves social preferences between A and B unchanged.

[4]According to Borda (Borda, 1781), "...we must assume that the degree of superiority which this voter gave A over B is the same as that he gave B over C. As candidate B is no more likely to be ranked in one particular place on the scale between A and C than in any other, we have no reason to say that the voter who ranked the candidates ABC wanted to place B nearer A than C or vice versa; no reason to say, that is, that he accorded the first more superiority over the second than he accorded the second over the third."

[5]As in the literature on social welfare spawned by Arrow's work, we abstract from the important problem of eliciting accurate information about group members' preferences, which is the subject of a related literature; see, for example,

approach the classical social welfare question from a *positive* perspective rather than a *normative* perspective.[6] Illuminating fundamental attitudes toward ordinal preference aggregation is important because those attitudes have potentially profound implications for public policy. For example, by adopting institutions that tend to deliver desirable outcomes according to the Condorcet criterion, a society will become susceptible to what Mill, in *On Liberty*, termed the "tyranny of the majority" (Mill, 1869). In contrast, by embracing institutions that tend to deliver desirable outcomes according to the Borda rule, a society will provide greater protections for minority populations, whether associated with geography, ideology, ethnicity, or religion.

Our analysis addresses four main questions. First, what rules and criteria do people actually follow when aggregating ordinal preferences? For instance, do they impose compromise solutions, or insist that the majority should prevail? Second are these rules stable, or do they vary from one context to another? To what extent do they reflect structural principles of preference aggregation, rather than contextual adaptations? Third, do people honor ordinal information, or do they try to make cardinal imputations before aggregating? Fourth, do common aggregation rules vary across cultures with divergent political and social traditions, and might such variation help to explain differences in policies?

Identifying the rules that govern ordinal aggregation is conceptually challenging. Even in simple social choice problems (e.g., with five people and three options), the set of possible mappings from preference profiles to best choices is astronomically large. We therefore proceed in four steps. First, drawing on the theoretical literature, we identify a reasonably large set of plausible aggregation rules. Each of these rules implies a distinctive *fingerprint* of implied best choices over the set of conceivable five-person three-option preference profiles. Second, we conduct an experiment in which subjects in the role of Social Planner make a series of decisions for other groups of subjects (Stakeholders). We assign each subject to a pre-specified rule using a Bayes classifier, which identifies the best match between each subject's empirical fingerprint and the theoretical fingerprints associated with the various rules. Third, we corroborate the classifications using a handful of discerning four-option social choice problems. Fourth, we use a clustering algorithm to determine whether our pre-specified rules omit empirically important possibilities.

In order to elicit subject-level empirical fingerprints, each Social Planner makes decisions for multiple preference profiles, knowing that any one of them may involve a real group of Stakeholders who care about the outcome. The Social Planners' decisions are of two types: assignment decisions (distributing five work tasks among five Stakeholders), and spending decisions (assigning a contribution on behalf of five Stakeholders to a single Swiss political party). We call these the *work domain* and

Gibbard (1973) and Satterthwaite (1975). While issues of manipulability inevitably arise in practice, they pertain to the constraints that appear in a mechanism design problem, rather than to the objective function, and it is the latter we seek to illuminate. Notably, Borda was primarily concerned with the social welfare question, as he is said to have described his preferred rule as intended for honest men (quoted in Black, 1958).

[6] It is arguable that, in representative democracies, policy makers ought to defer to citizens' judgments about appropriate criteria for preference aggregation. Under that view, positive analyses of social choice have important normative implications.

the *political domain*, respectively. Most of our analysis focuses on the work domain; we examine the political domain to evaluate context-sensitivity. Because we are interested in identifying structural aggregation preferences, we design the experiment to remove considerations arising from self-interest, paternalism (i.e., the tendency to ignore or discount Stakeholder judgments with which the Social Planner disagrees, as in Ambuehl et al., 2021a), and the potential incentive incompatibility of truthful preference revelation by Stakeholders.

In answer to the first question (which rules do people use), we find that the overwhelming majority of subjects behave as if they rely on *scoring rules*, which assign a score to each rank and select best options based on the total scores.[7] The two most common as-if aggregation criteria are the Borda rule (for which the score is linear in ranks), and *near-antipluralilty* (where antiplurality rule minimizes the number of last-place ranks).[8] A sizable majority ($> 60\%$) of subjects employ *strictly concave scoring rules*, of which antiplurality is an example. Substantively, these rules imply an even stronger preference for compromise than the Borda rule; technically, they have the property that improvements in low ranks are more important than improvements in high ranks. Condorcet (majoritarian) rules are relatively rare, as is the related concept of plurality rule, and likewise associated runoff criteria. Neither do people often gravitate toward supermajority or unanimity (Pareto) rules, even though those also provide minorities with varying degrees of protection. The classification's fit is excellent: empirical and theoretical fingerprints for assigned rules are remarkably similar. Analysis of discerning four-option profiles corroborates our conclusions concerning the prevalence of various rules. Clustering analysis identifies only one non-pre-specified rule of consequence ($> 2\%$ of subjects), and it differs from near-antiplurality on only one of 17 preference profiles.

In answer to the second question (stable structure versus contextual adaptations), we find that our classifications are highly predictive of choices out of sample, including across domains. This result reassures us that ordinal aggregation entails stable structural elements. This is not to say that the distribution of rules is the same in the work domain and the political domain. On the contrary, the differences between these distributions, though relatively small, are systematic and statistically significant, which points to a degree of context-specificity.

In answer to the third question, we find strong indications that subjects aggregate ordinal preferences based in part on inferences about cardinal utility. As a threshold matter, it is worth emphasizing that scoring rules provide latitude for injecting cardinal inferences into the social choice, whereas the Condorcet rule does not. We explore this question in two ways. First, we provide evidence on the validity of a choice axiom known as *Sen's α*, which states that the removal of an unchosen option from an opportunity set should not alter the selection from that set. We demonstrate that choices satisfy

[7] Consistent with this finding, Featherstone (2019) shows that, in the context of matching markets, policymakers often evaluate matches based on rank distributions (a special case of scoring rules), and sometimes tinker with the results of matching algorithms in attempts to improve the rank-distribution.

[8] Antiplurality is the scoring rule that assigns 0 to the last-place rank and 1 to all other ranks. It is thematically related to (but substantively distinct from) *last place aversion*, which has been found in some experiments (Kuziemko et al., 2014; Martinangeli and Windsteiger, 2020) but not others (Camerer et al., 2016).

Sen's α when the preference rankings provided to the Social Planner include the deleted item, but severely violate Sen's α when we also remove the item from the rankings. We infer that subjects likely draw inferences about the intensity of preferences from comparisons with options that are generally considered undesirable. Second, we examine the correlation between best-fit scoring parameters and the scoring parameters our Social Planners would use if they were money-metric utilitarians, given their elicited beliefs about Stakeholders' reservation valuations for first-, second-, and third-ranked choices. While the correlation corroborates the importance of cardinal inferences, further investigation suggests that subjects attach substantial domain-independent weight to the various ordinal ranks. In other words, they appear to deploy both cardinal and ordinal criteria.

To answer to the fourth question (comparisons across countries), we run supplemental experiments using general population samples, wherein social choices determine the allocation of a contribution over well-known charities. We find that the distributions of aggregation preferences in the U.S. and Sweden, countries with divergent political and social traditions, are remarkably similar, and both resemble the distribution for the student sample used in our main experiment.[9] Policy differences may therefore be attributable to other factors, such as beliefs, historical accidents, institutions, and/or equilibrium selection, as hypothesized by Alesina and Angeletos (2005). Nevertheless, we find suggestive evidence that the use of more concave scoring rules in experimental decisions correlates with a preference for electing compromise candidates.

Most broadly, our paper contributes to the literature on *positive welfare economics*, which uses empirical methods to determine how people evaluate the well-being of other individuals and groups (e.g. Andreoni et al. (2020); Almås et al. (2020); Ambuehl et al. (2021a); see Konow (2003); Gaertner (2009); Gaertner and Schokkaert (2012) for reviews). To our knowledge, only a handful of previous papers in this area have attempted to address the problem of ordinal preference aggregation from a positive perspective. The most closely related paper is Kara and Sertel (2005). Their analysis is confined to three rules (Condorcet, Borda, and "majoritarian compromise"), which they distinguish based on a few (four) preference profiles using hypothetical choices involving abstract options. In contrast, we examine real choices, use far more exhaustive lists of rules and preference profiles, estimate scoring parameters, use clustering analysis to detect omitted rules, test out-of-sample predictive accuracy, examine stability across domains and cultures,[10] investigate whether subjects try to make cardinal inferences, and provide evidence of external validity using general population samples.

Other positive work on ordinal preference aggregation addresses different questions than ours. The analysis in Weber (2017), which focuses on two-option choices, seeks to determine how subjects weigh the votes of delegates who represent groups of different sizes in assemblies such as the EU. Other related work elicits subjects' preferences over (five or fewer) voting procedures, rather than

[9] Ambuehl et al. (2021b) include an abridged version of the current experiment in a study of elected representatives in federal and state parliaments in Germany. They find similar qualitative results.

[10] Faravelli (2007) documents the importance of context in social decisions.

over outcomes (Engelmann and Grüner, 2017; Hoffmann and Renes, 2017; Engelmann et al., 2020).[11] In these experiments, it is up to the subjects to imagine what each rule might imply for any particular preference profile. Those inferences are often non-trivial, and it is possible that a subject would reject a seemingly appealing rule after learning what it implies.[12] In contrast, our subjects reveal their preferences over rules by making choices over explicit social outcomes. Choices over procedures rather than social outcomes may also implicate strategic considerations or procedural notions of justice and equity, which we intentionally remove from our study in order to address the classical problem of ordinal preference aggregation.

Our work is also related to the vast empirical literature on *other-regarding preferences*, which generally studies preference aggregation in settings where cardinal information (typically concerning money) is available (see, e.g., Fehr and Fischbacher, 2002; Cooper and Kagel, 2016, for reviews). The most closely related branch elicits preferences over the distribution of income (e.g. Epper et al., 2020; Fisman et al., 2021). Most of this literature is concerned with self-versus-other tradeoffs, rather than other-versus-other tradeoffs, although there are some exceptions. For example, Jackson and Yariv (2014) examine how subjects in the role of a Social Planner aggregate others' cardinal preferences intertemporally in settings where rates of discount differ. An alternative to the self-versus-other and other-versus-other paradigms examines social choice behind a veil of ignorance, as in Engelmann and Strobel (2004). Those experimental settings tend to induce utilitarian decisions (Bolton and Ockenfels, 2006; Jackson and Yariv, 2014).

This paper draws heavily on, and complements, the theoretical social welfare literature (Arrow et al., eds, 1991, 2010; Fishburn, 2015; Brandt et al., 2016). Most obviously, it documents the empirical relevance of various rules, but there are other points of contact. By examining the tendency for people to make cardinal inferences from ordinal information about preferences, we provide empirical context for a theoretical literature on utilitarian-optimal voting rules (de Laplace, 1812; Weber, 1978; Merrill III, 1984; Apesteguia et al., 2011; Boutilier et al., 2015; Pivato, 2016).[13] Moreover, the low prevalence of Condorcet rules suggests that most people do not accept Condorcet efficiency (the frequency with which a rule selects the Condorcet winner when one exists) as a normative principle, contrary to the position of Merrill III (1984) and others; see, for instance, Van Newenhizen (1992); Baharad and Nitzan (2003).

The remainder of the paper is organized as follows. We lay out our main strategy for detecting the use of specific social choice rules in Section 2. Section 3 details our experimental design. Section 4

[11] Another potential approach, which has been used in decision theory, is to ask subjects to endorse axioms rather than rules; see, e.g., Nielsen and Rehbeck (2020). Unfortunately, Arrow's theorem alerts us to the possibility that people might well endorse inconsistent axioms. In addition, social choice axioms are often complex and difficult for non-specialists to understand, and some common rules lack axiomatizations.

[12] Laslier (2012) reports the result of an informal vote by 22 voting theory experts on an ad hoc list of 18 aggregation rules. The winner is approval voting, followed by instant runoff. That vote considers rules comprehensively whereas this paper abstracts from strategic considerations to focus on the welfare (aggregation) question.

[13] Our conclusions concerning the tendency to infer cardinal values from ranks resonate with findings from a bargaining experiment by Herreiner and Puppe (2010), in which subjects received ordinal information concerning each others' preferences.

provides our classification results and all associated analyses. Section 5 describes our supplementary experiments involving general population samples. Section 6 provides some brief concluding remarks.

2 Conceptual framework

We are concerned with settings in which a decision maker, the *Social Planner*, must make a selection from a set of K social options, $\mathcal{A}$, on behalf of N *Stakeholders*. Each option has direct consequences for the Stakeholders, but the Social Planner is not materially affected by her decision. We assume throughout that each Stakeholder i has a linear preference ordering $\succsim_i$ over $\mathcal{A}$. To avoid technicalities we assume the orderings are strict (a property that is, in any case, generic). Before making a decision, the Social Planner learns the group's *preference profile*, $P = (\succsim_1, \ldots, \succsim_N)$. A *social choice rule* is a complete account of the Social Planner's choices for every preference profile; in other words, it is a mapping R from preference profiles into nonempty subsets of $\mathcal{A}$ (*best choices*).[14] A *resolute* rule admits no ties, in the sense that it maps every profile to single option. A rule is *irresolute* if at least one preference profile maps to a non-singleton set, indicating that there is more than one best choice.

Our objective is to determine which social choice rules people actually use when making decisions for others. The task of identifying these rules is challenging because the set of possibilities is astronomically large. To illustrate, consider a simple setting with 5 Stakeholders, where the Social Planner must choose among 3 options. In that case, the domain of any social choice rule satisfying *anonymity* (meaning that the rule treats all group members symmetrically) and *neutrality* (meaning that the rule treats all social options symmetrically) consists of 42 distinct preference profiles.[15] A social choice rule maps each of these profiles to a subset of the three options. Because there are seven such subsets, there are $7^{42} = 3.1 \times 10^{35}$ possible social choice rules for this simple environment. While most are unreasonable, 3.5×10^{28} exhibit no Pareto violations.[16]

Because the set of potential social choice rules is so large, our analysis proceeds in three steps. First, based on the literature, we identify a reasonably large collection of social choice rules encompassing the most plausible alternatives. Second, we classify Social Planners according to the pre-specified rules that most closely match their actual choices. Third, we deploy clustering analysis to determine whether our pre-specified rules omit empirically important alternatives.

[14] Arrow (1950) studied *social choice functions*, which map preference profiles into rankings over social alternatives. When A is finite, any social choice function F implies a corresponding social choice rule, which picks maximal elements according to F.

[15] We identified these profiles using an algorithm that enumerated all the alternatives and then eliminated ones that are redundant under anonymity and neutrality. It is also possible to arrive at the number of such profiles through a combinatorial argument. According to *The Hitchhiker's Guide to the Galaxy* by Douglas Adams, the number 42 is "The Answer to the Ultimate Question of Life, the Universe, and Everything," calculated by a supercomputer over 7.5 million year, but unfortunately no one knows the question. We are delighted to have resolved that mystery. You're welcome.

[16] The Pareto criterion rules out nothing for 27 profiles, one option for 12 profiles, and two options for 3 profiles. The number of rules without Pareto violations is therefore $7^{27} \times 3^{12}$.

In the remainder of this section, we summarize various social choice rules that our subjects may employ, and explain conceptually how we distinguish among them.

2.1 Possible social choice rules

There are in principle many ways to categorize social choice rules.[17] For our purposes, the most useful classification references the nature of the information the rule employs. The reason is that, in designing an experiment such as ours, one must be alert to the possibility that the method of presenting information concerning the preference profile P may nudge the Social Planner toward one type of social choice rule or another. Understanding the information requirements for various classes of rules helps ensure a neutral presentation.

As an example, consider the *Borda method*, which selects options that minimize the sum of ranks across all Stakeholders. To implement this rule, one must know, for each $A \in \mathcal{A}$ and $k \in \{1, ..., K\}$, the number of group members who rank option A k-th among the potential alternatives, denoted r_A^k. We call this distillation of P the *Borda data*.

An important class of social choice rules known as *scoring rules* use only the Borda data. A scoring rule establishes a *score vector*, $w = (w_1, w_2, \ldots, w_K)$ with $w_1 \geq \ldots \geq w_K$, and allocates w_1 points to each Stakeholder's top-ranked alternative, w_2 points to their second-ranked alternative, and so on. It then selects an option that maximizes the sum of the scores.[18] Taking $w_k = k$ for all k yields the Borda rule. *Plurality rule* emerges as a special case: when $w_1 > 0$ and $w_k = 0$ for $k > 1$, the scoring method maximizes first-place ranks. Alternatively, taking $w_k = w_1 > 0$ for $k < K$ and $w_K = 0$, the scoring method minimizes last-place ranks, a rule known as *negative voting* or *antiplurality*.

As a second example, consider the *Condorcet method*, also known as *pairwise majority rule*, which selects an option that majority-defeats all others (a *Condorcet winner*). To implement this rule, one must know, for each pair of options, $A, B \in \mathcal{A}$, the number of group members who rank A strictly above B, denoted v_{AB}. We call this distillation of P the *Condorcet data*.

