Condorcet Winners and Anscombe's Paradox Under Weighted Binary Voting

Carmel Baharav, Andrei Constantinescu and Roger Wattenhofer

Abstract

We consider voting on multiple independent binary issues. In addition, a weighting vector for each voter defines how important they consider each issue. The most natural way to aggregate the votes into a single unified proposal is *issue-wise majority* (IWM): taking a majority opinion for each issue. However, in a scenario known as *Ostrogorski's Paradox*, an IWM proposal may not be a Condorcet winner, or it may even fail to garner majority support in a special case known as *Anscombe's Paradox*.

We show that it is co-NP-hard to determine whether there exists a Condorcet-winning proposal even without weights. In contrast, we prove that the *single-switch* condition provides an Ostrogorski-free voting domain under identical weighting vectors. We show that verifying the condition can be achieved in linear time and no-instances admit short, efficiently computable proofs in the form of forbidden substructures. On the way, we give the simplest linear-time test for the *voter/candidate-extremal-interval* condition in approval voting and the simplest and most efficient algorithm for recognizing single-crossing preferences in ordinal voting.

We then tackle Anscombe's Paradox. Under identical weight vectors, we can guarantee a majority-supported proposal agreeing with IWM on strictly more than half of the overall weight, while with two distinct weight vectors, such proposals can get arbitrarily far from IWM. The severity of such examples is controlled by the maximum average topic weight $\tilde{w}_{max}$: a simple bound derived from a partition-based approach is tight on a large portion of the range $\tilde{w}_{max} \in (0, 1)$. Finally, we extend Wagner's rule to the weighted setting: an average majority across topics of at least $\frac{3}{4}$'s precludes Anscombe's paradox from occurring.

1 Introduction

There are numerous scenarios in which people must decide on a slate of binary issues and come up with a single outcome for each topic. When political parties form a platform, they must aggregate their base's opinions and provide a unified set of stances on numerous separate issues. Similarly, when voters head to the ballot box for local elections, they typically vote yes or no on a series of initiatives. The election results in one outcome for each individual topic. On a smaller scale, a group of flatmates might decide on a series of unrelated topics and generate a plan for living together. For example: Should the kitchen be cleaned once a week or twice a week? Should we get the red couch or the yellow couch?

The most natural way to decide the final outcome in all of these scenarios is to take the majority opinion on each individual topic and aggregate them into a unified party platform, legislative agenda, or roommate contract. However, this approach can yield a surprisingly undesirable outcome: a majority of the voters may actually be more unhappy with this result than if the opposite decision were made on every issue (known as *Anscombe's Paradox* [2]). How can this arise? Consider a setting with 5 voters and 3 independent binary issues. The following table illustrates the preferences of each voter on each of the 3 issues: $+1$ is in favor and -1 is against:

	Issue 1	Issue 2	Issue 3
v_1	+1	-1	-1
v_2	-1	+1	-1
v_3	-1	-1	+1
v_4	+1	+1	+1
v_5	+1	+1	+1

Now, assume each voter would only vote in favor of proposals that they agree with on more than half of the issues (in the paper, a voter will abstain when agreeing with a proposal on exactly half of the issues). Taking the majority on each topic yields the proposal $(+1, +1, +1)$. However, voters 1, 2, and 3 all disagree with a majority of this proposal. Therefore, if we posed this proposal for a vote, a majority of voters would vote against it. If, instead, we posed the opposite proposal $(-1, -1, -1)$, then voters 1, 2, and 3 would support it, and it would win the majority vote. Hence, in this scenario, the proposal comprising the *minority* opinion on each topic wins the majority vote, whereas the proposal comprising the majority opinions fails to get majority support.

An equivalent view on the previous scenario positions Anscombe's paradox in a broader context: instead of assuming a vote on a single proposal with people voting for/against it, let us assume that the vote happens between two competing proposals p and p' and each voter votes for whichever of p and p' agrees with their views on more topics, abstaining in case of equality. Seen as such, Anscombe's paradox is the situation where an *issue-wise majority* (IWM) proposal p loses the majority vote against $p' = \overline{p}$, defined as the opposite proposal of p. A less extreme variant of the paradox, known as Ostrogorski's paradox [35] happens when an IWM proposal p loses against some proposal p', not necessarily $\overline{p}$. Settling on the IWM proposal in such cases can lead to daunting situations where one of its opposers calls a final vote between p and p' that "surprisingly" unveils general dissatisfaction with what was otherwise a perfectly democratically chosen outcome.

Consequently, multi-issue aggregation mechanisms need to balance the tension between two majoritarian processes: majority on the individual topics and majority when proposals are compared to one another. In terms of the first, the chosen proposal should ideally stay somewhat close to IWM. In terms of the second, the chosen proposal should not be easily refuted by calling a vote against some other proposal.

Even when voters consider the issues to be of equal importance in their decision-making, we get paradoxical situations. However, in reality, voters rarely consider all issues to be equally important and often disagree on their importance; e.g., a Pew Research study from June 2023 indicated that in the United States, there were massive differences in perceived issue importance along partisan lines [36]. Some voting advice applications already attempt to account for personalized issue-importance, such as Smartvote [3]. Data from these applications can not only help assess how the current parties are aligning with the populace [5], it can also suggest potential new party platforms. Such pre-existing infrastructure to get data on both voter opinions and issue importance underscores the pertinence of issue weights to modeling this problem setting.

1.1 Our Contribution

We study the aggregation of opinions on multiple independent binary issues with respect to two measures of majoritarianism: agreement with issue-wise majority and success in pairwise proposal comparisons. Our analysis considers two weighting models: *external weights* and *internal weights*. In the former, the policy-maker sets a weight to each issue reflecting its relative importance, and voters use weighted agreement when comparing any two proposals. The latter is the same, but each voter is free to choose their own weighting vector. We use the "*unweighted setting*" to refer to the edge case where issues are equally-weighted.

1.1.1 Condorcet Winners

In the first part of the paper, we focus on the complexity of determining a *Condorcet-winning proposal*: a proposal that does not lose in a direct vote against any other proposal. Under external weights, we find that any Condorcet winner has to be an IWM proposal, while this does not extend to internal weights. However, even in the unweighted setting with an odd number of voters, where the IWM proposal is unambiguous, checking whether this proposal is a Condorcet winner is co-NP-hard (answering an open question in [14]).

An Ostrogorski-free domain. To mitigate this hardness result, it would be appealing to identify a large set of instances for which IWM proposals are Condorcet winners (i.e., Ostrogorski's paradox does not occur). If membership to this set could also be efficiently verified, this would allow for practically certifying "safe instances" where issue-wise majority is the right choice. We achieve this by the *single-switch* condition of Laffond and Lainé [26]: a preference matrix over ± 1 is single-switch if it admits a *single-switch presentation* — a way to permute and potentially negate some columns such that $+1$ entries on each row form a prefix or a suffix. They show that for the unweighted case, this condition implies that Ostrogorski's paradox does not occur. We extend and simplify their analysis to show that the same holds under external weights (but not always for internal weights). We then provide a linear-time algorithm for checking whether the preference matrix is single-switch and prove that no-instances admit short proofs of this fact in the form of small forbidden subinstances (that can also be identified in linear time by a black-box reduction to the recognition problem which we have not encountered before).

Secondary implications. Along the way, in this part, we make multiple secondary contributions: (i) we uncover an interesting topological connection: the set of single-switch presentations of a single-switch matrix can be compactly represented as the union of two mirror-image Möbius strips; (ii) our recognition algorithm for single-switch matrices proceeds by reducing to checking whether the columns of a matrix can be permuted so that the ones on each row form a prefix or a suffix — while a linear-time algorithm is known for this [15],[1] it relies on rather complex machinery — we instead give a much simpler direct algorithm with the same guarantees; (iii) our simpler algorithm can be adapted to yield the simplest and at the same time most efficient algorithm for checking the *single-crossing condition* in ranked social choice [17]. Similarly to the single-switch condition, the latter also admits a characterization in terms of small forbidden substructures [10], and finding such forbidden substructures can be achieved within the same time complexity using our black-box technique, a result which to the best of our knowledge is new.

1.1.2 Representative Majority-Supported Proposals

Settling on a Condorcet-winning proposal would be ideal, especially under external weights where such proposals are by default IWM proposals, but in the absence of Condorcet winners, a compromise is needed. In fact, the hardness of checking whether an IWM proposal is a Condorcet winner can be seen positively: it is computationally demanding to find the proposal that defeats it, so we need not fear a vote being called against the defeating proposal. Hence, it is reasonable to relax the demanding Condorcet condition: the chosen proposal should, at the least, not lose against its opposite — or, in the language of our first formulation of Anscombe's paradox above, should garner majority support. In the second part of the paper, we explore existence guarantees for majority-supported proposals that are as close as possible to an IWM proposal p_{IWM}. So far, this has been studied in the unweighted model [21, 14]: a weakly majority-supported proposal agreeing in strictly more than half of the issues with p_{IWM} exists and can be found in polynomial time, while achieving better guarantees is NP-hard. The word "weakly" can be dropped if majority is strict/unambiguous on at least one issue, i.e., some column

[1] Under the name of recognizing *voter/candidate-extremal-interval* preferences.

of the preference matrix has differing numbers of $+1$'s and -1's. We will be interested in the more complex weighted case.

External weights. We provide a matching guarantee to the unweighted case, showing that there always exists a weakly majority-supported proposal with strictly more than half the total weight in topics agreeing with p_{IWM}. Under a simple condition on certain higher-weight issues, we can also drop the word "weakly."

Internal weights. In sharp contrast, with as few as two different weight vectors, we construct families of instances where the distance between every weakly majority-supported proposal and the unique IWM proposal gets arbitrarily large. The severity of such examples is controlled by the maximum average topic weight $\tilde{w}_{max}$: we give a simple bound derived from a partition-based approach that is tight on a large portion of the range $\tilde{w}_{max} \in (0, 1)$.

More paradox-free instances. Finally, we generalize Wagner's Rule of Three-Fourths [39] for both external and internal weights: if the average weighted majority on the issues is at least $\frac{3}{4}$, then Anscombe's Paradox cannot occur. Without loss of generality, if $+1$ is a majority opinion on each topic, this translates to the total weight of $+1$'s in the preference matrix being at least $\frac{3}{4}$ of the total weight. A stronger condition precludes Ostrogorski's paradox under external weights: if on each column the relative weight of $+1$'s is at least $\frac{3}{4}$ of that column's total weight. This surprisingly simple check is a counterpart to the single-switch condition, once again giving a convenient characterization for a whole class of instances in which returning an IWM proposal is always a good choice.

1.2 Further Related Work

Variations on the question of how best to reach consensus on a series of issues have been studied thoroughly. We first go over models where all topics are considered equally important.

Approval voting is a popular mechanism that is frequently used for single-winner and multi-winner elections alike [7, 19]. Here, each participant indicates their approval for a subset of candidates. In contrast to our setting, not expressing the approval of a candidate does not give the same signal as voting for the "no" stance on an issue (which is a vote for the logical negation of the issue) [25].

Another related field of study is judgment aggregation, where a series of judges have viewpoints on multiple topics, but there is external logical consistency required between the topics [31]. As in our problem, a reasonable method of reaching consensus is to take the majority opinion on each topic. However, the outcome may fail to be logically consistent — this is the *Discursive Dilemma*, and can occur with as few as 3 judges and 3 topics [24]. There has been some investigation into conditions that avoid this paradox, like List's *unidimensional alignment* [30],[2] and other similar paradoxes under the name of *compound majority paradoxes* [33].

Our problem can also be viewed as a special instance of voting in combinatorial domains: multiple referenda with separable topics [9]. Multiple works explored generalizations of Anscombe's paradox and gave further impossibility results [4, 23], e.g., relating to the Pareto optimality of aggregation rules [34].

Significant work has also been done to characterize when such paradoxes *cannot* occur. Wagner proposed the Rule of Three-Fourths [39], preventing Anscombe's paradox, as well as a generalization [40]. Laffond and Lainé showed that if no two voters disagree on too many issues, then Anscombe's is prevented [27], and for *single-switch* preferences, Ostrogorski's does not occur [26].

The interested reader should also consult [1], which develops *Flexible Representative Democracy* as

[2]The unidimensional alignment condition might appear to closely resemble the single-switch condition, as it essentially requires that the transposed preference matrix be single-switch. However, this is not equivalent, as rows and columns play different roles — issue-wise majority aggregates along columns, not rows.

a hybrid between direct democracy (voting on each issue separately) and representative democracy (electing representatives who decide on behalf of the voters across all issues). In positioning their model, the authors discuss its relation to delegative (liquid) democracy and analyze paradoxical situations such as the Ostrogorski/Anscombe paradox, illustrating how representative systems can systematically diverge from majority preferences.

We now survey proposals to augment various voting systems with weights, allowing voters to express their degrees of interest or investment in the topics. Storable voting allows participants to delay using their vote in a given election, and accumulate votes to use in later elections that they have more stake in [12]. Quadratic voting proposes a somewhat similar system in which people are given an allotment of vote credits, and before a given election can buy a certain number of votes [28]. Both of these systems maintain that voters will use more votes for elections in which they feel strongly and believe they are likely to be pivotal in. Uckelman introduces a framework using goalbases to express cardinal (numeric) preferences over a combinatorial voting domain [38]. This, however, loses information by abstracting away the separability of issues: for us, the cardinal preferences are induced by the weighted Hamming distance. Lang also considers augmenting combinatorial voting with preference weights and provides several computational complexity results [29]. Satisfaction approval voting [8] modifies approval voting by spreading a voter's total weight equally over all of the candidates they approve of. Finally, there is recent interest in studying how voters have varying stakes in elections and how to accommodate these stakes to limit distortion [11, 20].

2 Model and Notation

For any non-negative integer m, write $[m] := \{1, \dots, m\}$. Given a real number x, write $sgn(x) \in \{-1, 0, 1\}$ for its *sign*. Note that for any two reals x, y, we have that $sgn(x \cdot y) = sgn(x) \cdot sgn(y)$.

We consider a setting with n voters and t independent, binary issues/topics. The decision space for each issue is $\mathbb{B} := \{\pm 1\}$. Each voter $i \in [n]$ is modeled as a dimension-t vector $v_i \in \mathbb{B}^t$ indicating for each issue $j \in [t]$ the opinion/preference $v_{i,j} \in \mathbb{B}$ of voter i on issue j. We call the matrix $\mathcal{P} = (v_{i,j})_{i \in [n], j \in [t]}$ the *preference profile*. We also write $\mathcal{P} = (c_1, \dots, c_t)$, where $c_1, \dots, c_t \in \mathbb{B}^n$ are the columns of the matrix.

For each issue $j \in [t]$, we are consistent with previous literature [39, 40, 21, 14] and define the *majority* $m_j \in [0, 1]$ on issue j to be the fraction of voters that prefer $+1$ on it; i.e., the number of $+1$'s in c_j, divided by n. If $m_j > 0.5$, then the *majority opinion* on issue j is $+1$; if $m_j < 0.5$, then it is -1, and if $m_j = 0.5$, then both $+1$ and -1 are majority opinions on issue j. Equivalently, if we write b_j for the sum of the entries in c_j (i.e., the column's ± 1-*balance*), a majority opinion on issue j is any $o \in \mathbb{B}$ satisfying $b_j \cdot o \geq 0$.

A *proposal* is a vector $p \in \mathbb{B}^t$ that consists of a decision for each issue. We write $\overline{p}$ for the *complement* of proposal p, which simply flips each bit of p; i.e., $\overline{p} = -p$. An *issue-wise majority (IWM)* is a proposal p where the decision on each topic is a majority opinion for the topic.

We study two weighting models: *external weights* and *internal weights*. In the former, an externally supplied vector of non-negative weights $w = (w_1, \dots, w_t)$ summing up to 1 is available, denoting the importance of each issue as seen collectively by the voters. The internal weights model generalizes this by having each voter $i \in [n]$ report an individual vector of weights $w_i = (w_{i,1}, \dots, w_{i,t})$; i.e., there need no longer be consensus on the importance of any fixed issue. For internal weights, we write W for the matrix with rows $w_1, \dots, w_n$. We call the *voting instance* the pair $\mathcal{I} = (\mathcal{P}, W)$ for internal weights and $\mathcal{I} = (\mathcal{P}, w)$ for external weights. We will also talk about the *unweighted* model, which is simply external weights with $w = (1/t, \dots, 1/t)$, and directly write $\mathcal{I} = \mathcal{P}$ for it. For the remainder of this section, we assume external weights — the internal weights model requires substantial additional notation so we postpone it to later on.

For any positive integer m, given two vectors $u, v \in \mathbb{B}^m$ and a vector of weights $w \in [0,1]^m$ with unit sum, we write $d_H(u, v, w) := \sum_{j=1}^{m} w_j \cdot \mathbb{I}(u_j \neq v_j)$ for the w-*weighted Hamming distance* between u and v. We omit the w argument when referring to the unweighted Hamming distance. For convenience, we write $\langle u, v\rangle_w := \sum_{j=1}^{m} w_j \cdot u_j \cdot v_j$ for the standard w-*weighted inner/dot-product*. One can easily show that $\langle u, v\rangle_w = 1 - 2 \cdot d_H(u, v, w)$.

Fix an instance $\mathcal{I} = (\mathcal{P}, w)$ in the external weights model. For each voter i with vote v_i we define their *individual* preference relation $\succcurlyeq_i$ between proposals. In particular, given two proposals $p, p' \in \mathbb{B}^t$, voter i weakly prefers p over p', written $p \succcurlyeq_i p'$, iff $d_H(v_i, p, w) \leq d_H(v_i, p', w)$. Note that this is equivalent to $\langle v_i, p\rangle_w \geq \langle v_i, p'\rangle_w \iff \langle v_i, p - p'\rangle_w \geq 0$. We write $\succ_i$ and $\approx_i$ for the strict and symmetric parts of $\succcurlyeq_i$, respectively. We define the *collective* preference relation $\succcurlyeq_{\mathcal{I}}$ between proposals: given two proposals $p, p' \in \mathbb{B}^t$, the voters collectively weakly prefer p over p', written $p \succcurlyeq_{\mathcal{I}} p'$, iff $|\{i \in [n] \colon p \succ_i p'\}| \geq |\{i \in [n] \colon p' \succ_i p\}|$. Note that this is equivalent to $\sum_{i=1}^{n} sgn(\langle v_i, p - p'\rangle_w) \geq 0$. We write $\succ_{\mathcal{I}}$ and $\approx_{\mathcal{I}}$ for the strict and symmetric parts of $\succcurlyeq_{\mathcal{I}}$, respectively. A proposal $p \in \mathbb{B}^t$ is a *Condorcet winner* if for any other proposal $p' \in \mathbb{B}^t$ we have $p \succcurlyeq_{\mathcal{I}} p'$.

