Trends in Computational Social Choice

8

Cite as: Dorothea Baumeister, Jörg Rothe, and Ann-Kathrin Selker. Strategic Behavior in Judgment Aggregation. In Ulle Endriss (editor), *Trends in Computational Social Choice*, chapter 8, pages 145–168. AI Access, 2017.

http://www.illc.uva.nl/COST-IC1205/Book/

CHAPTER 8

Strategic Behavior in Judgment Aggregation

Dorothea Baumeister, Jörg Rothe, and
Ann-Kathrin Selker

8.1 Introduction

Collective decision making is concerned with aggregating information reported by a number of individuals into a collective decision appropriately capturing the individual views as a whole. Examples include, most prominently, *preference aggregation in voting* (surveyed, e.g., by Zwicker, 2016, and Baumeister and Rothe, 2015) where voters express their preferences on the candidates and the collective decision is to select a winner; *argumentation frameworks* (surveyed, e.g., by Rahwan and Simari, 2009) where individuals express arguments on an issue that can attack or support each other and while the individuals may have different assessments of which arguments are valid or which attack which, a goal is to collectively decide which arguments to select according to certain criteria (e.g., conflict-freeness); *resource allocation* and *fair division* (surveyed, e.g., by Bouveret et al., 2016, Lang and Rothe, 2015, and Moulin, 2004) where agents have individual utilities for bundles of objects and the collective decision is to allocate the objects to agents so that social welfare is maximized or certain fairness conditions (e.g., envy-freeness) are satisfied; and *judgment aggregation* (previously surveyed, e.g., by Endriss, 2016, Baumeister et al., 2015b, Grossi and Pigozzi, 2014, List, 2012, and List and Puppe, 2009) where individual judges express possibly different opinions on whether some logically connected propositions are true or false and the collective decision is to find a joint judgment on their truth.

This chapter is devoted to judgment aggregation and will in particular focus on analyzing scenarios involving strategic behavior in this context. The beginnings of the field of judgment aggregation go back to the seminal work of Kornhauser and Sager (1986) who were the first to describe a situation that they called the *doctrinal paradox*.[1] For illustration, suppose three judges—Alyson, Bill, and Cadi—are going to adjudicate upon the guilt of their colleague, judge Don, who is accused of having accepted a bribe in a previous trial where he pronounced a verdict of not guilty of murder for an alleged mafia boss. In the present trial, judge Don

[1] As a more general variant, Pettit (2001) introduced the *discursive dilemma*; the differences between the two notions are discussed in detail by Mongin (2012).

is to be sentenced for five years in prison if and only if he is found guilty, first, of having taken a considerable amount of money from a close associate of the alleged mafia boss and, second, of having denied a relevant piece of evidence in court that would have entailed the death sentence for sure.

"To me it's crystal-clear," judge Alyson goes first. "$3000 is a considerable amount of money that was given to Don in an envelope when he thought no one were looking. And how can he *not* allow the knife with the mafia boss's fingerprints on it as a very relevant piece of evidence? It was found stuck in the victim's body, for goodness' sake! I conclude that Don has to go to prison."

"I do agree with your second point, Your Honor," says judge Bill slowly. "However, I disagree with your first point and, therefore, with your conclusion as well. Sure enough, $3000 sounds like a lot of money, but taking into account that it's *Canadian* dollars makes it much less sizeable. I wouldn't even speak of bribery here; it's just peanuts. And we cannot sentence Don to five years of prison for bribery if all he has taken is just peanuts."

"You can't be serious, Your Honor," judge Cadi now counters. "3000 bucks, Canadian or US, *is* a considerable amount of money and *cannot* go unpunished—*if* it indeed was used to bribe judge Don and to bias his judgment toward suppressing some relevant piece of evidence. However, I do agree with your conclusion that Don should not have to go to prison, because I do not consider this knife a relevant piece of evidence. May I remind you that the victim in fact was killed by machine gun fire? The body was completely perforated! I have no idea why this knife stuck in the body, but I do know for sure that it was not causing death and, hence, it was completely irrelevant for this trial."

Judge	Considerable amount?	Relevant evidence denied?	Is Don guilty?
Alyson	true	true	true
Bill	false	true	false
Cadi	true	false	false
Majority	true	true	false

Table 8.1: Doctrinal paradox

Table 8.1 shows the three individual judgments. Note that the proposition "Don is guilty" is logically equivalent to the conjunction of the propositions that "the amount is considerable" and "a relevant piece of evidence has been denied." Now, if we aggregate the three individual judgments by the majority rule, as shown in Table 8.1, we see that, even though the individual judgments are each logically consistent, we obtain a logically inconsistent collective judgment. That is why Kornhauser and Sager (1986) called it a paradox.

In Section 8.2, we will outline the basics of judgment aggregation and will discuss various judgment aggregation rules and their properties and the complexity of winner determination. The main part of this chapter is Section 8.3 where we will deal with strategic behavior in judgment aggregation, including manipulation, bribery, and control. In particular, we will give an overview of computational complexity results for the associated problems.

8.2 Foundations of Judgment Aggregation

In this section, we present the basics of judgment aggregation, introduce the preferences that judges may have about judgment sets as well as some common judgment aggregation rules and their properties, briefly mention some complexity-theoretic notions, and discuss the complexity of winner determination.

8.2.1 Basics

We briefly recall the basic notions of judgment aggregation, starting with the *formula-based framework*. Throughout this chapter, we will essentially use the notation of Endriss (2016), Baumeister et al. (2015b), and de Haan (2016b).

For a set PS of propositional variables, let $\mathcal{L}_{PS}$ denote the set of all propositional formulas that can be built from variables in PS by using the common boolean connectives (such as $\wedge$, $\vee$, $\neg$, $\Rightarrow$, and $\Leftrightarrow$). We use $\overline{\varphi}$ to denote the *complement of* φ, i.e., $\overline{\varphi} = \neg\varphi$ if φ is not negated, and $\overline{\varphi} = \psi$ if $\varphi = \neg\psi$. An *agenda* is a finite set $\Phi \subseteq \mathcal{L}_{PS}$ of *formulas* (or *issues* or *propositions*) without doubly negated formulas that is *closed under complementation* (i.e., $\overline{\varphi} \in \Phi$ for each $\varphi \in \Phi$). Every set $J \subseteq \Phi$ is called a *judgment set*. A judgment set J is said to be *complete* if $\varphi \in J$ or $\overline{\varphi} \in J$ for each $\varphi \in \Phi$, and J is said to be *consistent* if there exists a truth assignment such that each formula in J is true. Let $\mathcal{J}(\Phi)$ denote the set of complete and consistent judgment sets. For an agenda Φ, and a set $N = \{1, \ldots, n\}$ of *judges* (or *agents*), $\boldsymbol{J} = (J_1, \ldots, J_n) \in \mathcal{J}(\Phi)^n$ denotes their *profile of (individual) complete, consistent judgment sets*. If not stated otherwise, the presented examples and results will employ the formula-based framework.

The second framework often used in judgment aggregation is the *constraint-based framework*: The agenda $\Phi = \{\varphi_1, \ldots, \varphi_m, \neg\varphi_1, \ldots, \neg\varphi_m\}$ consists of a finite set of propositional variables and their negations and we have an integrity constraint Γ, i.e., a propositional formula over these variables that can be used to restrict the judgment sets we consider. A judgment set J is Γ-*consistent* if there exists a truth assignment such that each formula in the set and Γ are true. All other terms are defined accordingly. An overview of this framework is given, e.g., by de Haan (2016b).[2]

We say that two complete judgment sets, J and J', *agree on a proposition* $\varphi \in \Phi$ if either both contain φ or none of them does; otherwise, we say J and J' *disagree on* φ; and their *Hamming distance* $H(J, J')$ is the number of disagreements between J and J'. More generally, since we will also use the Hamming distance between two consistent, but not necessarily complete judgment sets J and J', $H(J, J')$ is defined as the number of positive issues occurring in exactly one of J and J' and its negation in the other. (One can also consider the *weighted Hamming distance*, denoted by $H_\omega(J, J')$ for a weight function $\omega : \Phi \to \mathbb{N}$ with $\omega(\varphi) = \omega(\overline{\varphi})$, where we sum up the weights of the corresponding issues instead of

[2] He defines also the formula-based framework so as to include an integrity constraint, not necessarily an element of $\mathcal{L}_{PS}$. Results whose proofs require this constraint will be marked. A detailed comparison of the formula-based and the constraint-based framework is due to Endriss et al. (2016). They compare the succinctness of both frameworks and explore the effect on computational problems.

counting them.) Define the Hamming distance between a profile $\boldsymbol{J}$ of consistent judgment sets and a consistent judgment set J' as $H(\boldsymbol{J}, J') = \sum_{J \in \boldsymbol{J}} H(J, J')$.

