

Handbook of Computational Social Choice

Edited by F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A.D. Procaccia

To appear as: U. Endriss. Judgment Aggregation. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A.D. Procaccia (eds.), *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

Contents

| | |
|--------------------------------------------------------------|-------------|
| 17 Judgment Aggregation | <i>page</i> |
| 17.1 Introduction | 1 |
| 17.2 Basics | 4 |
| 17.2.1 Formal Framework | 4 |
| 17.2.2 An Example: Simulating the Condorcet Paradox | 5 |
| 17.2.3 Axioms: Desirable Properties of Aggregation Rules | 6 |
| 17.2.4 A Simple Impossibility Theorem | 9 |
| 17.3 Aggregation Rules | 10 |
| 17.3.1 Quota-based Aggregation / Axiomatic Characterizations | 10 |
| 17.3.2 Distance-based Aggregation / Winner Determination | 12 |
| 17.3.3 Premise-based Aggregation / Strategic Manipulation | 14 |
| 17.4 Agenda Characterization Results | 17 |
| 17.4.1 Consistent Aggregation under the Majority Rule | 17 |
| 17.4.2 Existential Agenda Characterization | 18 |
| 17.4.3 Universal Agenda Characterization | 23 |
| 17.5 Related Frameworks | 25 |
| 17.5.1 Belief Merging | 25 |
| 17.5.2 Binary Aggregation | 26 |
| 17.5.3 Voting in Combinatorial Domains | 26 |
| 17.6 Applications in Computer Science | 27 |
| 17.7 Bibliographic Notes and Further Reading | 28 |
| <i>References</i> | 30 |

17

Judgment Aggregation

Ulle Endriss

17.1 Introduction

Social choice theory deals with the aggregation of information supplied by several individuals into a collective decision that appropriately reflects the views of the group as a whole. The most widely considered type of information is preference information. For example, in an election each voter supplies information regarding her preferences over the set of candidates and the voting rule in operation aggregates this information into the collective decision of which candidate wins the election. But the methodology of social choice theory may just as well be applied to other types of information, such as beliefs about whether certain statements are true or false. *Judgment aggregation* (JA from here on), the topic of this chapter, is an elegant formal framework for modeling this form of aggregation.

Let us begin with a famous example from the work of legal scholars Lewis A. Kornhauser and Lawrence G. Sager (Kornhauser and Sager, 1993). Suppose three judges together have to decide on a case regarding an alleged breach of contract. They will try to establish whether (a) the document in question really is a binding contract and whether (b) the promise given in that document really has been breached. Legal doctrine stipulates that the defendant is liable if and only if there have been both a contract and a breach. The three judges differ in their assessment regarding the two premises (and thus also regarding the conclusion):

| | <i>Contract?</i> | <i>Breach?</i> | <i>Liable?</i> |
|--------------|------------------|----------------|----------------|
| Judge Joe: | Yes | Yes | Yes |
| Judge Judy: | Yes | No | No |
| Judge Jules: | No | Yes | No |

What should be their collective decision regarding the defendant's liability? If they implement a majority vote on the conclusion (the rightmost column above), then they will find the defendant *not liable* (by a 2:1 majority). If instead they vote on the premises, they will have to accept that the contract was binding (by a 2:1 majority) and that it has been breached (again, by a 2:1 majority). In the latter case, legal

doctrine requires them to find the defendant *liable*. Thus, in the presence of this doctrine, two seemingly reasonable procedures lead to contradictory outcomes. This is known as the *doctrinal paradox*.