The Condorcet method is a special case of a *p-supermajority* rule, which declares option A better than option B if the fraction of Stakeholders who prefer A to B is at least p. The rule selects an option if it is not improvable according to this binary relation. Condorcet corresponds to the case in which p is the strict majority threshold, *Unanimity rule* (also know as the *Pareto rule*) to the case of $p = 1$, and *Supermajority rule* to intermediate cases. All p-supermajority rules use only the Condorcet data.

A well-known limitation of the Condorcet method is that it can give rise to cycles, and consequently does not yield a winner for certain preference profiles. The same issue arises for Supermajority rule,

[17] In practice, collective decision making employs many of the aggregation rules we consider. For instance, several national use multistage methods for presidential elections (see, e.g., Richie, 2004). Slovenian parliamentary elections employ the Borda count (Toplak, 2006), as do elections within the Irish Green Party (Baker, 2008). Swiss parliamentary elections incorporate disapproval votes (which are in the spirit of the antiplurality rule) (Portmann and Stojanović, 2019). The Debian Project uses a Condorcet extension (the Schulze method) for all decision-making by committees and plebiscites (Debian Project, 2016).

[18] Myerson (1995) provides an axiomatic characterization of the set of scoring rules.

but not for Unanimity rule. To apply Condorcet (or Supermajority) on an unrestricted domain, one must supplement the rule. A *Condorcet extension* selects a Condorcet winner when one exists, but otherwise employs some other criterion.[19] An example is the *top cycle* or *Smith set*, defined as the smallest collection of options that majority-defeat all options in the complement of the set. Many Condorcet extensions use only the Condorcet data. Examples include the top cycle, the *Minimax method*, *Copeland's rule*, the *Kemeny-Young method*, *Tideman's rule*, and the *Schulze method*.

Having distinguished between classes of rules according to information they employ, it is important to emphasize that the Borda data are not in general recoverable from the Condorcet data, and the Condorcet data are not in general recoverable from the Borda data. Accordingly, a presentation of preference profiles that makes either type of information more salient could potentially skew choices accordingly.

Some social choice rules require more complete information concerning preference profiles. For example, *Black's method* is the Condorcet extension that selects the Borda winner when a Condorcet winner fails to exist. It therefore relies on both the Condorcet data and the Borda data. However, that reliance is hierarchical, in the sense that the Condorcet data take precedence.

Multistage rules that winnow down the set of alternatives in a series of steps introduce almost limitless possibilities, inasmuch as they can, in principle, apply different criteria in each step. Even when the rule employs the same criterion round after round, the required information evolves. Consider, for example, the set of *runoff rules* based on scoring criteria, which repeatedly eliminate the alternative with the lowest score until only one option remains.[20] A well-known example is the *single transferrable vote rule*, which iteratively deletes options with the lowest plurality scores. Another is *Baldwin's rule*, which uses the Borda score. Such rules employ the Borda data for a sequence of shrinking menus. That information is not recoverable from the Borda data for the original menu. Still, a presentation that highlights the Borda data may encourage the use of such methods.

2.2 Main identification strategy

The core of our empirical analysis differentiates among social choice rules based on selections for problems involving 5 Stakeholders and 3 options. This focus offers three advantages.

First, 5-Stakeholder 3-option problems are the simplest settings that provide adequate scope for differentiation among a broad collection of rules. While the addition of Stakeholders or options can in principle permit greater discernment, it can also render social choice problems less cognitively manageable.

Second, with K options, the set of scoring rules is a $K-2$ parameter family. Consequently, when $K=3$, we can associate each scoring rule with a single parameter, s. To understand this point, note

[19] See Definition 2.8 in Brandt et al. (2016). The Campbell-Kelly theorem characterizes the set of Condorcet extensions (Theorem 2.3 in Brandt et al., 2016).

[20] For formal definitions and an axiomatic characterization, see Freeman et al. (2014).

that without loss of generality, we can assign a score of 1 to a Stakeholder's highest-ranked alternative, and a score of 0 to her lowest-ranked alternative. The parameter $s \in [0, 1]$ is then the score assigned to the middle alternative. The cases of $s = 0$, $s = 1/2$, and $s = 1$ correspond to the Plurality, Borda, and Negative Voting rules, respectively. The one-dimensionality of this class facilitates the interpretation of our results. In particular, the function relating ranks to weights is concave when $s > 1/2$, and convex when $s < 1/2$. Concavity (convexity) implies that replacing a third-place rank with a second-place rank is more (less) valuable than replacing a second-place rank with a first-place rank. Accordingly, concave scoring rules codify an aversion to giving Stakeholders their least favorite choices, while convex scoring rules codify an attraction to giving Stakeholders their favorite choices. These observations suggest that scoring rules may reflect inferences about cardinal utility,[21] a point to which we return in Section 4.6. They also suggest the interpretation of concavity as appreciation of compromise.

Third, with 5 Stakeholders and 3 options, we can in principle investigate choices exhaustively on the entire preference domain. As noted above, for social choice rules satisfying neutrality and anonymity, there are 42 distinct strict preference profiles. Eliciting a choice for every conceivable profile is therefore feasible.

In practice, many of the 42 preference profiles provide little or no discernment among rules. For example, if all Stakeholders have the same ranking, the best social choice is obvious. Including such problems lengthens the experiment, thereby risking the erosion of subjects' effort and attention, without adding significant value. Accordingly, we omit profiles that provide little or no differentiation among well-known rules. Our analysis is based on the 17 discerning profiles listed in Table 1. In Appendix A.1, we list the omitted profiles and exhibit their limited ability to distinguish among rules.

Table 1 also shows how the 17 profiles differentiate among familiar social choice rules. For example, if a subject uses a scoring rule, we can determine the parameter s from the choices they make for profiles 1 through 11. To understand this point, first consider profile 11. Because all subjects rank option B second, its score is $S_s(B) = 5s$. Four subjects rank option C first and one ranks it third, so its score is $S_s(C) = 4$. Option A is rank-dominated by C, so no scoring rule will select it. Therefore, the subject will choose option B if $S_s(B) > S_s(C)$, or equivalently, $s > 0.8$, and will choose option C if $s < 0.8$. If $s = 0.8$, the subject is indifferent between options B and C. Now consider profile 6, which differs from profile 11 only in that Stakeholder 2's preferences between A and C are reversed. Reasoning as before, we see that scoring rules select B over C when $s > 0.6$.

Profiles 1 through 11 constitute the set of all profiles for which there are exactly two options that are not weakly rank-dominated. For each, there is a threshold $\bar{s}$ at which the optimal choice switches, as in the cases of profiles 6 and 11. These thresholds divide the interval $[0, 1]$ into a sequence of subintervals with boundaries in the set $\mathcal{C} = \{0, \frac{1}{3}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, 1\}$. Based on decisions for profiles 1-11,

[21] Apesteguia et al. (2011) formalize this point.

Table 1: Three-alternative profiles.

Index	Profile	Rule predictions										
		Scoring cutoff	Scoring rules. $s \in$					Condorcet	Runoff rules, $s \in$			
		$\bar{s}$	$\{0\}$	$(0, \bar{s})$	$\{\bar{s}\}$	$(\bar{s}, 1)$	$\{1\}$		$[0, \frac{1}{3}]$	$(\frac{1}{3}, \frac{1}{2})$	$(\frac{1}{2}, 1)$	$\{1\}$
1*	A A C C C B B A A A C C B B B	1/3	C	C	{A,C}	A	A	C	C	C	C	{A,C}
2	A C C B B B A A A A C B B C C	1/3	{B,C}	B	{A,B}	A	A	A	B	A	A	{A,B}
3*	A A A C C B C C B B C B B A A	1/2	A	A	{A,C}	C	C	A	A	A	A	{A,C}
4*	A A B B B B C A A A C B C C C	1/2	B	B	{A,B}	A	A	B	B	B	B	{A,B}
5	A C C C B B B B B A C A A A C	1/2	C	C	{B,C}	B	B	C	C	C	C	{B,C}
6*	A A C C C B B B B B C C A A A	3/5	C	C	{B,C}	B	B	C	C	C	C	{B,C}
7	A C C C B B A B B A C B A A C	2/3	C	C	{B,C}	B	B	C	C	C	C	{A,B}
8*	A C C C B B B B B C C A A A A	2/3	C	C	{B,C}	B	B	C	C	C	C	{B,C}
9*	A A A B B B C C C C C B B A A	3/4	A	A	{A,C}	C	C	A	A	A	A	{A,C}
10	A C C C C B A A A A C B B B B	3/4	C	C	{A,C}	A	A	C	C	C	C	{A,C}
11*	A C C C C B B B B B C A A A A	4/5	C	C	{B,C}	B	B	C	C	C	C	{B,C}
12	A C C B B B A A C C C B B A A	0	-	-	{B,C}	C	C	{A,B,C}	B	B	B	{A,C}
13	A C C B B B A A C A C B B A C	1/2	{B,C}	{B,C}	{A,B,C}	A	A	{A,B,C}	B	B	{A,B,C}	{A,B,C}
14	A C C B B B A B C C C B A A A	0	-	-	{B,C}	{B,C}	{B,C}	B	B	B	B	{B,C}
15	A C C B B B A B A A C B A C C	0	-	-	{B,C}	B	{A,B}	B	B	B	B	{A,B}
16	A C C B B B A B C A C B A A C	0	-	-	{B,C}	B	B	B	B	B	B	{A,B}
17	A C C B B B B B A A C A A C C	0	-	-	{B,C}	B	B	B	B	B	B	{A,B}

Notes: Each profile is displayed as a 3×5-matrix. Columns correspond to Stakeholders and rows to preference ranks. A Stakeholder's first, second, and third-ranked alternatives are listed in the first, second, and third rows, respectively. For Condorcet-cyclical profiles, we indicate the set of options in the top-cycle. For decisions in the political domain, we only use the profiles indicated with an asterisk.

we can determine which interval contains the subject's scoring parameter.[22] In some cases, we can also distinguish choices based on scoring parameters at the boundaries of these intervals.[23]

As shown in Table 1, the remaining profiles provide further scope for differentiating among social choice rules. For example, because profiles 12 and 13 exhibit Condorcet cycles, all three options are best choices according to the top-cycle Condorcet extension, but not according to many other rules.

Each social choice rule generates an identifiable "fingerprint" of selections across the 17 preference profiles; see Figure 1. Each column in the figure represents a 5-Stakeholder 3-option preference profile, which we identify in the first panel. For example, the first column corresponds to the profile wherein two Stakeholders prefer A to B to C, while three prefer C to A to B. Each subsequent panel corresponds to a specified social choice rule; it shows the options that rule selects for each preference profile. Comparing the fingerprints across rules (panels) reveals both similarities and differences.

Our main classification results encompass 22 benevolent social choice rules.[24] We start with the 15 distinguishable rules that emerge from the scoring method. We can differentiate between $s = \frac{1}{2}$ (the Borda rule), 5 ranges of scoring parameters in the convex range (the most extreme of which corresponds to Plurality rule), and 10 ranges of strictly concave rules (the most extreme of which corresponds to Antiplurality rule). Our data can also differentiate among 4 social choice rules that emerge from the scoring runoff method. These rules correspond to values of s in the following ranges: $[0, \frac{1}{3}]$, $(\frac{1}{3}, \frac{1}{2})$, and $[\frac{1}{2}, 1)$, as well as $s = 1$.[25] Finally, we can distinguish three social choice rules that emerge from the p-supermajority top-cycle method. These rules correspond to values of p in the following ranges: $[\frac{1}{2}, \frac{3}{5}]$, $(\frac{3}{5}, \frac{4}{5}]$, and $(\frac{4}{5}, 1]$. Henceforth we adopt a slight abuse of terminology and call these rules Condorcet, Supermajority, and Unanimity, respectively. While our basic classification does not include other Condorect extensions, we conduct robustness analyses to determine whether this omission is material.

Because we distinguish between social choice rules based on their fingerprints, larger differences between fingerprints facilitate more reliable classifications. Figure 2 tabulates, for each pair of social choice rules, the number of preference profiles (out of the 17 we use for identification) for which their implications differ. Two patterns stand out. First, in the vast majority of cases, different rules have different implications for large numbers of profiles. Ignoring the diagonal, few of the entries in Panel A are less than 3, and many are greater than 8, indicating differences on more than half of profiles.

[22] As indicated in Table 1, profile 13 provides corroborative evidence as to whether $s \lesseqgtr \frac{1}{2}$, and the remaining profiles shed light on whether $s = 0$.

[23] When a scoring parameter coincides with an interval boundary $s \in \mathcal{C}$, the set of options selected for a given preference profile is the union of those selected with scoring parameter $s - \epsilon$ and those selected with scoring parameter $s + \epsilon$ for $\epsilon > 0$ sufficiently small. Accordingly, scoring rules with parameters $s \in \mathcal{C}$ are less resolute (produce more ties) than scoring rules with parameter $s \notin \mathcal{C}$.

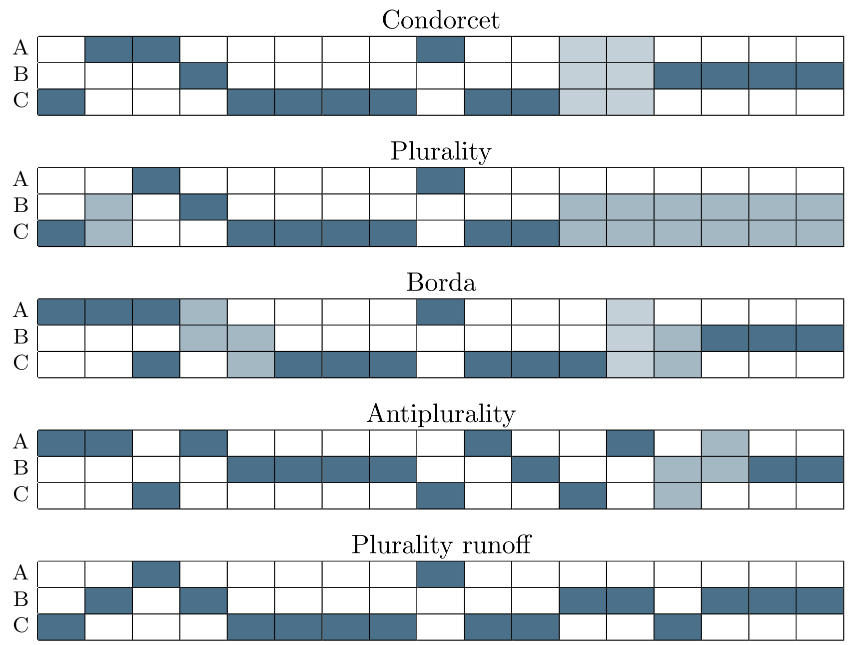
[24] We define a rule as *benevolent* if it satisfies the Pareto criterion.

[25] We implement scoring runoff rules as follows. In the first stage, we calculate the score associated with each option and drop the option with the lowest score. We then choose the majority-preferred option from the remaining alternatives. (Because this step involves at most two options, majority rules coincides with all scoring rules.) If two options are tied, we drop both of them. If all three options are tied, the runoff rule selects all of them. Three-way ties occur only for profile 13, and only for scoring runoff rules with $s \geq \frac{1}{2}$. We identify the set of distinguishable scoring runoff rules using a brute-force computer script.

Figure 1: Fingerprints of an example selection of social choice rules.

Preference profile

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$A > B > C$	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1
$A > C > B$	0	0	2	1	0	0	0	0	2	0	0	0	0	0	0	0	0
$C > A > B$	3	2	0	0	0	0	1	0	0	4	0	2	2	1	1	1	0
$C > B > A$	0	0	2	0	3	3	2	3	0	0	4	0	0	1	1	1	2
$B > C > A$	0	0	0	0	0	0	0	1	2	0	0	2	1	2	0	1	0
$B > A > C$	0	2	0	3	1	0	1	0	0	0	0	0	1	0	2	1	2



Notes: Each column corresponds to a three-alternative preference profile. Preference profiles are listed in the top row. The number in each cell in the top row indicates the number of Workers with the indicated preference ranking. Cells in the top row indicate the number of subjects with each of the preference orderings listed on the left of that row. Each cell in rows 2 to 5 is labeled 1 if the rule chooses the corresponding option and 0 if the rule does not choose the option. Non-resolute rules will choose more than one option for some profiles. Shades of blue indicate the number of tied options. The numbering of preference profiles corresponds to Table 1.

Second, three clusters of similar rules are visible. The first cluster consists of the strictly convex scoring rules numbered 2 to 4. The second cluster corresponds to the strictly concave scoring rules, numbered 6 to 15 (especially 6 to 8 and 12 to 15). Because these rules are members of the single family that is indexed by a continuous parameter s, adjacent rules tend to differ on small numbers of profiles, but more distant scoring rules are more easily differentiated. The third cluster consists of the scoring runoff rules with $s < 1$ (numbered 16 to 18). Observe also that the runoff rules, as well as the Condorcet rule (number 20), align more closely with convex scoring rules than with concave scoring rules. Hence, a preference for compromise may push subjects away from scoring runoff rules.

Figure 2: Distance between rules.