Besides the common choice of using classical propositional logic to formulate judgment aggregation settings (as we do throughout this chapter), Dietrich (2007) proposes a more general model that includes problems expressed in predicate, modal, or conditional logic and some multi-valued and fuzzy logics.

8.2.2 Preferences

We will model the strategic behavior of agents—either *internal* ones (who are judges themselves) or *external* ones (who from the outside seek to influence the result of a judgment aggregation procedure to their advantage)—who want to obtain a "better" outcome than before. Therefore, in order to measure the success of an attack, the agents need to rank the possible outcomes depending on their *desired set* J; desired sets will always be assumed to be contained in a complete and consistent set. However, given an agenda Φ with m positive issues, there are up to 2^m possible (complete and consistent) outcomes. That is why we need a compact way of representing agents' preferences over judgment sets, even if—as a consequence—we may lose information about their preferences.

We now define four preference types that were introduced by Dietrich and List (2007c) and later applied by Baumeister et al. (2015a,c). Here, we only define preferences over complete and consistent judgment sets. A *weak order* over $\mathcal{J}(\Phi)$ is a transitive and total binary relation $\succsim$ by which any two judgment sets in $\mathcal{J}(\Phi)$ can be compared with one another.

For each of the following preference types, an agent is said to be *indifferent between X and Y under this type*, denoted by $X \sim Y$, if $X \succsim Y$ and $Y \succsim X$, and to *strictly prefer X to Y under this type*, denoted by $X \succ Y$, if $X \succsim Y$ and not $Y \succsim X$.

Definition 8.1 (Preference types). *For an agenda Φ, let $X, Y \in \mathcal{J}(\Phi)$, let $J \subseteq \Phi$ be an agent's desired set (consistent but possibly incomplete), and let U_J be the set of all* unrestricted J-induced (weak) preferences*, i.e., the set of all weak orders $\succsim_J$ over $\mathcal{J}(\Phi)$ for which $X \sim_J Y$ whenever $X \cap J = Y \cap J$.*

We say that a weak order $\succsim_J \in U_J$ is a

1. top-respecting J-induced (weak) preference *if $X \succ_J Y$ whenever $X \cap J = J$ and $Y \cap J \neq J$, i.e., all we know is that the desired set J is a subset of this agent's most preferred judgment set;*
2. closeness-respecting J-induced (weak) preference *if $X \succ_J Y$ whenever $X \cap J \supset Y \cap J$, i.e., whenever X agrees with J on the same issues as Y with J and on at least one issue more than Y with J; and*
3. Hamming-distance-respecting J-induced (weak) preference *if $X \succsim_J Y$ whenever $H(X, J) \leqslant H(Y, J)$, i.e., whenever X and J disagree on at most as many issues as Y and J. (In the weighted case, $X \succsim_J Y \iff H_\omega(X, J) \leqslant H_\omega(Y, J)$.)*

While we learn (essentially) nothing from unrestricted J-induced preferences, top-respecting J-induced preferences tell us something about an agent's most preferred judgment set: namely, that it contains J. The same is true for closeness-respecting J-induced preferences, but for them we know in addition that judgment sets having additional agreements with J are preferred. This is

also the case for Hamming-distance-respecting J-induced preferences, which depend on the total number of disagreements and are the most restrictive preference type.

Let X and Y be complete and consistent judgment sets. We say that an agent with a (possibly incomplete) desired set J *possibly/necessarily weakly prefers* X *to* Y *under preference type* T if $X \succsim_J Y$ holds true in some/all J-induced weak orders of type T, and *possible/necessary preference of* X *to* Y *under type* T is defined analogously via $X \succ_J Y$. Since there is exactly one Hamming-distance-respecting J-induced weak order for each desired set J, the notions of possible and necessary preferences coincide for this type.

Example 8.1 (Preferences). *Consider the example from the introduction. Let* $N = \{A, B, C\}$ *denote the set of the three judges:* A*(lyson),* B*(ill), and* C*(adi). Let* $\Phi = \{a, \neg a, e, \neg e, g, \neg g\}$ *be the agenda, where* a *stands for* amount, e *for* evidence, *and* g *for* guilt *and where* g *is* $a \wedge e$.[3] *Further, let* $\boldsymbol{J} = (J_A, J_B, J_C) \in \mathcal{J}(\Phi)^3$ *with* $J_A = \{a, e, g\}$, $J_B = \{\neg a, e, \neg g\}$, *and* $J_C = \{a, \neg e, \neg g\}$ *be the profile of complete and consistent judgment sets (see Table 8.1). Now assume that judge Bill's desired set is* $J = \{e, \neg g\}$. *In addition to* J_A, J_B, *and* J_C, *the only possible complete and consistent judgment set for* Φ *is* $J_0 = \{\neg a, \neg e, \neg g\}$. *Since among these four sets, Bill's judgment set* J_B *is the only one containing* J, *Bill necessarily prefers* J_B *to all others (i.e., to* J_A, J_C, *and* J_0*) under top-respecting* J*-induced preferences. Assuming closeness-respecting* J*-induced preferences, Bill necessarily prefers* J_B *to* J_C *because* J_B *and* J *(of course) agree on the whole desired set* J, *whereas* J_C *and* J *only agree on a strict subset of* J. *However, Bill only possibly prefers* J_A *to* J_C *and he also possibly prefers* J_C *to* J_A*: Both judgment sets agree with* J *on different issues, so we do not know which set he actually prefers under closeness-respecting* J*-induced preferences. The situation is different when we assume Hamming-distance-respecting* J*-induced preferences. Since* $H(J_C, J) = 1 = H(J_A, J)$, *we know that—with respect to his desired set* J*—Bill is indifferent between these two judgment sets. Note that knowing that Bill is indifferent between these two judgment sets decisively differs from not knowing which of them he prefers to the other.*

8.2.3 Judgment Aggregation Rules and Their Properties

Having the individual judgment sets of the participating judges, a judgment aggregation rule is needed to reach a consensus. A *judgment aggregation rule* (or *procedure*) is a function F that maps any profile of judgment sets to a set of judgment sets, which we call the *(collective) outcome*. F is *complete* (*consistent*) if each $J \in F(\boldsymbol{J})$ is complete (consistent) for each profile $\boldsymbol{J} = (J_1, \ldots, J_n) \in \mathcal{J}(\Phi)^n$, and F is *resolute* if the outcome is always a singleton (and is *irresolute* otherwise). For resolute rules F, we write $F(\boldsymbol{J}) = J$ rather than $F(\boldsymbol{J}) = \{J\}$.

[3] This is an example of a conjunctive agenda. An agenda is *conjunctive* if it consists of premises $p_1, \ldots, p_k$, a conclusion of the form $p_1 \wedge \cdots \wedge p_k$, and their negations, and it is *disjunctive* if its conclusion is of the form $p_1 \vee \cdots \vee p_k$. Note that Dietrich and List (2007c) consider the conclusion to be just a variable c and add a "connection rule" $c \Leftrightarrow (p_1 \wedge \cdots \wedge p_k)$ or $c \Leftrightarrow (p_1 \vee \cdots \vee p_k)$ to the agenda. Note further that if we adapt our example to the constraint-based framework, g would be a propositional variable instead of the formula $a \wedge e$ and $g \Leftrightarrow (a \wedge e)$ would be the integrity constraint Γ.

The perhaps most intuitive way of judgment aggregation is the *(proposition-wise) majority rule* that was used in Table 8.1 to illustrate the doctrinal paradox. One way of circumventing the doctrinal paradox is to use a *premise-based* approach: Divide the agenda into premises and conclusions, then apply a rule on the premises and derive the outcome for the conclusions from the outcome for the premises. To generalize the majority rule, Dietrich and List (2007b) introduced the quota rules, and to guarantee complete and consistent outcomes, we use the premise-based approach and focus on the class of *uniform premise-based quota rules*. Under these rules, a premise is contained in the collective outcome if and only if the number of judges having it in their individual judgment sets exceeds the quota. The outcome for the conclusions can then be derived easily. Formally:

Definition 8.2 (Uniform premise-based quota rules). *Partition the agenda Φ into a set of premises Φ_p and a set of conclusions Φ_c, both closed under complementation, and partition Φ_p into sets Φ_1 and Φ_2 so that $\varphi \in \Phi_1$ if and only if $\overline{\varphi} \in \Phi_2$. (We assume that Φ_1 consists of all positive literals.) Let $|S|$ denote the cardinality of a set S and $\models$ the satisfaction relation. The* uniform premise-based quota rule with quota q, $0 \leqslant q < 1$ *and q rational, is a function mapping each profile $\boldsymbol{J} = (J_1, \ldots, J_n)$ over Φ to the collective outcome $UPQR_q(\boldsymbol{J}) = \Delta \cup \{\psi \in \Phi_c \mid \Delta \models \psi\}$, where $\Delta = \{\varphi \in \Phi_1 \mid |\{i \mid \varphi \in J_i\}| > nq\} \cup \{\varphi \in \Phi_2 \mid |\{i \mid \varphi \in J_i\}| \geqslant n(1-q)\}$.*

The special case $UPQR_{1/2}$ with an odd number of judges is simply called the premise-based procedure *(PBP, for short; see the work of Endriss et al., 2012).*

$UPQR_q$ is resolute and—if the agenda Φ is closed under propositional variables and the set of premises Φ_p consists of exactly all literals—$UPQR_q$ is also complete and consistent. In the following sections, we will assume that these restrictions hold. Note that a premise $\varphi \in \Phi_1$ is part of the collective outcome if and only if at least $\lfloor nq + 1 \rfloor$ judges accept it, whereas the outcome contains $\overline{\varphi}$ if and only if it is part of at least $\lceil n(1-q) \rceil$ judgment sets.