Now consider a more abstract rendering of the problem. Three judges have to assess the truth of three formulas of propositional logic: p ('contract'), q ('breach'), and their conjunction $p \wedge q$ (which we said was equivalent to 'liable'). This time we also include the result of applying majority voting to each formula:

| | p | q | $p \wedge q$ |
|-----------|-----|-----|--------------|
| Judge 1: | Yes | Yes | Yes |
| Judge 2: | Yes | No | No |
| Judge 3: | No | Yes | No |
| Majority: | Yes | Yes | No |

That is, we again obtain an unexpected, i.e., paradoxical, outcome: despite the fact that each individual judge provides a logically consistent set of judgments, the majority rule results in a judgment set that is inconsistent (there exists no assignment of truth values to propositional variables that would make p true, q true, and $p \wedge q$ false). So, our example demonstrates not merely a dilemma between premise-driven and conclusion-driven approaches to collective decision making, but rather a dilemma between a certain *responsiveness* to the views of decision makers (by respecting their majority decisions) and the *consistency* of collective decisions. This point was first made by the political philosopher Philip Pettit, who noted that this dilemma is not only relevant to analytical jurisprudence, but may strike whenever a group of people engage in a democratic decision making process involving several mutually dependent propositions (Pettit, 2001). Pettit introduced the term *discursive dilemma* for this problem, both to stress its relevance to the political discourse in general and to reflect the fact that we do not actually require the external legal doctrine from the original example to exhibit the problem.

Is there a way around this dilemma? Maybe the majority rule is the root of the problem and there are other methods of aggregation that can ensure a consistent outcome? In a seminal paper that introduced a formal framework for JA that permits us to ask and answer such questions, Christian List and Philip Pettit showed that this is not the case (List and Pettit, 2002): it is *impossible* to devise an aggregation rule that avoids the discursive dilemma—at least if we wish to maintain some of the most basic properties of the majority rule that, arguably, are fundamental features of any reasonable form of aggregation. These properties are *anonymity* ('treat all judges symmetrically'), *neutrality* ('treat all propositions symmetrically'), and *independence* ('base the collective decision regarding a given proposition only on the individual judgments regarding that same proposition'). We will formally state and prove this surprising result in Section 17.2.4.

The work of List and Pettit employs the *axiomatic method* commonly used in

economic theory, and specifically in social choice theory, e.g., to establish impossibility results in preference aggregation. This—together with the fact that JA is a natural framework in which to embed other frameworks of aggregation, specifically preference aggregation (see Section 17.2.2)—has triggered a sustained interest in JA among economic theorists. Their work has led to a deeper understanding of the circumstances under which it is either possible or impossible to perform consistent aggregation. In particular, these results clarify the role of the *agenda*, the set of propositions to be judged. For instance, it is easy to see that when the agenda consists solely of literals (i.e., propositional variables and their negations), then the majority rule will never produce an inconsistent outcome. We will review some representative examples of such results later on in Section 17.4.

Besides analytical jurisprudence, political philosophy, and economic theory, JA is also relevant to computer science, particularly to artificial intelligence (AI). For instance, JA suggests itself as a framework in which to study collective decision making in systems of autonomous software agents, given that logic is the preferred language in which to model the beliefs of a single such agent (we will discuss this and other applications of JA in computer science in Section 17.6). In fact, there are close connections between some of the work on *belief merging* in AI and the model of JA under consideration here (we will briefly comment on some of these connections in Section 17.5.1). Once computer scientists got interested in JA, this naturally led to a view of aggregation procedures as *algorithms* and, more generally, of reasoning about questions in JA as *computational problems*. We will adopt this perspective also in parts of this chapter and report, for instance, on the computational difficulty of recognizing whether an agenda is sufficiently simple to avoid all occurrences of the discursive dilemma for a given aggregation procedure.

The remainder of this chapter is organized as follows. Section 17.2 defines the *formal framework* of JA and, to exemplify the expressive power of the framework, shows how *preference aggregation* can be embedded into JA. The same section also introduces the most commonly used *axioms* in JA and then proves the basic *impossibility theorem* of List and Pettit mentioned earlier. Section 17.3 reviews three specific types of aggregators in some detail: *quota rules*, *distance-based aggregators*, and the *premise-based rule*. In the context of quota rules, we also discuss examples of the *axiomatic characterization* of aggregators; we use distance-based aggregation to exemplify the analysis of the complexity of *winner determination* in JA; and we review questions of *strategic manipulation* in the section dedicated to premise-based aggregation. Section 17.4 is devoted to *agenda characterization* results that clarify the extent to which instances of the discursive dilemma depend on the structural complexity of the agenda on which judges are asked to vote. Section 17.5 discusses related frameworks for collective decision making and Section 17.6 sketches possible applications in computer science. To improve readability, bibliographic references are kept to a minimum in the body of the chapter; such details are instead supplied in Section 17.7, which also provides pointers to further reading.