	Scoring, $s=0$	Scoring, $0<s<\frac{1}{3}$	Scoring, $s=\frac{1}{3}$	Scoring, $\frac{1}{3}<s<\frac{1}{2}$	Scoring, $s=\frac{1}{2}$	Scoring, $\frac{1}{2}<s<\frac{3}{5}$	Scoring, $s=\frac{3}{5}$	Scoring, $\frac{3}{5}<s<\frac{2}{3}$	Scoring, $s=\frac{2}{3}$	Scoring, $\frac{2}{3}<s<\frac{3}{4}$	Scoring, $s=\frac{3}{4}$	Scoring, $\frac{3}{4}<s<\frac{4}{5}$	Scoring, $s=\frac{4}{5}$	Scoring, $\frac{4}{5}<s<1$	Scoring, $s=1$	Runoff, $0\leq s\leq\frac{1}{3}$	Runoff, $\frac{1}{3}<s<\frac{1}{2}$	Runoff, $\frac{1}{2}<s<1$	Runoff, $s=1$	Condorcet	Supermajority	Unanimity
Scoring, $s=0$	0	5	6	6	10	10	11	11	13	13	15	15	16	16	16	7	7	7	16	7	15	17
Scoring, $0<s<\frac{1}{3}$	5	0	2	2	6	6	7	7	9	9	11	11	12	12	13	3	4	4	16	4	15	17
Scoring, $s=\frac{1}{3}$	6	2	0	2	6	6	7	7	9	9	11	11	12	12	13	5	5	5	14	5	15	16
Scoring, $\frac{1}{3}<s<\frac{1}{2}$	6	2	2	0	4	4	5	5	7	7	9	9	10	10	11	5	4	4	16	4	15	17
Scoring, $s=\frac{1}{2}$	10	6	6	4	0	4	5	5	7	7	9	9	10	10	11	8	7	6	12	6	14	15
Scoring, $\frac{1}{2}<s<\frac{3}{5}$	10	6	6	4	4	0	1	1	3	3	5	5	6	6	7	8	7	7	16	7	15	17
Scoring, $s=\frac{3}{5}$	11	7	7	5	5	1	0	1	3	3	5	5	6	6	7	9	8	8	15	8	15	17
Scoring, $\frac{3}{5}<s<\frac{2}{3}$	11	7	7	5	5	1	1	0	2	2	4	4	5	5	6	9	8	8	16	8	15	17
Scoring, $s=\frac{2}{3}$	13	9	9	7	7	3	3	2	0	2	4	4	5	5	6	11	10	10	15	10	15	17
Scoring, $\frac{2}{3}<s<\frac{3}{4}$	13	9	9	7	7	3	3	2	2	0	2	2	3	3	4	11	10	10	16	10	15	17
Scoring, $s=\frac{3}{4}$	15	11	11	9	9	5	5	4	4	2	0	2	3	3	4	13	12	12	14	12	16	16
Scoring, $\frac{3}{4}<s<\frac{4}{5}$	15	11	11	9	9	5	5	4	4	2	2	0	1	1	2	13	12	12	16	12	16	17
Scoring, $s=\frac{4}{5}$	16	12	12	10	10	6	6	5	5	3	3	1	0	1	2	14	13	13	15	13	17	17
Scoring, $\frac{4}{5}<s<1$	16	12	12	10	10	6	6	5	5	3	3	1	1	0	1	14	13	13	16	13	17	17
Scoring, $s=1$	16	13	13	11	11	7	7	6	6	4	4	2	2	1	0	15	14	14	15	14	17	17
Runoff, $0\leq s\leq\frac{1}{3}$	7	3	5	5	8	8	9	9	11	11	13	13	14	14	15	0	1	2	17	3	15	17
Runoff, $\frac{1}{3}<s<\frac{1}{2}$	7	4	5	4	7	7	8	8	10	10	12	12	13	13	14	1	0	1	17	2	15	17
Runoff, $\frac{1}{2}<s<1$	7	4	5	4	6	7	8	8	10	10	12	12	13	13	14	2	1	0	16	1	14	16
Runoff, $s=1$	16	16	14	16	12	16	15	16	15	16	14	16	15	16	15	17	17	16	0	16	16	13
Condorcet	7	4	5	4	6	7	8	8	10	10	12	12	13	13	14	3	2	1	16	0	13	15
Supermajority	15	15	15	15	14	15	15	15	15	15	16	16	17	17	17	15	15	14	16	13	0	4
Unanimity	17	17	16	17	15	17	17	17	17	17	16	17	17	17	17	17	17	16	13	15	4	0

Notes: This graph plots the set of 22 benevolent rules on both the horizontal and vertical axes. Each cell reports the number of profiles (out of the 17 used in the work domain) for which a given pair of rules differ from each other. We use the definition that two rules differ on a profile if they select a different subset of options (distance = 1); otherwise they do not differ on that profile (distance = 0).

3 Experimental design

3.1 Social choice problems

Overview We assign subjects to one of two roles: each *Social Planner* ('she') chooses alternatives that potentially affect groups of five *Stakeholders* ('he'). The only purpose of including Stakeholders in the experiment is to ensure that the Social Planners' decisions are consequential. Choice problems fall into two domains, task assignment (*work*) problems and budget allocation (*political*) problems. These domains relate, respectively, to mechanism design and political economy, two areas where ordinal preference aggregation is an important topic. For each domain, the Social Planner views a sequence of preference profiles and, in each instance, selects one of several alternatives. We match one out of every four Social Planners, selected at random, with a real group of Stakeholders. The actual preference profile for that group is among the ones that Social Planner considers, but is not identified as such. Although we only implement decisions that pertain to actual Stakeholder groups, from each Social Planner's perspective any one of her decisions could turn out to be a real choice.

Consequently, as long as she cares about the Stakeholders to some degree, she has an incentive to reveal her aggregation preferences truthfully for every preference profile she encounters. We focus primarily on the work domain, for which we employ the 17 profiles shown in Table 1. We examine the political domain, for which we employ a smaller set of profiles, to evaluate context-dependence.

Tasks for the work domain For the work domain, Stakeholders are *workers*, whom we recruit on Amazon Mechanical Turk. Each worker receives $15 for completing a single assigned task.[26] The compensation is sufficiently high to ensure that any attrition is non-systematic.

There are five work tasks. We choose tasks that resemble familiar activities for online workers, and that different workers might plausibly rank in different orders. The tasks are as follows: (i) *Image labeling.* The worker views a sequence of 400 images and identifies each by clicking a button. (ii) *Hate speech filtering.* The worker views 400 messages posted on twitter.com and indicates whether each includes hate speech such as racist or sexist statements. (iii) *Audio transcription.* The worker listens to a sequence of 400 words and, in each case, identifies the word by clicking a button. (iv) *Movie reviews classification.* The worker classifies 400 movie reviews according to whether they are positive or negative. (v) *Assigning apprentices to mentors.* The worker finds a pairwise stable match between five hypothetical apprentices and five hypothetical mentors knowing their preferences. The worker completes 20 rounds of this task.

Workers reveal their preferences over tasks in a preliminary session. After seeing a description of each task and trying it out to gain familiarity (except for the matching task, a single round of which some subjects find time-consuming), workers then rank the tasks from most to least preferred. We ensure incentive compatibility by informing workers that their rankings determine the task they perform with 5% probability, as follows: the computer randomly pre-selects two tasks, and workers perform the one they say they prefer. To preclude strategic reporting, we tell workers that some other process will determine their assigned tasks with 95% probability. Because we leave that process unspecified, workers have no information about the manner in which their own expressed preferences may factor into the alternative process. After the Social Planners make their choices, Workers complete their assigned tasks in a second session.

Social choice problems for the work domain A social alternative is a *task assignment*: it assigns each of the five tasks to one of the five workers in a group. Social Planners choose from menus of task assignments. For 3-option problems, menus consist of three randomly selected task assignments.

For the main portion of our experiment, Social Planners are students at the University of Zurich and the Swiss Federal Institute of Technology. (We examine US and Swedish general population samples in Section 5). At the beginning of the experiment, they acquaint themselves with the work tasks by reading descriptions and performing abbreviated versions (except for the matching task).

[26] Our survey automatically checks for correctness and requires the worker to continue until the entire task is completed correctly.

Their instructions describe the mTurk platform and provide information on the value of hourly worker compensation in our experiment.

A Social Planner proceeds through several rounds of decision making for her group. In each round, she observes an ordinal preference profile for the task assignments. One of these corresponds to her workers' actual preferences over task assignments, which we infer from their elicited preferences over tasks.[27] The Social Planner then chooses an assignment she considers best for the group. We also ask her to identify any alternative she considers just as good as the one she selects. While this expression of indifference has no consequences within our experiment, two considerations may mitigate the usual concerns about hypothetical choices: first, the question asks the Social Planner to report indifference that she presumably would have recognized when making the associated consequential choice moments prior; second, misrepresenting her indifference would not serve any other plausible objective (e.g., enhancement of social image).[28] However, as a precaution, we adopt two complementary approaches to identifying social choice rules, one of which uses the indifference data, and one which does not.

Recent research on paternalism shows that people tend to discount or even disregard preferences with which they disagree (Ambuehl et al., 2021a). That consideration is orthogonal to the focus of the current paper, which concerns the aggregation of ordinal preferences that are equally valid in the eyes of the Social Planner. To ensure that Social Planners cannot second-guess a worker's preferences (for example, by placing little weight on an expressed desire to complete the hate-speech filtering task), we limit the Social Planner's knowledge about task assignments. Specifically, we show the Social Planner a menu of abstract geometric symbols, explain that each represents a task assignment, and describe each worker's ranking of those assignments. However, we do not explain which worker performs which task in any given assignment.

In Section 2.1, we emphasized that different social choice rules use different data concerning preference orderings, and consequently that the presentation of preference profiles can potentially nudge Social Planners in one direction or another by highlighting particular information. Because a preference profile is an array of three-tuples of the form (option, worker, rank) that indicate the preference rank a particular worker assigns to a given option, there are three qualitatively different ways to display it in a two-dimensional table: rows and columns can be ranks and workers (in which case options appear in cells), ranks and options (in which case workers appear in cells), or workers and options (in which case ranks appear in cells).

We show the first possibility in Figure 3, which represents each worker as a vertical bar. Within each bar, the geometric symbols representing the assignments are ordered according to the worker's

[27] This information captures the workers' selfish preferences over task assignments. Any social preferences workers may have over task assignments might be interdependent, in the sense that each worker may wish to take other workers' preferences into account. To avoid such issues, we leave all interpersonal judgments to the Social Planner.

[28] An alternative procedure would be to (partially) incentivize indifference statements by randomizing among the pertinent options. We decided against that approach for two reasons. First, the procedure changes the nature of social alternatives by introducing lotteries over assignments, and thereby implicating workers' preferences over lotteries. Second, the social choice rules we consider do not specify random resolution. Adding that provision to a rule amounts to creating a non-standard extension.

preference, with the most preferred assignment on top. By clicking buttons, Social Planners can highlight or hide options. Highlighting an option makes the distribution of its ranks (the Borda data) readily apparent. When an option is hidden, the display repositions the remaining assignments onto two lines. This feature makes pairwise preference counts (the Condorcet data) readily apparent. Additionally, Social Planners can hide or rearrange the workers, either by dragging and dropping them, or by clicking a button to shuffle them randomly.

We show the second and third possibilities in Figures 4 and 5, respectively. With these presentations, the Borda data are arguably more salient, but Social Planners can easily access the Condorcet data by hiding options one at a time. Accordingly, we provide the same tools for exploration (hiding, highlighting, and rearranging).

Each Social Planner sees one and only one presentation format, which we select at random. We randomize symbols (representing options) and colors (representing workers) to ensure Social Planners do not conflate decisions across preference profiles.[29] We also randomize the positions of all alternatives and of all workers in each round.

In addition to the 17 preference profiles shown in Figure 1, each Social Planner who is matched to a real group of five workers views that group's actual preference profile and makes a choice, while other Social Planners view a randomly generated preference profile. Social Planners also make decisions for six four-option preference profiles and one two-option profile, which we randomly intermingle with the three-option profiles. Social Planners then view three final preference profiles, each of which rank either three or four alternatives, but they choose from menus that omit one of the alternatives. We provide more detail concerning all of these additional profiles and decisions in Sections 4.4 and 4.6. Altogether, Social Planners make choices for 28 preference profiles in the work domain.

The political domain To determine whether Social Planners apply consistent aggregation criteria across domains, we also present them with decisions involving political contributions. In this domain, Stakeholders (*citizens*) are voting-age members of the general Swiss population.[30] Social Planners, who are also voting-age Swiss nationals, direct a contribution of Fr. 30 (roughly $33.90 at the time of the experiment) to one of the five largest Swiss political parties, as measured by the number of members in the Swiss National Council.

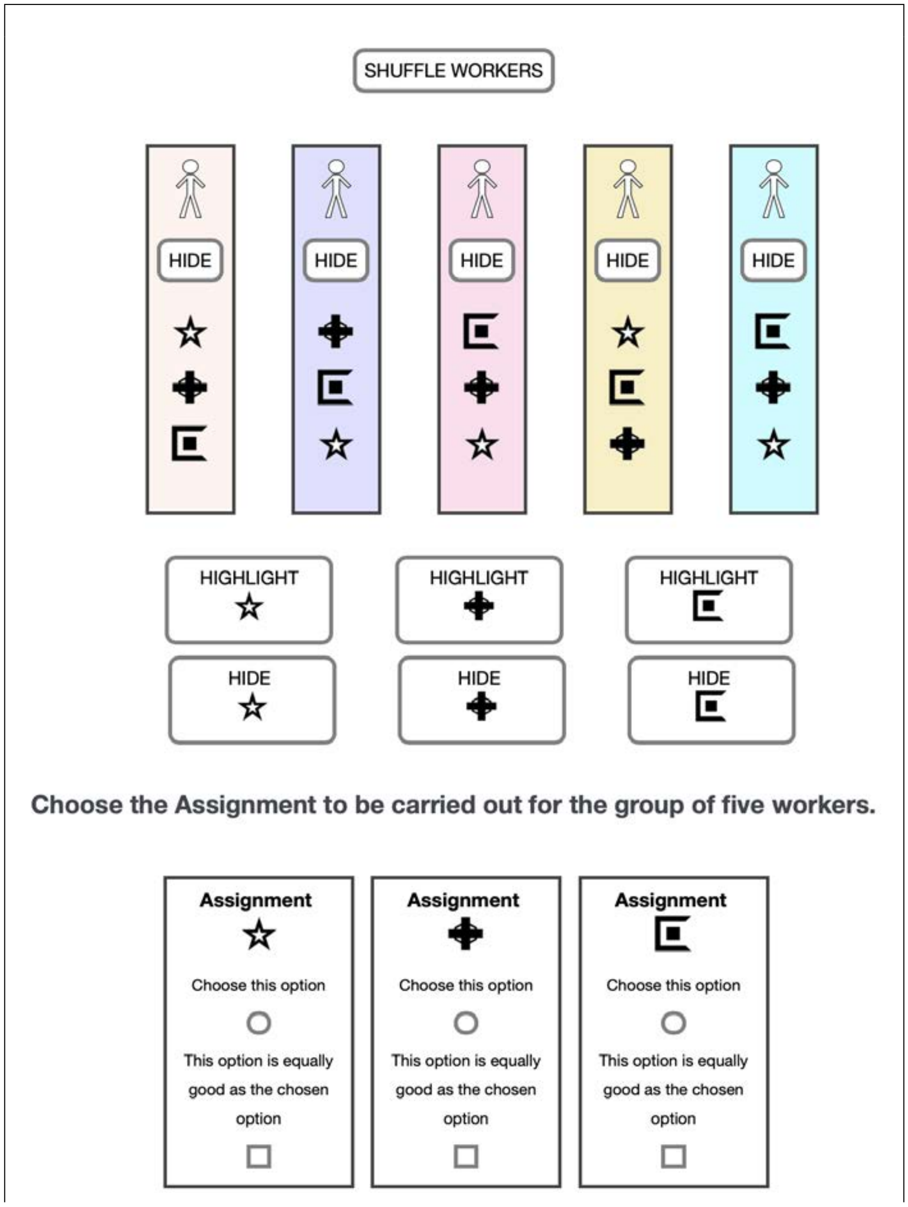
The procedures we use for political contributions are generally the same as for the work domain.[31] Here it is especially important to emphasize that, for any given preference profile, Social Planners

[29] For instance, this procedure prevents a Social Planner from deciding to favor the 'red worker' in one round because she disfavored that worker in the previous round.

[30] We recruited Stakeholders mostly through the survey company pollfish.com. We supplemented this sample by placing ads on Facebook and in the laboratory at the University of Zurich. Age and citizenship are self-reported.

[31] One exception involves the process for eliciting Stakeholders' preferences: citizens' expressed preferences determine the recipient of the donation with 2.5% probability, versus 5% probability for the analogous contingency in the work domain.

Figure 3: Social Planners' decision interface, version 1 (options in cells).