By contrast, in the *conclusion-based* approach, votes are taken only on the conclusions (e.g., by requiring that a conclusion $\psi \in \Phi_c$ is in the collective judgment set if a strict majority of judges have ψ in their individual judgment sets, and otherwise $\overline{\psi}$ is in the collective judgment set), and no collective judgments are made on the premises. An obvious disadvantage of conclusion-based procedures is that they always output incomplete collective judgment sets.

Another way of using majority to reach a consensus is to apply it sequentially. The input additionally contains a fixed order over the positive issues in the agenda. In each step, we check whether the current solution entails a solution for the next issue of the agenda, and if this is not the case, we apply the majority rule. Obviously, this always leads to complete and consistent outcomes, but the solution strongly depends on the chosen order, i.e., it is path-dependent (see, for example, the work by Dietrich and List, 2007b). Sequential variants of other judgment aggregation rules are defined analogously.

Yet another possibility of defining judgment aggregation rules is to consider distances between judgment sets and to choose those judgment sets that minimize the sum of the distances to the individual judgment sets. In voting theory, the method due to Kemeny (1959) also minimizes the sum of the distances to the votes to elect a winner. This approach has been transferred to judgment ag-

gregation by Pigozzi (2006) and further extended to the Prototype-Hamming rule by Miller and Osherson (2009): The Kemeny rule in judgment aggregation picks exactly the complete and consistent judgment sets closest to the given profile.[4]

Definition 8.3 (Kemeny rule). *Let Φ be an agenda. The* Kemeny rule *maps each profile $\boldsymbol{J}$ over Φ to the collective outcome $Kemeny(\boldsymbol{J}) = \operatorname{argmin}_{J \in \mathcal{J}(\Phi)} H(\boldsymbol{J}, J)$.*

Note that the majority outcome is consistent if and only if it coincides with the Kemeny outcome.[5] The Kemeny rule is complete, consistent, and irresolute.

Example 8.2. *Consider the setting in Example 8.1. Let $\Phi_p = \{a, \neg a, e, \neg e\}$ be the set of premises and let $\Phi_c = \{g, \neg g\}$ be the set of conclusions. Then $UPQR_{1/2}(\boldsymbol{J}) = J_A$, since a majority of judges accept a and e (and g is evaluated accordingly). On the other hand, $Kemeny(\boldsymbol{J}) = \{J_A, J_B, J_C\}$, since all three judgment sets have a Hamming distance of 4 to the profile $\boldsymbol{J}$, whereas the only other complete and consistent judgment set, $J_0 = \{\neg a, \neg e, \neg g\}$, has a Hamming distance of 5 to $\boldsymbol{J}$.*

While there is a large body of literature on specific voting rules in social choice theory, the early work in judgment aggregation has focused more on the study of impossibility results. More recently, further specific judgment aggregation rules have been introduced, for example, by Lang et al. (2011). They transfer minimization concepts from voting theory and logic-based knowledge representation and reasoning to define judgment aggregation rules that in some way minimize the part of a profile that has to be removed to reach a consensus. Lang et al. (2017) survey existing judgment aggregation rules, their properties, and the relations between them.

Besides consistency and completeness, many other properties of judgment aggregation rules have been studied, for example, by List and Pettit (2002) and Dietrich and List (2007c). We will focus on properties of resolute judgment aggregation rules only. A very basic property is the *universal domain* assumption. It requires that a rule's domain consists of all possible profiles of complete and consistent judgment sets, which is the case for the rules studied here. Another basic property is *anonymity*, which says that the order of the judges should have no influence on the collective outcome. A more demanding property is *independence*. A judgment aggregation rule is *independent* if for any two profiles with the same number of judges over the same agenda, if the individual judgments are the same for any given proposition, then the collective outcomes for both profiles should agree on this proposition. That is to say that the collective judgment regarding any proposition should be independent of the collective judgments on the remaining propositions. The *neutrality* property requires that if all judges have the same opinion on any two propositions, then the collective judgments on these propositions should also be the same. Unfortunately, List and Pettit (2002) show that if the agenda contains two literals and their conjunction, then no judgment

[4] This rule is also referred to as *median rule* by Nehring et al. (2011), *max-weight subagenda* by Lang and Slavkovik (2014), and *distance-based procedure* by Baumeister et al. (2015b) and Endriss et al. (2012), and Dietrich (2014) shows that it coincides with what he calls the *simple scoring rule*.

[5] The majority outcome is consistent if the profile $\boldsymbol{J}$ is *unidimensionally aligned* (List, 2003), i.e., if there is an alignment of the judges from left to right so that for each proposition φ, the judges accepting φ are to the left of the ones accepting $\overline{\varphi}$ (or vice versa).

aggregation rule always returning a complete and consistent collective outcome can simultaneously satisfy anonymity, neutrality, and independence.

Example 8.3. *Consider again the setting in Example 8.1, with the set of premises* $\Phi_p = \{a, \neg a, e, \neg e\}$ *and the set of conclusions* $\Phi_c = \{g, \neg g\}$*. When Cadi changes her mind on whether the amount is considerable, her new judgment set is* $J'_C = \{\neg a, \neg e, \neg g\} = J_0$*, so* $UPQR_{1/2}(\boldsymbol{J}') = J_B$ *for the modified profile* $\boldsymbol{J}'$*. But this violates independence, since the individual judgments on* g *do not change, but the collective judgment on* g *is not the same for both profiles:* $g \in J_A$ *but* $\neg g \in J_B$*.*

The *monotonicity* property informally says that a proposition should never be judged worse collectively because of receiving additional individual support: If a proposition is collectively accepted, but now some judge changes her mind from rejecting to accepting it while all other judges stick to their judgments of this proposition, then it should still be collectively accepted after this change. Dietrich and List (2007b) show that a class of judgment aggregation rules, the quota rules, can be characterized through the properties of anonymity, independence, and monotonicity. Many more characterization and impossibility results are known in judgment aggregation. For example, Dietrich and List (2007a) prove an analogue of the theorem of Arrow (1951—revised 1963) in judgment aggregation, and we will see more examples due to Dietrich and List (2007c) in Section 8.3.1 concerning strategic manipulation.

8.2.4 Winner Determination

Some desirable properties (such as completeness and consistency) of judgment aggregation rules have been described above. When used in multiagent systems with a large number of participating judges, computational aspects must also be taken into account. As judgment aggregation may be applied in security systems, it is extremely important that the collective outcome of a rule can be computed efficiently. This raises the question on the complexity of winner determination.

We assume the reader to be familiar with the basic notions of complexity theory, including the complexity classes P and NP and the notions of hardness and completeness for complexity classes. For more background on the relevant classes—namely, the classes $\Theta_2^p = \mathrm{P}_{||}^{\mathrm{NP}}$ (a.k.a. "parallel access to NP"), $\Sigma_2^p = \mathrm{NP}^{\mathrm{NP}}$, and $\Pi_2^p = \mathrm{coNP}^{\mathrm{NP}}$ that constitute the second level of the polynomial hierarchy and the parameterized class W[2]—we refer to the books by Rothe (2015) (Section 1.5), Rothe (2005) (Sections 5.2 and 5.3), Downey and Fellows (2013), and Chapter 11 of this book.

For the uniform premise-based quota rules, it has to be checked for every premise whether the quota is reached or not, which is obviously possible in polynomial time. Due to the agenda being closed under propositional variables, the collective outcome for the conclusions can then also be computed efficiently.

Unfortunately, this is not the case for the Kemeny rule in judgment aggregation. To study the computational complexity, an adequate decision problem has to be formulated. For irresolute rules F, Endriss et al. (2012) propose the following definition of F-WINNER-DETERMINATION: Given an agenda Φ, a profile $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$, and a subset $L \subseteq \Phi$, is there a $J^\star \subseteq \Phi$ with $L \subseteq J^\star$ such that $J^\star \in F(\boldsymbol{J})$?