For a voting instance $\mathcal{I}$, Ostrogorski's paradox occurs if some IWM proposal p_{IWM} is not a Condorcet winner, Anscombe's paradox occurs if for some IWM proposal p_{IWM} we have $\overline{p_{IWM}} \succ_{\mathcal{I}} p_{IWM}$, and the Condorcet paradox happens if there is no Condorcet-winning proposal.

3 Complexity of Determining a Condorcet Winner

In this section, we prove that it is co-NP-hard to determine whether an instance $\mathcal{I}$ admits a Condorcet-winning proposal, even in the unweighted setting with odd n:

Theorem 1. *Deciding whether an instance $\mathcal{I} = \mathcal{P}$ admits a Condorcet winner is co-NP-hard in the unweighted setting with odd n.*

This could be surprising given the following observation of Laffond and Lainé [26] for the unweighted model, which we extend to external weights:

Lemma 2. *Consider an external-weights instance $\mathcal{I}$ such that $p \in \mathbb{B}^t$ is a Condorcet winner for $\mathcal{I}$. Then, p is an IWM for $\mathcal{I}$.*

Proof. Assume the contrary, then there is an issue $j \in [t]$ such that $p_j \cdot b_j < 0$. Consider the proposal p^* obtained from p by flipping p_j. Then, $p^* \succ_{\mathcal{I}} p$, a contradiction. □

Lemma 2 shows that one can restrict the search space for Condorcet winners to IWM proposals. In the unweighted setting with odd n, there is a single such proposal, which we can assume without loss of generality to be $\mathbf{1} \in \mathbb{B}^t$. Nevertheless, even under these conditions, we will show that checking whether $\mathbf{1}$ is a Condorcet winner is co-NP-hard, or, equivalently, checking whether $\mathbf{1}$ is *not* a Condorcet winner is NP-hard. The latter occurs if and only if there is a proposal $p \in \mathbb{B}^t$ such that $p \succ_{\mathcal{I}} \mathbf{1}$, which, recall, means that strictly more voters $i \in [n]$ prefer $p \succ_i \mathbf{1}$ than $\mathbf{1} \succ_i p$. Hence, it suffices to prove that the following problem is NP-hard:

Problem "Major"
Input: Instance $\mathcal{I} = \mathcal{P}$ in the unweighted setting with odd n such that $\mathbf{1}$ is the issue-wise majority.
Output: Does there exist a proposal $p \in \mathbb{B}^t$ s.t. $p \succ_{\mathcal{I}} \mathbf{1}$?

To show its hardness, we need the following auxiliary problem:

Problem "Unanim"
Input: Voting instance $\mathcal{I} = \mathcal{P}$ in the unweighted setting.
Output: Does there exist a proposal $p \in \mathbb{B}^t$ s.t. $p \succ_i \mathbf{1}$ for all $i \in [n]$ (to be read "*p unanimously* defeats $\mathbf{1}$")?

Unanim is NP-hard [16, Theorem 2], but the proof in [16] is relatively complicated: we give a simpler one in Appendix A by noting the equivalence to choosing a subset of columns of $\mathcal{P}$ that sum up to a negative amount on each row (we also give a similar reformulation of Major for the interested reader).

Lemma 3. *Major is NP-hard.*

Proof. We reduce from the NP-hard problem Unanim. Consider an instance $\mathcal{I} = \mathcal{P}$ of Unanim with n voters. If there is an issue $j \in [t]$ disapproved by all voters in $\mathcal{P}$, then $\mathcal{P}$ is a yes-instance of Unanim: all voters prefer the proposal with $+1$ in all coordinates except the j-th to proposal $\mathbf{1}$. This case can be easily detected in polynomial time, so we henceforth assume the contrary.

We build an instance $\mathcal{I}' = \mathcal{P}'$ of Major from $\mathcal{P}$ by adding $n-1$ voters approving all issues. For $\mathcal{P}'$ to be a valid instance for Major we need that $2n-1$ is odd (which it is) and that $\mathbf{1}$ is the issue-wise majority. The latter holds because at least $n-1+1 = n$ voters approve of each issue: the $n-1$ added ones and at least one from the first n by our assumption. It remains to show that a proposal $p \in \mathbb{B}^t$ unanimously defeats $\mathbf{1}$ in $\mathcal{P}$ iff it majority-defeats $\mathbf{1}$ in $\mathcal{P}'$.

Assume $p \in \mathbb{B}^t$ unanimously defeats $\mathbf{1}$ in $\mathcal{P}$. Then, each of the first n voters in $\mathcal{P}'$ prefers p to $\mathbf{1}$. Since there are only $n-1 < n$ other voters in $\mathcal{P}'$, a majority of the voters in $\mathcal{P}'$ prefer p to $\mathbf{1}$.

Conversely, assume $p \in \mathbb{B}^t$ majority-defeats $\mathbf{1}$ in $\mathcal{P}'$. Clearly, $p \neq \mathbf{1}$ has to hold, so all of the $n-1$ added voters prefer $\mathbf{1}$ to p. To counteract this, since $p \succ_{\mathcal{I}'} \mathbf{1}$, the first n voters in $\mathcal{P}'$ must prefer p to $\mathbf{1}$, meaning that p unanimously defeats $\mathbf{1}$ in $\mathcal{P}$. □

For completeness, we put the pieces together to give a self-contained proof of Theorem 1 in Appendix A.

4 An Ostrogorski-free Domain

As we have seen, at least for external weights, a Condorcet-winning proposal has to be an issue-wise majority proposal. Yet, we proved that determining whether one of them is actually Condorcet-winning is co-NP-hard, even in the unweighted case with odd n, where there is only one such proposal to check. To mitigate this hardness result, it would be useful if we could identify a large set of instances for which IWM proposals are guaranteed to be Condorcet-winning, i.e., Ostrogorski's paradox does not occur. Laffond and Lainé [26] introduced the *single-switch* condition, which achieves exactly this goal for the unweighted setting. Furthermore, they showed that it is the most general condition preventing Ostrogorski's paradox among conditions that do not consider the multiplicities of the votes (i.e., conditions defining a *domain*) or whether a vote is negated or not (i.e., they only look at the set $\{\{v_i, \overline{v_i}\} \mid i \in [n]\}$ and not at how many times each v_i or $\overline{v_i}$ is repeated). In particular, if an instance in the unweighted model is not single-switch, then it is possible to add copies of some of the votes v_i (or their negations $\overline{v_i}$) so that some issue-wise majority proposal is not a Condorcet winner. Two important questions underpinning their condition are: (i) Does it still guarantee the existence of a Condorcet winner in the (at least externally) weighted setting? (ii) Is it possible to check whether it applies in polynomial time? If not, are there short proofs of this fact? In this section, we answer all these questions in the affirmative.

A preference profile (matrix) $\mathcal{P} = (c_1, \ldots, c_t)$ is *single-switch* (SSW) if we can flip (multiply by -1 all entries in) some columns and then permute the columns to get a new profile $\mathcal{P}'$ such that $+1$ entries on every row form either a prefix or a suffix, in which case we say that $\mathcal{P}'$ is an SSW presentation of $\mathcal{P}$. We

1	2	3	4	5	6
+1	+1	+1	-1	+1	-1
+1	-1	+1	+1	-1	+1
+1	+1	-1	-1	+1	-1

(a) Profile $\mathcal{P}$.

2	1	3	4	$\overline{5}$	6
+1	+1	+1	-1	-1	-1
-1	+1	+1	+1	+1	+1
+1	+1	-1	-1	-1	-1

(b) Single-switch presentation of $\mathcal{P}$.

Figure 1: The profile $\mathcal{P}$ in Fig. 1a is single-switch because its columns can be permuted and flipped as in Fig. 1b to ensure that ones on each row form a prefix or a suffix.

allow flipping no columns or leaving all columns in their original place. Intuitively, issues are arranged along a left-right axis. Left-wing voters approve a prefix of issues, with the length depending on their tolerance, while right-wing voters similarly approve a suffix of issues.[3] See Fig. 1 for an illustration of the notion. A voting instance $\mathcal{I}$ is single-switch if its preference profile $\mathcal{P}$ is single-switch.

4.1 For External Weights Single-Switch Prevents Ostrogorski's Paradox

We find that, assuming external-weights, the single-switch condition guarantees that all IWM proposals are Condorcet winners. To show this, we first show that every issue-wise majority proposal does not lose against its opposite, i.e., Anscombe's paradox does not occur. We do this by streamlining and adapting the argument in [26] (which was only for the unweighted model). Because the single-switch condition is closed under removing issues, the general statement then follows easily by noting that, under external weights, Ostrogorski's paradox happens if and only if there is a subset of issues inducing an instance where Anscombe's paradox happens. The details are deferred to Appendix B.1.

Theorem 4. *In the external-weights model, every issue-wise majority proposal of a single-switch instance is a Condorcet winner.*

4.2 Recognizing Single-Switch Profiles

The result in the previous section is particularly appealing: in the external-weights model, if the preferences are single-switch, any issue-wise majority proposal is a Condorcet winner. This bypasses our previous hardness result in the case of single-switch preferences. However, this is only useful provided one can quickly tell whether a given profile $\mathcal{P}$ is single-switch or not. In this section, we show that this can be determined in linear time, i.e., $O(nt)$. For yes-instances, our algorithm also determines an SSW presentation $\mathcal{P}'$ (implicitly also the permutation and flips used to obtain it). Given $\mathcal{P}'$, we also characterize the set of all SSW presentations as the union of two "orbits" around $\mathcal{P}'$ and its column-reversal. These orbits can be attractively interpreted topologically as two mirror-image Möbius strips. To begin, we need the following observation following easily from the case $n = 1$. See Appendix B.2 for the proof.

Lemma 5. *Consider a profile $\mathcal{P}$ admitting an SSW presentation $\mathcal{P}' = (c_1, \dots, c_t)$. Then, $\mathcal{P}'_r := (c_2, \dots, c_t, \overline{c_1})$ is also a SSW presentation of $\mathcal{P}$. Furthermore, any t (circularly) consecutive columns in $\mathcal{P}'' := (c_1, \dots, c_t, \overline{c_1}, \dots, \overline{c_t})$ form an SSW presentation of $\mathcal{P}$.*

Hence, any SSW presentation $\mathcal{P}'$ of a profile $\mathcal{P}$ corresponds to a set of $2t$ such presentations that we call the *orbit* $O_{\mathcal{P}'}$ of $\mathcal{P}'$. Formally, these are the $2t$ profiles that can be obtained by taking t (circularly) consecutive columns in $\mathcal{P}''$ in the above. Note that the orbits of any two SSW presentations either

[3] This shares similarities with several related concepts, such as single-peaked and single-crossing preferences. However, unlike most other notions, we allow issues to be flipped before ordering them, as they can be logically negated without changing meaning.

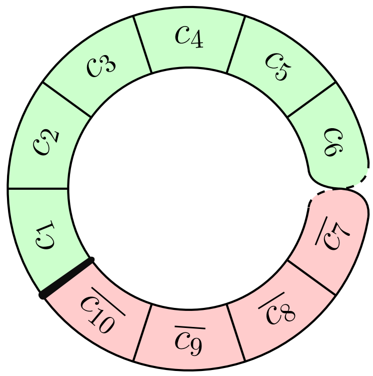

Figure 2: Möbius strip of orbit $O_{\mathcal{P}'}$ for $\mathcal{P}' = (c_1, \dots, c_{10})$. We start with a rectangular piece of paper of length 10 and write $(c_1, \dots, c_{10})$ on the (green) front side and $(\overline{c_1}, \dots, \overline{c_{10}})$ on the (red) backside. We then give the paper a length-wise half-turn and glue the endpoints (bold strip). This gives raise to a surface with a single continuous side.

coincide or are disjoint, so the set of all orbits partitions the set of SSW presentations of $\mathcal{P}$. Also, the $2t$ profiles in $O_{\mathcal{P}'}$ are pairwise distinct, which can be easily seen by considering the case $n = 1$, under which $\mathcal{P}''$ is circularly equivalent to a list of t minus ones followed by t ones. This reasoning additionally allows us to assign to each orbit a *representative*, namely the profile with all -1's on the first row:

Corollary 6. *Every orbit contains exactly one profile where the first row is all* -1*'s.*

Orbits can be understood through a topological lens: For the orbit $O_{\mathcal{P}'}$ of $\mathcal{P}' = (c_1, \dots, c_t)$ take an $n \times t$ rectangular piece of paper and write the columns $c_1, \dots, c_t$ on the front and $\overline{c_1}, \dots, \overline{c_t}$ on the back, such that for each $i \in [t]$, column c_i on the front aligns with column $\overline{c_i}$ on the back. Then, give the paper a length-wise half-twist and glue the left and right sides to form a surface known as a Möbius strip: see Fig. 2. Cutting along the width of the strip between *any* two columns recovers an $n \times t$ piece of paper with one SSW presentation on one side and its opposite on the other side. In high-level terms, each orbit is topologically a Möbius strip.

To check whether a profile $\mathcal{P}$ is single-switch, by Corollary 6, it suffices to check for presentations with all -1's in the first row: all other presentations are generated by the orbits of such presentations. There is a simple strategy to achieve this: flip columns in $\mathcal{P}$ to make the first row all -1's, and then check whether columns in the resulting profile can be permuted to ensure that ones on each row form a prefix or a suffix. This amounts to recognizing single-switch-no-flips profiles: A profile $\mathcal{P}$ is *single-switch-no-flips* (SSWNF) if its columns can be permuted to get a new profile $\mathcal{P}'$ such that $+1$ entries on every row form either a prefix or a suffix, in which case we say that $\mathcal{P}'$ is an SSWNF presentation of $\mathcal{P}$.

Recognizing single-switch-no-flips profiles. Telling whether a profile $\mathcal{P} = (c_1, \dots, c_t)$ is single-switch-no-flips can be achieved by appending a negated copy of $\mathcal{P}$ underneath [15] and running a solver for the *Consecutive Ones Problem* (C1P), which can be solved in $O(nt)$ time [6], implying the same about our problem. However, such solvers are complicated and notoriously error-prone: most available implementations fail on at least some edge cases [18]. Moreover, reducing to C1P does not utilize the additional structure present in our problem and hence does not shed light on the structure of all solutions, as we set out to do. We give a much simpler algorithm achieving the $O(nt)$ time-bound: Find an index x maximizing $d_H(c_1, c_x)$. Then, sort (using Counting Sort) the columns based on their Hamming distance from c_x to get a profile $\mathcal{P}' = (c'_1, \dots, c'_t)$ where $d_H(c_x, c'_i) \le d_H(c_x, c'_{i+1})$ for $i \in [t-1]$ (i.e., ties in Hamming distance can be broken arbitrarily). We claim that either $\mathcal{P}'$ is the unique SSWNF presentation of $\mathcal{P}$ (up to reversing the order of the columns), or there is no such presentation, so we can easily check in additional $O(nt)$ time whether the candidate solution works.

All required claims shown formally in Appendix B.2:

Theorem 7. *There is a* simple $O(nt)$ *algorithm computing (or deciding the inexistence of) an SSWNF presentation of a profile* $\mathcal{P}$*. Moreover, if it exists, this presentation is unique up to reversing column order.*

Appendix B.2 also provides a much ampler discussion of related work for this sub-problem, including the relation between our algorithm and previous algorithms for recognizing *single-crossing preferences*. As a bonus, it gives a similar simpler, more efficient algorithm for recognizing single-crossing preferences, running in time $O(nt\sqrt{\log n})$, improving state of the art [17, Algorithm 4].

Putting it together. To decide whether a profile $\mathcal{P}$ is single-switch, we flip columns in $\mathcal{P}$ to get a profile $\mathcal{P}'$ with only -1's in the first row and then use the algorithm in Theorem 7 to find an SSWNF presentation $\mathcal{P}''$ of $\mathcal{P}'$ (and hence also $\mathcal{P}$). If it exists, this presentation is unique up to column reversal, so we can also characterize the set of all SSW presentations of $\mathcal{P}$ by unioning the orbits of $\mathcal{P}''$ and its column-reversal. Note that these two orbits may coincide for pathological input profiles $\mathcal{P}$.

Theorem 8. *There is an* $O(nt)$ *algorithm computing (or deciding the inexistence of) an SSW presentation of a profile* $\mathcal{P}$*. If the algorithm returns a presentation* $\mathcal{P}''$*, let* $\mathcal{P}''_r$ *be* $\mathcal{P}''$ *with the order of the columns reversed, then the set of all SSW presentations of* $\mathcal{P}$ *is* $O_{\mathcal{P}''} \cup O_{\mathcal{P}''_r}$*.*

4.3 Forbidden Subprofiles Characterization of Single-Switch Preferences

Whenever the single-switch condition is not satisfied, it would be useful if there were a short proof of this fact: a small subprofile that is not single-switch. Formally, a profile/matrix $\mathcal{P}$ contains a profile/matrix $\mathcal{P}'$ as a *subprofile/submatrix* if we can remove (possibly zero) rows and columns from $\mathcal{P}$ to get $\mathcal{P}'$ up to permuting rows and columns. Note that existence is not immediate: there could exist arbitrarily large matrices not satisfying the condition but all of whose proper submatrices do. We show that this is not the case: either the condition holds, or there is a 3×4 or 4×3 submatrix witnessing that this is not the case, as in the following:

Theorem 9. *A profile* $\mathcal{P}$ *is single-switch if and only if it does not contain as a subprofile* $\mathcal{P}_1^a, \mathcal{P}_2^a$ *and any profile that can be obtained from them by flipping rows and columns:*

$$\mathcal{P}_1^a = \begin{bmatrix} -1 & -1 & -1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \end{bmatrix} \qquad \mathcal{P}_2^a = \begin{bmatrix} -1 & -1 & -1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ -1 & -1 & +1 \end{bmatrix}$$

We prove Theorem 9 in Appendix B.3 by combining a similar characterization for single-switch-no-flips profiles given in [37] (under the name *voter/candidate-extremal-interval* preferences) with our insight that to go to the no-flips version it suffices to make one row all -1's. Henceforth, we call the 3×4 and 4×3 preference profiles in the theorem above *forbidden subprofiles*. Then, the theorem says that $\mathcal{P}$ is single-switch if and only if it contains no forbidden subprofiles. Note how this implies that single-switch profiles are relatively rare: the probability that a random binary $n \times t$ matrix is single-switch tends to zero as n and t tend to infinity.