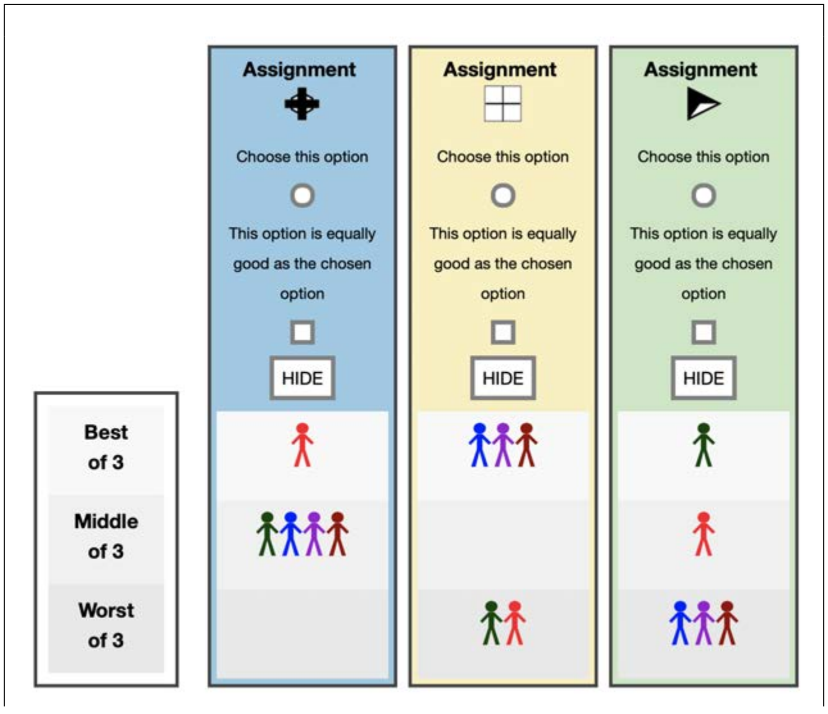


Notes: Subjects can drag and drop columns, click a button to shuffle columns, highlight choice options, hide choice options, and hide workers. If a subject hides an option, the remaining options are arranged on two rows regardless of their initial position. Symbols for options and colors for the bars representing each worker are randomly drawn each round. The order of workers is randomly drawn each round.

do not know which geometric symbol corresponds to which party, so they cannot impose their own political preferences.[32]

[32]To prevent Social Planners from drawing inferences about parties from preference distributions, we draw the group of five citizens from a sample that equally represents those who self-identify as politically left, right, and center. Social Planners are aware of the sample's composition.

Figure 4: Social Planners' decision interface, version 2 (individuals in cells)



Notes: Subjects can drag and drop columns, click a button to shuffle columns, hide choice options, highlight workers, and hide workers. If a subject hides an option, only two rows are shown; the remaining rows are labeled 'Best of 2' and 'Worst of 2.' Symbols for options and colors for worker symbols are randomly drawn each round. The order of options is randomly drawn each round.

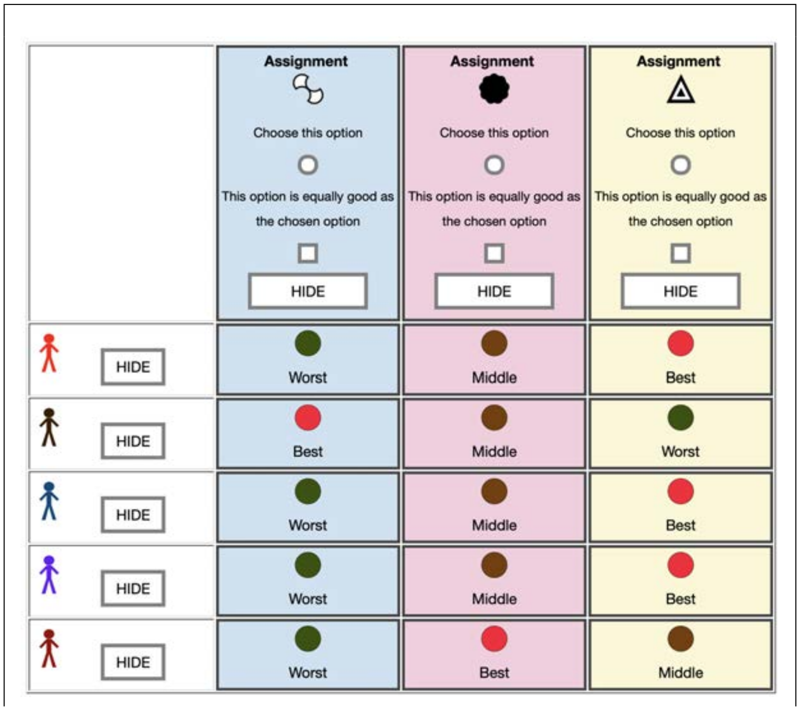
To keep the length of the experiment manageable for our subjects, we use a smaller set of preference profiles for the political domain. Over 12 rounds, Social Planners view, in random order, the starred preference profiles in Column 2 of Table 1, a selection of four four-option profiles (see Section 4.4), and one additional profile (either the actual profile for her assigned citizen group or a randomly generated profile, depending on whether we assign her to a group).

3.2 Additional elicitations

Beliefs about WTA / WTP For the work domain, we ask Social Planners to predict the average reservation wages for the tasks workers rank first, second, and third, knowing only that rankings involve three tasks randomly selected from the five possibilities.[33] To increase statistical power, Social Planners answer five (nearly) redundant versions of these questions, where each version specifies the

[33] Possible answers range from \$0 to \$10 in increments of \$0.50. Half of the subjects see these questions ordered by rank (first, second, third), and half see them ordered by reverse rank (third, second, first). Before subjects answer these questions, we remind them about the price and income level in the US and the standard wages of mTurk workers. The instructions for these tasks explain the concept of WTP using an example for which we randomize the elicited valuation (either \$5 or \$1).

Figure 5: Social Planners' decision interface, version 3 (ranks in cells)



Notes: Subjects can drag and drop columns, drag and drop rows, click a button to shuffle columns, click a button to shuffle rows, highlight preference ranks, hide choice options, and hide workers. If a subject hides an option, only two symbols, labeled 'best' and 'worst' are shown. The color of the symbols labeled 'best' and 'worst' differs depending on whether two or three options are displayed, to signify that the best (worst) option among two is not necessarily best (worst) among three. Symbols for options and colors for worker symbols are randomly drawn each round. The orders of workers and options are randomly drawn each round. Colors for the preference ranks are randomly drawn on the individual level but remain constant throughout the experiment.

city and state where the worker lives.[34] In the event one of these predictions ends up determining the Social Planner's payment, she receives Fr. 30 minus Fr. 3 for every dollar by which her prediction differs from the truth, which we assess using incentivized multiple decision lists in preliminary sessions involving a separate set of workers.[35]

For the political domain, we ask Social Planners to predict the average Swiss citizen's willingness to pay to trigger or prevent the donation workers rank first, second, and third. The details are the same as for the work domain, with a few small exceptions.[36]

[34] We use two sets of five city labels, randomly assigned, so we can determine whether the labels are consequential.

[35] Social Planners also indicate their own reservation price for completing each of the five tasks, but these responses are unincentivized.

[36] Possible predictions for the WTP to trigger the donation lie in the set {>15, 15, 13, 11, 9, 7, 5, 3, 1, -1, -3, -5 ,-7, -9, -11, -13, -15, <-15}. Over the five rounds, Social Planners make predictions for citizens with different first names rather than different cities of residence.

As we discuss subsequently, we use these responses to assess the hypothesis that Social Planners aggregate ordinal preferences based on implicit inferences about cardinal utility (as suggested in Apesteguia et al., 2011).

Risk preferences We elicit risk preferences using the method developed in Holt and Laury (2002). On each line of a multiple decision list, subjects choose between two lotteries. The first lottery pays Fr. 23 with probability p or Fr. 15 with probability $(1-p)$. The second pays Fr. 38 with probability p or Fr. 3 with probability $(1-p)$. The list includes 11 binary choices with $p \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$. In a second otherwise identical decision list, the first lottery pays either Fr. 20 or Fr. 13, while the second pays either Fr. 34 or Fr. 5.

Social preferences Subjects complete a multiple decision list involving the following pairs of alternatives: "*Increase each of the five workers' payoffs by $2. Decrease my own study payment by CHF X*," or "*Do not increase the group members' payoffs. Leave my own study payment unchanged.*" Each line employs a different value of X in the set $\{0, 1, 2, 3, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$. In an otherwise identical second list, the first option is *"Increase each of the five workers' payoffs by Fr. X (exchanged to USD). Decrease my own study payment by Fr.* $5X$," and the values of X lie in $\{0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2\}$. Notice that the first list implicates both efficiency and equity, while the second implicates only equity.

Psychological characteristics, knowledge of social choice theory, and demographics Subjects complete the four-item version of the Cognitive Reflection Test by Thomson and Oppenheimer (2016). For the last 143 subjects, we added four questions probing their knowledge of social choice theory.[37] Subjects also report a variety of personal characteristics.[38]

3.3 Additional details

Timing Table 2 provides an overview of the experiment's temporal structure. Social Planners perform ordinal preference aggregation tasks in parts A (work domain) and B (political domain). Within each domain, we display preference profiles in individually random order. We also randomize the order of Parts A and B, and include instructions about the decision interface in the part that appears first. To avoid nudging subjects to think about cardinal utility, we elicit risk aversion and beliefs about WTA/WTP in Part C, after the preference aggregation tasks are complete.

[37] Subjects report whether they have taken a class covering pertinent topics. We illustrate a three-option, three-citizen cyclical preference profile and ask subjects to name the paradox. We also ask subjects to name the Borda rule and Arrow's impossibility theorem based on verbal descriptions.

[38] These include: gender, age, field of study and degree level they are working towards, grades in university entrance exams in mathematics and in their first language, canton in which they completed their university entrance exam, their main language, whether they live with their parents, their number of siblings, monthly spending, religiosity, religion, political stance, and political party they voted for in the last election of the Swiss National Council.

Table 2: Schematic overview of the experiment.

Part 0: Initial instructions

Part A: Task assigment

1. Instructions concerning task assignment
2. Instructions about the interface that displays preference profiles
3. 25 task assignment decisions (intermingled)
4. 3 task assignment decisions with unavailable options (intermingled)

Part B: Donation to a political party

1. Instructions concerning the donation to a political party
2. 12 party donation decisions

Part C: Further elicitations

1. Beliefs
   - 5 rounds of incentivized belief elicitation about workers' reservation wages, followed by an unincentivized elicitation of own reservation wages for completing each of the five tasks
   - 6 rounds of incentivized belief elicitation about citizens' political preferences, followed by an unincentivized elicitation of own willingness to pay to trigger or prevent the donation to each of the political parties
2. Preferences
   - Risk preferences
   - Social preferences
3. Other characteristics
   - Demographic information
   - Cognitive Reflection Test
   - Knowledge about social choice theory

Notes: Each stage in Part C directly follows instructions concerning that stage. Half of the subjects proceed through the experiment in the order displayed. For the other half, Part B and Part A are interchanged. The latter subjects receive the instructions about the interface that displays preference profiles in Part B instead of in Part A. For those subjects, the two stages of Part C.1 are also interchanged.

Incentives A randomly selected decision (pertaining to WTP/WTA, risk preferences, or social preferences) from part C determines the Social Planner's own payment. As we have explained, social choices are incentivized in the sense that each one may be consequential for others, and we also incentivize the elicitation of workers' preferences.

Instructions and comprehension checks The full instructions for Social Planners, which we present on-screen, appear in Appendix E.1. The presentation requires subjects to try out each option in the preference display (e.g., hide and highlight). Subjects must pass two comprehension checks to continue with the study.[39] Each consists of nine questions. The first set concerns the preference displays and the second concerns the general decision environment. Subjects must answer all nine questions correctly to proceed. If they make a mistake, we ask them to review the instructions and try again, but we do not tell them which question(s) they missed.[40] We conduct the experiment English.[41]

4 Analysis

Our analysis focuses on 405 subjects in the role of Social Planner recruited from the joint subject pool of the University of Zurich and the Swiss Federal Institute of Technology. Subjects participated in 11 online sessions, supervised via video-conferencing software (Zoom), in January 2021. We restricted the sample to Swiss citizens by checking each participants' government-issued ID.

The median subject completed the session in 82 minutes and received Fr. 50.[42] In addition to the 405 subjects who completed the study, another 12 potentially eligible subjects started it.[43] Only five subjects failed to complete the experiment after presenting a valid ID. The implied attrition rate among potentially eligible subjects was therefore between 1.2% and 3%. The median age is 23. Among the subjects who were asked, 7% reported having taken a class that covered social choice theory, but only 1%, 3%, and 2% could correctly name Arrow's theorem, the Condorcet paradox, and the Borda rule, respectively. While our subject pool includes a higher proportion of women (61%) than men and skews towards the political left (70%, 15%, and 14% of subjects rate themselves as left, center, and right, respectively), Section 5 shows that similar results obtain in general population samples.

We structure our analysis as follows. Subsection 4.1 exhibits aggregate choice patterns. Subsection 4.2 presents our main classification results for the work domain. In Subsection 4.3, we use clustering methods to determine whether our classification omits any empirically important social choice rules. Subsection 4.4 provides corroborating evidence on classifications using discerning four-option profiles. Subsection 4.5 then investigates the extent to which our classification captures structural aggregation principles by comparing choices across domains and by evaluating out-of-sample predictive

[39] These are also listed in Appendix E.1.

[40] This feature prevents the subject from trying to pass the comprehension check by trial and error (the chance of passing either of the comprehension checks by chance is less than 0.2%).

[41] A good command of English is a curricular requirement for all students in our subject pool. The invitation emails mention that subjects must be fluent in English.

[42] At the time of the experiment, 1 Fr. = USD 1.13.

[43] An additional 7 subjects started the study but were not eligible.

performance. In Subsection 4.6, we investigate the extent to which aggregation rules reflect cardinal imputations. Throughout, we pool across the three methods of presenting preference profiles, but mention instances where results differ.

4.1 Aggregate choice patterns

Figure 6 shows choice frequencies for each of the 17 three-option profiles in the work domain. Structure is readily apparent, such as the near-unanimous decisions for profiles 2, 16, and 17. Because nearly all subjects choose the option that is both rank-dominant and a Condorcet winner in profiles 16 and 17, we can infer that they act benevolently toward their groups. Notably, disagreements are common for the score-identifying profiles 3 to 11, indicating heterogeneity in preferred aggregation criteria.

Figure 6: Distribution of choices in three-option problems

Profile	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Condorcet	C	A	A	B	C	C	C	C	A	C	C	ABC	ABC	B	B	B	B
Scoring																	
$\bar{s}$	1/3	1/3	1/2	1/2	1/2	3/5	2/3	2/3	3/4	3/4	4/5	-	1/2	0	0	0	0
$s < \bar{s}$	C	B[a)]	A	B	C	C	C	C	A	C	C	C[a)]	BC	-	-	-	-
$s > \bar{s}$	A	A	C	A	B	B	B	B	C	A	B	C	A	BC	B[b)]	B	B
Choice freq.																	
A	86	96	36	63	1	0	1	1	72	29	1	1	73	1	6	1	1
B	0	2	0	37	75	61	43	39	3	0	29	8	15	62	93	98	99
C	13	1	64	0	24	38	56	60	25	70	70	91	13	37	1	1	0

[a)] For $s = 0$, {B,C} is selected.
[b)] For $s = 1$, {A,B} is selected.

Notes: Darker shades of blue indicate higher frequencies. The numerical percentage appears within each shaded cell. The profile numbering is the same as for Table 1.

Patterns for profiles 1 and 2 anticipate our overall conclusions. If, just for the moment, we confine attention to scoring rules and the Condorcet top-cycle extension, we would infer that 86% of subjects use scoring rules with $s > \frac{1}{3}$ (because they choose A in profile 1), that only 2-3% of subjects use scoring rules with $s < \frac{1}{3}$ (because they choose B or C in profile 2), and that 10% of subjects use the Condorcet rule (because the fraction of A increases from 86% to 96% between profiles 1 and 2). The third-option comparisons inherent in scoring rules also point to widespread violations of Arrow's IIA.

Choices for profiles 12 and 13 corroborate the low prevalence of the Condorcet rule, in that we see little indication of indeterminacy despite the presence of cycles. For profile 12, 91% of subjects agree on option C, which is rank-dominant, and for profile 13, 73% of subjects select option A, which minimizes the number of last-place ranks.

Choices for profiles 3, 4, and 5 corroborate the prevalence of concave scoring rules, which select options C (64%), A (63%), and B (75%), respectively. In contrast, all convex scoring rules and the Condorcet rule select options A (36%), B (37%), and C (24%). Either pattern is consistent with the Borda rule, which is only partially resolute for these profiles.

4.2 Main classification

4.2.1 Classification procedures

Next we classify individual subjects according to the social choice rules they appear to use. We deploy two Bayes classification procedures (Hastie et al., 2001), which assign each subject to the rule with the greatest posterior probability conditional on her observed choices under distributional assumptions detailed below.[44]

The first Bayes classification procedure only relies on consequential choices (i.e., it excludes information on indifference). The data for each subject then consists of 17 options, one for each preference profile. We make the following four assumptions (analogously to Costa-Gomes et al., 2001): (i) The prior distribution over pre-specified rules is uniform. (ii) For each of the 17 outcomes, the subject follows her assigned rule with probability $1-\epsilon$, and uniformly randomizes over the three options with probability ϵ, where ϵ is distributed uniformly over $[0,1]$. (iii) Decision errors are independent across preference profiles.[45] (iv) When a rule is irresolute, the subject randomizes uniformly over the prescribed choices.[46] We assign each subject the rule R^* and noise level ϵ^* that maximize this posterior probability, $P(R,\epsilon|c)$, where c is their choice vector.[47] When more than one rule maximizes the Bayesian posterior, we assign the subject to a maximizing rule at random.