Theorem 8.1 (Endriss et al., 2012). *$Kemeny$-WINNER-DETERMINATION is Θ_2^p-complete.*

In addition to the decision problems, Endriss and de Haan (2015) study search problems for winner determination in judgment aggregation. Determining the complexity of the search problem, which outputs a collective outcome, is more useful for practical purposes than determining that of the decision problem, which merely gives a yes/no answer. For the Kemeny rule, an even more fine-grained complexity analysis is given by de Haan (2016a,b): the parameterized complexity with respect to five parameters and their combinations. He studies both the formula-based and the constraint-based framework, with the surprising result that even though classical complexity results are the same in both models, the parameterized complexity results differ. In addition to the above winner determination problem, Lang and Slavkovik (2014) determine the complexity of problems that ask whether the collective outcome satisfies some given property.

8.3 Strategic Behavior in Judgment Aggregation

We now survey various scenarios of strategic behavior in judgment aggregation, namely manipulation, bribery, and control, which have been intensively studied for elections in computational social choice (see Conitzer and Walsh, 2016, Faliszewski and Rothe, 2016, and Baumeister and Rothe, 2015), and we show how to transfer these models from preference to judgment aggregation.

8.3.1 Manipulation

Example 8.4. *The court is hiring new judges. Chief judge Zoe is on a business trip officially (even though, inofficially and undercover, she is meeting with the alleged mafia boss—who was just acquitted of murder—in the underworld bar "Angels from Hell"), leaving the hiring decision to her judges Alyson, Bill, Cadi, and Don (who at present is not yet on trial for having been bribed). According to the job description, a new judge is to be* hired *if and only if s/he has a proven track record and expertise in at least one of the areas this court is so renowned for:* drug trafficking offenses *(denoted by variable d),* financial crimes *(f),* large-scale frauds *(ℓ), and* organized crime *(o). That is, the (disjunctive) agenda contains the premises d, f, ℓ, and o, the conclusion $h = d \vee f \vee \ell \vee o$, and their negations.*

Elena, Felix, George, and Hillevi have applied for a job as a judge. The four judges in charge, using $UPQR_{2/3}$, quickly and unanimously agree on three of these candidates: Felix and George will be hired, but Hillevi fails. Elena's case, though, is not as clear. After listening to his co-judges' arguments and reasons, Don knows their judgment sets: $J_A = \{d, f, \neg\ell, \neg o, h\}$, $J_B = \{d, \neg f, \ell, \neg o, h\}$, and $J_C = \{\neg d, \neg f, \neg\ell, \neg o, \neg h\}$.

Looking at Elena's application papers, Don's truthful judgment would be the same as Cadi's, which would result in not hiring Elena because at least one of the premises must be accepted by at least three judges for her to be hired. However, he wouldn't be Don if he'd look only at papers! Indeed, Don is looking at Elena . . . and

suddenly he has an agenda of his own and changes his mind. What he sees is a beautiful young lady and, being an outcome-oriented person, he doesn't care about her expertise and track record; all that matters for him is the conclusion: He wants Elena to be hired! Rather than his truthful judgment set $J_D = J_C$, he thus reports the set $J_D^* = \{d, f, \neg\ell, \neg o, h\} = J_A$, just as Alyson. With three judges accepting d for Elena, instead of $UPQR_{2/3}(\boldsymbol{J})$ providing the collective outcome $\{\neg d, \neg f, \neg\ell, \neg o, \neg h\}$ for the truthful profile $\boldsymbol{J}$, we have the outcome $\{d, \neg f, \neg\ell, \neg o, h\}$ for the modified profile $\boldsymbol{J}^*$, which means that Elena will be hired.

"What?" baffled Cadi looks at Don with a reproachful glance. "Didn't we have the same opinion on Elena when we discussed her application?" Then, looking again at the glossy photograph on Elena's application folder and becoming suspicious, she adds, "Shame on you, Don! You are sexist and a manipulator!"

Dietrich and List (2007c) were the first to study manipulation and strategy-proofness in judgment aggregation. In particular, they introduced the preference types presented in Definition 8.1 so as to formulate a judgment aggregation analogue of the famous Gibbard–Satterthwaite Theorem from social choice theory, which is due to Gibbard (1973) and Satterthwaite (1975) and, roughly, says that no reasonable voting rule can be strategy-proof (i.e., were it to satisfy a number of reasonable conditions including strategy-proofness, it would be dictatorial).

Dietrich and List (2007c) define a resolute judgment aggregation rule F to be *strategy-proof* if for each profile $\boldsymbol{J} = (J_1, \ldots, J_n)$ of individual judgment sets, for each judge i, and for each preference relation induced by J_i according to one of the preference types in Definition 8.1, i weakly prefers the outcome $F(\boldsymbol{J})$ (resulting, in particular, from her truthful judgment set J_i) to any outcome $F(\boldsymbol{J}_{-i}, J_i^*) = F(J_1, \ldots, J_{i-1}, J_i^*, J_{i+1}, \ldots, J_n)$ (i.e., to any outcome of F on the profile identical to $\boldsymbol{J}$ except with J_i replaced by J_i^*) with a misreported judgment set J_i^*.

By contrast, they also define a preference-free notion of *nonmanipulability*: F is *manipulable at profile $\boldsymbol{J} = (J_1, \ldots, J_n)$ by individual judge i on proposition $\varphi \in \Phi$* if J_i disagrees with $F(\boldsymbol{J})$ on φ, but J_i agrees with $F(\boldsymbol{J}_{-i}, J_i^*)$ on φ for some misrepresented judgment set J_i^*. F is said to be *nonmanipulable* if F is not manipulable at any profile by any individual judge on any proposition in Φ.[6] The crucial difference between strategy-proofness and nonmanipulability is that the former notion is based on preferences and so expresses *incentives* of individual judges to misreport their judgment sets, whereas the latter notion is preference-free and thus merely captures the existence of an *opportunity* for individual judges to manipulate. Dietrich and List (2007c) provide the following characterization result.

Theorem 8.2 (Dietrich and List, 2007c). *Every resolute judgment aggregation rule satisfying universal domain is nonmanipulable if and only if it is independent and monotonic.*

In particular, for conjunctive and disjunctive agendas (as defined in Footnote 3), conclusion-based judgment aggregation is independent and monotonic

[6] More generally, Dietrich and List (2007c) define these notions on any subset of the agenda, which we will here neglect for simplicity. They also show that monotonicity in Theorem 8.2 can be replaced by a weaker form of monotonicity and the equivalence still holds true.

and therefore, by Theorem 8.2, nonmanipulable, whereas premise-based judgment aggregation rules such as PBP are not independent and thus are manipulable.[7] Dietrich and List (2007c) also provide an impossibility result for a large class of agendas, the so-called *path-connected agendas* that contain the conjunctive and disjunctive agendas: For them, a resolute judgment aggregation rule F satisfies universal domain, always outputs consistent and complete judgment sets, and is responsive[8] and nonmanipulable if and only if F is a dictatorship of some individual judge.[9]

This is the above-mentioned analogue of the Gibbard–Satterthwaite Theorem in judgment aggregation. However, being based on the preference-free concept of nonmanipulability, the above impossibility result does not take the judges' incentives into account. Using a game-theoretic approach, Dietrich and List (2007c) introduced the preference types stated in Definition 8.1 to model different motivations of the individual judges. Specifically, assuming *unrestricted* J-induced (weak) preferences (with J being the judge's desired set, i.e., the outcome that matters for this judge) means that this judge's preferences are not linked to her judgments in any systematic way. On the other hand, *top-respecting*, or even *closeness-respecting* or *Hamming-distance-respecting*, J-induced (weak) preferences model situations where judges want the collective judgments to be close to their own individual desired sets. Now, for each resolute judgment aggregation rule satisfying universal domain, strategy-proofness (as defined earlier) implies that judging truthfully is a weakly dominant strategy for every individual judge in a game-theoretic sense (see, e.g., the book chapter by Faliszewski et al., 2015).

Theorem 8.3 (Dietrich and List, 2007c). *Every resolute judgment aggregation rule satisfying universal domain is strategy-proof for closeness-respecting preferences if and only if it is nonmanipulable.*

From Theorems 8.2 and 8.3, we immediately have that strategy-proofness for closeness-respecting preferences is equivalent to simultaneously requiring independence and monotonicity. Another consequence is that we can replace "nonmanipulable" by "strategy-proof for closeness-respecting preferences" in the impossibility result for path-connected agendas stated above. Note that the implication from left to right in this characterization (i.e., if F satisfies these conditions then it is dictatorial) holds true for any preference type containing the closeness-respecting preferences (e.g., it also holds for top-respecting preferences), as strategy-proofness for this more general preference type implies strategy-proofness for closeness-respecting preferences and thus dictatorship. The other way round, the implication from right to left in this characterization (i.e., if F is dictatorial then it satisfies these conditions) holds for any preference type contained in the top-respecting preferences (e.g., it also holds for closeness-respecting preferences), for otherwise a dictatorship would not be strategy-proof (even though it is nonmanipulable).