Finding forbidden subprofiles. So far, we have seen that non-membership to the class of single-switch preferences admits short proofs, but can such proofs also be constructed efficiently? Given some no-instance, it is straightforward to determine which forbidden subprofiles occur in it in time $O(n^3t^4 + n^4t^3)$. In contrast, our recognition algorithm runs in time $O(nt)$, but does not identify a forbidden subprofile. We will now assume our $O(nt)$ recognition algorithm as a black box and show how to identify a forbidden subprofile for a given no-instance $\mathcal{P}$ in time $O(nt)$.

Let us first describe an $O(n^2t + nt^2)$ approach: one at a time, try to remove each row and each column of $\mathcal{P}$, i.e., $n + t$ removal attempts; if doing so makes the resulting profile a yes-instance, undo the removal, and otherwise let it persist. At the end, the ensuing no-instance $\mathcal{P}'$ is a subprofile of $\mathcal{P}$ whose proper subprofiles are yes-instances, so $\mathcal{P}'$ is a forbidden subprofile, completing the argument.

We now modify the previous idea to run in time $O(nt)$ by removing multiple rows/columns at a time. We will first only remove rows, and then, starting from the resulting profile, only columns. The reasoning for columns is entirely analogous, so we only describe the procedure for rows: partition the rows into 5 groups $G_1, \ldots, G_5$, each of size roughly $n/5$. Because all forbidden subprofiles are of size 3×4 or 4×3, any occurrence of a forbidden subprofile in $\mathcal{P}$ only uses rows from at most 4 of the 5 groups. Consequently, we can find a group G_i such that removing all rows in G_i from $\mathcal{P}$ keeps the property that $\mathcal{P}$ is a no-instance. Doing so requires at most 5 calls to the recognition algorithm, so it can be done in overall time $O(nt)$. Ignoring for brevity the cases where n is not divisible by 5, this reasoning shows how to reduce n to $4n/5$ in time $O(nt)$. Applying the same reasoning iteratively until n goes below 5 takes total time $O(nt)$ because the geometric series $\sum_{i=0}^{\infty}(4/5)^i$ converges.

Theorem 10. *Given a non-single-switch profile $\mathcal{P}$, a forbidden subprofile of $\mathcal{P}$ can be determined in time $O(nt)$.*

We note that the previous idea applies more broadly; e.g., for single-crossing preferences, which admit a characterization in terms of two small *forbidden subinstances* [10], our $O(nt\sqrt{\log n})$ recognition algorithm can be bootstrapped to also produce a forbidden subinstance for no-instances within the same time bound. A formal statement and more details can be found in Appendix B.3.

5 Anscombe's Paradox

When preferences are not single-switch, determining whether an IWM proposal is a Condorcet winner is co-NP hard. In light of this, we focus on the most diabolical subset of Ostrogorski paradox instances: those inducing Anscombe's paradox (where an IWM proposal is defeated by its complement, or, equivalently, an IWM proposal fails to get majority support). If Anscombe's paradox occurs, a natural question is: "How close can we get to any given IWM while still requiring that the proposal gets majority support?"

We first explore this question under external weights, i.e., in instances $\mathcal{I} = (\mathcal{P}, w)$ where all voters share the same, unit-sum weights vector w. Then, we introduce the necessary notation and study it for internal weights. Finally, we give a simple characterization of a broad swath of instances that avoid Anscombe's paradox entirely for internal weights. We assume throughout that $t > 1$, as Anscombe's paradox does not occur with one topic, and without loss of generality that $m_j \geq 0.5$ for all $j \in [t]$ (i.e., that +1 is a majority opinion on all topics).

Formally, some voter i supports (approves of) a proposal p if $d_H(v_i, p, w) < 1/2$, opposes (disapproves of) p if $d_H(v_i, p, w) > 1/2$, and is indifferent to p if $d_H(v_i, p, w) = 1/2$. A proposal is *strictly majority-supported* if more people support it than oppose it and *weakly majority-supported* if no more people oppose it than support it. Our definition of majority support matches [14] but differs from [21] (where indifferent voters count towards the proposal's support).

5.1 External Weights

In the unweighted case, it is straightforward to argue that for any IWM, there exists a weakly majority-supported proposal within distance $\leq \frac{1}{2} + \frac{1}{2t}$ because at least one proposal in every complementary pair $(p, \overline{p})$ gets weak majority support (and at least one pair satisfies the distance bound for both proposals). A slightly better guarantee of distance $< \frac{1}{2}$ holds by a more difficult proof [14, 21]. For external weights, the complementary pairs argument no longer gives a bound close to $\frac{1}{2}$ if no subset of topic weights sum up close to $\frac{1}{2}$. One may hope to reduce to the unweighted case by splitting topics into multiple equal-weight topics and use the $< \frac{1}{2}$ bound there, but the resulting majority-supported proposals may have different values for an original topic's clones, making it hard to translate to proposals in the original instance. Despite these setbacks, we surprisingly find that the $< \frac{1}{2}$ guarantee still holds for external weights. Our proof, deferred to Appendix C.1.2, simplifies and adapts the argument in [14]. We also guarantee *strict* majority support if there is a strict majority in at least one *relevant* topic, roughly meaning topics with high enough weight to be the tipping point in a vote (see Appendix C.1.1 for a formal definition).

Theorem 11. *For any $\mathcal{I} = (\mathcal{P}, w)$ and p_{IWM}, there is a weakly majority supported proposal p with $d_H(p, p_{IWM}, w) < 1/2$. If majority is strict in any relevant topic, "weak" can be replaced with "strict".*

To accommodate space constraints, our results for the internal weights model are deferred to Appendix C.2

6 Conclusion and Future Work

We explored how best to represent the will of voters on multiple, separable issues when optimizing for two potentially conflicting ideals: agreement with issue-wise majority and success in pairwise proposal comparisons. Additionally, we augmented previous multi-issue voting models to account for non-uniform and individualized issue importance. We demonstrated that determining whether an IWM is a Condorcet winner is co-NP hard, but provided an efficiently checkable condition under which Ostrogorski's paradox does not occur. We then examined instances where an IWM loses to the opposing proposal (i.e., Anscombe's paradox occurs) and showed how our two weighting models alter our ability to reconcile the two objectives. While we now have a rich understanding of the interaction of these two majoritarian ideals, one could optimize for different notions of representation in the proposal selection. It would be interesting to study variants of maximizing total voter "satisfaction" — the total weight voters have on topics that they agree with the final proposal on (a weighted version of an objective proposed in [21]). On the technical side, our work leaves open a number of interesting questions and gaps: (i) Our Theorem 11 for external weights is only existential. In contrast, in the unweighted setting, [14] also provide a polynomial-time method to derandomize the probabilistic argument. Extending this approach to the weighted setting appears generally more challenging but likely feasible in pseudo-polynomial time with slightly more involved techniques. (ii) Paper [14] also shows a hardness result for the unweighted case: telling whether a proposal achieves more agreement with an IWM than guaranteed by the probabilistic argument is NP-hard. It would be interesting to get a similar result for every fixed weights vector w. (iii) We have only succeeded in proving that our bounds in Theorem 27 are tight for some portion of the range $\tilde{w}_{max} \in (0, 1)$.

Acknowledgments. Carmel Baharav was supported by an ETH Excellence Scholarship. We thank Edith Elkind for the simpler proof of [16, Theorem 2] that we include in the appendix. We thank Tamio-Vesa Nakajima for helping typeset Fig. 2 and the many productive discussions surrounding the linear-time recognition algorithm for single-switch preferences (including noting Lemma 5). We also extend our gratitude to the attendees of AAMAS'25 for their thoughtful feedback and remarks. Finally, we are grateful to the anonymous reviewers for their insightful comments and suggestions that helped improve the quality of the paper.

References

[1] Ben Abramowitz and Nicholas Mattei. Flexible representative democracy. *Social Choice and Welfare*, 64(1):263–308, February 2025.

[2] Gertrude Elizabeth Margaret Anscombe. On frustration of the majority by fulfilment of the majority's will. *Analysis*, 36(4):161–168, 1976.

[3] Christine Benesch, Rino Heim, Mark Schelker, and Lukas Schmid. Do voting advice applications change political behavior? *The Journal of Politics*, 85(2):684–700, 2023.

[4] Jean-Pierre Benoît and Lewis A Kornhauser. Only a dictatorship is efficient. *Games and Economic Behavior*, 70(2):261–270, 2010.

[5] Sebastian Bensland, Karim Saleh, and Pietro Ronchetti. A machine learning analysis of the swiss political spectrum and candidate recommendation process. *ETH Zürich*, 2020. URL https://pub.tik.ee.ethz.ch/students/2020-FS/GA-2020-01.pdf.

[6] Kellogg S. Booth and George S. Lueker. Testing for the consecutive ones property, interval graphs, and graph planarity using pq-tree algorithms. *Journal of Computer and System Sciences*, 13(3):335–379, 1976. ISSN 0022-0000. doi: https://doi.org/10.1016/S0022-0000(76)80045-1. URL https://www.sciencedirect.com/science/article/pii/S0022000076800451.

[7] Steven J Brams and Peter C Fishburn. Approval voting. *American Political Science Review*, 72(3): 831–847, 1978.

[8] Steven J Brams and D Marc Kilgour. Satisfaction approval voting. *Mathematical and Computational Modeling: With Applications in Natural and Social Sciences, Engineering, and the Arts*, pages 273–298, 2015.

[9] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of computational social choice*. Cambridge University Press, 2016.

[10] Robert Bredereck, Jiehua Chen, and Gerhard J. Woeginger. A characterization of the single-crossing domain. *Social Choice and Welfare*, 41(4):989–998, 2013. ISSN 01761714, 1432217X. URL http://www.jstor.org/stable/42001441.

[11] Harry Brighouse and Marc Fleurbaey. Democracy and proportionality. *The Journal of Political Philosophy*, 18(2):137–155, 2010.

[12] Alessandra Casella. Storable votes. *Games and Economic Behavior*, 51(2):391–419, 2005.

[13] Timothy M. Chan and Mihai Pătraşcu. Counting inversions, offline orthogonal range counting, and related problems. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '10, pages 161–173, USA, 2010. Society for Industrial and Applied Mathematics. ISBN 9780898716986.

[14] Andrei Constantinescu and Roger Wattenhofer. Computing the best policy that survives a vote. *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS '23)*, 2023.

[15] Edith Elkind and Martin Lackner. Structure in dichotomous preferences. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 2019–2025. AAAI Press, 2015. ISBN 9781577357384.

[16] Edith Elkind, Abheek Ghosh, and Paul W. Goldberg. Complexity of deliberative coalition formation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 4975–4982, Jun. 2022.

[17] Edith Elkind, Martin Lackner, and Dominik Peters. Preference restrictions in computational social choice: A survey. *arXiv preprint arXiv:2205.09092*, 2022.

[18] Simon D. Fink, Matthias Pfretzschner, and Ignaz Rutter. Experimental comparison of pc-trees and pq-trees. *ACM J. Exp. Algorithmics*, 28, oct 2023. ISSN 1084-6654. doi: 10.1145/3611653. URL https://doi.org/10.1145/3611653.

[19] Peter C Fishburn. An analysis of simple voting systems for electing committees. *SIAM Journal on Applied Mathematics*, 41(3):499–502, 1981.

[20] Bailey Flanigan, Ariel D Procaccia, and Sven Wang. Accounting for stakes in democratic decisions. FORC'23 (non-archival), 2023.

[21] Robin Fritsch and Roger Wattenhofer. The price of majority support. *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS '22)*, 2022.

[22] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. 1979.

[23] Umberto Grandi. The common structure of paradoxes in aggregation theory. *arXiv preprint*

arXiv:1406.2855, 2014.

[24] Lewis A Kornhauser and Lawrence G Sager. The one and the many: Adjudication in collegial courts. *Calif. L. Rev.*, 81:1, 1993.

[25] Martin Lackner and Piotr Skowron. Approval-based committee voting. In *Multi-Winner Voting with Approval Preferences*, pages 1–7. Springer, 2022.

[26] Gilbert Laffond and Jean Lainé. Single-switch preferences and the ostrogorski paradox. *Mathematical Social Sciences*, 52(1):49–66, 2006.

[27] Gilbert Laffond and Jean Lainé. Unanimity and the anscombe's paradox. *Top*, 21:590–611, 2013.

[28] Steven P Lalley and E Glen Weyl. Quadratic voting: How mechanism design can radicalize democracy. In *AEA Papers and Proceedings*, volume 108, pages 33–37. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2018.

[29] Jérôme Lang. Logical preference representation and combinatorial vote. *Annals of Mathematics and Artificial Intelligence*, 42:37–71, 2004.

[30] Christian List. A possibility theorem on aggregation over multiple interconnected propositions. *Mathematical Social Sciences*, 45(1):1–13, 2003.

[31] Christian List. The theory of judgment aggregation: an introductory review. *Synthese*, 187(1): 179–207, 2012.

[32] Tomomi Matsui and Yasuko Matsui. A survey of algorithms for calculating power indices of weighted majority games. *Journal of the Operations Research Society of Japan*, 43(1):71–86, 2000.

[33] Hannu Nurmi. Compound majority paradoxes. In *Voting Paradoxes and How to Deal with Them*, pages 70–86. Springer Berlin Heidelberg, 1999. ISBN 978-3-662-03782-9.

[34] Ipek Özkal-Sanver and M Remzi Sanver. Ensuring pareto optimality by referendum voting. *Social Choice and Welfare*, 27:211–219, 2006.

[35] Douglas W Rae and Hans Daudt. The ostrogorski paradox: a peculiarity of compound majority decision. *European Journal of Political Research*, 4(4):391–398, 1976.

[36] Pew Research. Inflation, health costs, partisan cooperation among the nation's top problems, 2023. URL https://www.pewresearch.org/politics/2023/06/21/inflation-health-costs-partisan-cooperation-among-the-nations-top-problems/. Last accessed 19 September 2023.

[37] Zoi Terzopoulou, Alexander Karpov, and Svetlana Obraztsova. Restricted domains of dichotomous preferences with possibly incomplete information. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(6):5726–5733, May 2021. doi: 10.1609/aaai.v35i6.16718. URL https://ojs.aaai.org/index.php/AAAI/article/view/16718.

[38] Joel Uckelman. *More than the sum of its parts: compact preference representation over combinatorial domains*. University of Amsterdam, 2009.

[39] Carl Wagner. Anscombe's paradox and the rule of three-fourths. *Theory and Decision*, 15(3):303, 1983.

[40] Carl Wagner. Avoiding anscombe's paradox. *Theory and Decision*, 16(3):233, 1984.

Carmel Baharav
ETH Zürich
Zürich, Switzerland
Email: cbaharav@ethz.ch

Andrei Constantinescu
ETH Zürich
Zürich, Switzerland
Email: aconstantine@ethz.ch

Roger Wattenhofer
ETH Zürich
Zürich, Switzerland
Email: wattenhofer@ethz.ch

A Complexity of Determining a Condorcet Winner

In this section, we show that Unanim is NP-hard. This was previously known [16, Theorem 2], but the proof there is a more complicated reduction from Independent Set. In contrast, our proof proceeds by recasting the problem in terms of selecting a subset of column vectors whose sum is negative in all coordinates. This alternate view enables a more natural reduction from Exact Cover By 3-Sets. Afterward, we put all the pieces together to give a self-contained formal proof of Theorem 1. We conclude the section by providing a similar "choosing a subset of vectors" formulation for Major, which we believe could be interesting more broadly.

To prove that Unanim is NP-hard, we note that Unanim admits an elegant reformulation: write $\mathcal{P} = (c_1, \ldots, c_t)$ in terms of its issue/column vectors, then selecting a proposal $p \in \mathbb{B}^t$ amounts to choosing the subset of issues where p differs from $\mathbf{1}$. A voter $i \in [n]$ prefers $p \succ_i \mathbf{1}$ iff the sum of the selected vectors is strictly negative in coordinate i, so Unanim asks to select vectors such that the sum is negative in all n coordinates. Hence, Unanim is equivalent to the following problem, which we find of independent interest:

> **Problem "Negative-Sum-Subset" (NSS)**
> **Input**: Collection C of t dimension-n vectors over ± 1.
> **Output**: Does there exist $C' \subseteq C$ s.t $\mathbf{v} := \sum_{c \in C'} c$ is negative in all coordinates: $\mathbf{v}_i < 0$ for all $i \in [n]$?

Lemma 12. *UNANIM = NSS is NP-hard.*

Proof. We reduce from the problem Exact Cover By 3-Sets (X3C), which is well-known to be NP-hard [22]. An instance of X3C is given by a ground set S of size $3s$ and a set X of size-3 subsets of S (also called *triples*); it is a yes-instance if and only if there exists a subset $X' \subseteq X$ that forms an exact cover of S (i.e., the triples in X' cover each element of S exactly once); note that $|X'| = s$ must hold if so.

Consider an instance of X3C, we want to construct an instance of NSS such that the former is a yes-instance if and only if the latter one is.

We can assume without loss of generality that there is a triple $\{a, b, c\} \in X$ such that no other element of X contains any of a, b, or c. Indeed, if not, we can add three new elements to S and a set comprising them to X. The modified instance is a yes-instance if and only if the original instance is.