When limiting consideration to consequential choices, irresolute rules have a built-in advantage: their predictions more easily encompass actual choices. Our first procedure creates a countervailing disadvantage: irresolute rules receive less "credit" (in terms of the increment to the posterior probability) than resolute rules when both turn out to be consistent with a given choice.[48]

The second Bayes classification procedure employs subjects' indifference statements along with their consequential choices. The data for each subject then consists of 17 subsets of optimal options, one for each preference profile. We continue to impose assumptions (i) and (iii), along with a slightly modified version of assumption (ii) (when deviating from her rule, the subject randomizes uniformly

[44]Simulations show that maximum-likelihood approaches perform poorly, in large part because they exhibit extreme bias towards less resolute rules in our setting. In contrast, the Bayes classifier performs well in simulations (see Appendix B.2).

[45]With this type of independence assumption, the method is sometimes called the *Naïve* Bayes Classifier.

[46]Because our displays present alternatives and workers in random order (redrawn in each round), even positional criteria (such as always choosing the option on the left) will yield uniform distributions.

[47]See Appendix B.1 for the explicit derivation of $P(R,\epsilon|c)$. This procedure is known as the Maximum A Posteriori (MAP) decision rule. The robustness of Bayesian classifiers has been documented extensively (see, e.g., Webb, 2010).

[48]This observation reflects the more general principle of the Bayesian Occam's Razor (see MacKay, 2003).

over the seven subsets of options, rather than the three options). Under these assumptions, a rule maximizes the posterior probability if and only if it maximizes the number of profiles for which it predicts the correct subset. As with our first procedure, we break ties at random.

The two procedures complement each other. On the one hand, restricting attention to consequential choices may yield more reliable results. On the other hand, the indifference data provide pertinent information, particularly inasmuch as the average Social Planner expresses irresoluteness for 18.01% of the profiles (just over 3 of 17).[49] Information on indifference also allow us to dispense with assumption (iv).

We pre-specify the 22 benevolent social choice rules discussed in Section 2.2. We also include a malevolent versions each rule by inverting each Stakeholder's preference ranking before applying the rule. Accordingly, we consider a total of 44 possible aggregation rules.

For our first classification procedure, we absorb the following three scoring rules into neighboring intervals: $s = \frac{3}{5}$, $s = \frac{4}{5}$, and $s = 1$; we do the same with their malevolent counterparts. The reason is that we cannot separately identify these rules without using information on indifference. As shown in Figure 2, each differs from nearby scoring rules on exactly one profile. Moreover, for the single differentiating profile, each prescribes the union of the options selected when the scoring parameter is slightly larger, and when it is slightly smaller. Because our first procedure always favors resoluteness over irresoluteness when both possibilities are consistent with the same observation, it will never select $s \in \{\frac{3}{5}, \frac{4}{5}, 1\}$ over all nearby scoring rules. In contrast, because there are *two* profiles for which the scoring rule with $s = \frac{1}{3}$ is irresolute while its neighbors are resolute, that rule can rationalize certain choice patterns that are inconsistent with its neighbors. A similar observation holds for $s \in \{0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}\}$.

4.2.2 Classification results

Figure 7 displays our main classification results. Panel A, which relies on incentivized choices alone, is strikingly similar to Panel B, which also incorporates indifference data. The following salient features merit emphasis.

First, the two most common aggregation methods are the Borda rule and the near-antiplurality rule (with $4/5 < s < 1$). Together, these rules account for 30 to 40 percent of all subjects, depending on the classification procedure.

Second a clear majority of all subjects (62.5% in Panel A, and 61.5% in Panel B) follow *strictly concave scoring rules* with $\frac{1}{2} < s \leq 1$. In contrast, strictly *convex* scoring rules ($s < \frac{1}{2}$) are unpopular. Surprisingly few subjects follow plurality rule (1% in Panel A, and 0.25% in Panel B.)

Third, fewer than 5% of subjects choose consistently with the Condorcet rule, and virtually no subjects follow other p-supermajority rules (supermajority and unanimity). A small minority follow scoring runoff rules (4.7% in Panel A and 7.2% in Panel B).

[49] In comparison, the set of our benevolent pre-specified rules produce a tie for 23.8% of the profiles, on average. Focusing only on the Condorcet rule and the set of all scoring rules we can identify, this figure falls to 14.3%. Social Planners designate all options as equally good for only 1.54% of profiles.

Because our pre-specified rules contain only one Condorcet extension (top cycle), these classifications could in principle understate the prevalence of rules in the Condorcet class. We therefore enlarge the set of pre-specified rules to include *all conceivable* Condorcet extensions, and reclassify subjects using our first procedure.[50] We find that 10.6% of subjects use a Condorcet extension, a modest increase, some of which would occur by chance.[51] It is unlikely that the low prevalence of Condorcet reasoning is a statistical or experimental artifact, because the Condorcet rule coincides on a large number of profiles with convex scoring rules, which are also unpopular. We provide further corroboration of this finding based on four-option problems in section 4.4.

Fourth, we see little if any evidence that subjects are either malevolent or lazy. The fraction of subjects assigned to malevolent rules is de minimis. Lazy subjects would be inclined to select the most easily implementable rule. Plurality rule is arguably the least cognitively demanding alternative, followed by plurality runoff, yet the vast majority of subjects evidently find both of these rules unappealing. As we have noted, most subjects employ strictly concave scoring rules which, with the exception of antiplurality, require more nuanced and complex judgments.

We obtain similar results for each method of presenting preference profiles. The only notable difference across display formats is that Borda is more common than near-antiplurality with the first and third display formats, while the opposite is true for the second display format; furthermore, this difference is larger when we use the indifference data (see Appendix C.1).

Random benchmark We draw statistical inferences by comparing our classification to a random-choice benchmark. We construct the benchmark by generating 4,000 artificial subjects whose simulated selections for each profile are uniformly distributed across options. We generate indifference data by designating each of the unchosen options as equally good based on independent Bernoulli draws. We choose the Bernoulli probability to match the average size of the best-option sets according to the actual subjects. We then assign each simulated subject to a pre-specified social choice rule using both of our procedures. We construct the distribution of subjects assigned to each rule under the null hypothesis of random choice by drawing 1,000 bootstrap samples of 405 simulated subjects each.

Figure 7 plots, for each rule, the mean fraction of simulated subjects assigned to that rule, as well as the 1st and 99th percentiles of the corresponding distribution. The average fractions of simulated subjects assigned to a malevolent rule is 50.4% in panel A and 40.9% in panel B, which exceed the ranges of the figures.[52] Turning to benevolent rules, in panel A we see that the fraction of subjects

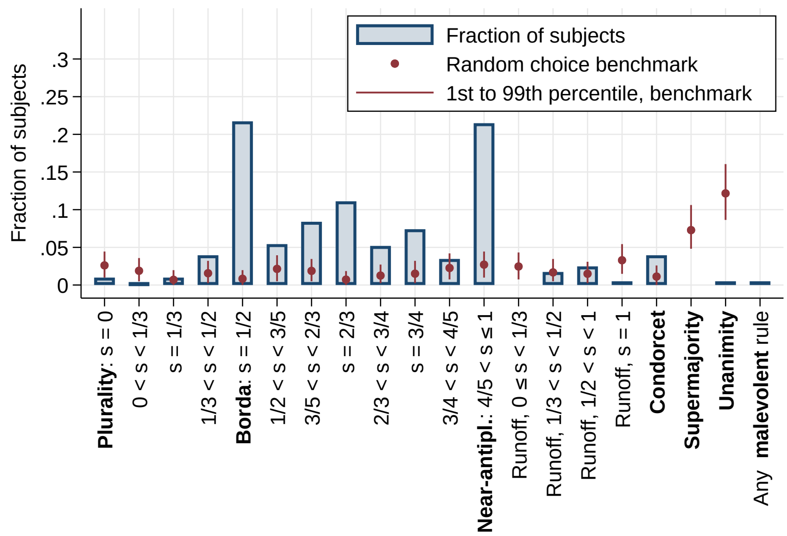
[50] Because Condorcet winners fail to exist for two of the 17 profiles, there are only nine resolute Condorcet extensions. As long as we include all of them, there is no need to add any irresolute Condorcet extensions, because our first procedure never assigns a subject to a less resolute rule that is equally consistent with the subject's choices (Bayesian Occam's Razor).

[51] The incremental subjects assigned to a Condorcet extension were originally assigned to a heterogeneous set of rules.

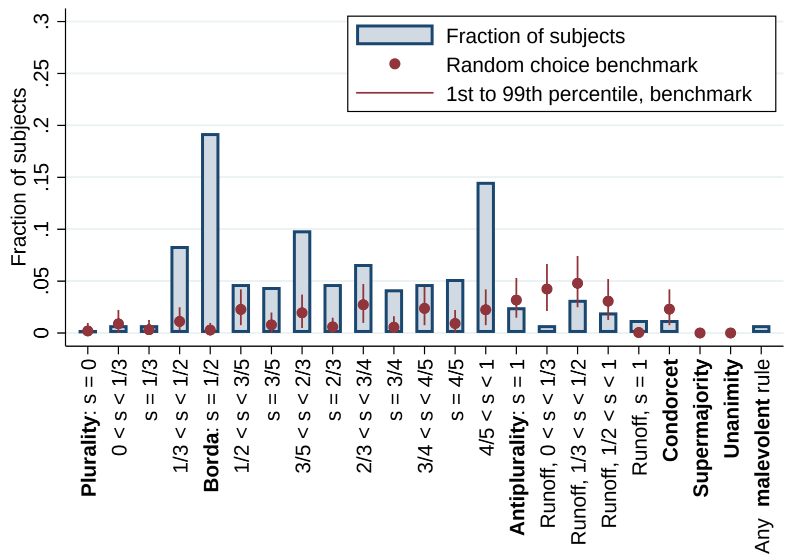
[52] Because our selection of profiles is not random, these benchmarks can deviate from 50%. To see this point, consider the example of profile 1. Only the choice of the rank-dominated alternative B is consistent with a malevolent scoring rule. Hence, one third of uniformly random choices would be classified as consistent with a malevolent scoring rule based on that profile alone. The same argument applies to each of our score-identifying profiles. These constitute a larger fraction of all profiles in the political domain than in the work domain.

Figure 7: Best fitting pre-specified rules

A. Using incentivized choice alone



B. Using incentivized choice and indifference data



Notes: Both panels show the frequency of subjects classified as following each pre-specified rule (blue bars). Each red dot indicates the fraction of 4,000 simulated subjects who randomize uniformly that the pertinent method classifies as a particular pre-specified type. From that sample of 4,000 simulated subjects, we draw 1,000 bootstrap samples of 405 subjects each to a construct a distribution of classifications. The red lines extend from the 1st to the 99th percentiles of that distribution. The random-choice benchmark for malevolent rules is not visible as it exceeds the range of the graph (50.4% and 40.9% in the work and political domains, respectively) .

assigned to every weakly concave scoring rule except $\frac{3}{4} < s < \frac{4}{5}$ exceeds the 99th percentile of the random benchmark. In contrast, the fractions of subjects assigned to the two most convex scoring rules ($s < \frac{1}{3}$) fall short of the 1st percentile of the random benchmark. While the fraction of subjects assigned to the Condorcet rule is small, it is larger than we would observe by chance, a conclusion for which we find corroboration in Section 4.4. Finally, scoring runoff rules, supermajority, and unanimity are no more prevalent (and in some cases significantly less prevalent) than for the random-choice benchmark. Similar conclusions generally follow from panel B, with the qualification that the random-choice simulations assign lower frequencies to scoring rules that coincide with the boundaries of distinguishable intervals.

Because our two classification procedures yield such similar results, for the sake of brevity we focus on the first procedure in subsequent sections.

Goodness of fit The average value of the estimated noise parameter ϵ^* provides a formal goodness-of-fit measure. Using incentivized choices alone, we obtain a mean noise parameter of 0.10, which signifies that the average subject chose randomly in just 1.7 of the 17 preference profiles. A whopping 42.0% of our subjects fit their assigned rule perfectly. For the remainder, the mean noise parameter is 0.16, equivalent to choosing randomly for 2.7 of the 17 preference profiles.[53] In contrast, the average estimated noise parameter for our simulated random-choice data is 0.46, and 0.57 when excluding simulated subjects classified as Supermajority or Unanimity (10 of 17 profiles).[54] When we use the indifference data, subjects' responses match the sets of options selected by their assigned rule for 14 of 17 profiles (85%). In contrast, for our random-choice benchmark, the corresponding number is 44%.[55] The fact that actual choices fit the rules dramatically better than random-choice simulations also indicates that the vast majority of our subjects paid attention and chose thoughtfully.

A more detailed picture of goodness of fit emerges from comparisons between the theoretical fingerprints associated with particular rules and the empirical fingerprints of the subjects assigned to those rules. The top half of Figure 8, panel A, plots the distribution of choices, by profile, for subjects classified as using the Borda rule. The bottom half of the same panel shows the Borda prescriptions. The fit is plainly tight. The highest choice frequency for any Borda-proscribed option is 16% (profile 6, option B). The second highest is 8% (profile 1, option C). The frequencies of all other Borda-proscribed options are at most 5%. While these subjects rarely depart from the Borda

[53] Appendix C.2 shows classifications of subjects based on the incentivized data split according to whether they fit their assigned rules perfectly or imperfectly. While the distributions are generally similar, Borda is more prevalent relative to near-antiplurality for perfectly fitting subjects. This pattern is mechanical: because Borda is less resolute than near-antiplurality, it has more scope for matching choice patterns perfectly.

[54] The mean estimated noise parameter in the random-choice data never equals its actual value (unity) because some random-choice sequences will always match some choice rules relatively well. The downward bias in the estimates of ϵ reflects overfitting.

[55] Alternatively, one can think of an empirical fingerprint as consisting of $3 \times 17 = 51$ binary variables, each of which indicates whether a given option is selected for a given profile. The average subject's choice coincides with the prediction of the best-fitting rule in 47 of these 51 cases (92%), which is far higher than the random-choice benchmark (65.9%).

Figure 8: Empirical fingerprints of classified subjects

A. Subjects classified as Borda

Empirical Choices

A	92	99	66	44	0	0	1	0	95	5	0	0	49	0	1	0	0
B	0	0	0	56	60	16	5	5	0	0	3	2	27	66	99	98	100
C	8	1	34	0	40	84	94	95	5	95	97	98	24	34	0	2	0

Rule prediction

A	100	100	50	50	0	0	0	0	100	0	0	0	33	0	0	0	0
B	0	0	0	50	50	0	0	0	0	0	0	0	33	50	100	100	100
C	0	0	50	0	50	100	100	100	0	100	100	100	33	50	0	0	0

B. Subjects classified as near-antiplurality

Empirical Choices

A	95	100	9	85	1	0	0	0	32	85	0	0	97	1	9	0	0
B	0	0	0	15	98	100	84	87	5	0	100	3	2	59	90	100	100
C	5	0	91	0	1	0	16	13	63	15	0	97	1	40	1	0	0

Rule prediction

A	100	100	0	100	0	0	0	0	0	100	0	0	100	0	0	0	0
B	0	0	0	0	100	100	100	100	0	0	100	0	0	50	100	100	100
C	0	0	100	0	0	0	0	0	100	0	0	100	0	50	0	0	0

Notes: In both panels, the top half displays the choices by subjects assigned to a certain rule. The bottom half displays the choices the rule predicts. The figures pertain to the classification procedure that does not use indifferent data. Darker shades of grey indicate higher frequencies. The numerical percentage appears within each shaded cell. For the theoretical rule, we take the frequency of each best option to be 50% for a two-way tie and 33% for a three-way tie.

rule, they do deviate somewhat from uniform resolution of ties (see, for example, profiles 3 and 14, where the proportions are one-third/two-thirds rather than half-half).

Panel B of Figure 8 provides analogous information for the near-antiplurality rule, which differs from Borda on 10 of the 17 profiles. The fit is also good, if slightly less tight. The highest choice frequency for any antiplurality-proscribed option is 32% (profile 9, option A). These subjects select four other proscribed options with frequencies ranging between 13% and 16% (see profiles 4, 7, 8, and 10), but they select each of the remaining 28 proscribed options with frequencies of 5% or less.

Overall, the 22 pre-specified benevolent social choice rules fit the data remarkably well given that there are more than 232 trillion (7^{17}) possible rules for this restricted set of 17 profiles, of which more than 18 trillion ($7^{14} \times 3^3$) are "reasonable" in the sense that they exhibit no Pareto violations. To further allay possible concerns about overfitting, we show in subsection 4.5 that our classification predicts well out of sample.

4.3 Is the classification complete?

To determine whether our classification analysis excludes empirically important rules, we look for clusters of subjects whose choices more closely resemble each others' than any of the pre-specified possibilities. Our approach is similar to that of Costa-Gomes and Crawford (2006).

To identify clusters, we use the k-modes clustering algorithm, which is an adaptation of the well-known k-means method to categorical data (Huang, 1998). The algorithm begins by arbitrarily selecting k subjects as initial cluster centers. Then it iterates two steps. First, each subject is assigned to the cluster for which the center matches her choices on the largest number of preference profiles. Second, cluster centers are updated: the new cluster center consists of the modal choice for subjects assigned to that cluster.[56] The algorithm terminates once the cluster centers stabilize. We use our pre-specified rules to create additional potential cluster modes that appear in all iterations. Because we do not use the indifference data for this exercise, we include every resolute version of each pre-specified rule. Following Costa-Gomes and Crawford (2006), if a subject is equidistant from a pre-specified rule and an endogenous cluster, we assign them to the pre-specified rule.