[7] However, when we consider an agenda restricted to only the premises, this is simply the majority rule, which is independent and monotonic and thus nonmanipulable.

[8] F is said to be *responsive* if for each *contingent* proposition $\varphi \in \Phi$ (which means that both $\{\varphi\}$ and $\{\overline{\varphi}\}$ are consistent), there are profiles $\boldsymbol{J}$ and $\boldsymbol{J}'$ such that $\varphi \in F(\boldsymbol{J})$ and $\varphi \notin F(\boldsymbol{J}')$.

[9] F is a *dictatorship of judge* i if i always dictates the collective outcome: $F(\boldsymbol{J}) = J_i$ for each $\boldsymbol{J}$.

For an agenda that is conjunctive or disjunctive, two special cases of closeness-respecting preferences are particularly important: outcome- and reason-oriented preferences. A judge with *outcome-oriented preferences* (such as Don in Example 8.4) is not interested in the premises; all he cares about is that the collective judgment on the conclusion matches his own (desired) judgment. A judge with *reason-oriented preferences*, by contrast, is not interested in the collective judgment on the conclusion; all she cares about is that the collective judgments on the premises match her own (desired) judgments, i.e., the reasons in support of the conclusion are what matters for her, more than the conclusion itself. While outcome-oriented preferences are often the better motivational assumption in economics, reason-oriented preferences better fit the arguments made in deliberative settings of democracy. Which kind of preference is appropriate of course depends on the situation and on the subjective goals of the agents involved.

Recall that conclusion-based judgment aggregation is nonmanipulable for conjunctive and disjunctive agendas and therefore, by Theorem 8.3, it is also strategy-proof for these agendas and closeness-respecting preferences, which immediately implies strategy-proofness for outcome- and reason-oriented preferences. However, Dietrich and List (2007c) show that, for a conjunctive or disjunctive agenda, while the premise-based procedure is not strategy-proof for outcome-oriented preferences, it is strategy-proof for reason-oriented preferences. They also show that for outcome-oriented preferences, premise- and conclusion-based judgment aggregation are strategically equivalent in a game-theoretic sense: For both rules and for each profile, there is a (weakly) dominant strategy profile in equilibrium yielding the same collective outcome on the conclusion.

From Theorems 8.2 and 8.3 we know that independence and monotonicity provide a criterion for nonmanipulability and for strategy-proofness for closeness-respecting preferences. However, what about judgment aggregation rules that are not independent or monotonic, such as the premise-based procedure? Can we at least provide some protection for them by showing that the manipulation problem is computationally intractable (i.e., NP-hard)?

Endriss et al. (2010, 2012) were the first to study strategic manipulation of judgment aggregation rules from a computational social choice perspective. They obtained the following result for PBP under Hamming-distance-respecting preferences. First, let us define the corresponding manipulation problem, denoted by PBP-H-MANIPULATION, as follows: Given an agenda Φ, a profile $\boldsymbol{J} = (J_1, \ldots, J_n)$ in $\mathcal{J}(\Phi)^n$, and a manipulator i, does there exist a judgment set $J_i^* \in \mathcal{J}(\Phi)$ such that $H(J_i, PBP(\boldsymbol{J}_{-i}, J_i^*)) < H(J_i, PBP(\boldsymbol{J}))$? That is, is it possible for the manipulator to report an insincere judgment set J_i^* such that the collective outcome under PBP is closer to her truthful judgment set J_i in terms of Hamming distance than the collective outcome under PBP if she had reported J_i itself?

Theorem 8.4 (Endriss et al., 2012). PBP-H-MANIPULATION *is* NP*-complete.*

Baumeister et al. (2013, 2014, 2015a) continued this study and obtained complexity results for manipulation with respect to the class of uniform premise-based quota rules (which, in particular, contains PBP) under unrestricted, top-respecting, closeness-respecting, and Hamming-distance-respecting preferences,

considering not only complete but also incomplete desired sets so as to capture, for instance, outcome- and reason-oriented preferences. These incomplete desired sets are not restricted to the premises or the conclusions, though; all they need to satisfy is that they can be consistently extended to the whole agenda. Preferences are then restricted to the issues occurring in the desired set.

Inspired by the notions, due to Konczak and Lang (2005), of possible and necessary winners from voting theory, Baumeister et al. (2015a) consider the notions of possible and necessary strategy-proofness in judgment aggregation. Noting that *necessary strategy-proofness* captures what Dietrich and List (2007c) call strategy-proofness (as defined earlier in this section), Baumeister et al. (2015a) introduce the other notion by defining a resolute judgment aggregation rule F to be *possibly strategy-proof for unrestricted/top-respecting/closeness-respecting weak preferences* (see Definition 8.1) if for each profile $\boldsymbol{J} = (J_1, \ldots, J_n)$, for each judge i, and for each preference relation induced by i's desired set J_i according to the corresponding preference type, i possibly weakly prefers (as defined after Definition 8.1) the undoctored outcome $F(\boldsymbol{J})$ to the outcome $F(\boldsymbol{J}_{-i}, J_i^*)$ resulting from any misrepresented judgment set J_i^*.[10] Clearly, (necessary) strategy-proofness implies possible strategy-proofness for each of these preference types.

Since $UPQR_q$ is independent and monotonic whenever the agenda contains only premises, it is (necessarily) strategy-proof for closeness-respecting preferences. However, $UPQR_q$ is not strategy-proof in general (and many other judgment aggregation rules aren't either). Therefore, Baumeister et al. (2015a) have studied the computational complexity of the corresponding decision problems. For example, given a preference type T, they define the problem $UPQR_q$-T-POSSIBLE-MANIPULATION as follows: Given an agenda Φ, a profile $\boldsymbol{J} = (J_1, \ldots, J_n) \in \mathcal{J}(\Phi)^n$, and a consistent (not necessarily complete) set $J \subseteq J_i$ desired by manipulator i, is there a judgment set $J_i^* \in \mathcal{J}(\Phi)$ such that i possibly prefers the outcome $UPQR_q(\boldsymbol{J}_{-i}, J_i^*)$ to the undoctored outcome $UPQR_q(\boldsymbol{J})$ under preference type T? $UPQR_q$-T-NECESSARY-MANIPULATION is defined analogously, except that the manipulator necessarily (not only possibly) prefers $UPQR_q(\boldsymbol{J}_{-i}, J_i^*)$ to $UPQR_q(\boldsymbol{J})$ under preference type T (where, for the reasons mentioned in Footnote 10, we omit "POSSIBLE" and "NECESSARY" in the problem name under Hamming-distance-respecting preferences and simply write $UPQR_q$-H-MANIPULATION).

Baumeister et al. (2015a) also define an exact variant of manipulation for uniform premise-based quota rules, denoted by $UPQR_q$-EXACT-MANIPULATION, to model situations where a manipulator wants to achieve not only a better (in terms of the preferences given in Definition 8.1) but a *best* outcome for her desired set (in the sense that everything she desires is actually contained in the collective outcome resulting from the manipulation): Given the same input as above, does there exist a set $J_i^* \in \mathcal{J}(\Phi)$ such that $J \subseteq UPQR_q(\boldsymbol{J}_{-i}, J_i^*)$? They obtained the following results for these problems.

Theorem 8.5 (Baumeister et al., 2015a). *Table 8.2 summarizes the results on the manipulation problems defined above for $UPQR_q$, q rational and $0 \leqslant q < 1$.*

[10] Since just one Hamming-distance-respecting weak preference order is induced by any given desired set, we simply use the term *strategy-proofness* for them, without distinguishing between possible and necessary strategy-proofness.

Preference type	POSSIBLE MANIPULATION with incomplete desired set	NECESSARY MANIPULATION with incomplete desired set	POSSIBLE MANIPULATION with complete desired set	NECESSARY MANIPULATION with complete desired set
Unrestricted	NPC	possibly sp	in P	possibly sp
Top-respecting	NPC	NPC	in P	possibly sp
Closeness-respecting	NPC	NPC	NPC[(a)]	possibly sp
H-respecting	NPC, W[2]-hard[(c)]		NPC[(b)], W[2]-hard[(c)]	
EXACT	NPC		sp	

[(a)] This result is due to Selker (2014).
[(b)] This result is due to Endriss et al. (2012) for the special case of the premise-based procedure.
[(c)] Parameterized by the number of changes in the premises of the manipulator's desired set.

Table 8.2: Results of Baumeister et al. (2015a) on manipulation for $UPQR_q$. Key: NPC means "NP-complete" and sp means "strategy-proof."