Now, to create an instance of NSS, we construct $|X| + s - 1$ dimension-$(3s + 1)$ vectors over ± 1. The first $3s$ coordinates correspond to the elements of the ground set. For each set $x \in X$, we construct a vector with -1 in coordinates that correspond to elements of x as well as in the last coordinate (in other coordinates, we have $+1$). We refer to these vectors as *set vectors*. In addition, we construct $s - 1$ vectors with -1 in the first $3s$ coordinates and $+1$ in the last coordinate; we refer to these vectors as *dummy vectors*.

We claim that this is a yes-instance of NSS if and only if the original instance of X3C was a yes-instance.

($\Leftarrow$) Suppose we started with a yes-instance of X3C, so let $X' \subseteq X$ be an exact cover. Recall that this implies $|X'| = s$. Take the s set vectors corresponding to X' as well as all $s - 1$ dummy vectors. Summing up these vectors, in the last coordinate, we get $s \cdot (-1) + (s - 1) \cdot (+1) = -1$. In every other coordinate, we have $s - 1$ entries -1 from the dummy vectors, one entry -1 from the vector corresponding to the triple that covers the respective element, and $s - 1$ entries $+1$ from the other selected set vectors, making for a total sum of $(s - 1) \cdot (-1) + 1 \cdot (-1) + (s - 1) \cdot (+1) = -1$. Hence, the resulting instance is a yes-instance of NSS.

($\Rightarrow$) Conversely, suppose the resulting instance is a yes-instance of NSS. Hence, pick some of the vectors so that the sum in each coordinate is negative. We will show that, among picked vectors, the set vectors correspond to an exact cover. Suppose we picked t set vectors. Then, we picked at most $t-1$ dummy vectors (as otherwise, the sum in the last coordinate would be non-negative). Note that this implies $t > 0$.

Assume for a contradiction that the set vectors among selected vectors do not correspond to a cover. Then, there is some coordinate among the first $3s$ in which all selected set vectors have $+1$. However, in that case, the sum of the vectors cannot be negative in that coordinate, as we have t entries $+1$ from set vectors and at most $t-1$ entries -1 from dummy vectors, a contradiction.

On the other hand, assume for a contradiction that the selected set vectors do correspond to a cover, but it is not an exact cover, so some element is covered more than once. In particular, each element is covered at least once, and at least one element element is covered more than once, so the sum of the sizes of the triples that the selected set vectors correspond to has to be at least $3s+1$, meaning that we had to pick at least $s+1 = \lceil \frac{3s+1}{3} \rceil$ set vectors. Now, recall that there is some element a (without loss of generality, the element corresponding to the first coordinate) that is only contained in one triple in X.[4] Then, in our chosen collection of vectors, we have at most one set vector with -1 in the first coordinate and, hence, at least s set vectors with $+1$ in this coordinate. Hence, considering only set vectors, the sum in this coordinate is at least $1 \cdot (-1) + s \cdot (+1) = s-1$. Hence, even if all $s-1$ dummy vectors were picked, the sum in this coordinate is at least $s-1+(s-1)\cdot(-1) = 0$, and hence non-negative, a contradiction.

Hence, the set vectors among picked vectors correspond to an exact cover, so the original instance of X3C is a yes-instance, completing the proof. □

We now give a self-contained formal proof of Theorem 1, restated below for convenience.

Theorem 1. *Deciding whether an instance $\mathcal{I} = \mathcal{P}$ admits a Condorcet winner is co-NP-hard in the unweighted setting with odd n.*

Proof. In the unweighted setting with odd n, checking for the existence of a Condorcet winner is equivalent to checking whether the (unique) issue-wise majority proposal is a Condorcet winner (by Lemma 2). It is enough to show hardness for the case where the issue-wise majority proposal is $\mathbf{1}$.[5] Hence, we want to show that the following problem is co-NP-hard: "Given a voting instance $\mathcal{I} = \mathcal{P}$ in the unweighted setting with odd n where $\mathbf{1}$ is the issue-wise majority proposal, determine whether $\mathbf{1}$ is a Condorcet winner." This is equivalent to showing that the 'negated' problem is NP-hard: "Given a voting instance $\mathcal{I} = \mathcal{P}$ in the unweighted setting with odd n where $\mathbf{1}$ is the issue-wise majority proposal, determine whether $\mathbf{1}$ is *not* a Condorcet winner." Not being a Condorcet winner is equivalent to there existing a proposal $p \in \mathbb{B}^t$ such that $p \succ_{\mathcal{I}} \mathbf{1}$. Hence, the negated problem is precisely Major, which is NP-hard by Lemma 3, completing the proof. □

We end by pointing out that it is also possible to formulate Major as a vector problem. We found it clearest not to do so in our proofs, but the problem resulting in doing so is elegant and might see other applications. Namely, the following problem is equivalent to Major, and hence NP-hard:

Problem
"More-Negative-Than-Positive-Sum-Subset"
Input: Collection C of t dimension-n vectors over ± 1, s.t. n is odd and each vector's entries sum up to a positive amount.
Output: Does there exist $C' \subseteq C$ s.t. $\mathbf{v} := \sum_{c \in C'} c$ is negative in more coordinates than it is positive: $\sum_{i \in [n]} sgn(\mathbf{v}_i) < 0$?

[4] In fact, there are three such elements, but we do not need this fact.

[5] This is, in fact, equivalent to the general case by negating/flipping issues where -1 is the majority opinion, but we do not need this distinction to prove hardness.

	e'_1	e'_2	e'_3	e'_4
e_0	-1	-1	-1	
e_1	+1	-1	-1	-1
e_2	+1	+1	-1	-1
e_3	+1	+1	+1	-1

Figure 3: Matrix in the proof of Lemma 13 for $t = 3$. The helper e_0 and e'_4 are depicted above/to the right.

B An Ostrogorski-free Domain

B.1 For External Weights Single-Switch Prevents Ostrogorski's Paradox

In this section, we prove that, under external weights, the single-switch condition guarantees that all IWM proposals are Condorcet winners, i.e., Ostrogorski's paradox does not occur. To this end, we first show that every issue-wise majority proposal does not lose against its opposite, i.e., Anscombe's paradox does not occur. This was already known for the unweighted model [26]. We, however, found the proof in [26] relatively tricky to parse: in part because of certain missing/unclear details and in part because of the relatively involved case distinctions in the second part of the proof. We adapt this proof to external weights and also simplify and streamline it, removing the need for case distinctions by noting symmetries and wisely manipulating *sgn* functions. This yields a shorter and clearer argument that highlights better where the assumption about our proposal being an IWM comes into play, which was not immediate from the original presentation.

Lemma 13. *Consider an external-weights single-switch instance $\mathcal{I} = (\mathcal{P}, w)$. Then, for any issue-wise majority proposal p, we have $p \succcurlyeq_{\mathcal{I}} \overline{p}$.*

Proof. Without loss of generality, assume that ones form a prefix or a suffix on every row of $\mathcal{P}$.[6] For each $1 \le i \le t$ define $e_i \in \mathbb{B}^t$ such that $e_{i,j} = +1$ if $j \le i$, and $e_{i,j} = -1$ otherwise. It follows that every row in $\mathcal{P}$ belongs to the set $\cup_{i=1}^{t} \{e_i, \overline{e_i}\}$. Given a vector $\mathbf{u} \in \mathbb{B}^t$ write $\#_{\mathcal{P}}(\mathbf{u}) := |\{i \in [n] \mid v_i = \mathbf{u}\}|$ for the number of voters in $\mathcal{P}$ whose vote is $\mathbf{u}$, in which case we say that those voters are of *type* $\mathbf{u}$. For $1 \le i \le t$, define $x_i := \#_{\mathcal{P}}(e_i) - \#_{\mathcal{P}}(\overline{e_i})$ to be the number of voters in $\mathcal{P}$ with vote e_i minus the number of voters in $\mathcal{P}$ with vote $\overline{e_i}$. We want to show that $p \succcurlyeq_{\mathcal{I}} \overline{p}$, which amounts to $\sum_{i=1}^{n} sgn(\langle v_i, p - \overline{p} \rangle_w) \ge 0$, which is equivalent to $\sum_{i=1}^{n} sgn(\langle v_i, p \rangle_w) \ge 0$. Since voters can only be of the $2t$ types, this is equivalent to:

$$\sum_{i=1}^{t} \left(\#_{\mathcal{P}}(e_i) \cdot sgn(\langle e_i, p \rangle_w) + \#_{\mathcal{P}}(\overline{e_i}) \cdot sgn(\langle \overline{e_i}, p \rangle_w) \right) \ge 0$$

Since $\langle \overline{e_i}, p \rangle_w = -\langle e_i, p \rangle_w$, this is the same as:

$$\sum_{i=1}^{t} x_i \cdot sgn(\langle e_i, p \rangle_w) \ge 0 \tag{1}$$

To prove this, we are given the fact that p is an issue-wise majority proposal. Recall that for $j \in [t]$ we defined $b_j := \sum_{i=1}^{n} v_{i,j} = \sum_{i=1}^{t} x_i \cdot e_{i,j}$, in which case what we know amounts to $b_j \cdot p_j \ge 0$ for all $j \in [t]$.

To begin the proof, note that the type votes $e_1, \dots, e_t$ form a $t \times t$ matrix with one row for each vote. Let $e'_1, \dots, e'_t$ be the columns of this matrix. For uniformity in the reasoning that follows, it will be helpful to define $e'_{t+1} := \overline{e'_1} = -\mathbf{1}$ and $e_0 := \overline{e_t} = -\mathbf{1}$. See Fig. 3 for an illustration. With these added

[6] Note that assuming this and, at the same time, that $\mathbf{1}$ is an issue-wise majority proposal would lose generality.

conventions, for any $j \in [t]$ we have that e'_j and e'_{j+1} differ only in coordinate j, which is $+1$ in e'_j and -1 in e'_{j+1}, and, moreover, e_j and e_{j-1} differ only in coordinate j, which is $+1$ in e_j and -1 in e_{j-1}. Let us also extend the definition of b to make b_{t+1} well-defined.

As a result, for all $i \in [t]$ we have that $b_i - b_{i+1} = 2 \cdot x_i$ and $\langle e_i - e_{i-1}, p\rangle_w = 2 \cdot w_i \cdot p_i$. Moreover, by definition, $b_{t+1} = -b_1$ and $\langle e_t, p\rangle_w = -\langle e_0, p\rangle_w$, which together imply that $b_{t+1} \cdot sgn(\langle e_t, p\rangle_w) = b_1 \cdot sgn(\langle e_0, p\rangle_w)$.

Armed as such, we substitute in Eq. (1) to get that we need to show that:

$$\sum_{i=1}^{t} \frac{b_i - b_{i+1}}{2} \cdot sgn(\langle e_i, p\rangle_w) \geq 0 \iff$$
$$\sum_{i=1}^{t} (b_i - b_{i+1}) \cdot sgn(\langle e_i, p\rangle_w) \geq 0 \iff$$
$$\sum_{i=1}^{t} b_i \cdot sgn(\langle e_i, p\rangle_w) - \sum_{i=1}^{t} b_{i+1} \cdot sgn(\langle e_i, p\rangle_w) \geq 0$$

The last term in the second sum is $b_{t+1} \cdot sgn(\langle e_t, p\rangle_w) = b_1 \cdot sgn(\langle e_0, p\rangle_w)$, so the second sum stays the same if we change its summation bounds from $(1, t)$ to $(0, t-1)$. Doing so and then rewriting in terms of $i+1$ instead of i, the second sum equals $\sum_{i=1}^{t} b_i \cdot sgn(\langle e_{i-1}, p\rangle_w)$. Combining the two sums and then factoring out the b_i yields:

$$\sum_{i=1}^{t} b_i \cdot (sgn(\langle e_i, p\rangle_w) - sgn(\langle e_{i-1}, p\rangle_w)) \geq 0$$

To show that this is true, we will just show that each term is non-negative. Consider a fixed $i \in [t]$, then we would like to show that $b_i \cdot (sgn(\langle e_i, p\rangle_w) - sgn(\langle e_{i-1}, p\rangle_w)) \geq 0$. This happens if and only if:

$$sgn(b_i) \cdot (sgn(\langle e_i, p\rangle_w) - sgn(\langle e_{i-1}, p\rangle_w)) \geq 0 \iff$$
$$sgn(b_i) \cdot sgn(\langle e_i, p\rangle_w) \geq sgn(b_i) \cdot sgn(\langle e_{i-1}, p\rangle_w) \iff$$
$$sgn(b_i \cdot \langle e_i, p\rangle_w) \geq sgn(b_i \cdot \langle e_{i-1}, p\rangle_w)$$

Because the sign function is monotonic, it hence suffices to show that $b_i \cdot \langle e_i, p\rangle_w \geq b_i \cdot \langle e_{i-1}, p\rangle_w \iff b_i \cdot \langle e_i - e_{i-1}, p\rangle_w \geq 0 \iff b_i \cdot (2 \cdot w_i \cdot p_i) \geq 0$, which is true since $b_i \cdot p_i \geq 0$. $\square$

The general statement that Ostrogorski's paradox does not occur will now follow easily by combining Lemma 13 with the following:

Lemma 14. *Ostrogroski's paradox occurs for an instance $\mathcal{I} = (\mathcal{P}, w)$ in the external-weights model if and only if Anscombe's paradox occurs on an instance $\mathcal{I}'$ obtained from $\mathcal{I}$ by removing (possibly zero) issues from $\mathcal{I}$ and renormalizing the weights to sum up to 1.*

Proof. We prove the two directions separately:

($\Leftarrow$) Assume Anscombe's paradox occurs on an instance $\mathcal{I}'$ obtained from $\mathcal{I}$ by removing a subset of issues $R \subseteq [t]$ from $\mathcal{I}$ and renormalizing the weights to sum up to 1. Then, by definition, there is an IWM proposal p'_{IWM} for $\mathcal{I}'$ such that $\overline{p'_{IWM}} \succ_{\mathcal{I}'} p'_{IWM}$. Complete p'_{IWM} into an IWM proposal p_{IWM} for $\mathcal{I}$ and define p_* to agree with p_{IWM} in topics in R and disagree in topics in $[t] \setminus R$. Then, Ostrogorski's paradox occurs for $\mathcal{I}$: proposal p_{IWM} is an IWM and $p_* \succ_{\mathcal{I}} p_{IWM}$ by construction because $\overline{p'_{IWM}} \succ_{\mathcal{I}'} p'_{IWM}$.

($\Rightarrow$) Assume Ostrogroski's paradox occurs for $\mathcal{I}$. Let p_{IWM} and p_* be such that p_{IWM} is an IWM proposal for $\mathcal{I}$ and $p_* \succ_{\mathcal{I}} p_{IWM}$. Define $R \subseteq [t]$ to be the set of topics in which p_* and p_{IWM} agree, and create $\mathcal{I}'$ from $\mathcal{I}$ by removing issues in R and renormalizing the weights to sum up to 1. Moreover, restrict p_{IWM} to topics in $[t] \setminus R$ to get an IWM proposal p'_{IWM} for $\mathcal{I}'$. Then, Anscombe's paradox occurs for $\mathcal{I}'$: proposal p'_{IWM} is an IWM and $\overline{p'_{IWM}} \succ_{\mathcal{I}'} p'_{IWM}$ by construction because $p_* \succ_{\mathcal{I}} p_{IWM}$. □

Theorem 4. *In the external-weights model, every issue-wise majority proposal of a single-switch instance is a Condorcet winner.*

Proof. Assume this was not the case and consider a single-switch instance $\mathcal{I}$ in the external-weights model for which there exists an issue-wise majority proposal that is not a Condorcet winner (equivalently, Ostrogorski's paradox occurs for $\mathcal{I}$). By Lemma 14, Anscombe's paradox occurs on an instance $\mathcal{I}'$ obtained from $\mathcal{I}$ by removing (possibly zero) issues from $\mathcal{I}$ and renormalizing the weights to sum up to 1. Since $\mathcal{I}$ is single-switch, so is $\mathcal{I}'$. Hence, Anscombe's paradox occurs in a single-switch instance, contradicting Lemma 13. □

B.2 Recognizing Single-Switch Profiles

In this section, we first prove Lemma 5, restated below for convenience. Then, we delve into the task of recognizing single-switch-no-flips profiles, providing an ample guided discussion of the relation between our simpler $O(nt)$ algorithm and related work, including the task of recognizing single-crossing preferences. As a bonus, we discuss a simpler and, at the same time, more efficient algorithm for recognizing single-crossing preferences, running in time $O(nt\sqrt{\log n})$.

Lemma 5. *Consider a profile $\mathcal{P}$ admitting an SSW presentation $\mathcal{P}' = (c_1, \ldots, c_t)$. Then, $\mathcal{P}'_r := (c_2, \ldots, c_t, \overline{c_1})$ is also a SSW presentation of $\mathcal{P}$. Furthermore, any t (circularly) consecutive columns in $\mathcal{P}'' := (c_1, \ldots, c_t, \overline{c_1}, \ldots, \overline{c_t})$ form an SSW presentation of $\mathcal{P}$.*

Proof. It is enough to show this for the case where $\mathcal{P}$ has $n = 1$ rows, in which case $c_1, \ldots, c_t \in \mathbb{B}$. Let us assume $c_1 = 1$, as the other case is analogous. Since $\mathcal{P}'$ is an SSW presentation of $\mathcal{P}$, let $1 \le k \le t$ be such that $\mathcal{P}'$ starts with k ones and the rest are -1's. As a result, $\mathcal{P}'_r$ by definition starts with $k-1$ ones, and the rest are -1's, implying that $\mathcal{P}'_r$ is an SSW presentation of $\mathcal{P}'$, and hence also of $\mathcal{P}$ by transitivity.

To get the part about $\mathcal{P}''$, apply the previous reasoning for $\mathcal{P}'_r$ repeatedly, each time taking the first column, negating it, and moving it to the end. Doing so $2t$ times leads back to the original presentation, and along the way, we get the advertised single-switch presentations. □

Recognizing single-switch-no-flips profiles. We now focus on deciding whether a given preference profile $\mathcal{P}$ is single-switch-no-flips. A first way to do so requires rather involved machinery, by reducing to the *Consecutive Ones Problem* (C1P). In the C1P problem, the input is an $n \times t$ matrix with ± 1 entries. The goal is to permute its columns so that the ones on each row form a consecutive interval. Solving C1P on a matrix with a negated copy of itself appended underneath corresponds to requiring a solution for the original matrix where not only ones are consecutive, but also -1's, meaning that ones form a prefix or a suffix on each row. C1P can be solved in $O(nt)$ time [6], hence giving an immediate solution to check whether $\mathcal{P}$ is single-switch-no-flips within the same time. However, linear-time C1P solvers are complicated and notoriously error-prone: most available implementations fail on at least some edge cases [18]. Moreover, reducing to C1P does not utilize the additional structure present in our problem and hence does not shed light on the structure of all solutions, as we set out to do. We note that this way of checking whether a profile is single-switch-no-flips has previously appeared in [15], where the authors use it to solve the equivalent problem of recognizing voter/candidate-extremal-interval preferences.

