Complexity Results and Algorithms for Manipulation and Bribery in Judgment Aggregation

Ari Conati[1], Andreas Niskanen[1], Ronald de Haan[2] and Matti Järvisalo[1,*]

[1]University of Helsinki, Finland
[2]University of Amsterdam, the Netherlands
ORCID (Ari Conati): https://orcid.org/0000-0003-3023-3991, ORCID (Andreas Niskanen): https://orcid.org/0000-0003-3197-2075, ORCID (Ronald de Haan): https://orcid.org/0000-0003-2023-0586, ORCID (Matti Järvisalo): https://orcid.org/0000-0003-2572-063X

Abstract. The study of limits of strategic behavior in collective decision making is a central topic in computational social choice. Focusing on judgment aggregation, we provide complexity results and algorithms for manipulation and bribery under various aggregation rules. Specifically, we show that manipulation and bribery are complete for the second level of the Polynomial Hierarchy and detail aggregation-rule-specific strong refinements for effective counterexample-guided abstraction refinement algorithms based on iterative calls to a maximum satisfiability solver for both manipulation and bribery. We provide an open-source implementation of the approach and empirically evaluate its performance on standard PrefLib datasets, showing that the strong refinement strategies developed in this work enable scaling up to solving more instances.

1 Introduction

A central topic in the AI research area of computational social choice [5, 30] is the study of limits of strategic behavior in collective decision making, see, e.g., [3, 10, 19, 16, 17]. Strategic behavior is often considered undesirable; for example, the ability to manipulate the outcome of a voting process is generally harmful. It is important to understand how difficult it is both in theory and practice to influence the results of social choice procedures by strategic behavior.

In this work, we study computational aspects of strategic behavior in the context of judgment aggregation (JA) [23, 18, 12]. Judgment aggregation offers a generic and well-established formal logical framework for modeling various settings where agents must reach joint agreements through aggregating the preferences, judgments, or beliefs of individual agents by social choice mechanisms. In terms of forms of strategic behavior, we focus on two central notions: *manipulation* [14, 4, 11, 6], i.e., the task of determining whether an individual can enforce their preferred group judgment by expresssing an insincere individual judgment, and *bribery* [4, 11, 6], where the task is to determine whether it is possible for an external party to enforce their preferred group judgment by bribing several individuals involved in the group decision process to express insincere individual judgments. In terms of judgment aggregation procedures (JA rules), we cover a wide selection of central rules, specifically, Kemeny [28, 26, 21, 27, 14, 13], Slater [26, 21, 27, 13], MaxHamming [21], Young [21], and Dodgson [26].

* Corresponding Author. Email: matti.jarvisalo@helsinki.fi

Until now, manipulation and bribery have been shown to be Σ_2^{p}-complete only under the Kemeny rule [11]. We considerably extend this result by establishing Σ_2^{p}-completeness for all of the considered JA rules. Thus, the problems of manipulation and bribery in JA both face a significant complexity barrier. Complementing the complexity results, we extend a recently outlined algorithmic approach for deciding manipulation and bribery [9], first outlined for the Kemeny rule, to cover the four other rules as well. The algorithmic approach is based on the general approach of counterexample-guided abstraction refinement (CEGAR) [7, 8], and makes iterative use of a maximum satisfiability (MaxSAT) solver [1]. A key aspect of any CEGAR approach is the strength of *refinements* applied at each iteration of the CEGAR loop. The purpose of a refinement strategy is to prune out from further consideration the most recent solution candidate that turns out to be a non-solution by additional constraints. The most basic form of refinement is a constraint that rules out only the single most recent solution candidate found. This basic refinement was proposed in [9]. However, by using such a basic refinement, the CEGAR approach effectively degenerates into an approach which naively enumerates potentially exponentially many solution candidates. Hence, to make the CEGAR approach to manipulation and bribery more effective in practice, *stronger* refinement strategies are needed. The stronger the refinement steps, i.e., the more solution candidates can be ruled out from subsequent search based on a single counterexample at each iteration, the fewer iterations can be expected to be needed for termination. In order to make the CEGAR approach to manipulation and bribery more efficient, we derive nontrivial aggregation-rule-specific strong refinements for each of the JA rules. Furthermore, the CEGAR approach originally outlined for the Kemeny rule was not previously implemented [9]. We provide an open-source implementation of the approach, extended to cover all of the considered JA rules as detailed in this work together with our strong refinements. We empirically evaluate the runtime performance of the implementation on standard PrefLib datasets, showing the benefits of the strong refinements.

2 Preliminaries

We recall judgment aggregation and aggregation rules [22, 15].

Judgment Aggregation Consider a set X of propositional variables (*issues*), and let $\neg X = \{\neg x \mid x \in X\}$. The set of literals

$\Phi = X \cup \neg X$ is the *agenda*. A *judgment set* $J \subseteq \Phi$ represents an individual opinion on the agenda. The judgment set J is *complete* if for all $x \in X$ either $x \in J$ or $\neg x \in J$, and Γ-*consistent* with respect to a propositional formula Γ if $\Gamma \wedge \bigwedge_{l \in J} l$ is satisfiable. Let $\mathcal{J}(\Phi, \Gamma)$ be the collection of all complete and Γ-consistent judgment sets.

Definition 1. *A judgment aggregation framework consists of a set* $I = \{1, \ldots, n\}$ *of* agents*, an* agenda Φ *over the set of issues* X*, a propositional formula* Γ *(the* integrity constraint *that may contain variables outside* X*), and a* profile $P = (J_i)_{i \in I}$ *of complete and* Γ*-consistent judgment sets* $J_i \in \mathcal{J}(\Phi, \Gamma)$ *representing the opinions of each agent* $i \in I$*.*

For a literal $l \in \Phi$, we denote by $N(P, l) = |\{i \in I \mid l \in J_i\}|$ the number of agents which support agenda item l. Similarly, we denote by $\Delta(P, l) = N(P, l) - N(P, \neg l)$ the difference in support of l and $\neg l$. The *majoritarian judgment set* $\mathcal{M}(P) = \{l \in \Phi \mid \Delta(P, l) > 0\}$ consists of agenda items supported by the majority of agents.

Connection to Other Variants of Judgment Aggregation In terms of the general applicability of our results, we note that various variations of the above definitions have been considered—see, e.g., [15] for an overview. As explained e.g. in [9], the "literal-based" framework definition we use is as expressive as "formula-based" frameworks in which propositional formulas are allowed as issues. In particular, a formula-based framework corresponds to a literal-based framework by (i) including for each formula φ an issue x_φ as a literal, and (ii) adding the constraint $(x_\varphi \leftrightarrow \varphi)$ to the integrity constraint. This is in fact a standard trick in SAT for taking a name (here issue x_φ) to represent a formula (here φ). As identified by Endriss et al. [15], all frameworks for judgment aggregation that have been considered can be divided into four classes depending on their expressivity (modulo polynomial-time translations). These four classes are based on whether or not the framework allows for: (i) additional variables that are not in one-to-one correspondence with the issues in the agenda, and (ii) separate input and output constraints. While we do not explicitly consider input and output constraints separately, our results straightforwardly extend to this case. Specifically, by employing SAT-based solvers, our algorithmic approach allows for enforcing integrity constraints seamlessly by expressing them in propositional logic. Our hardness results are presented for the case where additional variables are allowed; this is also the sake of representation, and the proofs can be straightforwardly adjusted to hold also for the case that does not allow for additional variables.

Aggregation Rules A *judgment aggregation rule* R maps each profile P to a collection of *collective judgment sets* $R(P) \subseteq \mathcal{J}(\Phi, \Gamma)$. We assume that the agenda Φ and the integrity constraint Γ are fixed. In general, most judgment aggregation rules aim to preserve Γ-consistency of collective judgment sets while remaining close to the majoritarian judgment set.

Slater maximizes the agreement with the majoritarian judgment set in terms of the number of agenda items. Formally, $\text{SLATER}(P) = \arg\max_{J \in \mathcal{J}(\Phi,\Gamma)} |J \cap \mathcal{M}(P)|$.

The next rules are based on minimizing a cost function defined using the Hamming distance between complete (and Γ-consistent) judgment sets J and J', denoted as $d(J, J') = |J \setminus J'| = |J' \setminus J|$.