Recently, de Haan (2017) studied the Kemeny rule with respect to exact manipulation and manipulation under Hamming-distance-respecting preferences, for both the weighted and the unweighted case. These problems are defined analogously to the corresponding $UPQR_q$ problems above (with an additionally given weight function for the case of the weighted Hamming distance), but—since the Kemeny rule is irresolute—they require that *each* set in the new outcome is preferred to *each* set in the old outcome. The complexity of these problems is stated in the following theorem, which remains valid in the constraint-based framework.

Theorem 8.6 (de Haan, 2017). *$Kemeny$-*EXACT-MANIPULATION*, $Kemeny$-H-*MANIPULATION*, and $Kemeny$-H_ω-*MANIPULATION *are Σ_2^p-complete.*[11]

Having studied manipulation by a single judge so far, a natural question is whether the situation changes when more than one judge tries to manipulate. In voting theory and computational social choice, this is referred to as *coalitional manipulation*, investigated, for instance, by Conitzer et al. (2007) in the context of computational complexity (see also Conitzer and Walsh, 2016). *Group manipulation in judgment aggregation*, introduced by Botan et al. (2016), studies the corresponding setting where a group of judges tries to coordinate a manipulative action in order to improve the result. In their model, preferences over judgments are modeled via the Hamming distance, and the goal is to minimize the sum of the Hamming distances between the manipulators' judgments and the outcome. Whenever no more than two agents try to manipulate, they show that a neutral and, as they call it, "unbiased" aggregation rule is group-strategy-proof if and only if it is independent and monotonic. This does no longer hold for a group of three or more manipulators, though. They also introduce a variant of group manipulation for "fragile coalitions," where manipulators fear that perhaps not all of them will indeed execute the manipulative action.

[11] His results require an integrity constraint even in the formula-based framework. Note that the results for exact manipulation and for manipulation under weighted-Hamming-distance-respecting preferences even hold for a singleton desired set, three judges, and a unidimensionally aligned profile.

Finally, we mention the work of Grossi et al. (2009) who initialize a study of situations where manipulation in judgment aggregation is not considered to be driven by malicious intent but to be "virtuous" and thus desirable: For example, to avoid an unpleasing inconsistent collective outcome, judges may have reason to report less preferred judgment sets, even though they may not be truthful.

8.3.2 Bribery

Example 8.5. *One year earlier, at the night right before the trial against the alleged mafia boss is opened, judge Don (who has been appointed to the jury) secretly meets with a high-rank mafioso in the "Angels from Hell" bar.*

"Tomorrow we will decide which witnesses to summon in the trial, which experts to appoint, and which evidence to allow or deny for the trial," Don explains. "The good news for your boss is that we couldn't find any witness still alive. The bad news is that his knife was found stuck in the victim's body with his fingerprints all over it, and the pathologist, Dr. Slitter, told me that he believes this knife indeed was causing death. That makes your boss the prime suspect."

"Don't worry about Slitter. We've kidnapped his wife and son, he'll testify whatever we want. Make sure he'll be appointed as expert. He'll say machine gun fire killed the sleazebag. And the guy with the gun, y'know, is m... wasn't arrested."

"OK," Don says, "then I do know what to do. I'm on the jury with Bill and Cadi. I'm uncertain about him, but Cadi will for sure deny the knife as evidence if she thinks it wasn't causing death, and so will I, which means we will outvote Bill in any case. And with the knife gone, which is the only piece of evidence linking your boss to the crime, he can look forward to a verdict of not guilty. So you can give me the money: $6000, as we said."

"$6000? That's way too much. Y'know, they've frozen all our bank accounts and I still have to pay some guys' salaries! I give you $4500."

"$6000 was the deal!"

"OK, I give you half of it now," and he passes him an envelope, "and you'll get the other half tomorrow after the trial if everything runs smoothly."

"All right," says Don delighted and takes the money. "Hey!" he suddenly snaps. "That is Canadian *money!" But the felon has already disappeared in the dark of the night.*

Inspired by the work on bribery in voting, due to Faliszewski et al. (2009a,b) and surveyed by Faliszewski and Rothe (2016) and Baumeister and Rothe (2015), Baumeister et al. (2015a, 2011) initiated the study of bribery in judgment aggregation, focusing on $UPQR_q$ and the Hamming distance. In particular, they define the problem $UPQR_q$-BRIBERY as follows: Given an agenda Φ, a profile $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$, a consistent (not necessarily complete) set $J \subseteq J' \in \mathcal{J}(\Phi)$ desired by the briber (an external agent), and a positive integer k (the briber's budget), is it possible to change up to k individual judgment sets in $\boldsymbol{J}$ such that for the resulting new profile $\boldsymbol{J}^*$ we have that $H(UPQR_q(\boldsymbol{J}^*), J) < H(UPQR_q(\boldsymbol{J}), J)$? They also consider a variant called $UPQR_q$-MICROBRIBERY (just as Faliszewski et al., 2009b, do in voting), which is defined analogously, except that the budget k is now a bound on the number of premise entries the briber can change in the given individual judgment sets in $\boldsymbol{J}$. And for both problems, they also consider an exact variant, called

$UPQR_q$-EXACT-BRIBERY and $UPQR_q$-EXACT-MICROBRIBERY, where the question in the problem is not whether the new profile is closer to the desired set according to the Hamming distance, but whether the desired set J is contained in the outcome $UPQR_q(\boldsymbol{J}^*)$ for the modified profile. For the special case of PBP, we have:

Theorem 8.7 (Baumeister et al., 2015a). PBP-BRIBERY, PBP-MICROBRIBERY, PBP-EXACT-BRIBERY, *and* PBP-EXACT-MICROBRIBERY *are* NP*-complete.*

Baumeister et al. (2015a) also study this problems in terms of parameterized complexity, showing that PBP-EXACT-BRIBERY is W[2]-hard when parameterized by the number of bribes. More generally, for an in-depth treatise of the complexity of both bribery and microbribery under top- and closeness-respecting preferences for $UPQR_q$, we refer to the work of Baumeister et al. (2015c).

Further, de Haan (2017) defines the corresponding bribery and exact bribery problems for the Kemeny procedure, again for weighted and unweighted Hamming distances and with the additional requirement that *each* set in the new outcome is preferred to *each* set in the old one. In the weighted case, the input additionally contains a weight function.

The following result also holds in the constraint-based framework.

Theorem 8.8 (de Haan, 2017). $Kemeny$-EXACT-BRIBERY *and* $Kemeny$-BRIBERY *(weighted and unweighted) are* Σ_2^p*-complete.*[12]

A field closely related to judgment aggregation is that of lobbying in multiple referenda introduced by Christian et al. (2007). This problems corresponds to a judgment aggregation problem where we have only logically unconnected premises in the agenda and some external agent tries to influence some voters in order to reach a desired outcome when evaluated according to the majority rule. Hence, this problem is very closely related to the bribery problems described in this section. Such lobbying problems (and generalizations thereof) have also been studied by Bredereck et al. (2014) and Binkele-Raible et al. (2014). Also related, though in a somewhat different context, is the work by Alon et al. (2015). In their setting, voters support or reject proposals. A ballot is *accepted* by a voter if she supports at least half of the proposals in it. The task is then to find a vote that is accepted either by all voters or by a majority of them.

8.3.3 Control

Example 8.6. *As Example 8.2 shows, Don will be found guilty if the judges use* $UPQR_{1/2}$ *to aggregate their judgments. However, this would be detrimental to chief judge Zoe (responsible, in particular, for appointing judges to cases at this court) and her secret lover, the alleged mafia boss, because it would lead to further investigations into the matter and would most certainly result in a retrial. Therefore, Zoe decides to publicly question the authority of the three appointed judges.*

[12] Again, his results require an integrity constraint even in the formula-based framework, and the results for $Kemeny$-EXACT-BRIBERY and $Kemeny$-BRIBERY in the weighted case even hold for a singleton desired set, three judges, and a unidimensionally aligned profile.

"This a matter of utmost importance," Zoe exclaims. "I rule that some of our most trusted judges should assist in forming a decision, since three judges are certainly not enough to handle this intricate and delicate problem!"

Zoe needs to achieve an addition of at least two judges that both agree on $\neg a$ or $\neg e$ so that at least one of a and e does no longer exceed the quota. Fortunately for her, possible candidates for these additional judges are Elena with judgment set $J_E = \{\neg a, \neg e, \neg g\}$, Felix with $J_F = \{a, e, g\}$, and George with $J_G = \{\neg a, e, \neg g\}$, so she appoints Elena and George as additional judges, saving both her lover and Don.