Throughout this chapter we shall assume familiarity with the very basics of propositional logic (see, e.g., van Dalen, 2013), particularly the notion of logical consistency. In a few selected places we furthermore assume familiarity with basic concepts from the theory of computational complexity (see, e.g., Arora and Barak, 2009, or the introductory chapter of this *Handbook*).

17.2 Basics

In this section we define the formal framework of JA, which originally was laid down by List and Pettit (2002) and since then has been further refined by several authors, notably Dietrich (2007).¹ We also sketch how to embed preference aggregation problems into JA, review the most important axioms encoding desirable properties of aggregators proposed in the literature, and discuss a basic impossibility theorem.

17.2.1 Formal Framework

Let \mathcal{L} be a set of propositional formulas built from a finite set of propositional variables using the usual connectives $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$, and the constants \top ('true') and \perp ('false'). For every formula φ , define $\sim\varphi$ to be the *complement* of φ , i.e., $\sim\varphi = \neg\varphi$ if φ is not negated, and $\sim\varphi = \psi$ if $\varphi = \neg\psi$ for some formula ψ . We write $\Delta \models \varphi$ in case formula φ is true whenever all formulas in the set Δ are true.

An *agenda* is a finite nonempty subset $\Phi \subseteq \mathcal{L}$ that does not contain any doubly-negated formulas and that is closed under complementation (i.e., if $\varphi \in \Phi$ then $\sim\varphi \in \Phi$).² For ease of exposition, we shall assume that Φ is *nontrivial* in the sense of including (at least) two logically independent formulas α and β (i.e., all of $\{\alpha, \beta\}$, $\{\alpha, \sim\beta\}$, $\{\sim\alpha, \beta\}$, and $\{\sim\alpha, \sim\beta\}$ are consistent).³ Some authors exclude the possibility of Φ including a tautology or a contradiction, but we do not make this assumption here. A *judgment set* J for Φ is a subset $J \subseteq \Phi$. For example, the discursive dilemma sketched in the introduction involved the agenda $\Phi = \{p, \neg p, q, \neg q, p \wedge q, \neg(p \wedge q)\}$. The judgment set of judge 3 was $J_3 = \{\neg p, q, \neg(p \wedge q)\}$, i.e., rather than labeling formulas φ with 'yes' and 'no', as we did earlier, we now either include φ or $\sim\varphi$ in the relevant judgment set.

We call a judgment set J *complete* if $\varphi \in J$ or $\sim\varphi \in J$ for all $\varphi \in \Phi$; we call it *complement-free* if for no $\varphi \in \Phi$ we have both $\varphi \in J$ and $\sim\varphi \in J$; and we call it *consistent* if there exists an assignment of truth values (*true* or *false*) to propositional variables under which all formulas in J are true. Note that every consistent set is complement-free, but the converse is not true. For instance, the

¹ The particular mode of exposition chosen here is closely based on the author's joint work with Umberto Grandi and Daniele Porello (Endriss et al., 2012).

² The reason for introducing the notion of complement is that it often simplifies presentation. For example, if $\varphi \in \Phi$, then the '*negation* of φ ' will only be in Φ if φ is not a negated formula, while we can speak of the '*complement* of φ ' without having to take any such precautions.

³ For example, this nontriviality condition is satisfied if α and β are propositional variables.

majority judgment set from our introductory example, $\{p, q, \neg(p \wedge q)\}$, is clearly complete and complement-free, but it is not consistent: if we set both p and q to be true, then the third formula $\neg(p \wedge q)$ necessarily comes out as false. Let $\mathcal{J}(\Phi)$ denote the set of all complete and consistent subsets of Φ .