Kemeny maximizes the agreement with the whole profile P, hence minimizing the sum of the Hamming distances to the judgment sets in P. Formally, $\text{KEMENY}(P) = \arg\min_{J \in \mathcal{J}(\Phi,\Gamma)} \sum_{i \in I} d(J, J_i)$.

MaxHamming minimizes the maximum Hamming distance to judgment sets in P, that is, $\text{MAXH}(P) = \arg\min_{J \in \mathcal{J}(\Phi,\Gamma)} \max_{i \in I} d(J, J_i)$.

We also consider rules based on modifying the input profile P in a minimum way so as to obtain a Γ-consistent majoritarian judgment set. Let $\mathcal{P}(\Phi, \Gamma)$ be the set of all profiles over Φ consisting of complete and Γ-consistent judgment sets. For a profile $P \in \mathcal{P}(\Phi, \Gamma)$ and a subset $I' \subseteq I$ of agents, we notate $P[I'] = (J_i)_{i \in I'}$.

Young selects those complete and Γ-consistent judgment sets which are obtained as supersets of majoritarian judgment sets of profiles from which the least possible number of judgment sets in P are removed. That is, $\text{YOUNG}(P)$ considers profiles $P^{\text{Y}} \in \mathcal{P}(\Phi, \Gamma)$ with $P^{\text{Y}} = P[I^{\text{Y}}]$ for some $I^{\text{Y}} \subseteq I$ for which $\mathcal{M}(P^{\text{Y}})$ is Γ-consistent, maximizing $|P^{\text{Y}}|$, and selects all $J \in \mathcal{J}(\Phi, \Gamma)$ with $J \supseteq \mathcal{M}(P^{\text{Y}})$.

Dodgson selects those complete and Γ-consistent judgment sets which are obtained as supersets of majoritarian judgment sets of profiles in which the least possible number of opinions are reverted. That is, $\text{DODGSON}(P)$ considers profiles $P^{\text{D}} \in \mathcal{P}(\Phi, \Gamma)$ with $P^{\text{D}} = (J_i^{\text{D}})_{i \in I}$ for which $\mathcal{M}(P^{\text{D}})$ is Γ-consistent, minimizing $\sum_{i \in I} d(J_i, J_i^{\text{D}})$, and selects all $J \in \mathcal{J}(\Phi, \Gamma)$ for which $J \supseteq \mathcal{M}(P^{\text{D}})$.

Manipulation and Bribery We recall manipulation and bribery as earlier defined for the Kemeny rule [11]. Let R be a judgment aggregation rule. Consider a judgment aggregation framework with a set of agents I, an agenda Φ, an integrity constraint Γ, and a profile $P = (J_1, \ldots, J_n) \in \mathcal{P}(\Phi, \Gamma)$, as well as an *outcome* $L \subseteq \Phi$ as input. The task in manipulation and bribery is to construct a modified profile $P_{\text{new}} \in \mathcal{P}(\Phi, \Gamma)$ so that the outcome L is guaranteed to be achieved in any collective judgment set $J \in R(P_{\text{new}})$. The problems differ in which kind of modifications are allowed. We consider bribery with corrupt judges.

In *manipulation*, an agent $m \in I$ can influence the collective judgment sets by specifying any judgment set J' (instead of J_m). A modified profile P_{new} is of the form $(J_1, \ldots, J_{m-1}, J', J_{m+1}, \ldots, J_n)$ with $J' \in \mathcal{J}(\Phi, \Gamma)$. In *bribery*, an external agent can influence the judgment sets of at most k agents among a given set $C \subseteq I$ of corrupt agents. A modified profile P_{new} differs from P by at most k judgment sets J_i with $i \in C$. The task is to find a modified profile P_{new} such that $L \subseteq J_{\text{new}}$ for all $J_{\text{new}} \in R(P_{\text{new}})$. The decision variant of the problem asks whether such a profile exists.

SAT and MaxSAT For a Boolean variable x there are two literals, x and $\neg x$. A clause C is a disjunction ($\vee$) of literals. A CNF formula F is a conjunction ($\wedge$) of clauses. We denote by $V(F)$ variables of F. An assignment $\tau \colon V(F) \to \{0, 1\}$ maps variables to 0 (false) or 1 (true), and extends to literals via $\tau(\neg x) = 1 - \tau(x)$, to clauses via $\tau(C) = \max\{\tau(l) \mid l \in C\}$, and to formulas via $\tau(F) = \min\{\tau(C) \mid C \in F\}$. The *Boolean satisfiability* problem (SAT) asks whether a given formula F is satisfiable (i.e., is there an assignment τ with $\tau(F) = 1$); if not, F is unsatisfiable. In the *(weighted partial) maximum satisfiability* problem (MaxSAT for short) [1] input consists of "hard" clauses F_{hard}, "soft" clauses F_{soft}, and a weight function w mapping each soft clause $C \in F_{\text{soft}}$ to an integer $w(C) \geq 0$. The task is to find an assignment τ which satisfies F_{hard} and minimizes the cost $c(\tau) = \sum_{C \in F_{\text{soft}}} w(C)(1 - \tau(C))$ incurred by not satisfying soft clauses.

3 Complexity Results

We start with our complexity results.

Theorem 1. *For each judgment aggregation rule* $R \in \{\text{KEMENY}, \text{SLATER}, \text{MAXH}, \text{YOUNG}, \text{DODGSON}\}$*, the problems of manipulation and bribery are* Σ_2^p*-complete. Hardness holds even when outcome* L *is restricted to* $|L| = 1$*.*

Figure 1: The profile P in the proof sketch for Theorem 1.

P	J_1	J_2	J_3	J_4	$\cdots$	J_{2n}	J_{2n+1}	$\cdots$	$\cdots$	J_{6n}
z_1	0	0	1	1	$\cdots$	1	1	1 0 0 0 0 0 0 0	$\cdots$	
z_1'	0	0	1	1	$\cdots$	1	0	0 1 1 0 0 0 0 0	$\cdots$	
z_2	0	0	1	1	$\cdots$	1	0	0 0 0 0 1 1 0 0	$\cdots$	
z_2'	0	0	1	1	$\cdots$	1	0	0 0 0 0 0 0 1 1	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$		$\ddots$	
w	1	0	1	1	$\cdots$	1	0	1 0 1 0 0 0 0 0	$\cdots$	

For intuition, we provide three proof sketches: one as a generic Σ_2^{p}-membership argument, one for Σ_2^{p}-hardness under YOUNG, and one for Σ_2^{p}-hardness under MAXH. Full proofs are available in an online appendix [31].

Proof: Σ_2^p-membership. We describe a polynomial-time nondeterministic algorithm with access to an NP oracle that decides the problem. The algorithm first guesses a modified profile P_{new} and checks that it adheres to the constraints for the respective problem. Checking that P_{new} is in $\mathcal{J}(\Phi, \Gamma)$ can be done using the NP oracle. For each judgment aggregation rule R that we consider, we know that the problem of deciding if there is some collective judgment set $J \in R(P_{\text{new}})$ such that $L \not\subseteq J$ is in the complexity class Θ_2^{p} [15]. Therefore, in (deterministic) polynomial time, using the NP oracle, we can check whether there is some collective judgment set $J \in R(P_{\text{new}})$ such that $L \not\subseteq J$—or conversely, whether $L \subseteq J$ for all $J \in R(P_{\text{new}})$. □

For KEMENY, Σ_2^{p}-completeness has been shown previously [11]. However, the reduction with slightly modified arguments provides hardness also for SLATER and DODGSON. We will next sketch a proof for hardness of manipulation for YOUNG and for MAXH.

Proof (sketch): Σ_2^p-hardness for YOUNG. We give a reduction from $\exists\forall$-QBF-SAT. Let $\varphi = \exists Z \forall Y \psi$ be a QBF with $Z = \{z_1, \ldots, z_n\}$. We construct an instance of the manipulation problem for YOUNG as follows. As issues, we consider the set $X = Z \cup \{\, z' \mid z \in Z \,\} \cup \{w\}$, where w and each x' is a fresh variable. We define the integrity constraint as $\Gamma = \bigwedge_{z \in Z}(z \oplus z')(\psi \to w)$. We let $L = \{w\}$, and we let $m = 3$. Finally, we let the profile P be as in Figure 1. (The judgment sets in the constructed profile P are not consistent with Γ. In the full proof we ensure that the profile is in fact in $\mathcal{P}(\Phi, \Gamma)$.)