It may also happen that Don will be found guilty when using the Kemeny rule (see Example 8.2). To ensure that the non-guilty verdict for her lover remains valid, Zoe has a backup plan up her sleeve for this case. If she succeeds in dropping the question of whether the amount of money taken by Don was considerable, then the new collective outcome (given the original profile) only consists of $J'_B = \{e, \neg g\}$ with a Hamming distance of 2 as opposed to the three disagreements of the other possible judgment sets (all restricted to the new agenda).

As this example shows, another possible influence on the judgment aggregation process is *control* where an external agent, commonly called the *chair* (or *"chief judge"* in the above example), is able to change the structure of the process. As surveyed by Faliszewski and Rothe (2016) and Baumeister and Rothe (2015), control in elections has been studied since the seminal paper by Bartholdi III et al. (1992) and has produced a vast number of results. Inspired by this work, Baumeister et al. (2012a,b) introduced this type of strategic behavior to judgment aggregation. In their scenarios, they focus on the judges: The chair is able (1) to add a certain number of judges to the given profile from another given profile, (2) to delete a certain number of judges, (3) to replace judges (which combines adding and deleting judges by the chair first deleting a number of judges and then adding the same number of judges from another given profile), or (4) to bundle the judges into groups so that every group of judges only decides over their own subset of issues in a partition of the agenda.[13] Here, we focus on only the first three control actions. For the resolute rules $UPQR_q$, *possible/necessary control by adding/deleting/replacing judges* asks,[14] given the chair's desired set, whether the chair possibly/necessarily prefers the new outcome to the old one under a certain preference type, and *exact control by adding/deleting/replacing judges* asks whether the desired set is a subset of the new outcome.

Theorem 8.9 (Baumeister et al., 2015c). *Possible and necessary control by adding and by deleting judges is* NP*-complete for $UPQR_{1/2}$ under closeness-respecting preferences, even when the desired set is complete.*

Theorem 8.10 (Baumeister et al., 2012a). *Exact control by adding and by deleting judges and control by adding and by deleting judges under Hamming-distance-respecting preferences are* NP*-complete for $UPQR_{1/2}$.*

[13] A variant of control by bundling judges is studied by Alon et al. (2013), yet with the bundling occurring on the issues. When several issues are bundled together, the judges have to decide whether to accept or reject the whole bundle. They show that the problems related to such bundling attacks are computationally hard when simple majority is used to aggregate the individual judgments.

[14] Since possible and necessary preferences coincide for the Hamming distance, we only consider *control by adding/deleting/replacing judges* for this preference type, dropping "*possible/necessary.*"

Baumeister et al. (2015c, 2012a) further show that Theorems 8.9 and 8.10 also hold for control by replacing judges, even for a rational quota q, $0 \leqslant q < 1$.

Recently, de Haan (2017) introduced two control scenarios that focus on the issues: The chair can change the agenda by adding an arbitrary number of issues to the agenda, or by deleting an arbitrary number of issues. He then asks whether the given desired set is contained in each set of the resulting outcome, i.e., he is interested in the exact control variant.

Note that the following result also holds for the constraint-based framework.

Theorem 8.11 (de Haan, 2017). *Exact control by adding and by deleting issues is Σ_2^p-complete for $Kemeny$.*[15]

Dietrich (2016) also studies how to influence the outcome via the agenda. An agenda is said to be *sensitive* if the collective outcome depends on the choice of propositions that are being aggregated. Three types of agenda-insensitivity are introduced and characterized axiomatically, along with an impossibility theorem.

8.4 Conclusions and Outlook

We have surveyed strategic behavior in judgment aggregation, focusing on both axiomatic characterizations and computational complexity and distinguishing the same strategic scenarios that are well known and have been extensively studied in computational social choice: manipulation, bribery, and control. While preference and judgment aggregation have a lot in common (specifically, the *preferences* of either voters or judges), there are also crucial differences where judgment aggregation parts company from voting—for instance, due to logical constraints on and dependencies between judgments and due to the need to compactly represent the judges' preferences. Still, we suspect that computational social choice will severely keep influencing the field of judgment aggregation and will continue to shape the future of this field. One issue that we consider particularly important for future work is to model judgment aggregation as a dynamic process over time. After all, judgment aggregation often *is* a dynamically evolving process, with new judges arriving and others departing or with the outcome heavily depending on the order in which propositions are considered.

While sequential variants of judgment aggregation rules have already been studied (e.g., by Dietrich and List, 2007b, as mentioned in Section 8.2.3), it would be very interesting to adapt other approaches from computational social choice to dynamic settings in judgment aggregation, such as the work by Tennenholtz (2004) on dynamic voting, the Stackelberg voting games studied by Desmedt and Elkind (2010) and Xia and Conitzer (2010), and the work by Parkes and Procaccia (2013) who use Markov decision processes to model evolving preferences. And modeling *strategic behavior* dynamically in judgment aggregation can also be inspired by the work of Hemaspaandra et al. (2014, 2012a, 2017a, 2012b, 2017b) on online manipulation, online candidate control, and online voter control in sequential elections.

[15] This theorem uses the formula-based framework with an additional integrity constraint.

Acknowledgments

This work was supported in part by the NRW Ministry for Innovation, Science, and Research, DFG grant RO 1202/15-1, and an SFF grant of HHU Düsseldorf.

Bibliography

N. Alon, D. Falik, R. Meir, and M. Tennenholtz. Bundling attacks in judgment aggregation. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, pages 39–45. AAAI Press, July 2013.

N. Alon, R. Bredereck, J. Chen, S. Kratsch, R. Niedermeier, and G. Woeginger. How to put through your agenda in collective binary decisions. *ACM Transactions on Economics and Computation*, 4(1):Article 5, 2015.

K. Arrow. *Social Choice and Individual Values*. John Wiley and Sons, 1951—revised 1963.

J. Bartholdi III, C. Tovey, and M. Trick. How hard is it to control an election? *Mathematical and Computer Modelling*, 16(8/9):27–40, 1992.

D. Baumeister and J. Rothe. Preference aggregation by voting. In J. Rothe, editor, *Economics and Computation. An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*, Springer Texts in Business and Economics, chapter 4, pages 197–325. Springer-Verlag, 2015.

D. Baumeister, G. Erdélyi, and J. Rothe. How hard is it to bribe the judges? A study of the complexity of bribery in judgment aggregation. In *Proceedings of the 2nd International Conference on Algorithmic Decision Theory (ADT)*, pages 1–15. Springer-Verlag *Lecture Notes in Artificial Intelligence #6992*, October 2011.

D. Baumeister, G. Erdélyi, O. Erdélyi, and J. Rothe. Bribery and control in judgment aggregation. In F. Brandt and P. Faliszewski, editors, *Proceedings of the 4th International Workshop on Computational Social Choice*, pages 37–48. AGH University of Science and Technology, Kraków, Poland, September 2012a.

D. Baumeister, G. Erdélyi, O. Erdélyi, and J. Rothe. Control in judgment aggregation. In *Proceedings of the 6th European Starting AI Researcher Symposium*, pages 23–34. IOS Press, August 2012b.

D. Baumeister, G. Erdélyi, O. Erdélyi, and J. Rothe. Computational aspects of manipulation and control in judgment aggregation. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT)*, pages 71–85. Springer-Verlag *Lecture Notes in Artificial Intelligence #8176*, November 2013.

D. Baumeister, G. Erdélyi, O. Erdélyi, and J. Rothe. Computational aspects of manipulation and control in judgment aggregation. In A. D. Procaccia and T. Walsh, editors, *Proceedings of the 5th International Workshop on Computational Social Choice*, Pittsburgh, USA, June 2014. Carnegie Mellon University.

D. Baumeister, G. Erdélyi, O. Erdélyi, and J. Rothe. Complexity of manipulation and bribery in judgment aggregation for uniform premise-based quota rules. *Mathematical Social Sciences*, 76:19–30, 2015a.

D. Baumeister, G. Erdélyi, and J. Rothe. Judgment aggregation. In J. Rothe, editor, *Economics and Computation. An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*, Springer Texts in Business and Economics, chapter 6, pages 361–391. Springer-Verlag, 2015b.

D. Baumeister, J. Rothe, and A. Selker. Complexity of bribery and control for uniform premise-based quota rules under various preference types. In *Proceedings of the 4th International Conference on Algorithmic Decision Theory (ADT)*, pages 432–448. Springer-Verlag *Lecture Notes in Artificial Intelligence #9346*, September 2015c.

D. Binkele-Raible, G. Erdélyi, H. Fernau, J. Goldsmith, N. Mattei, and J. Rothe. The complexity of probabilistic lobbying. *Discrete Optimization*, 11(1):1–21, 2014.

S. Botan, A. Novaro, and U. Endriss. Group manipulation in judgment aggregation. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 411–419. IFAAMAS, May 2016.