We search for clusters in the work domain fixing $k = 1, 2, 3, 5$, and 10. For $k = 1$, we run the algorithm 405 times, using each subject's choices as an initial cluster center once. For $k \geq 2$, we run the algorithm 1000 times, in each case randomly setting the initial cluster centers equal to the choices of k randomly selected subjects (excluding those who perfectly conform to pre-specified rules), and retaining the solution with the lowest total within-cluster distance.[57] We limit the pre-specified rules to all scoring rules and all Condorcet extensions, which have a total of 296 resolute components. We exclude unanimity because the resolute components (of which there are $2^3 \times 3^{14} > 38 \times 10^6$) encompass all choice patterns that are consistent with the Pareto principle. Similarly, we exclude supermajority

[56]In our setting, the modal choice is a vector specifying the most common selection for each preference profile.

[57]We use the Hamming distance, i.e. the number of profiles on which two choice sequences differ from each other.

because it is massively irresolute, so the number of resolute components is enormous. These exclusions are likely inconsequential given the small number of subjects assigned to these rules in section 4.2, and in any case the exclusion of an empirically important rule can only increase the fraction of subjects assigned to endogenous clusters. Other exclusions are attributable to redundancies.[58]

Table 5 displays the resulting endogenous clusters and the fractions of subjects assigned to each of them. For $k = 1$, just under 5% (19 of 405) of subjects are assigned to an endogenous cluster. For $k = 2$, the same cluster emerges plus a second that attracts under 2% of subjects (7 of 405). As we increase k further, these same two clusters remain, and the others encompass even fewer subjects. For $k = 10$, the smallest three clusters are degenerate (1 subject), indicating that we can find no other consequential similarities. Notably, the rule for the largest cluster is a one-profile deviation from near-antiplurality, from which it departs on profile 9 by selecting option A rather than C. We note that this is the only profile for which near-antiplurality selects an option ranked last by some Stakeholder, and not ranked first by any other Stakeholder.

4.4 Corroboration based on four-option social choice problems

Four-option profiles provide additional opportunities to distinguish cleanly between classes of social choice rules. Specifically, there are two four-option profiles, labeled "Condorcet-separating 1" and "Condorcet-separating 2" in Table 3 (numbered 18 and 19), for which the Condorcet winner is *rank-dominated*. These profiles distinguish between all Condorcet extensions and the entire class of *proper* scoring rules (ones that do not assign the same score to any two ranks) because the former must select the Condorcet winner while the latter cannot select a rank-dominated option.

We randomly intermingle both of these profiles with the three-option profiles in both domain blocks. While four-option problems are more cognitively demanding, recall that the experimental interface allows subjects to hide options, which makes it easy to identify Condorcet winners.

As shown in the left half of Table 4 (which pertains to the work domain), subjects choose the Condorcet winner roughly one-fifth of the time for each of these profiles. However, only 11.6% consistently choose the Condorcet winner for both profiles.[59] In contrast, because a negligible fraction of subjects choose option D, roughly 80% of the individual choices are consistent with a proper scoring rule. Moreover, for 69.6% of subjects, the pair of choices is consistent with some scoring rule in the following class: $[1, s_1, s_2, 0]$ where $s_1 = (2/3)^\gamma$, $s_2 = (1/3)^\gamma$, and $\gamma \in [0, \infty]$.[60]

In interpreting the preceding results, it is important to bear in mind that rules outside the Condorcet class can rationalize the selection of option A for profiles 18 and 19. As indicated in Table 3, plurality rule ($s_1 = s_2 = 0$) prescribes options A and B, while the plurality runoff rule prescribes

[58]For example, because plurality runoff always selects a plurality winner, the resolute components of the latter rule contain those of the former.

[59]This frequency, while relatively small, is significantly higher than would be observed by chance if the distributions of choices were independent across profiles, $0.205 \times 0.212 = 0.043$ (χ^2-test, $p < 0.001$).

[60]This class of rules is consistent with choosing B for both profiles, choosing C for both profiles, and (for a threshold value of γ) choosing B in one and C in the other.

Table 3: Four-alternative profiles.

Label	Index	Profile	Option selected by some proper scoring rule	Condorcet	Plurality (* = runoff)
Condorcet-separating 1	18	A A B B C B B A C D C C C D A D D D A B	B or C	A	{A*,B}
Condorcet-separating 2	19	A A B B C B B A D D C C C C A D D D A B	B or C	A	{A*,B}
Runoff-separating 1	22	A A B D D B B C A C C C D C B D D A B A	A or B or C	{A, B, C, D}	{A,D*}
Runoff-separating 2	23	A A C D D B B B A B C C D C C D D A B A	A or B or C	{A, B, C, D}	{A,D*}

Notes: Each preference profile is displayed as a 4×5-matrix; columns correspond to workers, rows correspond to preference ranks. A worker's r-ranked alternative is listed in the rth row. For Condorcet-cyclical profiles, we indicate the set of options in the top-cycle.

Table 4: Choices on class-separating profiles

Domain	Work				Politics			
Option	A	B	C	D	A	B	C	D
Condorcet winner	✓				✓			
Optimal for some scoring rule		✓	✓			✓	✓	
Condorcet-separating 1	0.205	0.694	0.099	0.002	0.188	0.746	0.064	0.002
Condorcet-separating 2	0.212	0.719	0.064	0.005	0.188	0.800	0.012	0.000
Consistent Condorcet	0.116				0.111			
Consistent Scoring	0.696				0.733			
Option	A	B	C	D	A	B	C	D
Plurality-runoff winner				✓				✓
Optimal for some scoring rule	✓	✓	✓		✓	✓	✓	
Runoff-separating 1	0.257	0.385	0.314	0.044	0.353	0.405	0.188	0.054
Runoff-separating 2	0.202	0.615	0.141	0.042	0.259	0.643	0.042	0.056
Consistent Runoff	0.012				0.028			
Consistent Scoring	0.884				0.930			

Notes: This table displays the fraction of subjects choosing each of the four options in each of the class-separating profiles. The first 262 subjects were not presented with the profile labelled Runoff-separating 2 in the political domain. The fraction of subjects consistently choosing in accordance with the plurality runoff rule in that domain is based on the remaining subjects.

Table 5: Clustering results

Profile	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Condorcet	C	A	A	B	C	C	C	C	A	C	C	ABC	ABC	B	B	B	B
Scoring																	
$\bar{s}$	$1/3$	$1/3$	$1/2$	$1/2$	$1/2$	$3/5$	$2/3$	$2/3$	$3/4$	$3/4$	$4/5$	-	$1/2$	0	0	0	0
$s < \bar{s}$	C	B$^{a)}$	A	B	C	C	C	C	A	C	C	C$^{a)}$	BC	-	-	-	-
$s > \bar{s}$	A	A	C	A	B	B	B	B	C	A	B	C	A	BC	B$^{b)}$	B	B
	Endogenous clusters, $k = 1$																
% subjects																	
4.69	A	A	C	A	B	B	B	B	A	A	B	C	A	B	B	B	B
Total: 4.69																	
	Endogenous clusters, $k = 2$																
% subjects																	
4.69	A	A	C	A	B	B	B	B	A	A	B	C	A	B	B	B	B
1.73	A	A	C	B	B	B	C	B	C	C	B	C	A	C	B	B	B
Total: 6.42																	
	Endogenous clusters, $k = 3$																
% subjects																	
4.69	A	A	C	A	B	B	B	B	A	A	B	C	A	B	B	B	B
1.73	A	A	C	B	B	B	C	B	C	C	B	C	A	C	B	B	B
0.25	A	A	C	A	A	A	B	A	B	A	A	B	A	A	C	B	A
Total: 6.67																	
	Endogenous clusters, $k = 5$																
% subjects																	
4.69	A	A	C	A	B	B	B	B	A	A	B	C	A	B	B	B	B
1.73	A	A	C	A	B	B	B	B	A	A	B	C	A	C	B	B	B
1.73	A	A	C	B	B	B	C	B	C	C	B	C	A	C	B	B	B
1.48	A	A	A	B	B	B	B	C	A	C	C	C	A	C	B	B	B
0.74	A	A	A	A	B	B	B	C	B	A	B	C	A	A	C	B	B
Total: 10.37																	
	Endogenous clusters, $k = 10$																
% subjects																	
4.69	A	A	C	A	B	B	B	B	A	A	B	C	A	B	B	B	B
1.73	A	A	C	B	B	B	C	B	C	C	B	C	A	C	B	B	B
1.48	A	A	A	B	B	B	B	C	A	C	C	C	A	B	B	B	B
1.23	A	A	A	B	B	B	C	C	A	C	C	C	A	B	B	B	B
0.99	A	A	C	A	B	B	C	C	C	C	C	C	A	C	B	B	B
0.74	A	A	C	A	B	B	C	B	A	C	C	C	C	B	B	B	B
0.74	A	A	C	A	B	B	B	C	C	C	B	C	A	B	B	B	B
0.25	A	A	C	A	A	A	B	A	B	A	A	B	A	A	C	B	A
0.25	B	C	B	B	A	B	C	C	C	B	A	A	B	A	C	A	B
0.25	A	C	A	B	C	C	B	B	A	A	B	C	A	B	C	C	B
Total: 12.35																	

Notes: This table shows the clusters which endogenously emerge in addition to our pre-specified rules from application of a k-modes algorithm, along with the fractions of subjects assigned to each such cluster. For ease of comparison, the top section of the table shows the choices of selected pre-specified rules.

A. Antiplurality rule ($s_1 = s_2 = 1$) also prescribes either A or B. One way to distinguish between Condorcet and plurality runoff on the one hand, and plurality and antiplurality on the other, is to examine the indifference data. Of those who select A for both profiles, 36.2% rate B as equally good, which suggests that more than a third of these selections reflect either plurality or antiplurality rule. It is also noteworthy that only 51.1% of these same subjects choose option C for profile 1 (consistent with both Condorcet and plurality rule), and also choose A, the Condorcet winner, rather than B or C, the plurality winners, for profile 2. Accordingly, choices for our four-option profiles imply that the fraction of subjects who implicitly follow some Condorcet rule in the work domain is on the order of 6% to 7%, which is consistent with our findings for three-option problems.

Profiles 22 and 23 in Table 3, which we also intermingled with the three-option profiles, leverage the principle of rank-dominance to provide clean separation between proper scoring rules and the plurality runoff rule. As shown in the left half of Table 4, subjects choose option D, the plurality runoff winner, roughly 5% of the time. Only 1.2% of subjects choose D for both profiles,[61] whereas 88.4% of subjects behave as if both choices reflect a single proper scoring rule from the family specified above.[62] Accordingly, only a small fraction of the Condorcet-consistent behavior for profiles 18 and 19 is likely attributable to use of the plurality runoff rule.

Four-option profiles such as 22 and 23 also enable more nuanced distinctions among rules. For example, they can provide a foundation for distinguishing concave scoring from a general aversion to lowest ranks; see Appendix C.3.

4.5 Contextual judgments or structural aggregation principles?

Next we investigate the extent to which our classification captures subjects' structural aggregation principles, as opposed to contextual judgments.

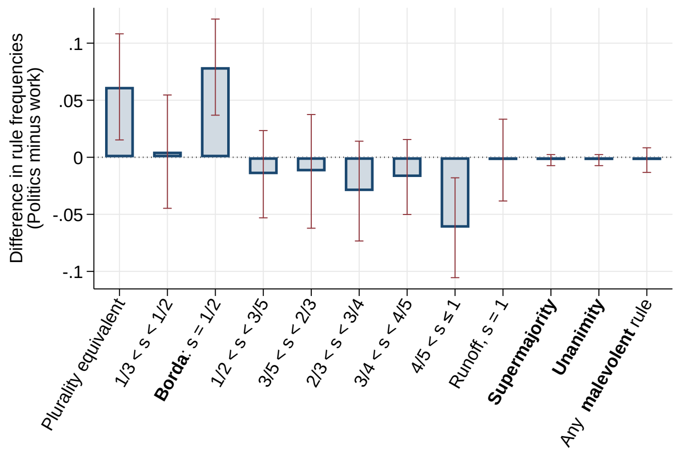
To the extent we have uncovered subjects' structural aggregation principles, we would expect our classifications to be stable across the work and political domains. Recall that, for the latter, we present subjects with a smaller number of preference profiles (7 instead of 17). While we can still identify scoring rules for which the parameter falls within the same intervals as before, we can no longer separately identify rules for which it lies at the boundaries of those intervals (i.e., points in $\mathcal{C}$).[63] Moreover, the fingerprints of scoring rules with $s \leq \frac{1}{3}$ (including plurality rule), the Condorcet rule, and all scoring runoff rules with $s < 1$ coincide for these seven profiles. For our current purposes, we will call this group the "plurality-equivalent rules."

[61] Though small, this fraction is significantly higher than would be observed by chance if the distributions of choices were independent across profiles (χ^2-test, $p < 0.001$).

[62] For these two profiles, every combination of choices in {A,B,C} is consistent with some scoring rule, except that a choice of option C in profile 23 implies $\gamma \geq 0.69$, and, accordingly, that C is also the best choice for profile 22.

[63] As explained in Section 4.2, consequential choice data can identify such scoring rules only if the two nearby scoring rules $s + \epsilon$ and $s - \epsilon$ differ on at least two preference profiles. Appendix A.2 shows the distance between all pairs of pre-specified rules in the political domain.

Figure 9: Comparison between work and political domains



Notes: Classifications are based only on consequential choices, and on the same set of seven preference profiles for both domains. The category 'plurality equivalent' includes scoring rules with $s \leq 1/3$, the Condorcet rule, and all scoring runoff rules with $s < 1$.

Figure 9 shows the differences in the resulting distributions of classifications for the work and political domains. Here we rely only on consequential choices, and we reclassify subjects in the work domain based on the same seven preference profiles to ensure comparability. Systematic differences are immediately apparent. Within the political domain, subjects are less likely to use strictly concave scoring rules, especially $\frac{4}{5} < s \leq 1$, and are more likely to use either the Borda rule or a plurality-equivalent rule. The difference is highly statistically significant (Wilcoxon signed-rank test, $p < 0.01$).[64] It bears emphasis, however, that these differences are of limited magnitude. The increase in the frequency of plurality-equivalent rules is on the order of 6%, and the decrease in the frequency of all strictly concave rules is 13.6%.

Even if aggregation rules are not completely stable across domains, they may still capture stable tendencies that reflect underlying structure. Notably, results for the four-option class-separating profiles in Table 4 exhibit strikingly little domain sensitivity.[65] It follows that people are generally attracted to scoring rules and generally averse to majoritarian criteria.

To evaluate the stability of aggregation preferences more comprehensively, we examine the out-of-sample and cross-domain predictive power of our classifications. For each evaluation, we designate a training set of preference profiles which we use to classify Social Planners based on their consequential choices, as well as a test set which we use to evaluate predictive success. We score predictions as follows. If the actual choice lies outside the predicted best-choice set, the score is zero. If it lies within

[64] We apply the test to the 400 (out of 405) subjects who are classified as following a scoring rule or a plurality-equivalent rule in both domains.

[65] While all subjects saw profile 23 in the work domain, only those in Sessions 8 - 11 saw it in the political domain.

the predicted best-choice set, the score is 1, $\frac{1}{2}$, or $\frac{1}{3}$ depending on whether there are 1, 2, or 3 best options. With this scoring system, if underlying choices are in fact uniformly random, the average score will be $\frac{1}{3}$ irrespective of the sizes of predicted best-choice sets. Averaging this score across all profiles in the test set, we obtain a measure of predictive accuracy that lies between 0 and 1.

We conduct four separate evaluations, two involving within-domain predictions, and two involving cross-domain predictions. For the within-domain predictions, we use leave-one-out-cross-validation (LOOCV): we designate one preference profile in the pertinent domain as the test set and use the other same-domain profiles as the training set. We repeat using each profile as the test set and average the predictive scores. For cross-domain predictions, we designate all profiles in one domain as the training set and all profiles in the other domain as the test set. Thus, the training set for the work domain consists of either 16 (within-domain) or 17 (cross-domain) profiles, while the training set for the political domain consists of either 6 (within-domain) or 7 (cross-domain) profiles.[66] Because of these differences, caution is warranted when comparing the various predictions to each other rather than to benchmarks.

As shown in Table 6, the average predictive accuracy scores for our four evaluations range from 0.758 (work to politics) to 0.853 (work to work). In all four cases, predictive performance is far superior to the random-choice benchmark (expected score of $\frac{1}{3}$).

Next we quantify the improvement in predictive performance that results from making appropriate assignments of subjects to pre-specified rules, rather than merely from pre-specifying a set of rules that generally coincide with reasonable tendencies. To this end, we offer three alternative benchmarks. For the first, we make predictions based on a uniform random assignment of each subject to one of the 44 pre-specified rules. The second benchmark is identical, except that we restrict these assignments to the 22 benevolent rules. The third refines the second by randomizing based on the estimated distribution of subjects across rules, rather than uniformly. There are two versions of the third benchmark, which differ according to whether we use the estimated distribution for the work domain or the political domain. In each case, we perform the procedure using 1000 bootstrap draws of our sample, and report both the mean and the 99th percentile of the resulting score.