The main idea behind why this reduction works is the following. The majority outcome $\mathcal{M}(P)$ of P is not consistent with the constraint Γ. To get a strict majority outcome that is consistent by deleting as few as possible judgment sets from P is to delete exactly $2n$ sets among $J_{2n+1}, \ldots, J_{6n}$—and to do this in a way that for each $z_i \in Z$ either two judgment sets are deleted that both contain z_i or two judgment sets are deleted that both contain z_i'—and deleting one of J_1, J_2. The possible ways of achieving consistency with Γ by deleting a minimum number of voters are in correspondence with all possible truth assignments to the variables in Z. Moreover, whether the choice of J_1, J_2 yields a consistent majority outcome depends on the choice of sets among $J_{2n+1}, \ldots, J_{6n}$: if that choice corresponds to a truth assignment α such that $\forall Y \psi[\alpha]$ is true, then only deleting J_2 leads to consistency (and the collective outcome includes w), and otherwise either choice of J_1 or J_2 leads to consistency (and the collective outcome may or may not include w).

The manipulator can then change J_3 to a judgment set that in the same way corresponds to a truth assignment α to the variables in Z. After doing this, the ways of deleting a minimum number of judgment sets to yield a consistent majority outcome are restricted to those corresponding to the truth assignment α. Thus, the manipulator can only enforce that $L = \{w\} \subseteq J^*$ for all $J^* \in \text{YOUNG}(P_{\text{new}})$ by changing J_3 to a judgment set that corresponds to a truth assignment α such that $\forall Y \psi[\alpha]$ is true. This is possible if and only if the original QBF is true. □

Proof (sketch): Σ_2^p-hardness for MAXH. We give a reduction from $\exists\forall$-QBF-SAT. Let $\varphi = \exists Z \forall Y \psi$ be a QBF with $Z = \{z_1, \ldots, z_n\}$. We construct an instance of the manipulation problem for MAXH as follows. As issues, we consider the set $X = Z \cup \{\, v_i, v_i', z_i' \mid z_i \in Z \,\} \cup \{w\}$, where w and each z_i', v_i, v_i' is a fresh variable. We define the integrity constraint Γ as follows.

$$\Gamma = \begin{array}{l} \Big(\bigwedge_{z_i \in Z}(\neg z_i \wedge \neg z_i' \wedge \neg v_i \wedge \neg v_i')\Big) \vee \\ \Big(\bigwedge_{z_i \in Z}((z_i \oplus z_i') \wedge \neg v_i \wedge \neg v_i') \wedge (\psi \to w)\Big) \vee \\ \Big(\bigwedge_{z_i \in Z}((z_i \oplus z_i') \wedge (v_i \oplus v_i'))\Big) \end{array}$$

We let $L = \{w\}$ and $m = 1$. Finally, we let the profile P consist of two judgment sets $J_1 = J_2 = \{\, \neg\chi \mid \chi \in \Phi \,\}$.

The main idea of this reduction is the following. The manipulator can change their judgment set J_1 in two general ways, corresponding to the second and third disjunct of Γ. If they change their judgment set to something J' that satisfies the second conjunct, both (the original) J_1 and J' will be outcomes, as they achieve minimax Hamming distance to the profile, and as $w \notin J_1$, this will not achieve the manipulator's goal. If the manipulator changes their judgment set to something J'' that satisfies the third disjunct, the collective outcome(s) that achieve minimax Hamming distance to the profile are judgment sets J^* that agree with J'' on all issues z_i, z_i' and that set all issues v_i, v_i' to false—these have Hamming distance n or $n+1$ to J_2 and to J'', whereas all other judgment sets have Hamming distance at least $n+2$ to some set in the profile. By choosing a set J'' that corresponds to a truth assignment α for Z such that $\forall Y \psi[\alpha]$ is true, the manipulator can enforce that only the judgment set J^* that includes w is the collective outcome. In fact, this is the case if and only if J'' corresponds to a truth assignment α for Z such that $\forall Y \psi[\alpha]$ is true. □

For each rule R, Σ_2^{p}-hardness for manipulation straightforwardly carries over to the problem of bribery via reducing manipulation to bribery by setting $C = \{J_m\}$ and letting $k = 1$.

4 CEGAR for Manipulation and Bribery

Turning to algorithms, we extend a recently outlined approach to manipulation and bribery in the form of a MaxSAT-based CEGAR algorithm [9]. The approach was originally outlined for the specific case of the Kemeny rule, using a very basic refinement strategy for ruling solution candidates out of further consideration one-by-one. Here we extend the approach to cover all five considered JA rules. This requires detailing rule-specific abstraction formulas and counterexample checks; we detail these extensions in Section 5. Furthermore, we improve on the basic refinements by deriving stronger, rule-specific refinement strategies for the considered JA rules. The stronger refinements (Section 6) have the potential of significantly reducing the number of iterations needed for terminating the CEGAR approach.

We first describe the CEGAR algorithm for manipulation and bribery as Algorithm 1, following [9] but with the JA rule given as input for generality. An *abstraction* is initialized as a MaxSAT instance $(F_{\text{hard}}, F_{\text{soft}}, w)$ (line 1) whose optimal solutions each correspond to a modified profile P_{new} and a collective judgment set $J \in R(P_{\text{new}})$. Issues X correspond directly to variables in the MaxSAT instance and

Algorithm 1 MaxSAT-based CEGAR for manipulation (M) and bribery (B). **Input**: Problem variant $S \in \{M, B\}$, judgment aggregation rule R, agenda Φ, integrity constraint Γ, profile $P \in \mathcal{P}(\Phi, \Gamma)$, outcome $L \subseteq \Phi$.

1: $(F_{\text{hard}}, F_{\text{soft}}, w) \leftarrow \textsc{Abstraction}_{S,R}(\Phi, \Gamma, P)$
2: **while true do**
3: $\quad(c^*_{\text{abs}}, \tau_{\text{abs}}) \leftarrow \textsc{MaxSAT}(F_{\text{hard}} \wedge \bigwedge_{l \in L} l, F_{\text{soft}}, w)$
4: $\quad$**if** $c^*_{\text{abs}} = +\infty$ **then return false**
5: $\quad P_{\text{new}} \leftarrow \textsc{Profile}(\tau_{\text{abs}})$, $J^*_{\text{abs}} \leftarrow \textsc{Judgment}(\tau_{\text{abs}})$
6: $\quad F_{\text{cex}} \leftarrow F_{\text{hard}} \wedge \textsc{Fix}_S(P_{\text{new}}) \wedge \bigvee_{l \in L} \neg l$
7: $\quad(c^*_{\text{cex}}, \tau_{\text{cex}}) \leftarrow \textsc{MaxSAT}(F_{\text{cex}}, F_{\text{soft}}, w)$
8: $\quad$**if** $c^*_{\text{cex}} > c^*_{\text{abs}}$ **then return** $P_{\text{new}}, J^*_{\text{abs}}$
9: $\quad F_{\text{hard}} \leftarrow F_{\text{hard}} \wedge \textsc{Refine}_{S,R}(c^*_{\text{abs}}, \tau_{\text{cex}})$

therefore any constraints on J, including Γ, are directly expressed as hard clauses over X. However, additional constraints on X do not preserve the collective judgment sets in general.

We iteratively solve this MaxSAT instance with the additional requirement that the outcome L is in a collective judgment set J (line 3). If there is no solution, we return false, since then there is no modified profile with a collective judgment set including the outcome (line 4). Otherwise, we extract a candidate modified profile P_{new} and a candidate collective judgment set J^*_{abs} from the optimal solution τ_{abs} (line 5). Since we enforced the outcome L as hard clauses, $J^*_{\text{abs}} \in R(P_{\text{new}})$ does not necesssarily hold. Also, we need to check whether *every* judgment set in $R(P_{\text{new}})$ contains L. Both of these checks can be done via a single additional MaxSAT call. We fix the candidate profile P_{new} as unit clauses $\textsc{Fix}_S(P_{\text{new}})$, add a constraint enforcing that L is *not* included in a collective judgment set J, and solve the resulting MaxSAT instance (lines 6–7). If the cost c^*_{cex} of this instance exceeds the cost c^*_{abs} of the abstraction, every judgment set in P_{new} includes L, so we return P_{new} (line 8). Otherwise, the candidate profile P_{new} and its collective judgment set $J^*_{\text{cex}} \in R(P_{\text{new}})$ corresponding to τ_{cex} is a *counterexample*, since $J^*_{\text{cex}} \not\supseteq L$.[1] We thus *refine* the abstraction, i.e., add constraints based on c^*_{abs} and τ_{cex} which rule out one or more modified profiles in which the outcome is guaranteed not to be achieved (line 9). For correctness it suffices to rule out the counterexample modified profile.