A second way to recognize single-switch-no-flips profiles reduces to the problem of recognizing *single-crossing preferences*. In this problem, the input consists of a set A, of size denoted by n, and a list

$(\succ_1, \ldots, \succ_t)$ of linear orders over A. The goal is to permute this list to obtain a new list $(\succ'_1, \ldots, \succ'_t)$ such that for any $a, a' \in A$ with $a \neq a'$ the set $\{j \in [t] : a \succ'_j a'\}$ forms a prefix or a suffix. The reduction is not too difficult: start with a preference profile $\mathcal{P}$ and define $A = \cup_{i \in [n]} \{a_i^0, a_i^1\}$ and a list $(\succ_1, \ldots, \succ_t)$ of linear orders over A as follows. For each $j \in [t]$, order $\succ_j$ ranks the elements in A as $\{a_1^0, a_1^1\} \succ_j \{a_2^0, a_2^1\} \succ_j \ldots \succ_j \{a_n^0, a_n^1\}$, breaking the tie inside each bracket as follows: for each $i \in [n]$, rank $a_i^0 \succ_j a_i^1$ if $v_{i,j} = 1$, and $a_i^1 \succ_j a_i^0$ otherwise. One can check that permutations of the list with the required property correspond to permutations of the columns of $\mathcal{P}$ such that the ones on each row form a prefix or a suffix. Recognizing single-crossing preferences can be achieved in $O(nt \log n))$ time [17, Algorithm 4], meaning that our problem also can. In contrast to the C1P approach, this algorithm has a reasonable implementation. Moreover, by the standard fact that, when it exists, the single-crossing permutation is unique up to reversal, we get that the SSWNF presentation is unique up to reversal whenever it exists (we will give a self-contained proof later on, so we omit the details here for brevity).

The previous approach can be modified to run in time $O(nt)$ by identifying and adapting its super-linear components. Most prominently, [17, Algorithm 4] computes $O(t)$ times the Kendall Tau distance between certain orders $\succ_i$ and $\succ_j$ with $i, j \in [t]$, which is defined as the number of pairs of elements on which $\succ_i$ and $\succ_j$ disagree: $d_{KT}(\succ_i, \succ_j) := \{(a, a') \in A^2 : a \succ_i a' \text{ and } a' \succ_i a\}|$. This is done in time $O(n \log n)$ by finding the number of inversions of a permutation. However, for our particular construction of the orders $(\succ_1, \ldots, \succ_t)$, disagreements between orders can only occur on pairs of the form (a_k^0, a_k^1) with $k \in [n]$, and the number of such disagreements is precisely the Hamming distance $d_H(c_i, c_j)$, which can be computed in time $O(n)$. The algorithm also sorts a list of $O(t)$ integers with values bounded by $O(n^2)$. To get the right time complexity, this is done depending on whether n or t is larger, either in time $O(t \log t)$ or using Counting Sort in time $O(n^2 + t)$. We note that using Radix Sort would have sufficed to make this $O(n + t)$ without the need for a case distinction.[7] For our usage, the values in the list are instead bounded by $O(n)$, so Counting Sort suffices directly (this is because Kendall Tau distances can be quadratic in n, while Hamming distances only linear in n). This completes the required modifications. We note, moreover, that their algorithm proceeds in two stages: first, a candidate list $(\succ'_1, \ldots, \succ'_t)$ is determined, and then it is checked whether it satisfies the single-crossing condition. Moreover, should a solution exist, the candidate list is the unique one up to reversing the list. The second stage is, perhaps surprisingly, the more difficult one to achieve efficiently, and its correctness proof is the subtler part of the argument. For our purposes, however, the first stage suffices since, given the candidate solution, it is easy to check whether ones on each row form a prefix or a suffix in additional time $O(nt)$.

It is possible to give a self-contained $O(nt)$ algorithm for our problem following the outline above (without going through the reduction to single-crossing preferences). However, the resulting algorithm is still arguably not the simplest. Instead, in the following we present a simpler, $O(nt)$ direct algorithm for recognizing single-switch-no-flips profiles $\mathcal{P} = (c_1, \ldots, c_t)$. The algorithm combines insights from [17, Section 4.2] with a simple new observation. We defer further elaboration on the connection with single-crossing preferences until the end of the section for clarity. Our algorithm proceeds as follows: First, find an index x maximizing $d_H(c_1, c_x)$. Then, sort (using Counting Sort) the columns based on their Hamming distance from c_x to get a profile $\mathcal{P}' = (c'_1, \ldots, c'_t)$ where $d_H(c_x, c'_i) \leq d_H(c_x, c'_{i+1})$ for $i \in [t-1]$ (i.e., ties in Hamming distance can be broken arbitrarily). We claim that either $\mathcal{P}'$ is the unique SSWNF presentation of $\mathcal{P}$ (up to reversing the order of the columns), or there is no such presentation. The algorithm runs in $O(nt)$ time. We prove all required claims in the following theorem:

Theorem 7. *There is a* simple $O(nt)$ *algorithm computing (or deciding the inexistence of) an SSWNF presentation of a profile $\mathcal{P}$. Moreover, if it exists, this presentation is unique up to reversing column order.*

[7] With this small modification, and if counting inversions is performed with the $O(n\sqrt{\log n})$ algorithm of Chan and Pătrașcu [13], single-crossing preferences can be recognized in the better time $O(nt\sqrt{\log n})$. We do not give the details here in order not to dilute the message. Instead, we discuss a simpler $O(nt\sqrt{\log n})$ algorithm at the end of the section based on similar ideas.

Proof. We first repeat the full algorithm briefly:

1. Say $\mathcal{P} = (c_1, \ldots, c_t)$;
2. Find an index x maximizing $d_H(c_1, c_x)$;
3. Sort (using Counting Sort) the columns based on their Hamming distance from c_x to get a profile $\mathcal{P}' = (c'_1, \ldots, c'_t)$ where $d_H(c_x, c'_i) \le d_H(c_x, c'_{i+1})$ for $i \in [t-1]$;
4. Check whether ones form a prefix or a suffix on each row of $\mathcal{P}'$. If yes, return $\mathcal{P}'$, else return "not single-switch-no-flips."

Each step of the algorithm runs in time $O(nt)$, so it attains the required time-bound. It remains to show that: (i) if no solution exists, then this is correctly reported; (ii) if $\mathcal{P}^* = (c^*_1, \ldots, c^*_t)$ is an arbitrary SSWNF presentation of $\mathcal{P}$, then the algorithm returns either $\mathcal{P}^*$ or its reverse (which note further implies that the solution is unique up to reversal). Part (i) follows immediately from the last step in the algorithm. To show part (ii), we will need an auxiliary claim:

Claim 15. *Assume $1 \le i < j < k \le t$, then $d_H(c^*_i, c^*_j) \le d_H(c^*_i, c^*_k)$ with equality iff $c^*_j = c^*_k$. Similarly, $d_H(c^*_i, c^*_k) \le d_H(c^*_j, c^*_k)$ with equality iff $c^*_i = c^*_j$.*

Proof. We will only prove the first part, as the second follows analogously. First, observe that for any $u, v \in \mathbb{B}^n$ we can write $d_H(u, v) = \sum_{\ell=1}^{n} \mathbb{I}(u_\ell \ne v_\ell)$. Hence, it suffices to show that for each $1 \le \ell \le n$ we have $\mathbb{I}(c^*_{i,\ell} \ne c^*_{j,\ell}) \le \mathbb{I}(c^*_{i,\ell} \ne c^*_{k,\ell})$ and then sum up those inequalities. This amounts to showing that $c^*_{i,\ell} \ne c^*_{j,\ell} \implies c^*_{i,\ell} \ne c^*_{k,\ell}$. Assume the contrary: $c^*_{k,\ell} = c^*_{i,\ell} \ne c^*_{j,\ell}$. This means that on row ℓ of $\mathcal{P}^*$ columns $i < j < k$ read either $+1, -1, +1$ or $-1, +1, -1$, which either way means that ones do not form a prefix or a suffix on this row, contradicting that $\mathcal{P}^*$ is an SSWNF presentation of $\mathcal{P}$. To get the equality case, note that we have shown $d_H(c^*_i, c^*_j) \le d_H(c^*_i, c^*_k)$ by summing inequalities, so equality occurs iff all summed inequalities are tight; i.e., for all $1 \le \ell \le n$ we have $\mathbb{I}(c^*_{i,\ell} \ne c^*_{j,\ell}) = \mathbb{I}(c^*_{i,\ell} \ne c^*_{k,\ell})$, which amounts to $c^*_{i,\ell} \ne c^*_{j,\ell} \iff c^*_{i,\ell} \ne c^*_{k,\ell}$, and in turn to $c^*_{i,\ell} = c^*_{j,\ell} \iff c^*_{i,\ell} = c^*_{k,\ell}$, which holds iff $c^*_{j,\ell} = c^*_{k,\ell}$. As a result, equality occurs iff for all $1 \le \ell \le n$ we have $c^*_{j,\ell} = c^*_{k,\ell}$, i.e., $c^*_j = c^*_k$. □

Armed as such, we can show that the column c_x maximizing $d_H(c_1, c_x)$ found by the algorithm satisfies $c_x \in \{c^*_1, c^*_t\}$. To see this, let y be such that $c_1 = c^*_y$, then any column c_x that maximizes $d_H(c_1, c_x)$ equals a column c^*_z that maximizes $d_H(c^*_y, c^*_z)$. By Claim 15, it follows that $c_z \in \{c^*_1, c^*_t\}$, so $c_x \in \{c^*_1, c^*_t\}$.

In the following, we will show that if $c_x = c^*_1$, then the output of the algorithm is $\mathcal{P}' = \mathcal{P}^*$, and if $c_x = c^*_t$, then $\mathcal{P}' =$ the reverse of $\mathcal{P}^*$, completing the proof.

The equality parts of Claim 15 give that two columns in $\mathcal{P}^*$ have the same Hamming distance from c^*_1 (or c^*_t), if and only if they are equal. Since $c_x \in \{c^*_1, c^*_t\}$, this means that two columns in $\mathcal{P}$ have the same Hamming distance from c_x if and only if they are equal.[8]

The algorithm constructs $\mathcal{P}'$ by ordering the columns of $\mathcal{P}$ in non-decreasing order of Hamming distance from c_x. By the previous, equal Hamming distances correspond to identical columns, so $\mathcal{P}'$ as defined by us to satisfy $d_H(c_x, c'_i) \le d_H(c_x, c'_{i+1})$ for $i \in [t-1]$ is actually unique no matter the tie-breaking for equal distances.[9]

If $c_x = c^*_1$, note that $(c'_1, \ldots, c'_t) = (c^*_1, \ldots, c^*_t)$ satisfies $d_H(c_x, c'_i) \le d_H(c_x, c'_{i+1})$ for $i \in [t-1]$. By the uniqueness of $(c'_1, \ldots, c'_t)$, this means that $(c^*_1, \ldots, c^*_t)$ is the output of our algorithm, i.e., $\mathcal{P}' = \mathcal{P}^*$.

If $c_x = c^*_t$, the reasoning is analogous, leading to $\mathcal{P}' =$ the reverse of $\mathcal{P}^*$. □

[8] Note that this hinges on our assumption that $\mathcal{P}$ admits $\mathcal{P}^*$ as an SSWNF presentation (and would be false in general).
[9] Fact which again is only true because in this part of the proof we assumed that $\mathcal{P}$ is single-switch-no-flips.

Better recognition for single-crossing preferences. Our $O(nt)$ algorithm for recognizing single-switch-no-flips profiles can be easily modified to recognize single-crossing preferences: Consider an input consisting of a set A and a list $(\succ_1, \dots, \succ_t)$ of linear orders over A. First, find an index x maximizing $d_{KT}(\succ_1, \succ_x)$. Then, sort (using Radix Sort) the list according to the Kendall Tau distance from $\succ_x$ to get a list $(\succ'_1, \dots, \succ'_t)$ where $d_{KT}(\succ_x, \succ'_i) \leq d_{KT}(\succ_x, \succ'_{i+1})$ for $i \in [t-1]$ (i.e., ties can be broken arbitrarily). The resulting list is either the unique permutation of the input list witnessing the single-crossing property (up to reversal), or there is no such permutation. After identifying this candidate solution, one checks whether the single-crossing property is satisfied using the second stage of [17, Algorithm 4], namely [17, Algorithm 2]: the solution is valid if and only if $d_{KT}(\succ'_1, \succ'_i) + d_{KT}(\succ'_i, \succ'_{i+1}) = d_{KT}(\succ'_1, \succ'_{i+1})$ for all $2 \leq i < t$. If the $O(n\sqrt{\log n})$ algorithm of Chan and Pătraşcu [13] is used for counting permutation inversions, the previous yields a simpler, and at the same time, more efficient algorithm for recognizing single-crossing preferences, running in time $O(nt\sqrt{\log n})$. The proof of correctness of this algorithm follows the same outline as the proof of Theorem 7.[10] We also note that our algorithm is, in fact, a more efficient implementation of [17, Algorithm 3]. The latter tries all $O(t)$ options for $\succ'_1$ and, for each one, sorts by Kendall Tau distance from $\succ'_1$ to get a candidate solution, which is then verified as in our case using [17, Algorithm 2]. Our improvement was to notice that $\succ'_1$ can be determined efficiently without trying out all options for it, hence removing a linear factor from the time complexity. In contrast, instead of efficiently determining $\succ'_1$, [17, Algorithm 4] takes a more intricate approach.

Theorem 16. *Single-crossing preferences can be recognized in time $O(nt\sqrt{\log n})$, including producing a witnessing permutation for yes-instances (which is the unique solution up to reversal).*

B.3 Forbidden Subprofiles Characterization of Single-Switch Preferences

In this section, we prove Theorem 9, which establishes the forbidden subprofiles characterization of single-switch profiles. Afterward, we give further details about how our recognition algorithm for single-crossing preferences in Theorem 16 can be bootstrapped to also produce a forbidden subinstance without sacrificing runtime, similarly to the proof of Theorem 10.

To begin with proving Theorem 9, a profile $\mathcal{P}$ is single-switch if and only if $\mathcal{P}'$ is single-switch-no-flips, where $\mathcal{P}'$ is the profile obtained from $\mathcal{P}$ by flipping columns so that the first row is all -1's. As a result, it suffices to understand how short proofs of non-membership look for the class of single-switch-no-flips preferences. To this end, [37] considered the profiles:

$$\mathcal{P}_1 = \begin{bmatrix} +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \end{bmatrix} \quad \mathcal{P}_2 = \begin{bmatrix} +1 & -1 & -1 \\ -1 & +1 & -1 \\ -1 & -1 & +1 \end{bmatrix}$$

and showed that $\mathcal{P}'$ is single-switch-no-flips if and only if it does not contain as subprofiles $\mathcal{P}_1, \mathcal{P}_2$ and the profiles obtainable from them by flipping any subset of rows. In total, one can check that this leads to 5 non-equivalent profiles $\mathcal{P}_1, \dots, \mathcal{P}_5$, where $\mathcal{P}_1, \mathcal{P}_2$ are the ones above and $\mathcal{P}_{2+i}$ for $1 \leq i \leq 3$ is $\mathcal{P}_2$ with the first i rows flipped. The original presentation lists the 5 profiles explicitly, but since both single-switch and single-switch-no-flips preferences are closed under flipping rows, we find our account cleaner. For the interested reader, we note that the result of Terzopoulou et al. [37] can also be recovered from [10], where it is shown that non-membership to the class of single-crossing preferences is always witnessed by one of two small subinstances (this can be done using as a lens the reduction to single-crossing preferences recognition in Appendix B.2).

Lemma 17. *A profile $\mathcal{P}$ is single-switch if and only if $\mathcal{P}'$ does not contain as a subprofile any of $\mathcal{P}_1, \dots, \mathcal{P}_5$. Here $\mathcal{P}'$ denotes the profile obtained from $\mathcal{P}$ by flipping columns so that the first row is all -1's.*

[10] This is no accident: one can reduce from recognizing single-crossing preferences to recognizing single-switch-no-flips profiles by introducing a voter for each pair of distinct elements in A. This incurs a quadratic computational cost but suffices to recover correctness.