Let $N = \{1, \dots, n\}$ be a set of $n > 1$ *judges* (or *individuals*, or *agents*). For ease of exposition, we shall assume that n is odd. We often refer to subsets $C \subseteq N$ as *coalitions* and we write $\bar{C} := N \setminus C$ for the complement of C . A *profile* is a vector of judgment sets $\mathbf{J} = (J_1, \dots, J_n) \in \mathcal{J}(\Phi)^n$, one for each judge. We write $N_\varphi^{\mathbf{J}} := \{i \in N \mid \varphi \in J_i\}$ for the set of judges accepting the formula φ in profile \mathbf{J} .

A (resolute) *judgment aggregation rule* (or *aggregator* for short) for agenda Φ and judges $N = \{1, \dots, n\}$ is a function $f : \mathcal{J}(\Phi)^n \rightarrow 2^\Phi$ mapping every profile into a single (collective) judgment set (2^Φ denotes the powerset of Φ).⁴ Note that the resulting judgment set need not be complete and consistent, but the individual sets in the profile are always assumed to have these properties. An example of an aggregator is the (strict) *majority rule* $f_{\text{maj}} : \mathbf{J} \mapsto \{\varphi \in \Phi \mid |N_\varphi^{\mathbf{J}}| > \frac{n}{2}\}$, which accepts a formula φ if more than half of the individual judges do.

We conclude this review of the formal framework with two technical definitions. First, we occasionally require a means of measuring how dissimilar two judgment sets are. The *Hamming distance* $H(J, J')$ of two complete and complement-free judgment sets J and J' is the number of non-negated agenda formulas on which they differ. That is, $H(J, J') := |J \setminus J'| = |J' \setminus J|$. Second, to better understand the sources of inconsistency in aggregation we require a means of abstracting away from formulas that do not contribute to an observed inconsistency. We call an inconsistent set X *minimally inconsistent* if every proper subset of X is consistent.

17.2.2 An Example: Simulating the Condorcet Paradox

To demonstrate the versatility of JA, let us briefly sketch how we can use it to simulate the standard framework of preference aggregation. Suppose three *voters* each express a strict preference order (i.e., a complete, antisymmetric, and transitive binary relation) over a set of *alternatives* $A = \{a, b, c\}$:

| | |
|--------|---------------------|
| Ann: | $a \succ b \succ c$ |
| Bob: | $c \succ a \succ b$ |
| Chloé: | $b \succ c \succ a$ |

If we try to aggregate these individual preferences using the majority rule (now in the sense of accepting $x \succ y$ if at least two of the three individuals do), then we obtain a cycle: $a \succ b \succ c \succ a$ (two out of three voters prefer a over b , and so forth). This is the classical *Condorcet paradox* (McLean and Urken, 1995).

⁴ *Irresolute* aggregators, which allow for ties between several collective judgment sets in the outcome, while also of some interest, are studied less frequently (but see, e.g., Lang and Slavkovik, 2013).

Now construct a JA scenario as follows. Let \mathcal{L} be the propositional language built from the propositional variables $\{p_{a>b}, p_{a>c}, p_{b>a}, p_{b>c}, p_{c>a}, p_{c>b}\}$. Let Φ consist of all literals in \mathcal{L} as well as these formulas (and their complements):

- $p_{a>b} \leftrightarrow \neg p_{b>a}$, $p_{a>c} \leftrightarrow \neg p_{c>a}$, $p_{b>c} \leftrightarrow \neg p_{c>b}$
- $p_{a>b} \wedge p_{b>c} \rightarrow p_{a>c}$, and similarly for all other permutations of a, b, c

Observe how, when we interpret $p_{a>b}$ as ‘I consider a being preferable over b ’ and so forth, we get the first group of formulas above to encode the fact that our preference order should be complete and antisymmetric, while the second group expresses that it should be transitive. Let us call Φ the *preference agenda* (for three alternatives). Now consider a consistent and complete profile for judges Ann, Bob, and Chloé, in which all three of them accept the positive formulas listed above. Such a profile is fully determined once we fix each judge’s stance on $p_{a>b}$, $p_{b>c}$, and $p_{a>c}$:

| | $p_{a>b}$ | $p_{b>c}$ | $p_{a>c}$ |
|-----------|-----------|-----------|-----------|
| Ann: | Yes | Yes | Yes |
| Bob: | Yes | No | No |
| Chloé: | No | Yes | No |
| Majority: | Yes | Yes | No |