Proposition 1. *Let* $\textsc{Abstraction}_{S,R}(\Phi, \Gamma, P)$ *be a MaxSAT instance whose optimal solutions* τ *map to a modified profile* $\textsc{Profile}(\tau)$ *and a collective judgment set* $\textsc{Judgment}(\tau) \in R(\textsc{Profile}(\tau))$. *If* $\textsc{Refine}_{S,R}(c^*_{abs}, \tau_{cex})$ *excludes exactly the counterexample profile* $\textsc{Profile}(\tau_{cex})$*, Algorithm 1 terminates, and returns a modified profile* P_{new} *if and only if for all* $J \in R(P_{new})$ *it holds that* $L \subseteq J$.

However, as we will detail, for each problem variant and judgment aggregation rule, it is possible to derive strong refinements that rule out additional modified profiles where the outcome is not included in every collective judgment set and still maintain correctness.

5 Abstractions and Counterexample Checks

Next, we detail MaxSAT encodings for the abstraction $\textsc{Abstraction}_{S,R}(\Phi, \Gamma, P)$, encoding profile modifications specific to the problem variant S and collective judgment sets specific to the judgment aggregation rule R. The encodings build on the MaxSAT encodings for outcome determination [9], producing MaxSAT instances whose optimal solutions are in a one-to-one correspondence with collective judgment sets. Specifically, in manipulation and bribery collective judgment sets are computed from a modified profile—which is encoded in our MaxSAT instance—instead of a fixed profile provided as input as in outcome determination. We hence extend the MaxSAT encodings from [9] to take the (nondeterministic) profile modifications into account. The extension for the case of Kemeny was described in [9]. Here we provide analogous encodings for the Slater, MaxHamming, Young, and Dodgson rules. In each of the encodings, issues X are directly included as MaxSAT variables to represent a collective judgment set under R, and Γ is included as hard clauses to ensure Γ-consistency.

5.1 Manipulation

In manipulation, an agent $m \in I$ can specify any judgment set $J' \in \mathcal{J}(\Phi, \Gamma)$. We represent this judgment set using Boolean variables m_x for each $x \in X$. We enforce Γ-consistency of J' with the hard clauses $\Gamma[x \mapsto m_x \mid x \in X]$. Given an optimal solution τ, a modified profile $P_{\text{new}} = (J_1, \ldots, J_{m-1}, J', J_{m+1}, \ldots, J_n)$ is obtained by setting $J' = \{x \in \Phi \mid \tau(m_x) = 1\} \cup \{\neg x \in \Phi \mid \tau(m_x) = 0\}$. For the counterexample check, this profile is fixed via unit clauses $\textsc{Fix}_M(P_{\text{new}}) = \bigwedge_{x \in J'} m_x \wedge \bigwedge_{\neg x \in J'} \neg m_x$. We continue by detailing rule-specific encodings of the abstraction. For the following, we let $P_{-m} = P[I \setminus \{m\}]$ be the profile consisting of judgment sets without the manipulator.

Kemeny. We first recall the MaxSAT encoding for the abstraction for the case of Kemeny as described in [9]. Collective judgment sets minimize the sum of Hamming distances to the individual judgment sets. Disregarding the manipulator, the objective function is represented by including for each agent $i \in I \setminus \{m\}$ a soft clause (l) with unit weight for each $l \in J_i$. Equivalently, for each $l \in \Phi$, we include a soft clause (l) with weight $N(P_{-m}, l)$. To account for the manipulator, we include unit-weight soft clauses $(m_x \rightarrow x)$ and $(\neg m_x \rightarrow \neg x)$ for each $x \in X$ representing the cost incurred due to the judgment set of the manipulator.

We now detail analogous MaxSAT encodings for the abstraction under the Slater, MaxHamming, Young, and Dodgson rules.

Slater. Collective judgment sets maximize the agreement to the majoritarian judgment set of the modified profile. The majoritarian judgment set cannot be influenced by the manipulator on issues $x \in X$ where the support of x and $\neg x$ differs by more than 1. Thus, for $l \in \mathcal{M}(P_{-m})$ with $\Delta(P_{-m}, l) \geq 2$, we include a unit-weight soft clause (l). To account for the judgment set of the manipulator, we define $X^{=} = \{x \in X \mid \Delta(P_{-m}, x) = 0\}$, $X^{+} = \{x \in X \mid \Delta(P_{-m}, x) = +1\}$, and $X^{-} = \{x \in X \mid \Delta(P_{-m}, x) = -1\}$. For issues in $x \in X^{=}$, the manipulator decides single-handedly whether x or $\neg x$ is included in the majoritarian judgment set. This is encoded with unit-weight soft clauses $(m_x \rightarrow x)$ and $(\neg m_x \rightarrow \neg x)$. For issues $x \in X^{+}$, the manipulator can remove x from $\mathcal{M}(P_{-m})$, encoded via the unit-weight soft clause $(m_x \rightarrow x)$. Symmetrically, for $x \in X^{-}$, the manipulator can remove $\neg x$ from $\mathcal{M}(P_{-m})$, encoded as the unit-weight soft clause $(\neg m_x \rightarrow \neg x)$.

MaxHamming. We use additional variables a_x for each $x \in X$, with $a_x = 1$ iff the manipulator agrees with the collective judgment set, encoded as $\bigwedge_{x \in X}(a_x \leftrightarrow (x \leftrightarrow m_x))$. We also also use variables p_k for each $k = 1, \ldots, |X|$, where $p_k = 1$ if the Hamming distance of the collective judgment set to at least one judgment set—either J_i for $i \in I \setminus \{m\}$ or J'—of the modified profile is at least k. This is expressed for each $k = 1, \ldots, |X|$

[1] If $c^*_{\text{cex}} = c^*_{\text{abs}}$, then $L \subseteq J^*_{\text{abs}} \in R(P_{\text{new}})$ and $J^*_{\text{cex}} \in R(P_{\text{new}})$. But we are to find P_{new} s.t. all collective judgment sets include L.

as $\left(\bigvee_{i\in I\setminus\{m\}}\left(\sum_{l\in J_i}\neg l\geq k\right)\vee\left(\sum_{x\in X}\neg a_x\geq k\right)\right)\rightarrow p_k$. The maximum Hamming distance is minimized via unit-weight soft clauses $(\neg p_k)$ for each $k=1,\ldots,|X|$.

Young. Additional variables y_i for $i\in I$ indicate which agents are included in the modified profile P^{Y} under the Young rule. The unit-weight soft clauses (y_i) for $i\in I$ then maximize the number of included agents. What remains is to ensure that variables $x\in X$ are set according to the strict majority of the modified profile. To account for the presence of the manipulator, we use variables v_x for each $x\in X$, set to 1 iff the manipulator is in the modified profile and supports issue x, encoded via $\bigwedge_{x\in X}(v_x\leftrightarrow(y_m\wedge m_x))$. For a fixed number of Young agents k and for each $x\in X$ and $k=1,\ldots,n$, the constraint

$$\left(\sum_{i\in I}y_i\leq k\wedge\sum_{\substack{i\in I\setminus\{m\}\\ x\in J_i}}y_i+v_x\geq\lfloor k/2\rfloor+1\right)\rightarrow x$$

forces $x=1$ if a strict majority supports x. Symmetrically,

$$\left(\sum_{i\in I}y_i\geq k\wedge\sum_{\substack{i\in I\setminus\{m\}\\ x\in J_i}}y_i+v_x\leq\lceil k/2\rceil-1\right)\rightarrow\neg x$$

sets x to 0 if a strict majority supports $\neg x$.