This already gives short proofs for non-membership, but for our purposes, we would like a characterization in terms of the subprofiles of $\mathcal{P}$, not of $\mathcal{P}'$. This can be easily achieved given the previous: consider a non-single-switch profile $\mathcal{P}$, then, by the previous, $\mathcal{P}'$ contains one of $\mathcal{P}_1, \ldots, \mathcal{P}_5$, say $\mathcal{P}_i$. Note that $\mathcal{P}_i$ has no row that is all -1's, so $\mathcal{P}_i$ is, in fact, a subprofile of $\mathcal{P}'$ without its first row (which is all -1's). As a result, if we define $\mathcal{P}_i^a$ to be $\mathcal{P}_i$ with a row of -1's appended to the top, then $\mathcal{P}'$ also contains $\mathcal{P}_i^a$ as a subprofile. Given how $\mathcal{P}'$ was obtained from $\mathcal{P}$ by flipping a subset of columns, $\mathcal{P}$ contains a version of $\mathcal{P}_i^a$ with accordingly-flipped columns. Namely, if we let Π_i^a be the set of profiles that can be obtained from $\mathcal{P}_i^a$ by flipping any subset of columns, then $\mathcal{P}$ contains some $\mathcal{X} \in \Pi_i^a$ as a subprofile. Moreover, $\mathcal{X}$ is not single-switch, as $\mathcal{X}' = \mathcal{P}_i^a$ and $\mathcal{P}_i^a$ is not single-switch-no-flips (here $\mathcal{X}'$ denotes $\mathcal{X}$ with columns flipped to make its first row all -1's). As a result, we get the following:

Lemma 18. *A profile $\mathcal{P}$ is single-switch if and only if it contains no profile $\mathcal{X} \in \cup_{i=1}^5 \Pi_i^a$ as a subprofile.*

This result can be compressed into a more elegant form by leveraging the closure of single-switch preferences under flipping rows and columns and the way $\mathcal{P}_3, \ldots, \mathcal{P}_5$ were obtained from $\mathcal{P}_2$, as follows, which is precisely Theorem 9, restated below for convenience:

Theorem 9. *A profile $\mathcal{P}$ is single-switch if and only if it does not contain as a subprofile $\mathcal{P}_1^a, \mathcal{P}_2^a$ and any profile that can be obtained from them by flipping rows and columns:*

$$\mathcal{P}_1^a = \begin{bmatrix} -1 & -1 & -1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \end{bmatrix} \qquad \mathcal{P}_2^a = \begin{bmatrix} -1 & -1 & -1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ -1 & -1 & +1 \end{bmatrix}$$

Finding forbidden subinstances of *single-crossing* preferences. Our idea to use a fast black-box recognition algorithm to bootstrap a fast algorithm for finding a forbidden subprofile extends beyond our usage for single-switch preferences in Theorem 10. The same idea can be used for the class of single-crossing preferences, where non-membership to the class is witnessed by one of two small *forbidden subinstances* with $(t, n) \in \{(4, 4), (3, 6)\}$, as shown in [10]. Given our $O(nt\sqrt{\log n})$ recognition algorithm in Theorem 16, we can apply very similar reasoning to the proof of Theorem 10 (except now we need to split into $6 + 1 = 7$ groups) to get the following result, which, to the best of our knowledge, is new, even if we were to replace our improved time bound with that of the previously-known fastest algorithm, namely [17, Algorithm 4]:

Theorem 19. *Given a no-instance of the problem of recognizing single-crossing preferences, a forbidden subinstance can be determined in time $O(nt\sqrt{\log n})$.*

C Anscombe's Paradox

C.1 External Weights

This section of the appendix centers on Theorem 11. In Appendix C.1.1, we formalize the definition of relevant topics and discuss how to determine the set of relevant topics for a given instance $\mathcal{I} = (\mathcal{P}, w)$ efficiently. Appendix C.1.2 then proves Theorem 11. Throughout this section of the appendix, we use the notation $w(S) = \sum_{j \in S} w_j$ for any $S \subseteq [t]$.

C.1.1 Relevant Topics

A subset of the topics, $T \subseteq [t]$, is a *minimal topic group* under weight vector w if $w(T) > \frac{1}{2}$ and for all $j \in T$ we have that $w(T \setminus \{j\}) \leq \frac{1}{2}$. Then we call a topic $j \in [t]$ a *relevant topic* under w if it is in some minimal topic group under w. We denote the set of all relevant topics under w as R_w.

In proving Theorem 25 (see Appendix C.1.2), we define $B_m = \{p \in \mathbb{B}^t \colon \langle p, p_{IWM} \rangle_w > 0\}$ and select a proposal from it uniformly at random. Note that we can equivalently write $B_m = \{p \in \mathbb{B}^t \colon d_H(p, p_{IWM}, w) < 1/2\}$. We then claim that the only topics $j \in [t]$ such that $\Pr(p_j = (p_{IWM})_j) - \Pr(p_j \neq (p_{IWM})_j) > 0$ are relevant topics. We prove this claim below.

Claim 20. *For p selected uniformly at random from B_m, $\Pr(p_j = (p_{IWM})_j) > \frac{1}{2}$ if and only if $j \in R_w$.*

Proof. Fix some $j \in [t]$. We denote the number of proposals with $+(p_{IWM})_j$ for j and $-(p_{IWM})_j$ for j in B_m by N_+, N_- respectively.

($\Leftarrow$) Assume $j \in R_w$. Consider the bijection $q : \mathbb{B}^t \to \mathbb{B}^t$ that flips the jth entry of the proposal. If $d_H(p, p_{IWM}, w) < 1/2 - w_j$ and $p_j = (p_{IWM})_j$ then $p \in B_m$ and so is $q(p)$, so these cancel each other out when comparing $N_+ - N_-$. Note too that if $p_j = -(p_{IWM})_j$ and $p \in B_m$ then certainly $q(p)$ is as well. So we only have to consider the case where $p = (p_{IWM})_j, p \in B_m$ but $d_H(p, p_{IWM}, w) \geq 1/2 - w_j$. We know such a case must exist because $j \in R_w$. So, this means j is in some minimal topic group, T. Let p be the proposal where all topics in T are set to their value in p_{IWM}, and all remaining topics are set to their value in $-p_{IWM}$. By definition of minimal topic group, we have that $w(T) \leq 1/2 + w_j$, so $d_H(p, p_{IWM}, w) = 1 - w(T) \geq 1/2 - w_j$. Then $d_H(q(p), p_{IWM}, w) \geq 1/2$, so $q(p) \notin B_m$. So we have that $N_+ > N_-$. Hence $\Pr(p_j = (p_{IWM})_j) > 1/2$.

($\Rightarrow$) Assume $\Pr(p_j = ((p_{IWM})_j) > 1/2$. We know that all $p \in B_m$ with $p_j = -(p_{IWM})_j$ have $q(p) \in B_m$, so this means there must exist some $p \notin B_m$ with $p_j = -(p_{IWM})_j$ such that $q(p) \in B_m$. Find the p with the fewest indices agreeing with p_{IWM} such that this holds. Then the set of topics that $q(p)$ agrees with p_{IWM} on forms a minimal topic group for w. To see this, let T be the set of topics that $q(p)$ agrees with p_{IWM} on. We know that $w(T) > 1/2$ and $w(T \setminus \{j\}) \leq 1/2$ because $p \notin B_m, q(p) \in B_m$. Take any other $k \in T, k \neq j$. Assume for sake of contradiction that $W(T \setminus k) > 1/2$. Then we could have taken $p' \notin B_m$ to be equal to p at all indices except flipped for k. Then $p' \notin B_m$ and $q(p') \in B_m$, but p' has one fewer index agreeing with p_{IWM} than p. This contradicts the minimality in our selection of p. Hence T is a valid minimal topic group, and thus $j \in R_w$. $\square$

As Theorem 11 shows, relevant topics are useful for determining whether a proposal with majority support and with weighted agreement of $> 1/2$ with the issue-wise majority exists. How, then can we discern whether specific topics are relevant? This is essentially equivalent to determining whether or not a voter is a dummy voter in a weighted majority game. While this can be shown to be NP-hard with respect to general input w [32], we have the constraint that all elements of w are bounded in $[0, 1]$. We say that w has *polynomial precision* if all of its elements can be expressed as rational numbers with a common denominator that is polynomial in n and t. If additionally we assume that w has polynomial precision then we can give a polynomial time algorithm for determining the set of relevant topics. In order to do this we show the following claim:

Claim 21 (Relevance Monotonicity)**.** *Being a relevant topic is a monotonic property with respect to topic weight.*

Proof. Fix topics i, j such that $w_j \geq w_i$. We want to show that if i is a relevant topic then j is also a relevant topic. Assume i is relevant, then there exists some $S \subseteq [t] \setminus \{i\}$ such that $w(S) \leq 1/2$ but $w(S \cup \{i\}) > 1/2$. We have two cases:

1. $\mathbf{j} \notin \mathbf{S}$. In this case we can reuse S as proof of j's relevance. We have that $w(S) \leq 1/2$ and $w(S \cup \{j\}) = w(S) + w(j) \geq w(S) + w(i) = w(S \cup \{i\}) > 1/2$.

2. $\mathbf{j} \in \mathbf{S}$. In this case we define $S' = (S \setminus \{j\}) \cup \{i\}$ (we just swap in i for j). Then we have that $w(S') = w(S) - w(j) + w(i) \leq w(S) \leq 1/2$. We also can see that $w(S' \cup \{j\}) = w(S \cup \{i\}) > 1/2$.

Therefore, we have shown that if i is relevant, then j is also relevant. This is equivalent to saying that if j is irrelevant, i is irrelevant as well. □

Claim 21 implies that there is a "lowest weight relevant topic" such that all topics with weight less than it are irrelevant, and all topics with greater weight must be relevant. In order to find the lowest weight relevant topic we can run binary search over the topics sorted by weight. For each topic choice j, we can run knapsack on the remaining topics to look for a set such that $\frac{1}{2} - w_j < w(S) \leq 1/2$ as proof of its relevance. Assuming that all weights are integral, which we can achieve by temporarily scaling all of the weights and the respective bounds up, then there is an exact algorithm for knapsack that runs in polynomial time with respect to $|w| = t$ and the max weight item. As we only have to scale up the weights a polynomial factor with respect to nt (due to our assumption on their precision), the max weight item is also polynomial in nt. Therefore, the whole algorithm together is polynomial in nt. As any preference profile input has size at least $n \times t$ (just considering number of entries), this is a polynomial size operation with respect to the size of our overall problem input.

C.1.2 Existence of Representative Non-Losing Proposals in the External Weights Setting

In this section, we prove Theorem 11, restated here:

Theorem 11. *For any $\mathcal{I} = (\mathcal{P}, w)$ and p_{IWM}, there is a weakly majority supported proposal p with $d_H(p, p_{IWM}, w) < 1/2$. If majority is strict in any relevant topic, "weak" can be replaced with "strict".*

At a high level, our proof structure is the following (the same structure as used in [14]): For every voter, we define two bijective relations between proposals. Then, we piece these together to construct a third relation that "swaps" a proposal's weighted similarity to the voter with its weighted similarity to p_{IWM}. We then define two quantities that we take the expectation of. One can easily be shown to be non-negative, and we use the defined relations for each voter to show the equality of these expectations. The non-negativity of the second expectation then implies the existence of a majority-supported proposal, $p \in \mathbb{B}^t$, with $d_H(p_{IWM}, p, w) < 1/2$.

C.1.3 Structure-Preserving Maps

We define three different proposal transformations for each given voter. Fix some voter v and let $B^* \subseteq \mathbb{B}^t$ be the set of proposals p such that $\langle p_{IWM}, p \rangle_w \neq 0$ and $\langle v, p \rangle_w \neq 0$. We partition proposals in B^* into four categories: $T_{i,j}$ where $i, j \in \{-1, +1\}$ and $i = sgn(\langle p, p_{IWM} \rangle_w)$ and $j = sgn(\langle v, p \rangle_w)$. We construct the following "mask" for use in our transformations: $v \odot p_{IWM}$ where $\odot$ is the elementwise (Hadamard) product. This mask has $+1$ for topics on which v and p_{IWM} agree, and -1 for topics on which they disagree. Then we define two transformations $f_v^+, f_v^- : B^* \to B^*$ as follows:

$$f_v^+(p) = p \odot (v \odot p_{IWM}), \quad f_v^-(p) = p \odot -(v \odot p_{IWM})$$

f_v^+ flips a proposal on indices where v and p_{IWM} disagree, while f_v^- flips a proposal on indices where v and p_{IWM} agree. Note that the Hadamard product is commutative and for any vector $m \in \mathbb{B}^t, m \odot m = \mathbf{1}$. Additionally, for any vectors $a, b, c \in \mathbb{B}^t$ we have that $\langle a, b \odot c \rangle_w = \langle a \odot b, c \rangle_w$. From the first two properties, we can immediately see that f_v^+, f_v^- are both self-inverse and hence bijective.

Lemma 22. *f_v^+ maps proposals of type $T_{i,j}$ to proposals of type $T_{j,i}$ for $i, j \in \{0, 1\}$. Moreover, for any $p \in B^*$ we have that $\langle v, f_v^+(p) \rangle_w = \langle p, p_{IWM} \rangle_w$.*

Proof. Fix some $p \in B^*$. Then:

$$\begin{aligned}\langle v, f_v^+(p)\rangle_w = \langle v, p \odot v \odot p_{IWM}\rangle_w = \langle v \odot p \odot v, p_{IWM}\rangle \\ = \langle p, p_{IWM}\rangle_w\end{aligned}$$

As $f_v^+(p) \in B^*$, using the shown equality with one more application of f_v^+ and the fact that f_v^+ is self-inverse gives us that $\langle f_v^+(p), p_{IWM}\rangle_w = \langle v, f_v^+(f_v^+(p))\rangle_w = \langle v, p\rangle_w$. Hence, applying f_v^+ has the effect of swapping a proposal's weighted agreement with v with its weighted agreement with p_{IWM}. Therefore, f_v^+ maps $T_{i,j}$ to $T_{j,i}$. □

Lemma 23. *f_v^- maps proposals of type $T_{i,j}$ to proposals of type $T_{(-j),(-i)}$ for $i, j \in \{-1, +1\}$. For any $p \in B^*$ we have that $\langle v, f_v^-(p)\rangle_w = -\langle p_{IWM}, p\rangle_w$.*

Proof. Fix some $p \in B^*$. Then:

$$\begin{aligned}\langle v, f_v^-(p)\rangle_w = \langle v, p \odot -(v \odot p_{IWM})\rangle_w = \langle v \odot p \odot -v, p_{IWM}\rangle_w \\ = -\langle p, p_{IWM}\rangle_w\end{aligned}$$

Then again, as $f_v^-(p) \in B^*$, using the shown equality with one more application of f_v^- and the fact that its self-inverse gives us that $-\langle f_v^-(p), p_{IWM}\rangle_w = \langle v, f_v^-(f_v^-(p))\rangle_w = \langle v, p\rangle_w$. Hence we have that f^- swaps and negates a proposal's weights agreement with v with its weighted agreement with p_{IWM}. Therefore it indeed maps $T_{i,j}$ to $T_{(-j),(-i)}$. □

Now we combine our two bijective maps into a single map $f_v \colon B^* \to B^*$ defined as follows:

$$f_v(p) = \begin{cases} f_v^+(p) & \text{if } p \text{ is of type } T_{-1,-1} \text{ or } T_{+1,+1} \\ f_v^-(p) & \text{if } p \text{ is of type } T_{-1,+1} \text{ or } T_{+1,-1} \end{cases}$$

It follows from f_v^+ and f_v^- being self-inverse and Lemmas 22 and 23 that f_v is also self-inverse.

Corollary 24. *f_v maps proposals of type $T_{i,j}$ to proposals of type $T_{i,j}$. For any proposal p of type $T_{-1,-1}$ or $T_{+1,+1}$ we have $\langle v, f_v(p)\rangle_w = \langle p, p_{IWM}\rangle_w$ and for any proposal p of type $T_{-1,+1}$ or $T_{+1,-1}$ we have that $\langle v, f_v(p)\rangle_w = -\langle p, p_{IWM}\rangle_w$.*

Proof. This follows directly from the definition of f_v as well as Lemmas 22 and 23. □

C.1.4 Thought Experiments

Now we will detail our two quantities of interest through two thought experiments and show that their expectations are equivalent and non-negative. We define a subset of $\mathbb{B}^t$ denoted as $B_m = \{p \in \mathbb{B}^t \colon \langle p, p_{IWM}\rangle_w > 0\}$. Equivalently, for all $p \in B_m$, $d_H(p, p_{IWM}, w) < 1/2$. As w is externally fixed, this is some known set. We also note that the number of +1s minus the number of -1s for a given topic j in the preference profile can be written as $n \cdot m_j - n(1 - m_j) = n(2m_j - 1)$. Finally, let R_w be the set of all relevant topics under weight vector w (for a discussion of relevant topics, see Appendix C.1.1).

Thought Experiment 1. Our first thought experiment keeps a global counter X initialized to 0 and samples a proposal $p \in B_m$ uniformly at random. For each voter $i \in [n]$, we add $\langle v_i, p\rangle_w$ to X – equivalently, for each voter we go topic by topic and add w_j to X for each topic j voter i agrees with p on, and subtract w_j from X for each topic j voter i disagrees with p on.

Theorem 25. $\mathbb{E}[X] \geq 0$. *If there exists some $j \in R_w$ such that $m_j > 0.5$ then $\mathbb{E}[X] > 0$.*

Proof. We can write X in terms of variables $X_{i,j}$ that take on 1 if $v_{ij} = p_j$ and -1 otherwise as $X = \sum_{i=1}^{n} \sum_{j=1}^{t} w_j X_{i,j}$. Then we can evaluate $\mathbb{E}[X]$ more easily:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^{n} \sum_{j=1}^{t} w_j \mathbb{E}[X_{i,j}] = \sum_{j=1}^{t} w_j \sum_{i=1}^{n} \mathbb{E}[X_{i,j}] \\ &= \sum_{j=1}^{t} w_j \cdot n(2m_j - 1)(\Pr(p_j = +1) - \Pr(p_j = -1))\end{aligned}$$

We arrive at the last line because if $p_j = +1$ then $\sum_{i=1}^{n} X_{i,j} = \sum_{i \in [n]:\, v_{i,j}=+1} 1 + \sum_{i \in [n]:\, v_{i,j}=-1} -1 = n(2m_j - 1)$ and similarly if $p_j = -1$ then $\sum_{i=1}^{n} X_{i,j} = -n(2m_j - 1)$. Note that we can rewrite the sum with only the terms where there is a strict majority of +1 (so $m_j > 0.5$), because all terms where $m_j = 0.5$ evaluate to 0:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{\substack{j \in [t] \\ m_j > 0.5}} w_j n(2m_j - 1)(\Pr(p_j = +1) - \Pr(p_j = -1)) \\ &= \sum_{\substack{j \in [t] \\ m_j > 0.5}} w_j n(2m_j - 1)(\Pr(p_j = (p_{IWM})_j) - \Pr(p_j \neq (p_{IWM})_j))\end{aligned}$$

We arrive at the last line by noting that for all j with $m_j > 0.5$, $(p_{IWM})_j$ must be $+1$ as it is the unique majority for that issue.