This corresponds directly to the Condorcet paradox. If we translate back to preference aggregation, then $\{p_{a>b}, p_{b>c}, \neg p_{a>c}, \dots\}$, which is part of the judgment set returned by the majority rule, corresponds to the cycle $a > b > c > a$. If we stay in JA, then that same set becomes inconsistent once we add $p_{a>b} \wedge p_{b>c} \rightarrow p_{a>c}$ (which must also be part of the majority outcome, as it is accepted by all judges).

17.2.3 Axioms: Desirable Properties of Aggregation Rules

We have seen that using the majority rule can lead to problems. It does not meet all of our requirements. But what are those requirements? Below is a list of properties of aggregators that are intuitively appealing. In the jargon of social choice theory, such desirable properties are called *axioms*.

- An aggregator f is (propositionwise) *unanimous* if $\varphi \in J_i$ for all $i \in N$ entails $\varphi \in f(\mathbf{J})$, for all $\varphi \in \Phi$ and all $\mathbf{J} \in \mathcal{J}(\Phi)^n$. That is, if every individual judge accepts φ , then so should the collective represented by f .⁵
- An aggregator f is *anonymous* if $f(\mathbf{J}) = f(J_{\pi(1)}, \dots, J_{\pi(n)})$, for all $\mathbf{J} \in \mathcal{J}(\Phi)^n$ and all permutations $\pi : N \rightarrow N$. That is, f should treat all judges the same.
- An aggregator f is *neutral* if $\varphi \in J_i \Leftrightarrow \psi \in J_i$ for all $i \in N$ entails $\varphi \in f(\mathbf{J}) \Leftrightarrow \psi \in f(\mathbf{J})$, for all $\varphi, \psi \in \Phi$ and all $\mathbf{J} \in \mathcal{J}(\Phi)^n$. That is, f should treat all formulas

⁵ Another common formulation of unanimity requires only $f(J, \dots, J) = J$ for all $J \in \mathcal{J}(\Phi)$.

the same (if φ and ψ are accepted by the same judges, then the collective must accept either both or neither of them).⁶

- An aggregator f is *independent* if $\varphi \in J_i \Leftrightarrow \varphi \in J'_i$ for all $i \in N$ entails $\varphi \in f(\mathbf{J}) \Leftrightarrow \varphi \in f(\mathbf{J}')$, for all $\varphi \in \Phi$ and all $\mathbf{J}, \mathbf{J}' \in \mathcal{J}(\Phi)^n$. That is, whether we accept φ should only depend on the pattern of individual acceptances of φ .
- An aggregator f is *monotonic* if $\varphi \in J'_i \setminus J_i$ entails $\varphi \in f(\mathbf{J}) \Rightarrow \varphi \in f(\mathbf{J}_{-i}, J'_i)$, for all $i \in N$, all $\varphi \in \Phi$, all $\mathbf{J} \in \mathcal{J}(\Phi)^n$, and all $J'_i \in \mathcal{J}(\Phi)$.⁷ That is, if a collectively accepted formula φ is accepted by an additional judge i (switching her judgment set from J_i to J'_i), then φ should still get accepted.

Neutrality and independence together are also known as *systematicity*.⁸ Note that we do not claim, or even wish, that every aggregator satisfies *all* of these axioms. They merely are strong candidates to consider when drawing up a list of requirements. The majority rule, however, clearly satisfies all of them.

Our remaining axioms express that we would like to see properties such as consistency be *lifted* from the individual to the collective level:

- An aggregator f is *complete* if $f(\mathbf{J})$ is complete for all $\mathbf{J} \in \mathcal{J}(\Phi)^n$.
- An aggregator f is *complement-free* if $f(\mathbf{J})$ is complement-free for all $\mathbf{J} \in \mathcal{J}(\Phi)^n$.
- An aggregator f is *consistent* if $f(\mathbf{J})$ is consistent for all $\mathbf{J} \in \mathcal{J}(\Phi)^n$.