Dodgson. Additional variables $d_{i,x}$ for each $i\in I$ and $x\in X$ express the Dodgson-modified profile P^{D}, with $d_{i,x}=1$ iff $x\in J_i^{\text{D}}$. To ensure Γ-consistency of each J_i^{D}, we use the hard clauses $\Gamma[x\mapsto d_{i,x}\mid x\in X]$ for each $i\in I$. The constraints $\left(\sum_{i\in I}d_{i,x}\geq\lfloor n/2\rfloor+1\rightarrow x\right)\wedge\left(\sum_{i\in I}d_{i,x}\leq\lceil n/2\rceil-1\rightarrow\neg x\right)$ ensure that x (resp. $\neg x$) is included in the collective judgment set if it is supported by the strict majority of P^{D}. For each $i\in I\setminus\{m\}$ and $x\in X$, either $(d_{i,x})$ or $(\neg d_{i,x})$ is included as a unit-weight soft clause according to whether J_i supports or rejects issue x. To account for the manipulator, for each $x\in X$ we use unit-weight soft clauses $(m_x\rightarrow d_{m,x})$ and $(\neg m_x\rightarrow\neg d_{m,x})$.

Proposition 2. *For each $R\in\{$KEMENY, SLATER, MAXH, YOUNG, DODGSON$\}$, optimal solutions τ of the MaxSAT instance* $\text{ABSTRACTION}_{M,R}(\Phi,\Gamma,P)$ *correspond exactly to optimal modified profiles $P_{new}=(J_1,\ldots,J_{m-1},J',J_{m+1},\ldots,J_n)$ with $J'=\{x\in\Phi\mid\tau(m_x)=1\}\cup\{\neg x\in\Phi\mid\tau(m_x)=0\}$, and collective judgment sets $J\in R(P_{new})$ via $J=\tau\cap\Phi$.*

5.2 Bribery

In bribery, judgment sets of up to k agents from C can be modified. To represent these judgment sets, for each $i\in C$ and $x\in X$, we declare a Boolean variable $c_{i,x}$. We ensure Γ-consistency of these judgment sets via the hard clauses $\Gamma[x\mapsto c_{i,x}\mid x\in X]$ for each $i\in C$. To represent which agents are bribed, we declare a Boolean variable b_i for each $i\in C$. The bribery bound is encoded as the cardinality constraint $\sum_{i\in C}b_i\leq k$. Further, we enforce that if an agent is not bribed their judgment set is not changed via $\neg b_i\rightarrow\bigwedge_{x\in J_i}c_{i,x}\wedge\bigwedge_{\neg x\in J_i}\neg c_{i,x}$. Given an optimal solution τ, a modified profile P_{new} is obtained by replacing for each $i\in C$ the judgment set J_i with $J_i'=\{x\in\Phi\mid\tau(c_{i,x})=1\}\cup\{\neg x\in\Phi\mid\tau(c_{i,x})=0\}$. For the counterexample check, this profile is fixed using unit clauses $\text{FIX}_B(P_{\text{new}})=\bigwedge_{i\in C}\left(\bigwedge_{x\in J_i'}c_{i,x}\wedge\bigwedge_{\neg x\in J_i'}\neg c_{i,x}\right)$. For the following, we let $P_{-C}=P[I\setminus C]$ be the profile consisting of judgment sets of non-corrupt agents.

Kemeny. We again first recall the MaxSAT encoding for the abstraction for the case of Kemeny as described in [9]. Each term of the Kemeny objective for non-corrupt agents are encoded via a unit-weight soft clause (l) with weight $N(P_{-C},l)$ for each $l\in\Phi$. For each agent $i\in C$ and $x\in X$, we include unit-weight soft clauses $(c_{i,x}\rightarrow x)$ and $(\neg c_{i,x}\rightarrow\neg x)$ to represent the rest of the terms.

Our analogous MaxSAT encodings for the abstraction under the Slater, MaxHamming, Young, and Dodgson rules are as follows.

Slater. To represent agreement with the majoritarian judgment set of the modified profile, we need to calculate it within the encoding (similarly as for standard judgment aggregation under the Dodgson rule). The sum $S(P,x)=\sum_{i\in C}c_{i,x}+N(P_{-C},x)$ represents the support of issue x in the modified profile. For each $x\in X$, we include constraints $q_x\rightarrow((S(P,x)\geq\lfloor n/2\rfloor+1\rightarrow x)\wedge(S(P,x)\leq\lfloor n/2\rfloor-1\rightarrow\neg x))$ to enforce that if $q_x=1$, the collective judgment set includes x and $\neg x$ according to the majoritarian judgment set. Unit-weight soft clauses (q_x) for each $x\in X$ encode the Slater objective function.

MaxHamming. We use additional variables $a_{i,x}$ for each $i\in C$ and $x\in X$, with $a_{i,x}=1$ iff agent i agrees with the collective judgment set, encoded as $\bigwedge_{i\in C}\bigwedge_{x\in X}(a_{i,x}\leftrightarrow(x\leftrightarrow c_{i,x}))$. For $k=1,\ldots,|X|$, the formula $\left(\bigvee_{i\in I\setminus C}\left(\sum_{l\in J_i}\neg l\geq k\right)\vee\bigvee_{i\in C}\left(\sum_{x\in X}\neg a_{i,x}\geq k\right)\right)\rightarrow p_k$ forces $p_k=1$ if the Hamming distance of the collective judgment set to J_i for $i\in I\setminus C$ or J_i' for $i\in C$ is at least k. Distance is minimized via unit-weight soft clauses $(\neg p_k)$.

Young and Dodgson. For the Young rule, to account for corrupt agents $i\in C$, we declare variables $v_{i,x}$ for each issue x, and include hard clauses $v_{i,x}\leftrightarrow(y_i\wedge c_{i,x})$. The constraints are the same as for manipulation, with the exception that the summation $\sum_{i\in I\setminus C,x\in J_i}y_i+\sum_{i\in C}a_{i,x}$ represents the support of issue x. For the Dodgson rule, the constraints are identical to ones used for manipulation. We include unit-weight soft clauses $(d_{i,x})$ or $(\neg d_{i,x})$ depending on whether J_i for $i\in I\setminus C$ supports $x\in X$ or not. To cover corrupt agents, we include unit-weight soft clauses $(c_{i,x}\rightarrow d_{i,x})$ and $(\neg c_{i,x}\rightarrow\neg d_{i,x})$ for each $i\in C$ and $x\in X$.

Proposition 3. *For each $R\in\{$KEMENY, SLATER, MAXH, YOUNG, DODGSON$\}$, optimal solutions τ of the MaxSAT instance* $\text{ABSTRACTION}_{B,R}(\Phi,\Gamma,P)$ *correspond exactly to optimal modified profiles $P_{new}=(J_i')_{i\in I}$ with $J_i'=\{x\in\Phi\mid\tau(c_{i,x})=1\}\cup\{\neg x\in\Phi\mid\tau(c_{i,x})=0\}$ for $i\in C$ and $J_i'=J_i$ for $i\notin C$, and collective judgment sets $J\in R(P_{new})$ via $J=\tau\cap\Phi$.*

6 Refinement Strategies

We next detail strong refinement strategies for excluding several invalid candidate profiles based on counterexample checks. A simple and correct—yet inefficient—refinement, as proposed in [9] for the case of Kemeny, is to exclude exactly the modified profile P_{new} via adding the clause $\neg\text{FIX}_S(P_{\text{new}})$. However, by inspecting the counterexample $(c^*_{\text{cex}},\tau_{\text{cex}})$ we obtain stronger refinement constraints. At this point (line 9), we have a modified profile P_{new} and an optimal collective judgment set J^*_{abs} under the constraint $L\subseteq J^*_{\text{abs}}$ from the candidate solution $(c^*_{\text{abs}},\tau_{\text{abs}})$ to the abstraction. Assuming this modified profile via $\text{FIX}_S(P_{\text{new}})$, we negated the constraint $L\subseteq J^*_{\text{abs}}$, obtaining a solution $(c^*_{\text{cex}},\tau_{\text{cex}})$ with $c^*_{\text{cex}}\leq c^*_{\text{abs}}$. Now $J^*_{\text{cex}}=\tau_{\text{cex}}\cap\Phi$ is a counterexample judgment set, as $J^*_{\text{cex}}\in R(P_{\text{new}})$ and $L\not\subseteq J^*_{\text{cex}}$.