Now we observe that $\Pr(p_j = (p_{IWM})_j) \geq \Pr(p_j \neq (p_{IWM})_j)$ for all $j \in [t]$. Fix some $j \in [t]$. Then for any $p \in B_m$ such that $p_j \neq (p_{IWM})_j$, we know that there exists a $p' \in B_m$ that matches p on all entries except j, where $p'_j = (p_{IWM})_j$. p' is indeed in B_m because $\langle p', p_{IWM} \rangle_w = \langle p, p_{IWM} \rangle_w + 2w_j > 1/2$. Note that this mapping from p to p' is injective, so we have that there are at least as many proposals in B_m with a $(p_{IWM})_j$ for issue j as there are with a $-(p_{IWM})_j$. Hence, when selecting a proposal from B_m uniformly at random, the probability that it has a $(p_{IWM})_j$ for issue j is at least the probability that it has a $-(p_{IWM})_j$.

We also claim that $\Pr(p_j = (p_{IWM})_j) > \Pr(p_j \neq (p_{IWM})_j)$ if and only if $j \in R_w$. To see a proof of this claim, please refer to Appendix C.1.1 Claim 20. Therefore, we know that for all $j \notin R_w$, $\Pr(p_j = (p_{IWM})_j) - \Pr(p_j \neq (p_{IWM})_j) = 0$. Then we can rewrite our expectation only in terms of the relevant topics:

$$\mathbb{E}[X] = \sum_{\substack{j \in R_w \\ m_j > 0.5}} w_j n(2m_j - 1)(\Pr(p_j = (p_{IWM})_j) - \Pr(p_j \neq (p_{IWM})_j))$$

For any relevant topic j, w_j is strictly positive (otherwise j would not be in a minimal topic group) and as just mentioned $\Pr(p_j = (p_{IWM})_j) - \Pr(p_j \neq (p_{IWM})_j)$ is strictly positive as well. $n(2m_j - 1)$ must be positive in all of our terms because we only consider $j \in R_w$ such that $m_j > 0.5$. Therefore, every term in the sum is positive. If no relevant topic has $m_j > 0.5$ then $\mathbb{E}[X] = 0$. Otherwise, we have that there exists at least one term left in the sum, and $\mathbb{E}[X] > 0$. $\square$

Thought Experiment 2. For our second thought experiment we again sample $p \in B_m$ uniformly at random and maintain a global counter Y initialized to 0. Each voter i compares v_i with p. If they approve of p then they add $\langle p, p_{IWM} \rangle_w$ to Y, and if they disapprove of p then they subtract $\langle p, p_{IWM} \rangle_w$ from Y. If they are neutral (so $d_H(v_i, p, w) = 1/2$) then they leave Y unchanged.

Theorem 26. $\mathbb{E}[X] = \mathbb{E}[Y]$

Proof. First we write $Y = \sum_{i\in[n]} Y_i$ where Y_i is $\langle p, p_{IWM}\rangle_w$ if voter i approves of p, $-\langle p, p_{IWM}\rangle_w$ if voter i disapproves of p, and 0 if voter i is neutral. Then $\mathbb{E}[Y] = \sum_{i\in[n]} \mathbb{E}[Y_i]$ by linearity of expectation. From the first thought experiment, we have that $\mathbb{E}[X] = \sum_{i\in[n]} \mathbb{E}[\langle v_i, p\rangle_w]$ by definition of X. Hence, to show that $\mathbb{E}[X] = \mathbb{E}[Y]$, it suffices to show that for all $i \in [n]$ we have that $\mathbb{E}[Y_i] = \mathbb{E}[\langle v_i, p\rangle_w]$.

Fix some voter $i \in [n]$. Let B_+ and B_- be the sets of proposals in B_m that voter i approves of and disapproves of, respectively. Note that if i is neutral about B_m then $\langle v_i, p\rangle_w = 0$. Then we have that:

$$\begin{aligned}
\mathbb{E}[\langle v_i, p\rangle_w] &= |B_m|^{-1} \sum_{p\in B_m} \langle v_i, p\rangle_w \\
&= |B_m|^{-1} \left(\sum_{p\in B_+} \langle v_i, p\rangle_w + \sum_{p\in B_-} \langle v_i, p\rangle_w \right)
\end{aligned}$$

Then, because f_{v_i} is self-inverse, we can write:

$$\begin{aligned}
\mathbb{E}[\langle v_i, p\rangle_w] &= |B_m|^{-1} \left(\sum_{p\in B_+} \langle v_i, f_{v_i}(f_{v_i}(p))\rangle_w + \sum_{p\in B_-} \langle v_i, f_{v_i}(f_{v_i}(p))\rangle_w \right) \\
&= |B_m|^{-1} \left(\sum_{p\in B_+} \langle f_{v_i}(p), p_{IWM}\rangle_w - \sum_{p\in B_-} \langle f_{v_i}(p), p_{IWM}\rangle_w \right)
\end{aligned}$$

In the last line we use the fact that for $p \in B_+$, $p \in T_{+1,+1}$ because $p \in B_m$ so $\langle p, p_{IWM}\rangle_w > 0$ and $p \in B_+$ implies that $\langle v_i, p\rangle_w > 0$. By Corollary 24, we know that $f_{v_i}(p)$ is then also in $T_{+1,+1}$, and hence that $\langle v_i, f_{v_i}(f_{v_i}(p))\rangle_w = \langle f_{v_i}(p), p_{IWM}\rangle_w$. Similarly, we have that for $p \in B_-$, $p \in T_{+1,-1}$ because $p \in B_m$ so $\langle p, p_{IWM}\rangle_w > 0$ and $p \in B_-$ implies that $\langle v_i, p\rangle_w < 0$. Then $f_{v_i}(p)$ is also in $T_{+1,-1}$, so $\langle v_i, f_{v_i}(f_{v_i}(p))\rangle_w = -\langle f_{v_i}(p), p_{IWM}\rangle_w$. As f_{v_i} is a bijection on both B_+ and B_-, summing over a function of $f_{v_i}(p)$ for all $p \in B_+(B_-)$ is equivalent to summing over that function of p for all $p \in B_+(B_-)$:

$$\begin{aligned}
\mathbb{E}[\langle v_i, p\rangle_w] &= |B_m|^{-1} \left(\sum_{p\in B_+} \langle p, p_{IWM}\rangle_w - \sum_{p\in B_-} \langle p, p_{IWM}\rangle_w \right) \\
&= \mathbb{E}[Y_i]
\end{aligned}$$

Therefore, we have shown that $\mathbb{E}[X] = \mathbb{E}[Y]$. □

Finally, we can prove Theorem 11:

Proof. By Theorem 26, we know that $\mathbb{E}[Y] = \mathbb{E}[X]$, and by Theorem 25 we have that this quantity is always non-negative and is strictly positive when in there exists some relevant topic j such that $m_j > 0.5$. We call this condition the strict majority guarantee. Then there exists some proposal $p \in B_m$ such that $Y \geq 0$, or $Y > 0$ if we have the strict majority guarantee.

We now rewrite Y in a way that makes the connection to majority-supported proposals more explicit. We can write $Y = \langle p, p_{IWM}\rangle_w \cdot ((\#\text{ voters supporting } p) - (\#\text{ voters opposing } p))$. As $p \in B_m$, we know that $\langle p, p_{IWM}\rangle_w > 0$. Hence we have that:

$$\begin{aligned}
(\#\text{ voters supporting } p) - (\#\text{ voters opposing } p) &= \frac{Y}{\langle p, p_{IWM}\rangle} \\
&\geq 0
\end{aligned}$$

where the inequality is strict if we have the strict majority guarantee. As the number of approving voters is at least the number of disapproving voters, p is weakly majority-supported. With the strict majority guarantee, $Y > 0$ so we have that the number of approving voters is strictly greater than the number of disapproving voters, and p is strictly majority-supported. $\square$

C.2 Internal Weights: High Level Overview of Results

We now explore a model where individuals can have unique weight vectors, expressing not only diverse preferences on issue outcomes but also differing opinions on relative topic importance.

Internal Weights Model. In the internal weights model, an instance $\mathcal{I} = (\mathcal{P}, W)$ consists of a preference profile $\mathcal{P}$ and a weight profile W with rows $w_1, \ldots, w_n$ where each weight vector w_i corresponds to voter i, is non-negative, and sums to 1. The *average weight vector* is defined as $\tilde{w} := \frac{1}{n}\sum_{i=1}^{n} w_i$. Zero entries in the average weight vector correspond to issues that no voters placed any weight on (and hence can be ignored). We assume no such topics exist without loss of generality. We define the *majority* for a given topic j to be $m_j := \frac{1}{n\tilde{w}_j}\sum_{i=1}^{n} w_{i,j} \cdot \mathbb{I}(v_{i,j} = +1)$. This is the fraction of voter weight placed on that issue that prefers $+1$. Note that this agrees with our previous definition for external weights (where it was just the fraction of voters that prefer $+1$ on that topic). The average majority for a given preference profile is defined as $\tilde{m} := \sum_{j=1}^{t} \tilde{w}_j m_j$. This naturally weights consensus on issues proportionally to how important those issues are to the population.

Under external weights, we could give a constant upper bound (Theorem 11) on the minimum distance of some majority-supported proposal from an IWM, independent of the weight profile. As we will see in Theorems 27 and 28, the severity of Anscombe's Paradox under *internal* weights is closely related to the maximum average topic weight $\tilde{w}_{max}$ (the maximum entry in $\tilde{w}$). Formally, we will upper bound the worst-case *IWM distance* g_ℓ for instances with maximum average topic weight $\tilde{w}_{max} = \ell \in (0, 1)$ and selections of p_{IWM} for the instance:

$$g_\ell := \max_{\substack{\mathcal{I}=(\mathcal{P},W),\ p_{IWM} \\ s.t. \tilde{w}_{max}=\ell}} \left(\min_{p \text{ weakly majority-supported}} d_H(p, p_{IWM}, \tilde{w}) \right)$$

We first give a simple upper bound on g_ℓ for $\ell \in (0, 1)$ derived from a partition-based algorithm. Surprisingly, we then show that this seemingly weak upper bound is tight for a large portion of the range $\tilde{w}_{max} \in (0, 1)$. Our lower-bound constructions more strongly imply the existence of instances where *all* weakly majority-supported proposals are far from *all* IWM's. Fig. 4 provides a summary of the bounds we give on $g_{\tilde{w}_{max}}$.

Partition-based upper bounds. Theorem 27 guarantees both the existence of reasonable majority-supported proposals and provides an algorithm to efficiently recover them.

Theorem 27. *We have the following upper bounds on g_ℓ:*

- *If $\ell \in (0, 1/3)$, then $g_\ell \leq 1/2 + \ell/2$;*
- *If $\ell \in [1/3, 1/2]$, then $g_\ell \leq 1 - \ell$;*
- *If $\ell \in (1/2, 1)$, then $g_\ell \leq \ell$.*

In each case, we can compute a weakly majority-supported proposal p with $d_H(p, p_{IWM}, \tilde{w})$ at most the given bound in polynomial time.

The full proof is deferred to Appendix C.3.1, but the intuition is as follows: for any proposal, either it or its complement will get weak majority support (potentially both), and for any p_{IWM}, $d_H(\overline{p}, p_{IWM}, \tilde{w}) =$

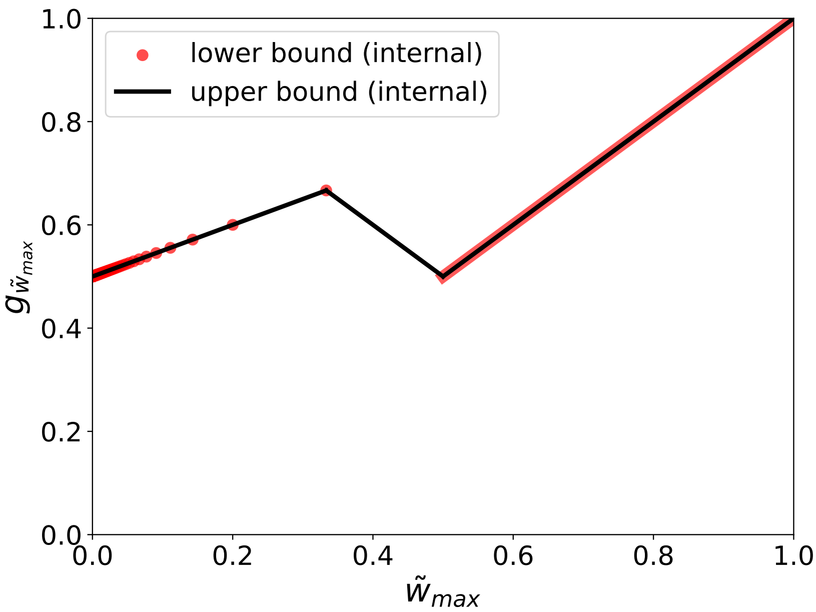

Figure 4: A summary of our bounds on $g_{\tilde{w}_{max}}$.

$1 - d_H(p, p_{IWM}, \tilde{w})$. Therefore, we construct p that keeps $\max\{d_H(\overline{p}, p_{IWM}, \tilde{w}), d_H(p, p_{IWM}, \tilde{w})\}$ small. This is ultimately equivalent to the partition optimization problem with the t entries in the average weight vector as inputs. Our bounds are constructive and give the pair $(p, \overline{p})$ achieving the bound.

Lower bounds. By definition, $g_\ell \leq 1$, so the upper bounds in Theorem 27 might seem fairly weak. However, in Theorem 28, we show that they are actually tight for many values of ℓ. This implies that when $\tilde{w}_{max}$ is large, $g_{\tilde{w}_{max}}$ can get arbitrarily close to 1.

Theorem 28. *The following lower bounds for g_ℓ hold:*

- *If $\ell = 1/(2k+1)$ with $k \in \mathbb{Z}_{\geq 0}$, then $g_\ell \geq 1/2 + \ell/2$;*
- *If $\ell \in (1/2, 1)$, then $g_\ell \geq \ell$.*

We conjecture that the upper bounds in Theorem 27 are tight for the remaining values of ℓ, but leave this to future work. The proof of Theorem 28 is deferred to Appendix C.3.2, but we provide the construction for $\ell \in (1/2, 1)$ and some intuition here. In the instance below, we choose x large enough such that $\tilde{w}_{max} = \ell$ and the first issue holds a strict majority of the weight for all voters. There are x copies of the first voter, and $x + 1$ copies of the second.

$$\mathcal{P} = \begin{matrix} (x) \times \\ (x+1) \times \end{matrix} \begin{bmatrix} +1 & +1 \\ -1 & +1 \end{bmatrix} \quad \mathcal{W} = \begin{matrix} (x) \times \\ (x+1) \times \end{matrix} \begin{bmatrix} \frac{x+1}{x} \cdot \ell & 1 - \frac{x+1}{x} \cdot \ell \\ \frac{x}{x+1} \cdot \ell & 1 - \frac{x}{x+1} \cdot \ell \end{bmatrix}$$

In this instance, all voters are essentially "single-issue voters" on the first topic, but the second type of voters split their weight slightly more evenly between the two topics. $+1$ is the weighted majority opinion on the first topic, but any proposal with $+1$ for that topic will not get majority support because voters of the second type will oppose it. Notably, $\mathbf{1}$ is the unique IWM in our constructions, implying there is no majority-supported proposal close to *any* IWM.

Theorem 28 quashes any hope of improving on Theorem 27 and proving a similar result to the external weights setting (where $g_\ell < 1/2$ held for any weights profile). Once voters can have distinct weight vectors, increasing $\tilde{w}_{max}$ can make the distance between all majority-supported proposals and IWM proposals arbitrarily large. We conclude this section by characterizing a group of voting instances in which Anscombe's Paradox will not occur.

Condition precluding Anscombe's Paradox. We find that generalizations of Wagner's Rule of Three-Fourths hold in both the external and internal weights settings:

Theorem 29. *If $\tilde{m} \geq 3/4$ then Anscombe's paradox will not occur. Additionally, if $m_j \geq 3/4$ for all $j \in [t]$ in the external weights setting, then Ostrogorski's paradox will not occur.*

Our proof (deferred to Appendix C.4) follows Wagner's original proof strategy of counting agreement with an IWM in an instance in two ways: column-wise and row-wise, but is modified to account for weights. We get the second part of our claim by using the fact that, under external weights, Ostrogorski's paradox occurs if and only if there is a subset of issues inducing an instance where Anscombe's paradox occurs.

C.3 Internal Weights: Non-Losing Proposals

In this section of the appendix, we provide proofs and further discussion of the results presented in Appendix C.2. In Appendix C.3.1 we provide intuition and proof for our upper bounds on g_ℓ in Theorem 27, and in Appendix C.3.2 we provide proof for our lower bounds on g_ℓ in Theorem 28. Finally in Appendix C.4 we prove the generalized Rule of Three-Fourths. Throughout this section of the appendix, we use the notation that $w(S) = \sum_{j \in S} w_j$ for any $S \subseteq [t]$.

C.3.1 Efficiently Finding Reasonable Non-Losing Proposals in Individual Weight Setting

As discussed in the body, for any proposal $p \in \mathbb{B}^t$, at least one of $p, \overline{p}$ is weakly majority-supported. Therefore, one way to get a simple upper bound on g_ℓ is to construct p given any p_{IWM} and take $\max\{d_H(p, p_{IWM}, \tilde{w}), d_H(\overline{p}, p_{IWM}, \tilde{w})\}$. Hence, to get a tighter upper bound, we would like to construct p such that this quantity is small. As $d_H(\overline{p}, p_{IWM}, \tilde{w}) = 1 - d_H(p, p_{IWM}, \tilde{w})$, this problem corresponds to the partition problem, because we want to partition the topics into two sets such that their average weight sums are as close as possible. We formalize this intuition below.

Theorem 27. *We have the following upper bounds on g_ℓ:*

- *If $\ell \in (0, 1/3)$, then $g_\ell \leq 1/2 + \ell/2$;*
- *If $\ell \in [1/3, 1/2]$, then $g_\ell \leq 1 - \ell$;*
- *If $\ell \in (1/2, 1)$, then $g_\ell \leq \ell$.*

In each case, we can compute a weakly majority-supported proposal p with $d_H(p, p_{IWM}, \tilde{w})$ at most the given bound in polynomial time.

Proof. Fix some instance $\mathcal{I} = (\mathcal{P}, W)$ with average weight vector $\tilde{w}_{\max}$. We will case on $\tilde{w}_{max}$ and construct a subset of topics, T, for later use in creating the proposal complement pair of interest.