Consistency and completeness together are often referred to as *collective rationality*. Observe that the majority rule is both complete and complement-free (at least in case n is odd), but that we have seen that it is *not* consistent.

It is useful to reformulate the independence axiom as follows: f is independent if (and only if) for every $\varphi \in \Phi$ there exists a family of sets of judges $\mathcal{W}_\varphi \subseteq 2^N$ such that for all $\mathbf{J} \in \mathcal{J}(\Phi)^n$ it is the case that $\varphi \in f(\mathbf{J})$ if and only if $N_\varphi^\mathbf{J} \in \mathcal{W}_\varphi$. We call \mathcal{W}_φ the set of *winning coalitions* for φ . That is, under an independent aggregator we only need to look at the coalition of judges accepting φ to be able to decide whether φ should be collectively accepted (and those coalitions with the power of getting φ accepted are what we call its winning coalitions). Understanding the structure of the space of winning coalitions provides a crucial key to understanding the dynamics of JA. For now, we pin down some basic facts about it:

Lemma 1 (Winning coalitions) *Let f be an independent aggregator and, for every formula $\varphi \in \Phi$, let $\mathcal{W}_\varphi \subseteq 2^N$ be the corresponding family of winning coalitions, i.e., $\varphi \in f(\mathbf{J}) \Leftrightarrow N_\varphi^\mathbf{J} \in \mathcal{W}_\varphi$ for all $\mathbf{J} \in \mathcal{J}(\Phi)^n$. Then the following are all true:*

- (i) *f is unanimous if and only if $N \in \mathcal{W}_\varphi$ for all $\varphi \in \Phi$.*

⁶ For some very simple agendas, such as $\Phi = \{p, \neg p\}$, all aggregators are vacuously neutral (because $p \in J_i \Leftrightarrow \neg p \in J_i$ is never true), even though, intuitively, they do not all treat all formulas the same. To exclude such pathological cases, we have made the assumption that the agenda is nontrivial.

⁷ Here (\mathbf{J}_{-i}, J'_i) denotes the profile we obtain when we replace J_i in \mathbf{J} by J'_i .

⁸ To be precise, any statement regarding the equivalence of axioms must be made w.r.t. a specific class of agendas. The equivalence mentioned here holds for nontrivial agendas.

- (ii) f is anonymous if and only if \mathcal{W}_φ is closed under equinumerosity, i.e., if and only if $C \in \mathcal{W}_\varphi$ and $|C| = |C'|$ entail $C' \in \mathcal{W}_\varphi$ for all $C, C' \subseteq N$ and all $\varphi \in \Phi$.
- (iii) f is neutral if and only if $\mathcal{W}_\varphi = \mathcal{W}_\psi$ for all $\varphi, \psi \in \Phi$.
- (iv) f is monotonic if and only if \mathcal{W}_φ is upward closed, i.e., if and only if $C \in \mathcal{W}_\varphi$ and $C \subseteq C'$ entail $C' \in \mathcal{W}_\varphi$ for all $C, C' \subseteq N$ and all $\varphi \in \Phi$.
- (v) f is complement-free if and only if \mathcal{W}_φ does not contain complementary coalitions, i.e., if and only if $C \notin \mathcal{W}_\varphi$ or $\overline{C} \notin \mathcal{W}_\varphi$ for all $C \subseteq N$ and all $\varphi \in \Phi$.
- (vi) f is complete if and only if \mathcal{W}_φ is maximal, i.e., if and only if $C \in \mathcal{W}_\varphi$ or $\overline{C} \in \mathcal{W}_\varphi$ for all $C \subseteq N$ and all $\varphi \in \Phi$.

Proof. These claims follow immediately from the relevant definitions. Only the case of neutrality requires closer inspection. Clearly, $\mathcal{W}_\varphi = \mathcal{W}_\psi$ for all $\varphi, \psi \in \Phi$ implies the symmetric treatment of all formulas and thus neutrality in the technical sense. For the converse, we make use of our assumption that Φ contains two formulas α and β such that all of $\{\alpha, \beta\}$, $\{\alpha, \sim\beta\}$, $\{\sim\alpha, \beta\}$, and $\{\sim\alpha, \sim\beta\}$ are consistent. So let f be an aggregator that is independent and neutral.