For Kemeny, Slater, and MaxHamming, our refinement strategies are based on comparing the cost of J^*_{cex} in an arbitrary candidate modified profile to the cost of the abstraction c^*_{abs}. Specifically, we can rule out additional modified profiles where the cost of J^*_{cex} is at most c^*_{abs}, since the optimality of J^*_{abs} guarantees that c^*_{abs} is a lower bound on the cost of any collective judgment set $J^*\supseteq L$ across all possible profiles P'_{new} allowed by the problem variant.

For Young, refinement constraints are based on the fact that in any candidate profile P'_{new}, removing the same agents to obtain a Young

modified profile incurs $c^*_{\text{cex}} \leq c^*_{\text{abs}}$ cost. If, in addition, the majoritarian judgment sets of the corresponding Young profiles remain the same, J^*_{cex} remains a counterexample in P'_{new}. For Dodgson, we reason about the number of modifications needed to obtain the same Dodgson modified profile from P'_{new}. As long as at most c^*_{abs} modifications need to be performed, J^*_{cex} remains a counterexample arising from the same modified profile. We now outline the specific refinement strategies for each of the problem variants covered in this work. For the following, assume that J^*_{cex} is a valid counterexample with cost $c^*_{\text{cex}} \leq c^*_{\text{abs}}$ (i.e., $J^*_{\text{cex}} \in R(P_{\text{new}})$).

6.1 Manipulation

In manipulation, a modified profile P_{new} is represented using the judgment set indicated by the manipulator J', encoded using variables m_x for each issue $x \in X$. In order to exclude further modified profiles where the cost c^*_{new} of J^*_{cex} is at most c^*_{abs}, for the Kemeny, Slater, and MaxHamming rules, our goal is to represent $c^*_{\text{new}} \not\leq c^*_{\text{abs}}$ using these variables.

Kemeny. The cost of J^*_{cex} in a modified profile is determined by the Hamming distances to judgment sets in P_{-m}, which remains constant with respect to the input and J^*_{cex}, and the Hamming distance to the judgment set J' indicated by the manipulator. To ensure that the new cost of J^*_{cex} is greater than c^*_{abs}, the manipulator must indicate a judgment set with sufficient distance to J^*_{cex}.

Slater. The cost of J^*_{cex} is determined by the disagreement with the majoritarian judgment set. In turn, the judgment set J' indicated by the manipulator only affects the cost of a counterexample for *swing issues* where the majoritarian judgment set is determined by J'. For a profile P', let $\Phi_{\text{swing}}(P') = \{l, \neg l \in \Phi \mid \Delta(P', l) = 0\} \cup \{l \in \Phi \mid \Delta(P', l) = 1\}$. The cost of the counterexample on fixed issues of P_{-m}, defined for a profile P' via $\Phi_{\text{fixed}}(P') = \{l \in \Phi \mid \Delta(P', l) \geq 2\}$, is constant. To ensure that the cost of J^*_{cex} exceeds c^*_{abs}, the manipulator must indicate a judgment set with sufficient disagreement to J^*_{cex} on issues in $\Phi_{\text{swing}}(P_{-m})$.

MaxHamming. The manipulator can dictate the cost of J^*_{cex} by indicating a judgment set which disagrees with J^*_{cex} on more issues than any J_i for $i \neq m$. Hence, the manipulator must disagree on enough issues to bring the cost of J^*_{cex} beyond c^*_{abs}.

Proposition 4. *Let $J'' \in \mathcal{J}(\Phi, \Gamma)$ be a candidate manipulator judgment set, and $P'_{new} = (J_1, \ldots, J_{m-1}, J'', J_{m+1}, \ldots, J_n)$ the corresponding modified profile. If*
for $R =$ KEMENY: $d(J^*_{cex}, J'') \leq c^*_{abs} - \sum_{i \in I \setminus \{m\}} d(J^*_{cex}, J_i)$;
for $R =$ SLATER:
$|(J'' \cap \Phi_{swing}(P_{-m})) \setminus J^*_{cex}| \leq c^*_{abs} - |\Phi_{fixed}(P_{-m}) \setminus J^*_{cex}|$:
for $R =$ MAXH: $d(J^*_{cex}, J'') \leq c^*_{abs}$;
then there exists $J \in R(P'_{new})$ with $L \not\subseteq J$.

By Proposition 4, we obtain the following refinement constraints for the Kemeny, Slater, and MaxHamming rules, respectively:

$$\sum_{x \in J^*_{\text{cex}}} \neg m_x + \sum_{\neg x \in J^*_{\text{cex}}} m_x \geq c^*_{\text{abs}} - \sum_{i \in I \setminus \{m\}} d(J^*_{\text{cex}}, J_i) + 1,$$

$$\sum_{\substack{x \in J^*_{\text{cex}} \\ \neg x \in \Phi_{\text{swing}}(P_{-m})}} \neg m_x + \sum_{\substack{\neg x \in J^*_{\text{cex}} \\ x \in \Phi_{\text{swing}}(P_{-m})}} m_x \geq c^*_{\text{abs}} - |\Phi_{\text{fixed}}(P_{-m}) \setminus J^*_{\text{cex}}| + 1,$$

$$\text{and} \sum_{x \in J^*_{\text{cex}}} \neg m_x + \sum_{\neg x \in J^*_{\text{cex}}} m_x \geq c^*_{\text{abs}} + 1.$$

Young. If $J^*_{\text{cex}} \supseteq \mathcal{M}(P^{\text{Y}})$ where P^{Y} does not include the manipulator's judgment set J', we can return false and terminate immediately. That is, no matter what the manipulator does, J^*_{cex} is a valid solution obtained by including the same $c^*_{\text{cex}} \leq c^*_{\text{abs}}$ judgment sets in the modified profile (removing the manipulator's judgment set). If J^*_{cex} is instead obtained from the majoritarian judgment set of $P^{\text{Y}} \cup \{J'\}$, the manipulator must indicate some judgment set J'' where $\mathcal{M}(P^{\text{Y}} \cup \{J''\}) \not\subseteq J^*_{\text{cex}}$. Otherwise, the counterexample J^*_{cex} is still a valid solution which can be obtained by removing the same judgment sets to form the modified profile $P^{\text{Y}} \cup \{J''\}$.

Dodgson. Recall that J^*_{cex} is a superset of the majoritarian judgment set of a modified profile P^{D} obtained by reverting some subset of the opinions indicated in the profile P_{new}. In this case, the manipulator must indicate a judgment set such that, in order to form the same profile P^{D}, additional modifications are required. In particular, if the Hamming distance to $(J')^{\text{D}}$ is sufficiently low, the same modified profile P^{D} can still be obtained with at most c^*_{abs} modifications.

Proposition 5. *Let $J'' \in \mathcal{J}(\Phi, \Gamma)$ be a candidate manipulator judgment set, and $P'_{new} = (J_1, \ldots, J_{m-1}, J'', J_{m+1}, \ldots, J_n)$ the corresponding modified profile. If*
for $R =$ YOUNG:
$J^*_{cex} \supseteq \mathcal{M}(P^{\text{Y}} \cup \{J'\})$, and J^*_{cex} and J'' agree on swing issues with respect to P^{Y}, i.e., $J^*_{cex} \cap \Phi_{swing}(P^{\text{Y}}) = J'' \cap \Phi_{swing}(P^{\text{Y}})$;
for $R =$ DODGSON: $d((J')^{\text{D}}, J'') \leq c^*_{abs} - \sum_{i \in I \setminus m} d(J^{\text{D}}_i, J_i)$;
then there exists $J \in R(P'_{new})$ with $L \not\subseteq J$.