S. Bouveret, Y. Chevaleyre, and N. Maudet. Fair allocation of indivisible goods. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 12, pages 284–310. Cambridge University Press, 2016.

R. Bredereck, J. Chen, S. Hartung, S. Kratsch, R. Niedermeier, O. Suchý, and G. Woeginger. A multivariate complexity analysis of lobbying in multiple referenda. *Journal of Artificial Intelligence Research*, 50:409–446, 2014.

R. Christian, M. Fellows, F. Rosamond, and A. Slinko. On complexity of lobbying in multiple referenda. *Review of Economic Design*, 11(3):217–224, 2007.

V. Conitzer and T. Walsh. Barriers to manipulation in voting. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 6, pages 127–145. Cambridge University Press, 2016.

V. Conitzer, T. Sandholm, and J. Lang. When are elections with few candidates hard to manipulate? *Journal of the ACM*, 54(3):Article 14, 2007.

R. de Haan. Parameterized complexity results for the Kemeny rule in judgment aggregation. In U. Grandi and J. Rosenschein, editors, *Proceedings of the 6th International Workshop on Computational Social Choice*, Toulouse, France, June 2016a.

R. de Haan. Parameterized complexity results for the Kemeny rule in judgment aggregation. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI)*, pages 1502–1510. IOS Press, August/September 2016b.

R. de Haan. Complexity results for manipulation, bribery and control of the Kemeny judgment aggregation procedure. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, May 2017. To appear.

Y. Desmedt and E. Elkind. Equilibria of plurality voting with abstentions. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC)*, pages 347–356. ACM Press, June 2010.

F. Dietrich. A generalised model of judgment aggregation. *Social Choice and Welfare*, 28(4):529–565, 2007.

F. Dietrich. Scoring rules for judgment aggregation. *Social Choice and Welfare*, 42(4):873–911, 2014.

F. Dietrich. Judgment aggregation and agenda manipulation. *Games and Economic Behavior*, 95:113–136, 2016.

F. Dietrich and C. List. Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, 29(1):19–33, 2007a.

F. Dietrich and C. List. Judgment aggregation by quota rules: Majority voting generalized. *Journal of Theoretical Politics*, 19(4):391–424, 2007b.

F. Dietrich and C. List. Strategy-proof judgment aggregation. *Economics and Philosophy*, 23(3):269–300, 2007c.

R. Downey and M. Fellows. *Parameterized Complexity*. Springer-Verlag, 2nd edition, 2013.

U. Endriss. Judgment aggregation. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 17, pages 399–426. Cambridge University Press, 2016.

U. Endriss and R. de Haan. Complexity of the winner determination problem in judgment aggregation: Kemeny, Slater, Tideman, Young. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 117–125. IFAAMAS, May 2015.

U. Endriss, U. Grandi, and D. Porello. Complexity of winner determination and strategic manipulation in judgment aggregation. In V. Conitzer and J. Rothe, editors, *Proceedings of the 3rd International Workshop on Computational Social Choice*, pages 139–150. Universität Düsseldorf, Düsseldorf, Germany, September 2010.

U. Endriss, U. Grandi, and D. Porello. Complexity of judgment aggregation. *Journal of Artificial Intelligence Research*, 45:481–514, 2012.

U. Endriss, U. Grandi, R. de Haan, and J. Lang. Succinctness of languages for judgment aggregation. In *Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 176–185. AAAI Press, April 2016.

P. Faliszewski and J. Rothe. Control and bribery in voting. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 7, pages 146–168. Cambridge University Press, 2016.

P. Faliszewski, E. Hemaspaandra, and L. Hemaspaandra. How hard is bribery in elections? *Journal of Artificial Intelligence Research*, 35:485–532, 2009a.

P. Faliszewski, E. Hemaspaandra, L. Hemaspaandra, and J. Rothe. Llull and Copeland voting computationally resist bribery and constructive control. *Journal of Artificial Intelligence Research*, 35:275–341, 2009b.

P. Faliszewski, I. Rothe, and J. Rothe. Noncooperative game theory. In J. Rothe, editor, *Economics and Computation. An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*, Springer Texts in Business and Economics, chapter 2, pages 41–134. Springer-Verlag, 2015.

A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41(4):587–601, 1973.

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

D. Grossi, G. Pigozzi, and M. Slavkovik. White manipulation in judgment aggregation. In *Proceedings of the 21st Benelux Conference on Artificial Intelligence (BNAIC)*, October 2009.

E. Hemaspaandra, L. Hemaspaandra, and J. Rothe. Controlling candidate-sequential elections. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI)*, pages 905–906. IOS Press, Aug. 2012a.

E. Hemaspaandra, L. Hemaspaandra, and J. Rothe. Online voter control in sequential elections. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI)*, pages 396–401. IOS Press, Aug. 2012b.

E. Hemaspaandra, L. Hemaspaandra, and J. Rothe. The complexity of online manipulation of sequential elections. *Journal of Computer and System Sciences*, 80(4):697–710, 2014.

E. Hemaspaandra, L. Hemaspaandra, and J. Rothe. The complexity of controlling candidate-sequential elections. *Theoretical Computer Science*, 2017a. To appear. DOI: 10.1016/j.tcs.2017.03.037.

E. Hemaspaandra, L. Hemaspaandra, and J. Rothe. The complexity of online voter control in sequential elections. *Journal of Autonomous Agents and Multi-Agent Systems*, 2017b. To appear. DOI: 10.1007/s10458-016-9349-1.

J. Kemeny. Mathematics without numbers. *Dædalus*, 88:571–591, 1959.

K. Konczak and J. Lang. Voting procedures with incomplete preferences. In *Proceedings of the Multidisciplinary IJCAI-05 Workshop on Advances in Preference Handling*, pages 124–129, July/August 2005.

L. Kornhauser and L. Sager. Unpacking the court. *Yale Law Journal*, 96(1): 82–117, 1986.

J. Lang and J. Rothe. Fair division of indivisible goods. In J. Rothe, editor, *Economics and Computation. An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*, Springer Texts in Business and Economics, chapter 8, pages 493–550. Springer-Verlag, 2015.

J. Lang and M. Slavkovik. How hard is it to compute majority-preserving judgment aggregation rules? In *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI)*, pages 501–506. IOS Press, 2014.

J. Lang, G. Pigozzi, M. Slavkovik, and L. van der Torre. Judgment aggregation rules based on minimization. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, pages 238–246. ACM Press, July 2011.

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

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

C. List. The theory of judgment aggregation: An introductory review. *Synthese*, 187(1):179–207, 2012.

C. List and P. Pettit. Aggregating sets of judgments: An impossibility result. *Economics and Philosophy*, 18(1):89–110, 2002.

C. List and C. Puppe. Judgment aggregation. In P. Anand, P. Pattanaik, and C. Puppe, editors, *The Handbook of Rational and Social Choice*, chapter 19, pages 457–482. Oxford University Press, 2009.

M. Miller and D. Osherson. Methods for distance-based judgment aggregation. *Social Choice and Welfare*, 32(4):575–601, 2009.

P. Mongin. The doctrinal paradox, the discursive dilemma, and logical aggregation theory. *Theory and Decision*, 73(3):315–355, 2012.

H. Moulin. *Fair Division and Collective Welfare*. MIT Press, 2004.

K. Nehring, M. Pivato, and C. Puppe. Condorcet admissibility: Indeterminacy and path-dependence under majority voting on interconnected decisions. Technical Report MPRA 32434, University of Munich, 2011.

D. Parkes and A. D. Procaccia. Dynamic social choice with evolving preferences. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, pages 767–773. AAAI Press, July 2013.

P. Pettit. Deliberative democracy and the discursive dilemma. *Philosophical Issues*, 11(1):268–299, 2001.

G. Pigozzi. Belief merging and the discursive dilemma: An argument-based account to paradoxes of judgment. *Synthese*, 152(2):285–298, 2006.

I. Rahwan and G. Simari, editors. *Argumentation in Artificial Intelligence*. Springer, 2009.

J. Rothe. *Complexity Theory and Cryptology. An Introduction to Cryptocomplexity*. EATCS Texts in Theoretical Computer Science. Springer-Verlag, 2005.

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-Verlag, 2015.

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

A. Selker. Manipulative Angriffe auf Judgment-Aggregation-Prozeduren. Master's thesis, Heinrich-Heine-Universität Düsseldorf, Institut für Informatik, Düsseldorf, Germany, September 2014.

M. Tennenholtz. Transitive voting. In *Proceedings of the 5th ACM Conference on Electronic Commerce (EC)*, pages 230–231. ACM Press, July 2004.

L. Xia and V. Conitzer. Stackelberg voting games: Computational aspects and paradoxes. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, pages 697–702. AAAI Press, July 2010.

W. S. Zwicker. Introduction to the theory of voting. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 2, pages 23–56. Cambridge University Press, 2016.