First, consider two formulas φ and ψ for which both $\{\varphi, \psi\}$ and $\{\sim\varphi, \sim\psi\}$ are consistent and take any coalition $C \in \mathcal{W}_\varphi$. We can construct a profile \mathbf{J} with $N_\varphi^{\mathbf{J}} = N_\psi^{\mathbf{J}} = C$. From C being winning for φ , we get $\varphi \in f(\mathbf{J})$. From neutrality we then get $\psi \in f(\mathbf{J})$, and from independence $C \in \mathcal{W}_\psi$. Thus, from $\{\varphi, \psi\}$ and $\{\sim\varphi, \sim\psi\}$ being consistent we get $\mathcal{W}_\varphi = \mathcal{W}_\psi$, so in particular $\mathcal{W}_\alpha = \mathcal{W}_{\sim\alpha} = \mathcal{W}_\beta = \mathcal{W}_{\sim\beta}$.

Note that α is contingent (i.e., it is neither a tautology nor a contradiction). Consider any other formula φ that is also contingent. If both $\{\varphi, \alpha\}$ and $\{\sim\varphi, \sim\alpha\}$ are consistent, then $\mathcal{W}_\varphi = \mathcal{W}_\alpha$ and we are done. So suppose it is not the case that both $\{\varphi, \alpha\}$ and $\{\sim\varphi, \sim\alpha\}$ are consistent. Observe that when we negate one of the formulas in an inconsistent set of two contingent formulas, we obtain a consistent set. Thus, if $\{\varphi, \alpha\}$ is inconsistent, then both $\{\sim\varphi, \alpha\}$ and $\{\varphi, \sim\alpha\}$ are consistent. And also in case $\{\sim\varphi, \sim\alpha\}$ is inconsistent, both $\{\sim\varphi, \alpha\}$ and $\{\varphi, \sim\alpha\}$ are consistent. Thus, in either case we get $\mathcal{W}_\varphi = \mathcal{W}_{\sim\alpha}$. To summarize, so far we have shown that $\mathcal{W}_\varphi = \mathcal{W}_\psi$ for any two contingent formulas φ and ψ in the agenda.

If φ is a tautology, then there exists a profile $\mathbf{J} \in \mathcal{J}(\Phi)^n$ with $N = N_\varphi^{\mathbf{J}} = N_\alpha^{\mathbf{J}}$. Thus, by neutrality and independence, $N \in \mathcal{W}_\varphi \Leftrightarrow N \in \mathcal{W}_\alpha$. As no coalition other than the grand coalition N can ever accept φ , it does not matter which other coalitions belong to \mathcal{W}_φ and we may simply assume that $\mathcal{W}_\varphi = \mathcal{W}_\alpha$. The proof for contradictions φ (using \emptyset in place of N) is analogous. \square

As a first simple example of the power of Lemma 1, let us see how we can use it to prove that every aggregator f that is neutral, independent, monotonic, and complete is also unanimous: From neutrality and independence we get that f is fully defined by a single family \mathcal{W} of winning coalitions. By completeness, either $N \in \mathcal{W}$ (in which case we are done) or $\overline{N} = \emptyset \in \mathcal{W}$. But also in this latter case, by monotonicity and due to $\emptyset \subseteq N$, we get $N \in \mathcal{W}$, i.e., f is unanimous.

17.2.4 A Simple Impossibility Theorem

We conclude our run through the basic theory of JA with a simple impossibility theorem. It shows that it is not just the majority rule that fails to reliably produce consistent outcomes, but that we cannot get consistency for any aggregator that satisfies what would appear to be rather innocent axiomatic requirements. This is the original impossibility theorem of JA, due to List and Pettit, which we had already mentioned in the introduction.