For all four evaluations and all benchmarks, the average predictive accuracy score exceeds the 99th percentile of the benchmark's distribution by a wide margin. To be sure, merely pre-specifying a set of rules that generally coincide with reasonable tendencies accounts for a sizable portion of the improvement in predictive accuracy relative to the random-choice benchmark. However, the gain from making appropriate assignments of individual subjects to specific rules is considerable. To illustrate, focus on the most demanding benchmark for the work domain (3a), for which the mean benchmark score is 0.762. If all of our individual-level assignments were correct, we would obtain a score of 0.921. This number represents the theoretical maximum for the average score, given the

[66] For the purpose of predicting from the political domain to the work domain, when the worker's classification encompasses more than one of the 22 pre-specified rules (such as the plurality-equivalent rules), we select one of them at random.

Table 6: Out-of-sample predictive power of the Bayesian classification

	(1)	(2)
Dependent variable	Fraction of correct predictions (weighted by resoluteness)	
Test domain	Work	Politics
Predictions		
Training domain		
Work	0.853	0.758
Politics	0.812	0.759
Benchmarks		
1. Uniform, all prespecified rules		
Mean	0.445	0.398
99th percentile	0.485	0.440
2. Uniform, non-malevolent prespecified rules		
Mean	0.720	0.651
99th percentile	0.744	0.686
3.a Estimated rule frequencies, work domain		
Mean	0.762	0.653
99th percentile	0.779	0.680
3.b Estimated rule frequencies, political domain		
Mean	0.696	0.625
99th percentile	0.716	0.652

Notes: Within-domain predictions are based on leave-out-one cross-validation.

overall distribution of pre-specified rules; it is less than 1.0 because the rules are partially irresolute. An average score of 0.853 therefore implies that the individual-level assignments achieve 57% (i.e., $(0.853 - 0.762)/(0.921 - 0.762)$) of the maximum possible gain in predictive accuracy over a baseline that randomly scrambles those assignments.

Based on the strong out-of-sample predictive performance of our classifications, we conclude that assigned rules capture the essence of subjects' actual aggregation criteria. While the shift in the distributions of selections between the two domains points to a degree of context-specificity, the accuracy of the cross-domain predictions reassures us that ordinal aggregation also entails stable structural elements.

4.6 Do Social Planners make cardinal inferences?

Why are scoring rules so popular? One possibility is that people are comfortable with the concept of cardinal utility, so they use the ordinal information they receive to make cardinal inferences before aggregating and making a selection (in the spirit of Apesteguia et al., 2011). This hypothesis is consistent with the context-sensitivity of scoring rules documented in section 4.5, but that finding may have other explanations. In this section, we provide additional evidence that corroborates the cardinal inference hypothesis.

Informational interventions Suppose a decision maker chooses option A from the set $\{A, B, C\}$. According to the choice axiom known as Sen's α, if we remove option C from the opportunity set, the chooser will still select A. A decision maker who draws (cardinal) inference about an option from the entire choice set will violate this axiom. In the current context, there are two distinct ways to remove C: we can continue to inform the Social Planner about the Stakeholders' rankings of all three options even though C is no longer available (variant 1), or we can limit this information to the rankings of A and B (variant 2). Imagine that, in the original problem, the Social Planner makes cardinal inferences about the attractiveness of options A and B from their rankings relative to C and applies a Samuelson-Bergson social welfare function. In variant 1, that information remains available, so the Planner should respect Sen's α. However, in variant 2, that information is no longer available, so depending on the rankings, we should see violations of Sen's α.[67]

Each column of Table 7 provides a separate test of the cardinal-inference hypothesis based on this design. For column (1), the "Baseline" profile (part A) consists of three alternatives. Option C is obviously inferior, and is almost never chosen. For the "Option removed, rank information retained" profile (part B), we remove option C from the menu but continue to display rankings that include it. Choice frequencies are similar and the differences are statistically insignificant, so we do not reject Sen's α. For the "Option and rank information removed" profile (part C) we remove option C from the menu and from the Stakeholders' rankings. Choice frequencies change dramatically, and we resoundingly reject Sen's α. The natural explanation is that Stakeholders generally regard C as a bad option. The fact that one of them thinks B is worse than C, whereas none think A is worse than C, leads most Social Planners to choose A over B. But when that information is removed, nearly all Social Planners select B based on the majority preference relationship.

Results for the other two columns in Table 7 are qualitatively similar. In column (2), where we remove option A from a three-alternative profile, we reject Sen's α even when information about the rankings of A remains available, but the choice frequencies do not change dramatically, especially compared with the outcome when we also remove A from the rankings. In column (3), where we remove option D from a four-option profile, the choice frequencies barely change when information about the rankings of D remain available, but change dramatically when they are unavailable.

[67]The hypothesized failure of Sen's α would suggest a corresponding failure of Arrow's Independence of Irrelevant Alternatives (IIA): if withholding information on the ranking of an unavailable option affects choice, then presumably changing its ranking will do likewise. We observe such violations, for instance, by comparing choices for profiles 1 and 3. According to Arrow's IIA, the choice frequency of option C in profile 1 should equal the choice frequency of option A in profile 3. As Figure 6 shows, the empirical frequencies are 13% and 36%, respectively, which differ from each other at $p < 0.001$.

Table 7: Effects of removing alternatives

	(1)	(2)	(3)
A. Baseline			
Profiles	A A B B B B C A A A C B C C C	A C C C B B B B B A C A A A C	A A B B C B B A C D C C C D A D D D A B
Choice distribution	A: 0.630, B: 0.365, C: 0.005	A: 0.012, B: 0.748, C: 0.240	A: 0.205, B: 0.694, C: 0.099, D: 0.002
B. Option removed, rank information retained			
Profiles	A A B B B B ~~C~~ A A A ~~C~~ B ~~C~~ ~~C~~ ~~C~~	~~A~~ C C C B B B B B ~~A~~ C ~~A~~ ~~A~~ ~~A~~ C	A A B B C B B A C ~~D~~ C C C ~~D~~ A ~~D~~ ~~D~~ ~~D~~ A B
Choice distribution	A: 0.573, B: 0.427	B: 0.642, C: 0.358	A: 0.217, B: 0.674, C: 0.109
C. Option and rank information removed			
Profiles	A A B B B B B A A A	B C C C B C B B B C	A A B B C B B A C A C C C A B
Choice distribution	A: 0.007, B: 0.993	B: 0.007, C: 0.993	A: 0.622, B: 0.368, C: 0.010
D. p-values			
A vs. B	0.531	0.007	1.000
B vs. C	0.000	0.000	0.000
A vs. C	0.000	0.000	0.000

Notes: The profile in column 1 of Panel C is profile 14 of Table 1. In Panel A, the profile in column 3 is profile 18 from Table 3, and the profiles in the first and second columns are profiles 4 and 5 from Table 1, respectively. Columns 1 and 2 of Panel C both display the distribution of choices that subjects made for the single two-alternative profile they encountered. All decisions concern the work domain. p-values are based on two-sample Kolmogorov-Smirnov tests for equality of distributions.

Synthetic money-metric scoring parameters While the preceding findings are generally consistent with the cardinal-inference hypothesis, they do not explicitly document reliance on cardinal information. The next part of our analysis fills that gap. On the assumption that Social Planners are money-metric utilitarians, we use their stated beliefs about Stakeholders' reservation valuations for first-ranked, second-ranked, and third-ranked options to construct synthetic money-metric scoring parameters. We then ask whether the scoring parameters that rationalize actual choices are related to these synthetic utilitarian versions.

Formally, let $u_r^{i,d}$ denote Social Planner i's belief about the average Stakeholder's reservation valuation for his r-ranked alternative, for rank $r \in \{1, 2, 3\}$ and domain $d \in$\{work, politics\}.[68] The synthetic utilitarian scoring rule employs the score vector $[1, \tilde{s}_d, 0]$, where

$$\tilde{s}_d = \frac{u_2^{i,d} - u_3^{i,d}}{u_1^{i,d} - u_3^{i,d}}.$$

We examine the relation between the synthetic utilitarian scoring parameter $\tilde{s}$ and subjects' actual best-fitting scoring parameter s. In the work domain, we exclude 9 subjects who say they believe the average reservation wage is lower for the second-ranked option than for the first-ranked option, or lower for the third-ranked option than for the second-ranked option, on the grounds that they are likely inattentive or confused; in the political domain we drop 6 subjects. Thus, $\tilde{s}_d \in [0, 1]$. To avoid selection effects, we assign all subjects to their best-fitting scoring rules. We use OLS regression, using interval midpoints whenever best-fit scoring parameters are interval-identified. All regressions include fixed effects for preference profile presentation modes and for the order in which the subject made decisions about the work domain and the political domain.

The regression in Column 1 of Table 8 pools observations across the work and political domains. It includes a dummy variable for domain and clusters standard errors on the subject level. The coefficient of the synthetic scoring rule is positive and statistically significant. It remains positive when we run the regression separately for the two domains, but it is statistically significant only for the political domain. A potential explanation for this difference is that we measure the synthetic scoring parameter with less noise in the political domain because subjects are more familiar with political attitudes than with inclinations to perform various tasks, and consequently there is less attenuation of the coefficient.

Columns 4 to 6 add controls for Social Planners' risk attitudes and altruism, both measured as the percentile rank of the subject's average switching point for the two pertinent multiple decision lists. The coefficient of risk aversion is positive, as one would expect if Social Planners maximize the expected utility of a Stakeholder assessed behind a "veil of ignorance" concerning which Stakeholder has which ranking (in which case greater risk aversion implies greater concavity of the scoring rule), but the effect is weak. We also find that more altruistic Social Planners tend to use more concave scoring rules.

Next we ask whether the observed differences in scoring parameters across regimes is at least partially attributable to differences in cardinal inferences. In Column 7, we regress the difference between best-fit scoring rules across regimes for each subject on the corresponding difference between

[68] We assume that the Social Planner's inferences about positional money-metric utility do not vary from one profile to another, and consequently elicit these beliefs only once.

the synthetic scoring rules. The coefficient of interest remains positive and statistically significant. Indeed, this estimate closely resembles its counterparts in columns 1 and 4.

Table 8: Relation between best-fitting scoring parameters and beliefs about reservation prices.

VARIABLE	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Estimated scoring parameter s						
Domain							
Work	✓	✓		✓	✓		✓
Politics	✓		✓	✓		✓	✓
Differenced (cross-domain)							✓
Scoring parameter $\tilde{s}$	0.168**	0.104	0.210**	0.147**	0.080	0.193*	0.137**
implied by beliefs	(0.067)	(0.076)	(0.104)	(0.067)	(0.077)	(0.106)	(0.069)
Risk aversion %-rank				0.063*	0.049	0.076	
				(0.033)	(0.034)	(0.047)	
Altruism %-rank				0.080**	0.065*	0.094*	
				(0.035)	(0.034)	(0.051)	
Political domain	-0.133***			-0.134***			
	(0.012)			(0.012)			
Observations	795	396	399	795	396	399	390
Subjects	405	396	399	405	396	399	390

Notes: Parameters estimated with OLS using midpoint values of best-fit score. All regressions control for the type of preference profile presentation and for whether the political domain was displayed before or after the work domain. Regressions include all subjects with monotonic beliefs about reservation prices. Regressions include subjects with multiple switches in the multiple decision lists used to elicit risk preferences and altruism. For these subjects we set the corresponding variable to the mean of the values among the other subjects, and we include two indicator variables for multiple switching points, one for each characteristic. Standard errors in columns 1 and 4 are clustered by subject.

Having shown that differences in cardinal inferences contribute to the observed differences in scoring rules across regimes, we now provide evidence that subjects also attach domain-independent weights to the various ordinal ranks. In other words, they appear to deploy a blend of cardinalism and ordinalism. In Column 1 of Table 9, we use an OLS regression to measure the subject-level correlation between synthetic scoring rules in the political domain and the work domain, conditional on presentation mode and order. While we acknowledge that these measures are noisy and that noise would attenuate the measured correlation, the absence of any correlation suggests that the true cardinal inferences are unrelated across regimes. In Column 2, we use a similar regression to measure the (conditional) subject-level correlation between best-fit scoring rules in the two domains. It is positive and highly significant despite the apparent absence of any correlation in cardinal inferences. Thus, best-fit scoring rules appear to reflect some common factor that is unrelated to cardinal inferences.

Table 9: Relation between synthetic and best-fitting scoring parameters across domains.

VARIABLES	(1)	(2)
	Scoring parameter in political domain	
	Synthetic	Best-fitting
Scoring parameter in work domain		
Synthetic	-0.074	
	(0.058)	
Best-fitting		0.638***
		(0.064)
Observations	390	390
R^2	0.012	0.261

Notes: OLS regressions. Both equations control for the type of preference profile presentation and for whether the political domain was displayed before or after the work domain. Regressions include all subjects with monotonic beliefs about reservation prices.

5 General population samples

In this section, we ask whether our main conclusions extend to general population samples. We also test whether the general public in countries with divergent social and political traditions, the United States and Sweden (Alesina and Glaeser, 2004), use similar or divergent criteria when aggregating ordinal preferences, and we explore external validity by asking whether ordinal aggregation criteria among the general public are related to attitudes toward political processes. Our cross-cultural comparison can potentially help explain why different nations gravitate toward different types of policies. As Alesina and Angeletos (2005) point out, policies may diverge either because different cultures have fundamentally different preferences, or because of beliefs, historical accidents, institutions, and/or equilibrium selection.

In these supplemental experiments, each social choice entails the allocation of $20 (in the U.S.) or SEK170 (in Sweden) to one of four charities: Doctors without Borders, Unicef, Oxfam, and the International Fund for Animal Welfare.[69] These organizations are well known in both countries and represent diverse causes that have broad appeal across the political spectrum. We recruited 712 Swedish and 805 U.S. voting-age citizens through Dynata and Lucid to serve as Social Planners. Each Social Planner aggregates the preferences of same-country Stakeholders, recruited through pollfish. All Social Planners see the same reduced set of profiles we used for the political domain in our main experiment, as well as either the actual profile for a group of Stakeholders or a randomly generated profile. Social Planners know that one of the preference profiles they consider corresponds to real Stakeholders with 10% probability. We use abridged instructions that carefully explain the preference

[69] At the time of the study, these amounts were roughly equivalent according to market exchange rates ($1 $\approx$ SEK 8.50).

displays (see Appendix E.2). Subjects must pass an abbreviated comprehension check to participate.[70] See Appendix D for additional implementation details and demographic summary statistics.

To make our samples more representative of the respective general populations, we weight observations as follows. For the US sample, we use the 2018 General Social Survey (Smith et al., 2019) to generate weights for 16 population categories defined by (i) gender (male, female), (ii) race (white, black, hispanic, other), and (iii) political party preference (Democrat, Republican). For the Swedish sample, we use data from Statistics Sweden (2018) to generate weights for 12 categories defined by (i) gender, and (ii) political party preference (Left Party, Social Democratic Party, Green Party, Centre Party, Moderate Party, Sweden Democrats).[71]

Because these data require us to categorize subjects based on seven three-option profiles rather than seventeen, it is important to bear in mind the following limitations. First, categorizations based on fewer choices are more sensitive to one or two noisy selections. Second, because it is harder to detect irresoluteness with fewer profiles when using only consequential choices, certain scoring rules such as Borda are more difficult to distinguish from neighboring rules. Third, some rules become entirely indistinguishable. In particular, the Condorcet rule, plurality rule, and plurality runoff all have the same implications for the reduced set of three-option profiles. In light of the first two issues, we rely on seven-profile classifications mainly to make comparisons across subject pools. The distributions themselves should be taken with a grain of salt, as there are systematic differences between the seven- and seventeen-profile classifications for our student population in the work domain.

In light of the third issue, we begin by discussing the four-option class-separating profiles, which cleanly differentiate between proper scoring rules, Condorcet rules, and plurality runoff, as in Section 4.4 (recall Table 3). Only 8.0% of U.S. subjects and 5.8% of Swedish subjects consistently choose the Condorcet winner for both Condorcet-separating profiles. Moreover, 24.4% (U.S.) and 14.6% (Sweden) of those subjects express indifference between options for at least one of these profiles, which suggests that they follow plurality rule. That leaves roughly 5% to 6% of subjects in the Condorcet category. Likewise, small minorities (3.1% of US subjects and 1.7% of Swedish subjects) choose the plurality runoff winner in both runoff-separating profiles. All of these results closely parallel our findings for the student sample.