- If $\tilde{w}_{\max} \in (0, 1/3)$, construct a set $S \subseteq [t]$ by starting with $S = [t]$ and then removing topics until we cannot remove anymore without the total weight going below $1/2$. We know that $\tilde{w}(S) \leq 1/2 + \tilde{w}_{max}$, as otherwise we could remove another topic without the weight dipping below $1/2$ (as all topics have weight at most $\tilde{w}_{max}$). If $\tilde{w}(S) \leq 1/2 + \tilde{w}_{max}/2$ then let $T = S$. Otherwise, we have that $\tilde{w}(S) > 1/2 + \tilde{w}_{max}/2$. Take any topic $j \in S$ and let $S' := S \setminus \{j\}$

and $S'' = [t] \setminus S'$. Then $\tilde{w}(S') = \tilde{w}(S) - \tilde{w}_j$ and $\tilde{w}(S'') = t - \tilde{w}(S')$. By construction of S we know that $\tilde{w}(S') < 1/2$. We can also lower bound it by using our lower bound on $\tilde{w}(S)$ and upper bounding $\tilde{w}_j$ by $\tilde{w}_{max}$: $\tilde{w}(S') \geq 1/2 + \tilde{w}_{max}/2 - \tilde{w}_{max} = 1/2 - \tilde{w}_{max}/2$. Then we have that $\tilde{w}(S'') \geq 1/2$ and $\tilde{w}(S'') \leq 1/2 + \tilde{w}_{max}/2$. Then let $T = S''$. Hence, in all cases $1/2 \leq \tilde{w}(T) \leq 1/2 + \tilde{w}_{max}/2$.

- If $\tilde{w}_{max} \in [1/3, 1/2]$, let $T = [t] \setminus \{j_{max}\}$ where $j_{max} \in [t]$ is the index of some maximum weight topic. Then $1/2 \leq \tilde{w}(T) = 1 - \tilde{w}_{max}$.
- If $\tilde{w}_{max} \in (1/2, 1)$ then let $T = \{j_{max}\}$. Then $1/2 \leq \tilde{w}(T) = \tilde{w}_{max}$.

We construct a proposal p as follows: for all $j \in T$, set $p_j = (p_{IWM})_j$, and for all $j \in [t] \setminus T$ set $p_j = -(p_{IWM})_j$. Then we have that p agrees with p_{IWM} on all topics in T and disagrees with p_{IWM} on all other topics. Then $d_H(p, p_{IWM}, \tilde{w}) = 1 - \tilde{w}(T)$ and $d_H(\overline{p}, p_{IWM}, \tilde{w}) = \tilde{w}(T)$. We know that at least one of these is weakly majority-supported. Hence,

$$\begin{aligned} \min_{p \text{ weakly majority-supported}} d_H(p, p_{IWM}, \tilde{w}) &\leq \max\{\tilde{w}(T), 1 - \tilde{w}(T)\} \\ &= \tilde{w}(T) \end{aligned}$$

We arrive at the second line because $\tilde{w}(T) \geq 1/2$. This bound holds for any selection of $\mathcal{I}$ with the same maximum average weight and for any selection of p_{IWM} for $\mathcal{I}$. Hence, $g_{\tilde{w}_{max}} \leq \tilde{w}(T)$, which gives us the desired bounds. Additionally, either p or $\overline{p}$ are weakly majority-supported and have the desired distance from p_{IWM}. Determining T, constructing $p, \overline{p}$, and checking their support takes time $O(nt)$ altogether. □

We note that if one wants to find a weakly majority-supported proposal with distance as close as possible to $1/2$ from a designated p_{IWM} using the proposal complement pair technique, this can also be done in polynomial time with an additional assumption on the average weight vector. Specifically, we assume that W has *polynomial precision*, so all of its elements can be expressed as rational numbers with a common denominator that is polynomial in n and t. We first scale up all elements of $\tilde{w}$ by the common denominator, so that they are all integers. By our assumption, we know that their scaled-up counterparts are polynomial in n and t. We then use the standard reduction from partition to knapsack and run the pseudo-polynomial DP for knapsack. Finally, we check which one of p or $\overline{p}$ (or both) has weak majority support.

C.3.2 Arbitrarily Low Agreement with the IWM

We provide the full proof for Theorem 28 here:

Theorem 28. *The following lower bounds for g_ℓ hold:*

- *If $\ell = 1/(2k+1)$ with $k \in \mathbb{Z}_{\geq 0}$, then $g_\ell \geq 1/2 + \ell/2$;*
- *If $\ell \in (1/2, 1)$, then $g_\ell \geq \ell$.*

Proof. We will actually show something slightly stronger than just upper bounding g_ℓ in this proof. In both ranges of ℓ, our constructions will be such that $\mathbf{1}$ is the *unique* IWM for the instance. Hence, the fact that any non-losing proposal is far away from $\mathbf{1}$ not only implies the bound on g_ℓ, it also implies that there are instances where any non-losing proposal is far away from *any* IWM for that instance.

Small ℓ. Fix some $k \in \mathbb{N}^+$ and let $\ell = 1/(2k+1)$. We illustrate the construction below for $k = 1$ for clarity and then describe the generalization to larger k.

$$\mathcal{P} = \begin{array}{r} 5\times \\ 5\times \\ 5\times \\ 4\times \end{array} \begin{bmatrix} -1 & +1 & +1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \end{bmatrix} \qquad \mathcal{W} = \begin{array}{r} 5\times \\ 5\times \\ 5\times \\ 4\times \end{array} \begin{bmatrix} 3/5 & 1/5 & 1/5 \\ 1/5 & 3/5 & 1/5 \\ 1/5 & 1/5 & 3/5 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

The generalization is as follows: we have $t = 2k + 1$ topics and $t + 1$ types of voters. Denote type i's preference and weight vectors as v^i and w^i respectively. There are $2t - 1$ copies of each of the first t types, and $t + 1$ copies of the last type of voter. The voters of type $t + 1$ prefer the all-ones vector and consider all issues to be equally important: $v^{t+1} = +\mathbf{1}$ and $w^{t+1} = (1/t, 1/t, \ldots, 1/t)$. Voters of type $i \in \{1, \ldots, t\}$ are single-issue voters on issue i and prefer the negative outcome, although they do place some importance on the other issues:

$$v^i_j = \begin{cases} -1 & \text{if } i = j \\ +1 & \text{o.w.} \end{cases} \qquad w^i_j = \begin{cases} \frac{t}{2t-1} & \text{if } i = j \\ \frac{1}{2t-1} & \text{o.w.} \end{cases}$$

We say that they are single-issue voters because they vote for a proposal if and only if it agrees with their position on that issue. Note that by symmetry of the weight matrix, all topics have the same weight in the average weight vector. As there are $t = 2k+1$ topics, $\tilde{w}_j = 1/(2k+1) = \ell$ for all $j \in [t]$. Therefore, ℓ is indeed $\tilde{w}_{max}$ of this profile. Additionally, all of the weights are nonnegative, and every weight vector type sums to 1, as $\frac{t}{2t-1} + (t-1) \cdot \frac{1}{2t-1} = 1$.

Now we show that $+\mathbf{1}$ is the unique IWM for this preference and weight profile. For any given topic, the total weight on $+1$ is $(t-1) \cdot (2t-1) \cdot \frac{1}{2t-1} + (t+1) \cdot \frac{1}{t}$, and the weight on -1 is $(2t-1) \cdot \frac{t}{2t-1} = t$. We have that $(t-1) \cdot (2t-1) \cdot \frac{1}{2t-1} + (t+1) \cdot \frac{1}{t} = t + \frac{1}{t} > t$. Therefore, $+1$ is the strict majority opinion on all topics.

Fix any $p \in \mathbb{B}^t$ such that $d_H(p, \mathbf{1}, \tilde{w}) < 1/2 + \ell/2$. Given our construction, this means that at least $k+1$ of the $2k+1$ topics are set to $+1$ in the proposal. Note that for all $i \in [t]$ such that $p_i = +1$, all voters of type i vote no on p. As there are at least $k+1$ of these indices, we have that at least $k+1$ types of voters vote against p. This corresponds to at least $(2t-1) \cdot (k+1)$ voters. The proposal can then get at most all of the remaining votes, which amount to $(2t-1) \cdot k + (t+1)$ votes. We have that $(2t-1)(k+1) = (2t-1)k + (2t-1) > (2t-1)k + (t+1)$ because $t \geq 3$. Therefore, p receives strictly more votes against it than for it. Hence, any non-losing proposal must have distance at least $1/2 + \ell/2$ from $\mathbf{1}$.

Big $\boldsymbol{\ell}$. Fix some $\ell \in (1/2, 1)$. We will construct a preference and weight profile, $\mathcal{P}, \mathcal{W}$, such that $\tilde{w}_{max} = \ell$, $\mathbf{1}$ is the sole IWM, and any non-losing proposal p has $d_H(p, \mathbf{1}, \tilde{w}) \geq \ell$. Let $x \in \mathbb{N}^+$ be such that $x > \max\left\{\frac{\ell}{1-\ell}, \frac{1}{2\ell-1}\right\}$. Note that both of the denominators are strictly positive because of our bounds on ℓ.

$$\mathcal{P} = \begin{array}{r} x\times \\ x+1\times \end{array} \begin{bmatrix} +1 & +1 \\ -1 & +1 \end{bmatrix} \qquad \mathcal{W} = \begin{array}{r} x\times \\ x+1\times \end{array} \begin{bmatrix} \frac{x+1}{x} \cdot \ell & 1 - \frac{x+1}{x} \cdot \ell \\ \frac{x}{x+1} \cdot \ell & 1 - \frac{x}{x+1} \cdot \ell \end{bmatrix}$$

First we show that all elements of $\mathcal{W}$ are in $[0, 1]$. It suffices just to show that this is true for the weights on the first issue, as every row sums to 1 and there are only two issues. More specifically, we will show all voters place weight in the range $(0.5, 1)$ on the first topic. First we upper bound:

$$\frac{\ell(x+1)}{x} < 1 \iff \ell x + \ell < x \iff \ell < x(1-\ell) \iff \frac{\ell}{1-\ell} < x$$

where the final inequality holds by our definition of x. Additionally,

$$\frac{\ell x}{x+1} > 0.5 \iff 2\ell x > x+1 \iff x > \frac{1}{2\ell-1}$$

where again the final inequality holds by our definition of x. Putting everything together we have that:

$$0.5 < \frac{\ell x}{x+1} \leq \frac{\ell(x+1)}{x} < 1$$

Not only does this confirm that $\mathcal{W}$ is a valid weight profile, it also informs us that both types of voters in this scenario are single issue voters on the first issue — they vote for a proposal if and only if it agrees with their preference on the first issue. Additionally, it gives us that $\tilde{w}_1 = \tilde{w}_{max}$. We have that

$$\begin{aligned}\tilde{w}_1 &= \left(x \cdot \frac{\ell(x+1)}{x} + (x+1) \cdot \frac{\ell x}{x+1}\right) \frac{1}{2x+1} \\ &= \frac{\ell(x+1) + \ell x}{2x+1} = \ell\end{aligned}$$

Hence, $\tilde{w}_{max} = \ell$ as desired.

Next, we show that $\mathbf{1}$ is the sole IWM for this profile. Clearly $+1$ is the unanimous majority opinion for the second topic. For the first topic, we have that $+1$ is the unique majority opinion if there is strictly more total weight on $+1$ than on -1 for that topic:

$$x \cdot \frac{\ell(x+1)}{x} > (x+1) \cdot \frac{\ell x}{x+1} \iff x+1 > x$$

Therefore, $+1$ is the strict majority for both issues and as such $\mathbf{1}$ is the unique IWM for this profile.

Finally, we claim that any non-losing proposal p must have $p_1 = -1$. To see this, recall that all voters are single-issue voters on the first topic. All $x+1$ voters with -1 as their preference for the first topic would vote against p if $p_1 = +1$. As they form a majority of voters, the proposal would lose. Therefore, $d_H(p, \mathbf{1}, \tilde{w}) \geq \tilde{w}_1 = \ell$.

It is also worth noting that this preference profile is single-switch, but Ostrogorski's paradox happens: $(-1, +1) \succ_{\mathcal{I}} (+1, +1)$. This instance highlights that the single-switch condition *does not* help for the internal weights setting. □

C.4 Internal Weights: Condition Precluding Anscombe's Paradox

Here we include the full proof for our generalized Rule of Three-Fourths, Theorem 29:

Theorem 29. *If $\tilde{m} \geq 3/4$ then Anscombe's paradox will not occur. Additionally, if $m_j \geq 3/4$ for all $j \in [t]$ in the external weights setting, then Ostrogorski's paradox will not occur.*

Proof. Fix an instance $\mathcal{I} = (\mathcal{P}, W)$ in the internal weights setting such that $\tilde{m} \geq 3/4$. Notice if W has identical rows, then this is equivalent to the external weights setting. We assume without loss of generality that the IWM proposal we are interested in verifying gets weak majority support is $\mathbf{1}$. This is indeed without loss of generality because if our original p_{IWM} has a -1 for some topic, j, we know that $m_j = 0.5$ (as we assume for all of Section 5 that $m_j \geq 0.5$, so for -1 to be a majority, the column must be exactly split). We can then flip all entries in that column of the preference profile — this is equivalent to having voters express their preferences on the negated version of the issue. Then $(p_{IWM})_j = +1$ as well because its decision on the *negated* version of issue j is the opposite of its former decision. Moreover, the fraction of weight on $+1$ in that column is still 0.5. So all issue majorities are unchanged, and hence the average majority is also unchanged. Therefore, assume $p_{IWM} = \mathbf{1}$.

We first define a variable W_{ones} counting the total weight placed on $+1$ in a preference profile, and show that if $W_{\text{ones}} \geq \frac{3n}{4}$ then $\mathbf{1}$ is weakly majority-supported (hence Anscombe's paradox does not occur). Then we will show that $W_{\text{ones}} = n \cdot \tilde{m}$.

Let $W_{\text{ones}} := \sum_{i=1}^{n} 1 - d_H(v_i, \mathbf{1}, w_i)$. We claim that if

$$W_{\text{ones}} \geq \left(\left\lfloor \frac{n}{2} \right\rfloor + 1\right)\left(\frac{1}{2}\right) + \left\lfloor \frac{n-1}{2} \right\rfloor$$

then $\mathbf{1}$ is weakly majority-supported. Assume for the sake of contradiction that this is not the case. For any voter i that opposes $\mathbf{1}$, we have that $d_H(v_i, \mathbf{1}, w_i) > 1/2$. As $\mathbf{1}$ is not even weakly majority-supported, we know that more than half of the voters (at least $\lfloor n/2 \rfloor + 1$) oppose $\mathbf{1}$. The remaining at most $n - \lfloor n/2 \rfloor - 1 = \lfloor (n-1)/2 \rfloor$ voters still must have a non-negative distance from $\mathbf{1}$. Hence we can upper bound W_{ones}:

$$\begin{aligned}
W_{ones} &= \sum_{\substack{i \in [n] \\ i \text{ opposes } \mathbf{1}}} 1 - d_H(v_i, \mathbf{1}, w_i) + \sum_{\substack{i \in [n] \\ i \text{ supports } \mathbf{1}}} 1 - d_H(v_i, \mathbf{1}, w_i) \\
&< \sum_{\substack{i \in [n] \\ i \text{ opposes } \mathbf{1}}} 1/2 + \sum_{\substack{i \in [n] \\ i \text{ supports } \mathbf{1}}} 1 \\
&\leq \left(\left\lfloor \frac{n}{2} \right\rfloor + 1\right)\left(\frac{1}{2}\right) + \left\lfloor \frac{n-1}{2} \right\rfloor
\end{aligned}$$

Hence, we have that $W_{\text{ones}} < \left(\left\lfloor \frac{n}{2} \right\rfloor + 1\right)\left(\frac{1}{2}\right) + \left\lfloor \frac{n-1}{2} \right\rfloor$, a contradiction. We now upper bound the RHS as follows:

$$\left(\left\lfloor \frac{n}{2} \right\rfloor + 1\right)\left(\frac{1}{2}\right) + \left\lfloor \frac{n-1}{2} \right\rfloor \leq \left(\frac{n}{2} + 1\right)\frac{1}{2} + \left(\frac{n-1}{2}\right) = \frac{3n}{4}$$

Hence, if $W_{\text{ones}} \geq \frac{3n}{4}$ then our previous condition is satisfied, and the issue-wise majority is non-losing.

Now we'll show the claimed relationship between $\tilde{m}$ and W_{ones}:

$$\begin{aligned}
n \cdot \tilde{m} &= n \sum_{j=1}^{t} \tilde{w}_j m_j = n \sum_{j=1}^{t} \tilde{w}_j \left(\frac{1}{n\tilde{w}_j} \sum_{i=1}^{n} w_{i,j} \cdot \mathbb{I}(v_{i,j} = +1)\right) \\
&= \sum_{j=1}^{t} \sum_{i=1}^{n} w_{i,j} \cdot \mathbb{I}(v_{i,j} = +1) = W_{\text{ones}}
\end{aligned}$$

Therefore we have that $W_{\text{ones}} \geq \frac{3n}{4} \iff \tilde{m} \geq \frac{3}{4}$.

To prove the second claim of the theorem, we fix a new instance $\mathcal{I}' = (\mathcal{P}, w)$ in the external weights model such that $m_j \geq 3/4$ for all $j \in [t]$. Therefore, $\mathbf{1}$ is the unique IWM. Assume for sake of contradiction that Ostrogorski's paradox occurs, so there is some proposal $p \neq \mathbf{1}$ such that $p \succ_{\mathcal{I}'} \mathbf{1}$. Then we know by Lemma 14 that there exists a sub-instance $\mathcal{I}''$ in which Anscombe's paradox occurs, where $\mathcal{I}''$ is obtained by restricting $\mathcal{I}'$ to some subset of issues $T \subseteq [t]$ and renormalizing the external weight vector. Note that the majorities on the topics in T are unchanged from the original profile. Hence, the average majority in $\mathcal{I}'$ is at least $3/4$, because each individual issue majority is at least $3/4$. This is a contradiction to the claim proven above. Therefore, Ostragorski's paradox does not occur in $\mathcal{I}'$. $\square$