Again, we directly obtain the following refinement constraints for Young and Dodgson rules, respectively:

$$\bigvee_{\substack{x \in J^*_{\text{cex}} \\ \neg x \in \Phi_{\text{swing}}(P^{\text{Y}})}} \neg m_x \vee \bigvee_{\substack{\neg x \in J^*_{\text{cex}} \\ x \in \Phi_{\text{swing}}(P^{\text{Y}})}} m_x,$$

$$\sum_{x \in (J')^{\text{D}}} \neg m_x + \sum_{\neg x \in (J')^{\text{D}}} m_x \geq c^*_{\text{abs}} - \sum_{i \in I} d(J^{\text{D}}_i, J_i) + 1.$$

6.2 Bribery

In bribery, a modified profile consisting of judgment sets J'_i for each $i \in C$ is encoded via variables $c_{i,x}$ for each corrupt agent $i \in C$ and issue $x \in X$. The refinement strategies for bribery generalize the ideas underlying refinement strategies for manipulation.

Kemeny. The cost of J^*_{cex} can be broken down to a constant part arising from P_{-C} and the sum of Hamming distances to judgment sets J'_i of corrupt agents $i \in C$. For the cost of J^*_{cex} to exceed c^*_{abs}, this sum must be sufficiently high.

Slater. We generalize the concept of swing issues to bribery as follows. Given our bribery budget k and any $l \in \Phi$, we can revert at most $\min(k, N(P[C], l))$ judgments of agents $i \in C$ to $\neg l$. For a profile P, we denote by $\Phi_{k\text{-swing}}(P) = \{l, \neg l \in \Phi \mid 0 \leq \Delta(P, l) \leq 2 \cdot \min(k, N(P[C], l))\}$, and $\Phi_{k\text{-fixed}}(P) = \{l \in \Phi \mid \Delta(P, l) > 2 \cdot \min(k, N(P[C], l))\}$. For the cost of J^*_{cex} to exceed c^*_{abs}, the strict majority needs to disagree with J^*_{cex} on sufficiently many swing issues.

MaxHamming. Similarly to manipulation, any agent $i \in C$ can potentially increase the cost of J^*_{cex}. In particular, the disagreement between J^*_{cex} and some bribed agent must exceed c^*_{abs}, or otherwise J^*_{cex} remains a counterexample.

Proposition 6. *For each $i \in C$, let $J''_i \in \mathcal{J}(\Phi, \Gamma)$ be a judgment set of a bribed agent, and P'_{new} be the corresponding profile where, for each $i \in C$, J_i is replaced by J''_i. If*

for $R = \text{KEMENY}$*:* $\sum_{i \in C} d(J^*_{cex}, J''_i) \leq c^*_{abs} - \sum_{i \in I \setminus C} d(J^*_{cex}, J_i)$*;*
for $R = \text{SLATER}$*:*
$|(\mathcal{M}(P'_{new}) \cap \Phi_{k\text{-}swing}(P)) \setminus J^*_{cex}| \leq c^*_{abs} - |\Phi_{k\text{-}fixed}(P) \setminus J^*_{cex}|$*;*
and for $R = \text{MAXH}$*:* $d(J^*_{cex}, J''_i) \leq c^*_{abs}$ *for all* $i \in C$*,*
then there exists $J \in R(P'_{new})$ *with* $L \not\subseteq J$*.*

For the Slater rule, the sums $S(P, x) = \sum_{i \in C} c_{i,x} + N(P_{-C}, x)$ represent the support of issue x in the encoding of the abstraction. The additional variables p_x and n_x for each $x \in X$ represent whether x or $\neg x$ are included in the majoritarian judgment set of the modified profile, encoded via equivalences $p_x \leftrightarrow S(P, x) \geq \lfloor n/2 \rfloor + 1$ and $n_x \leftrightarrow S(P, x) \leq \lceil n/2 \rceil - 1$. We obtain the refinements

$$\sum_{i \in C} \left(\sum_{x \in J^*_{\text{cex}}} \neg c_{i,x} + \sum_{\neg x \in J^*_{\text{cex}}} c_{i,x} \right) \geq c^*_{\text{abs}} - \sum_{i \in I \setminus C} d(J^*_{\text{cex}}, J_i) + 1,$$

$$\sum_{x \in J^*_{\text{cex}} \cap \Phi_{k\text{-swing}}(P)} n_x + \sum_{\neg x \in J^*_{\text{cex}} \cap \Phi_{k\text{-swing}}(P)} p_x \geq c^*_{\text{abs}} - |\Phi_{k\text{-fixed}}(P) \setminus J^*_{\text{cex}}| + 1,$$

and $\bigvee_{i \in C} \left(\sum_{x \in J^*_{\text{cex}}} \neg c_{i,x} + \sum_{\neg x \in J^*_{\text{cex}}} c_{i,x} \geq c^*_{\text{abs}} + 1 \right)$

for the Kemeny, Slater, and MaxHamming rules, respectively.

Young. Suppose $J^*_{\text{cex}} \supseteq \mathcal{M}(P^{\text{Y}})$ where P^{Y} is a subprofile of P_{new} for agents I^{Y}. If we use the same agents I^{Y} in a new modified profile, and the majoritarian judgment set remains the same, J^*_{cex} remains a counterexample. Furthermore, the majoritarian judgment set can be restricted to swing issues.

Dodgson. If we bribe agents so that the same modified profile can be acquired with a sufficiently low number of Dodgson modifications, then J^*_{cex} remains a counterexample.

Proposition 7. *For each* $i \in C$*, let* $J''_i \in \mathcal{J}(\Phi, \Gamma)$ *be a judgment set of a bribed agent, and* P'_{new} *be the corresponding profile where, for each* $i \in C$*,* J_i *is replaced by* J''_i*. If*
for $R = \text{YOUNG}$*:*
$J^*_{cex} \supseteq \mathcal{M}(P^{\text{Y}})$ *with* $P^{\text{Y}} = P_{new}[I^{\text{Y}}]$ *for* $I^{\text{Y}} \subseteq I$*, and* $J^*_{cex} \cap \Phi_{k\text{-}swing}(P[I^{\text{Y}}]) \supseteq \mathcal{M}(P'_{new}[I^{\text{Y}}]) \cap \Phi_{k\text{-}swing}(P[I^{\text{Y}}])$*;*
for $R = \text{DODGSON}$*:*
$\sum_{i \in C} d((J'_i)^{\text{D}}, J''_i) \leq c^*_{abs} - \sum_{i \in I \setminus C} d(J^{\text{D}}_i, J_i)$*;*
then there exists $J \in R(P'_{new})$ *with* $L \not\subseteq J$*.*

For the profile $P[I^{\text{Y}}]$ restricted to Young agents, the sums $S^Y(P[I^{\text{Y}}], x) = \sum_{i \in C \cap I^{\text{Y}}} c_{i,x} + N(P[(I \setminus C) \cap I^{\text{Y}}], x)$ represent the support of $x \in X$. We rule out the judgment sets in Proposition 7 with the following constraints for Young and Dodgson, respectively:

$$\bigvee_{x \in J^*_{\text{cex}} \cap \Phi_{k\text{-swing}}(P[I^{\text{Y}}])} \left(S^Y(P[I^{\text{Y}}], x) \leq \lceil (n - c^*_{\text{cex}})/2 \rceil - 1 \right) \vee$$

$$\bigvee_{\neg x \in J^*_{\text{cex}} \cap \Phi_{k\text{-swing}}(P[I^{\text{Y}}])} \left(S^Y(P[I^{\text{Y}}], x) \geq \lfloor (n - c^*_{\text{cex}})/2 \rfloor + 1 \right) \text{ and}$$

$$\sum_{i \in C} \left(\sum_{x \in (J'_i)^{\text{D}}} \neg c_{i,x} + \sum_{\neg x \in (J'_i)^{\text{D}}} c_{i,x} \right) \geq c^*_{\text{abs}} - \sum_{i \in I \setminus C} d(J^{\text{D}}_i, J_i) + 1.$$

7 Empirical Evaluation

We implemented the algorithms for manipulation and bribery on top of SATcha [9]. The implementation is available in open source [31].

Table 1: Manipulation, strong v simple refinements.