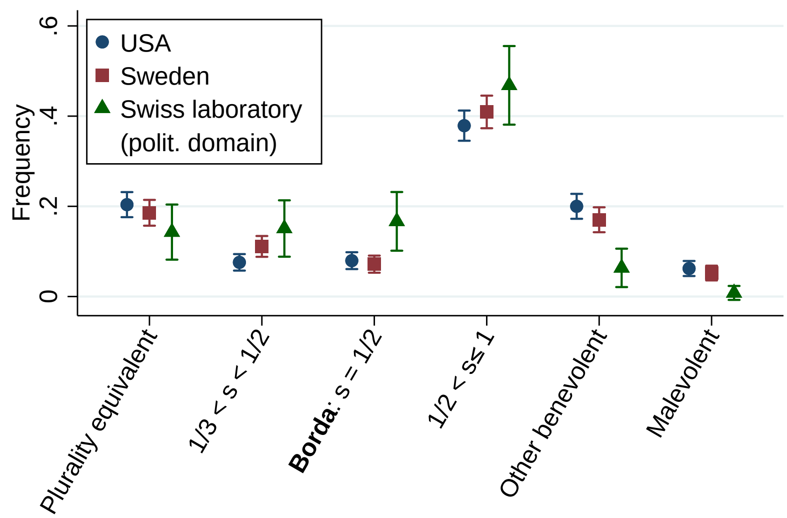
Next, using the Bayes classifier, we assign each subject to the best-fitting rule based on their consequential choices for the seven three-option profiles. Figure 10 displays the results. To facilitate comparisons across samples, we consolidate several categories (strictly concave scoring rules, all benevolent alternatives aside from scoring rules, and all malevolent rules). The figure shows classification frequencies for the U.S. sample, the Swedish sample, and the Swiss student sample (political domain). When comparing the results for the student sample to either of the general population samples, one

[70] Subjects recruited through Dynata could proceed in spite of failing the comprehension check. We exclude such observations from our data.

[71] For both countries, these are the most recent data that include the requisite information for each combination of characteristics.

should bear in mind that the domains are similar but not identical. Because we have found that aggregation rules vary somewhat across domains (Section 4.5), we would not expect the classifications to match perfectly.

Figure 10: Classification of general population samples to pre-specified rules



Notes: Bayes classifications based on consequential choices for the seven three-option preferences profiles indicated with stars in Table 1. General population observations weighted to make the samples representative with respect to gender, party preference, and race (US only).

Differences in classification frequencies between the Swedish and U.S. general population samples are remarkably small and statistically insignificant. Hence, the fundamental aggregation preferences of U.S. and Swedish citizens are extremely similar, which suggests that they do not contribute to policy divergences. The general population distributions also resemble the distribution for the student sample despite the difference in domains.[72] Like the low prevalence of Condorcet rules (noted above), the popularity of strictly concave scoring rules proves to be a robust phenomenon. Notably, the elevated frequencies of “Other benevolent rules” are attributable primarily to the antiplurality runoff rule, which is closely related to the strictly concave alternatives. The modestly higher frequencies of malevolent rules suggest that our general population samples may include slightly higher fractions of subjects who did not take the tasks seriously. Similarly, the slightly elevated use of plurality rule, the least cognitively demanding alternative, could be attributable to greater laziness.

Finally, we ask whether our measures of aggregation preference correlate with self-reported attitudes toward political processes. Our survey presents subjects with two hypothetical candidates for

[72]As expected, the Borda rule appears to be less popular when we classify fewer subjects based on seven profiles rather than seventeen. When we also employ the indifference data to classify subjects based on seven profiles, Borda frequencies increase substantially.

political leadership of the nation. It describes Candidate 1 as polarizing: "*Most citizens either love him or hate him. There is hardly anyone who does not have a strong opinion. If candidate 1 were elected, some citizens would be exhilarated, many others would be devastated, and nobody would be indifferent.*" We describe Candidate 2 as a compromise alternative: "*While he is nobody's greatest favorite, most citizens would be ok with candidate 2. If he were elected, nobody would be exhilarated, nobody would be devastated.*"[73] We ask subjects "*Which candidate better represents the will of the citizens of the nation?*" In addition, subjects indicate their level of agreement or disagreement with each of the following two statements: "*The political system should strive for compromise solutions that everyone can live with even if the result is nobody's absolute favorite*," and "*What the majority wants is right for a country, even if that makes some citizens suffer.*" We construct an index of preference for compromise policies as follows: we assign values 0, $\frac{1}{3}$, $\frac{2}{3}$, and 1 to the responses 'strongly disagree', 'disagree', 'agree', and 'strongly agree' for the first of these two questions, invert the scores for the second, and average responses across them.

Table 10: Relation between behavioral measures of preference for compromise and self-reports

VARIABLES	(1) Prefers compromise candidate	(2)	(3) Prefers compromise solutions	(4)
Mean of the dep. var.	0.801		0.686	
Scoring parameter	0.077*	0.070*	0.035	0.022
	(0.042)	(0.042)	(0.056)	(0.055)
Demographic controls		✓		✓
Subjects	1,404	1,404	1,404	1,404

Notes: OLS regressions. All regressions control for sample provider and display fixed effects. Demographic controls include nationality, gender, age, marital status, percentile rank of respondents' education within the sample in their respective country, indicators for being unemployed and for being part-time employed, as well as the percentile rank of income within the sample for the respondents' country.

Table 10 examines the relation between these self-reported attitudes and our behavioral measure of aggregation preference. For Columns 1 and 2, the dependent variable is a binary indicator of the preference for the compromise candidate; in Columns 3 and 4, it is our measure of preference for compromise policies. Each column reports an OLS regression pooling over U.S. and Swedish subjects. The main independent variable is the subject's scoring parameter, which we compute by restricting the Bayes classification to scoring rules and assigning interval midpoints. In all cases, we control for sample provider and display fixed effects. While the results are not strong, they are nevertheless suggestive. As expected, those who deploy more concave scoring rules tend to prefer

[73] We randomize the order of the descriptions as well as the candidate labels.

greater compromise in political processes. The relationship is significant at the 10% level in Columns 1 and 2, and statistically insignificant but directionally consistent in Columns 3 and 4.[74]

6 Conclusion

Our objective in this paper has been to understand the judgments people make when they aggregate others' ordinal preferences. We find that the overwhelming majority of subjects behave as if they rely on *scoring rules*. The Borda and antiplurality rules are the most common alternatives, and a sizeable majority of subjects uses strictly concave scoring rules, indicating a pronounced preference for compromise over majoritarian solutions. Plurality rule, Condorcet rules, supermajority, unanimity, and various runoff rules are relatively rare. The classification's fit is excellent, and clustering analysis reveals no major omissions from our list of pre-specified rules. We find systematic and significant differences in the distributions of rules between the work domain and the political domain, but these differences are of limited magnitude. Because our classifications are highly predictive of choices out of sample, including across domains, we infer that ordinal aggregation also entails stable structural elements. While subjects act as if they attach substantial domain-independent weight to the various ordinal ranks, we also find strong indications that subjects aggregate ordinal preferences based in part on inferences about cardinal utility, thus deploying a blend of cardinalism and ordinalism. Supplemental experiments show that the distributions of aggregation preferences in the U.S. and Sweden, countries with divergent political and social traditions, are remarkably similar, and both resemble the distribution for the student sample used in our main experiment. Even so, there is suggestive evidence that the use of more concave scoring rules in experimental decisions correlates with a preference for electing compromise candidates.

Our analysis suggests many potential directions of inquiry for future work. The mere fact that people rely to some degree on cardinal inferences when presented with ordinal information, and treat money as a cardinal index when presented with monetary payoffs, does not imply that they are money-metric welfarists. In the spirit of Roberts (1980), it would be useful to understand which types of cardinal comparisons people find meaningful, and which (if any) they find arbitrary. Do they ignore ordinal information when cardinal information is available, or do they use both? Are money-metric measures of well-being compelling when consequences are non-monetary? Another potential line of inquiry would investigate the relationships between preferences over rules revealed by choices over outcomes, preferences over rules revealed by choices over rules (as in Engelmann and Grüner, 2017; Hoffmann and Renes, 2017; Engelmann et al., 2020), and preferences over rules implied by approval of axioms (in the spirit of Nielsen and Rehbeck, 2020 and others). It would also be of interest to investigate how awareness of manipulability influences rule selection.

[74] Ambuehl et al. (2021b) offer corroborative evidence for external validity: German elected representatives of more centrist political parties use more concave scoring rules.

References

Alesina, Alberto and Edward Glaeser, *Fighting poverty in the US and Europe: A world of difference*, Oxford University Press, 2004.

_ **and George-Marios Angeletos**, "Fairness and redistribution," *American Economic Review*, 2005, *95* (4), 960–980.

Almås, Ingvild, Alexander W Cappelen, and Bertil Tungodden, "Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians?," *Journal of Political Economy*, 2020, *128* (5), 1753–1788.

Ambuehl, Sandro, B Douglas Bernheim, and Axel Ockenfels, "What Motivates Paternalism? An Experimental Study," *American Economic Review*, 2021, *111* (3).

_ **, Sebastian Blesse, Philip Dörrenberg, Christoph Feldhaus, and Axel Ockenfels**, "The Social Welfare Functions of Elected Representatives," *unpublished*, 2021.

Andreoni, James, Deniz Aydin, Blake Barton, B Douglas Bernheim, and Jeffrey Naecker, "When Fair Isn't Fair: Understanding Choice Reversals Involving Social Preferences," *Journal of Political Economy*, 2020, *128* (5).

Apesteguia, Jose, Miguel A Ballester, and Rosa Ferrer, "On the justice of decision rules," *The Review of Economic Studies*, 2011, *78* (1), 1–16.

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

_ , *Social choice and individual values*, John Wiley & Sons, 1951.

_ **, Amartya Sen, and Kotaro Suzumura, eds**, *Handbook of social choice and welfare*, Vol. 1, Elsevier, 1991.

_ **,** _ **, and** _ **, eds**, *Handbook of social choice and welfare*, Vol. 2, Elsevier, 2010.

Baharad, Eyal and Shmuel Nitzan, "The Borda rule, Condorcet consistency and Condorcet stability," *Economic Theory*, 2003, *22* (3), 685–688.

Baker, John, "Election of the Green Party Cathaoirleach, 2007," *Irish Political Studies*, 2008, *23* (3), 431–440.

Black, Duncan, *The theory of committees and elections*, Springer, 1958.

Bolton, Gary E and Axel Ockenfels, "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: comment," *American Economic Review*, 2006, *96* (5), 1906–1911.

Borda, JC, "Memoire sur les Elections an Scrutin," *Paris, France: Academie Royale de Science*, 1781.

Boutilier, Craig, Ioannis Caragiannis, Simi Haber, Tyler Lu, Ariel D Procaccia, and Or Sheffet, "Optimal social choice functions: A utilitarian view," *Artificial Intelligence*, 2015, *227*, 190–213.

Brandt, Felix, Vincent Conitzer, Ulle Endriss, Jerome Lang, and Ariel D. Procaccia, *Handbook of Computational Social Choice*, Cambridge University Press, 2016.

Camerer, Colin F, Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan et al., "Evaluating replicability of laboratory experiments in economics," *Science*, 2016, *351* (6280), 1433–1436.

Condorcet, Marquis de, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, L'Imprimerie Royale, 1785.

_ , *On the constitution and the functions of provincial assemblies*, Vol. 1994 of *Oeuvres de Condorcet. Paris.* 1788.

Cooper, David J and John H Kagel, "Other-regarding preferences," *The Handbook of Experimental Economics*, 2016, *2*, 217.

Costa-Gomes, Miguel A and Vincent P Crawford, "Cognition and behavior in two-person guessing games: An experimental study," *American Economic Review*, 2006, *96* (5), 1737–1768.

Costa-Gomes, Miguel, Vincent P Crawford, and Bruno Broseta, "Cognition and behavior in normal-form games: An experimental study," *Econometrica*, 2001, *69* (5), 1193–1235.

de Laplace, Pierre-Simon, "Leçons de Mathématiques, donnés à l'École Normale en 1795," *Journal de l'École Polytechnique*, 1812, *2* (17-18).

Debian Project, "Constitution for the Debian Project, v1.7," *https://www.debian.org/devel/constitution*, 2016.

Engelmann, Dirk and Hans Grüner, "Tailored Bayesian Mechanisms: Experimental Evidence from Two-Stage Voting Games," *Unpublished Manuscript*, 2017.

_ **and Martin Strobel**, "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments," *American Economic Review*, 2004, *94* (4), 857–869.

_ **, Hans Peter Gruener, Timo Hoffmann, and Alex Possajennikov**, "Minority Protection in Voting Mechanisms: Experimental Evidence," *Unpublished Manuscript*, 2020.

Epper, Thomas, Ernst Fehr, and Julien Senn, "Other-regarding preferences and redistributive politics," *University of Zurich, Department of Economics, Working Paper*, 2020, (339).

Faravelli, Marco, "How context matters: A survey based experiment on distributive justice," *Journal of Public Economics*, 2007, *91* (7-8), 1399–1422.

Featherstone, Clayton R, "Rank efficiency: Investigating a widespread ordinal welfare criterion," *unpublished*, 2019.

Fehr, Ernst and Urs Fischbacher, "Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives," *The Economic Journal*, 2002, *112* (478), C1–C33.

Fishburn, Peter C, *The theory of social choice*, Princeton University Press, 2015.

Fisman, Raymond, Pamela Jakiela, and Shachar Kariv, "The distributional preferences of Americans, 2013-2016," *Unpublished Manuscript*, 2021.

Freeman, Rupert, Markus Brill, and Vincent Conitzer, "On the axiomatic characterization of runoff voting rules," in "Twenty-Eighth AAAI Conference on Artificial Intelligence" 2014.

Gaertner, Wulf, *Distributive justice: An overview of experimental evidence*, Oxford University Press, 2009.

_ **and Erik Schokkaert**, *Empirical social choice: questionnaire-experimental studies on distributive justice*, Cambridge University Press, 2012.

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

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, *The elements of statistical learning*, Vol. 1, Springer series in statistics New York, 2001.

Herreiner, Dorothea K and Clemens Puppe, "Inequality aversion and efficiency with ordinal and cardinal social preferences—An experimental study," *Journal of Economic Behavior & Organization*, 2010, *76* (2), 238–253.

Hoffmann, Timo and Sander Renes, "Flip a coin or vote: Choosing Group Decision Rules," *Unpublished Manuscript*, 2017.

Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *American Economic Review*, 2002, *92* (5), 1644–655.

Huang, Zhexue, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data Mining and Knowledge Discovery*, 1998, *2* (3), 283–304.

Jackson, Matthew O and Leeat Yariv, "Present bias and collective dynamic choice in the lab," *American Economic Review*, 2014, *104* (12), 4184–4204.

Kara, Ayça E Giritligil and Murat R Sertel, "Does majoritarian approval matter in selecting a social choice rule? An exploratory panel study," *Social Choice and Welfare*, 2005, *25* (1), 43–73.

Konow, James, "Which is the fairest one of all? A positive analysis of justice theories," *Journal of Economic Literature*, 2003, *41* (4), 1188–1239.

Kuziemko, Ilyana, Ryan W Buell, Taly Reich, and Michael I Norton, ""Last-place aversion": Evidence and redistributive implications," *The Quarterly Journal of Economics*, 2014, *129* (1), 105–149.

Laslier, Jean-François, "And the loser is... plurality voting," in "Electoral systems," Springer, 2012, pp. 327–351.

MacKay, David JC, *Information theory, inference and learning algorithms*, Cambridge University Press, 2003.

Martinangeli, Andrea FM and Lisa Windsteiger, "Last word not yet spoken: a reinvestigation of last place aversion with aversion to rank reversals," *Experimental Economics*, 2020, pp. 1–21.

Merrill III, Samuel, "A comparison of efficiency of multicandidate electoral systems," *American Journal of Political Science*, 1984, pp. 23–48.

Mill, John Stuart, *On liberty*, Longmans, Green, Reader, and Dyer, 1869.

Myerson, Roger B, "Axiomatic derivation of scoring rules without the ordering assumption," *Social Choice and Welfare*, 1995, *12* (1), 59–74.

Newenhizen, Jill Van, "The Borda method is most likely to respect the Condorcet principle," *Economic Theory*, 1992, *2* (1), 69–83.

Nielsen, Kirby and John Rehbeck, "When Choices are Mistakes," *Available at SSRN 3481381*, 2020.

Pivato, Marcus, "Asymptotic utilitarianism in scoring rules," *Social Choice and Welfare*, 2016, *47* (2), 431–458.

Portmann, Lea and Nenad Stojanović, "Electoral discrimination against immigrant-origin candidates," *Political Behavior*, 2019, *41* (1), 105–134.

Richie, Robert, "Instant runoff voting: what Mexico (and others) could learn," *Election Law Journal*, 2004, *3* (3), 501–512.

Roberts, Kevin, "Social choice theory and the informational basis approach," in Christopher Morris, ed., *Amartya Sen*, Cambridge University Press, 2005, chapter 5, pp. 115–137.

Roberts, Kevin WS, "Interpersonal comparability and social choice theory," *The Review of Economic Studies*, 1980, pp. 421–439.

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

Smith, Tom W., Michael Davern, Jeremy Freese, and Stephen L. Morgan, *General Social Surveys, 1972-2018*, Chicago: NORC, 2019.

Statistics Sweden, "The Party Preference Survey in May 2018," https://www.scb.se/publication/34439 2018.

Thomson, Keela S and Daniel M Oppenheimer, "Investigating an alternate form of the cognitive reflection test," *Judgment and Decision Making*, 2016, *11* (1), 99.

Toplak, Jurij, "The parliamentary election in Slovenia, October 2004," *Electoral Studies*, 2006, *25* (4), 825–831.

Webb, Geoffrey I, "Naïve Bayes," *Encyclopedia of machine learning*, 2010, *15*, 713–714.

Weber, Matthias, "Choosing the rules: Preferences over voting systems for assemblies of representatives," *Journal of Economic Behavior & Organization*, 2017.

Weber, Robert J, "Comparison of public choice systems," *Cowles Foundation for Research in Economics, Yale University*, 1978.