	Strong			**Simple**		
Rule	#slv	#true	#false	#slv	#true	#false
Kemeny	**89**	**61**	**28**	61	58	3
Slater	**169**	**66**	**103**	56	50	6
MaxHam.	**44**	**29**	**15**	32	25	7
Young	**241**	40	**201**	43	40	3
Dodgson	**55**	**45**	**10**	46	43	3

Table 2: Bribery with strong refinements ($k = 1$).

	p = 0.2			**p = 0.5**		
Rule	#slv	#true	#false	#slv	#true	#false
Kemeny	75	63	12	59	50	9
Slater	169	71	98	160	94	66
MaxHam.	42	32	10	43	40	3
Young	235	75	160	156	94	62
Dodgson	72	68	4	73	70	3

We use the incremental MaxSAT solver UWrMaxSAT [29], and encode cardinality constraints via the iterative totalizer CNF encoding [2, 24] from PySAT [20]. The experiments were run on 2.40GHz Intel Xeon Gold 6148 CPUs and 381-GB memory using a 30-minute time and 16-GB memory limit for each instance. We use the 405 judgment aggregation instances from [9] based on PrefLib [25] preference aggregation instances. Note that the PrefLib datasets are from different scenarios and are therefore nonuniform. Hence, algorithm runtimes on the instances are not expectedly dictated by the absolute size parameters of the individual datasets. For the benchmarks, under manipulation the first voter is arbitrarily designated the manipulator, and their desired outcome is that their preferred candidate is most preferred in all optimal collective judgment sets under the given judgment aggregation rule. For bribery, the outcome is chosen in the same way based on the first voter's preferences, who is removed from the profile, and the remaining voters are each designated corrupt with probability 0.2 or 0.5, of which at most one can be bribed ($k = 1$).

Results for manipulation using strong and simple (blocking only the judgment set chosen by the manipulator in the previous iteration) refinements are summarized in Table 1. The number of instances solved (#slv) is noticeably higher with the strong refinements techniques detailed in Section 6; the difference is most noticeable on instances proven impossible to manipulate (#false), where using the simple refinement strategy is prohibitively inefficient for all but very small instances. Note that the "false" instances intuitively require ruling out all non-solutions, i.e. exhausting the entire search space, while solutions to "true" instances may be found earlier "by luck". The results show that the strength of the strong refinements is witnessed particularly on the "negative" instances. The impact of strong refinements in the number of instances solved is particularly pronounced under the Young and Slater rules. Results for bribery using strong refinement are shown in Table 2. The instances again are easiest to solve under Slater and Young, owing to the strength of the respective refinement constraints under these rules.

8 Conclusions

We provided new complexity results and algorithms for manipulation and bribery as two central forms of strategic behavior in judgment aggregation. The second-level completeness results extend earlier results to cover various central aggregation rules. The strong refinements developed for the MaxSAT-based CEGAR approach, as first implemented and evaluated here, allow for solving significantly more PrefLib instances. For further work, both the complexity results and the algorithmic approach can expectedly be extended to other forms of strategic behavior such as control.

Acknowledgements

Work financially supported by Academy of Finland (grants 347588 and 356046). The authors thank the Finnish Computing Competence Infrastructure (FCCI) for computational and data storage resources.

References

[1] F. Bacchus, M. Järvisalo, and R. Martins. Maximum satisfiability. In *Handbook of Satisfiability - Second Edition*, volume 336 of *FAIA*, pages 929–991. IOS Press, 2021.

[2] O. Bailleux and Y. Boufkhad. Efficient CNF encoding of Boolean cardinality constraints. In *CP*, volume 2833 of *LNCS*, pages 108–122. Springer, 2003.

[3] J. J. Bartholdi III, C. A. Tovey, and M. A. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3): 227–241, 1989.

[4] D. Baumeister, G. Erdélyi, O. J. Erdélyi, and J. Rothe. Complexity of manipulation and bribery in judgment aggregation for uniform premise-based quota rules. *Math. Soc. Sci.*, 76:19–30, 2015.

[5] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016. ISBN 9781107446984.

[6] R. Bredereck and J. Luo. Complexity of manipulation and bribery in premise-based judgment aggregation with simple formulas. *Inf. Comput.*, 296:105128, 2024.

[7] E. M. Clarke, O. Grumberg, S. Jha, Y. Lu, and H. Veith. Counterexample-guided abstraction refinement for symbolic model checking. *J. ACM*, 50(5):752–794, 2003.

[8] E. M. Clarke, A. Gupta, and O. Strichman. SAT-based counterexample-guided abstraction refinement. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, 23(7):1113–1123, 2004.

[9] A. Conati, A. Niskanen, and M. Järvisalo. SAT-based judgment aggregation. In *AAMAS*, pages 1412–1420. IFAAMAS, 2023.

[10] V. Conitzer, T. Sandholm, and J. Lang. When are elections with few candidates hard to manipulate? *J. ACM*, 54(3):14, 2007. doi: 10.1145/1236457.1236461. URL https://doi.org/10.1145/1236457.1236461.

[11] R. de Haan. Complexity results for manipulation, bribery and control of the Kemeny judgment aggregation procedure. In *AAMAS*, pages 1151–1159. ACM, 2017.

[12] U. Endriss. Judgment aggregation. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. Procaccia, editors, *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[13] U. Endriss. Judgment aggregation with rationality and feasibility constraints. In *AAMAS*, pages 946–954. IFAAMAS / ACM, 2018.

[14] U. Endriss, U. Grandi, and D. Porello. Complexity of judgment aggregation. *J. Artif. Intell. Res.*, 45:481–514, 2012.

[15] U. Endriss, R. de Haan, J. Lang, and M. Slavkovik. The complexity landscape of outcome determination in judgment aggregation. *J. Artif. Intell. Res.*, 69:687–731, 2020.

[16] P. Faliszewski, E. Hemaspaandra, and L. A. Hemaspaandra. How hard is bribery in elections? *J. Artif. Intell. Res.*, 35:485–532, 2009. doi: 10.1613/JAIR.2676. URL https://doi.org/10.1613/jair.2676.

[17] Z. Fitzsimmons, E. Hemaspaandra, A. Hoover, and D. E. Narváez. Very hard electoral control problems. In *AAAI*, pages 1933–1940. AAAI Press, 2019.

[18] D. Grossi and G. Pigozzi. *Judgment Aggregation: A Primer*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2014.

[19] E. Hemaspaandra, L. A. Hemaspaandra, and J. Rothe. Anyone but him: The complexity of precluding an alternative. *Artif. Intell.*, 171(5-6): 255–285, 2007. doi: 10.1016/J.ARTINT.2007.01.005. URL https://doi.org/10.1016/j.artint.2007.01.005.

[20] A. Ignatiev, A. Morgado, and J. Marques-Silva. PySAT: A Python toolkit for prototyping with SAT oracles. In *SAT*, volume 10929 of *LNCS*, pages 428–437. Springer, 2018.

[21] J. Lang, G. Pigozzi, M. Slavkovik, and L. W. N. van der Torre. Judgment aggregation rules based on minimization. In *TARK*, pages 238–246. ACM, 2011.

[22] J. Lang, G. Pigozzi, M. Slavkovik, L. van der Torre, and S. Vesic. A partial taxonomy of judgment aggregation rules and their properties. *Soc. Choice Welf.*, 48(2):327–356, 2017.

[23] C. List. The theory of judgment aggregation: an introductory review. *Synth.*, 187(1):179–207, 2012.

[24] R. Martins, S. Joshi, V. M. Manquinho, and I. Lynce. Incremental cardinality constraints for MaxSAT. In *CP*, volume 8656 of *LNCS*, pages 531–548. Springer, 2014.

[25] N. Mattei and T. Walsh. PrefLib: A library for preferences http://www.preflib.org. In *ADT*, volume 8176 of *LNCS*, pages 259–270. Springer, 2013.

[26] M. K. Miller and D. N. Osherson. Methods for distance-based judgment aggregation. *Soc. Choice Welf.*, 32(4):575–601, 2009.

[27] K. Nehring, M. Pivato, and C. Puppe. The Condorcet set: Majority voting over interconnected propositions. *J. Econ. Theory*, 151:268–303, 2014.

[28] G. Pigozzi. Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. *Synth.*, 152(2): 285–298, 2006.

[29] M. Piotrów. UWrMaxSat: Efficient solver for MaxSAT and pseudo-Boolean problems. In *ICTAI*, pages 132–136. IEEE, 2020.

[30] J. Rothe, editor. *Economics and Computation, An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*. Springer texts in business and economics. Springer, 2016. ISBN 978-3-662-47903-2.

[31] SATcha. A SAT-based system for judgment aggregation, 2024. URL https://bitbucket.org/coreo-group/satcha.