Theorem 2 (List and Pettit, 2002) *No aggregator for an agenda of the form $\Phi \supseteq \{p, q, p \wedge q\}$ can be anonymous, neutral, independent, complete, and consistent.*

Proof. Let f be an aggregator for an agenda Φ with $\{p, q, p \wedge q\} \subseteq \Phi$. For the sake of contradiction, assume f is anonymous, neutral, independent, complete, and consistent. By Lemma 1 and the first three properties, if two formulas φ and ψ are accepted by the same number of judges, then f must accept either both or neither of them. Now consider a profile \mathbf{J} in which $\frac{n-1}{2}$ judges accept both p and q ; one judge accepts p but not q ; one judge accepts q but not p ; and the remaining $\frac{n-3}{2}$ judges accept neither p nor q .⁹ Observe that for this profile we get $|N_p^{\mathbf{J}}| = |N_q^{\mathbf{J}}| = |N_{\neg(p \wedge q)}^{\mathbf{J}}| = \frac{n+1}{2}$. So we must accept either all of p , q , and $\neg(p \wedge q)$, or none of them. The former immediately clashes with consistency. If we take the latter route, completeness forces us to accept all of $\neg p$, $\neg q$, and $p \wedge q$, which again leads to a clash with consistency. \square

Theorem 2 is a negative result. Still, it can help us to pinpoint how we need to relax our requirements to make consistent aggregation possible after all.

One approach is to weaken the axioms. For instance, it is easy to define certain (trivial) aggregators that meet all but one of our requirements: the *dictatorship* that always returns a copy of the judgment set of the first judge only violates anonymity, while a *constant* aggregator that always returns some fixed judgment set in $\mathcal{J}(\Phi)$ only violates neutrality. Of course, such aggregators are of little practical interest. In Section 17.3 we discuss some more practically useful aggregation rules, which account for above impossibility by violating independence or completeness.

A second approach is to restrict the range of agendas on which we require our aggregator to perform satisfactorily. Section 17.4 will be devoted to results that identify, for a given (class of) aggregator(s), the precise class of agendas for which consistent aggregation is possible. This also takes care of a somewhat unsatisfactory feature of Theorem 2, namely its restriction to an overly specific class of agendas (those including p , q , and $p \wedge q$).

A third approach, finally, is to not restrict the agenda, but to instead assume

⁹ Note how we make use of our general assumption that n is odd. If n is even, we obtain a much more fundamental impossibility: there can be no anonymous, neutral, complete and *complement-free* aggregator even for simple agendas such as $\Phi = \{p, \neg p, q, \neg q\}$. To see this, consider what to do in case the number of judges accepting p equals that accepting $\neg p$.

that not every possible profile will be encountered in practice. We shall not discuss such *domain restrictions* in any detail here. They are similar to domain restrictions studied in voting theory, but arguably lack the intuitive appeal of, say, Black's *single-peakedness*, a domain restriction for preference profiles avoiding paradoxical election outcomes (Black, 1948). The most powerful result along these lines in JA is due to Dietrich and List (2010). They showed that, if a profile is *value-restricted* in the sense that for every minimally inconsistent (non-singleton) subset X of the agenda Φ there exist two formulas $\varphi_X, \psi_X \in X$ such that no individual accepts both φ_X and ψ_X , then the outcome of the majority rule will be consistent. The proof is immediate: For the sake of contradiction, assume the majority outcome is inconsistent. Then it must include some minimally inconsistent set X with associated formulas φ_X and ψ_X , each of which must have been accepted by a (strict) majority. But then, by the pigeonhole principle, at least one judge must have accepted both of them, which contradicts the assumption of value restriction.

17.3 Aggregation Rules

The majority rule is but one method of aggregation. In this section we introduce three families of aggregators (quota rules, distance-based rules, and premise-based rules), and for each of them discuss to what extent they allow us to address the deficit of the majority rule highlighted by the discursive dilemma. For each family, we then exemplify a specific research direction in JA: the axiomatic characterization of aggregators, the complexity of winner determination, and strategic manipulation.

17.3.1 Quota-based Aggregation / Axiomatic Characterizations

Under a *quota rule* we accept a given formula φ if the number of individual judges accepting φ reaches a certain threshold. Formally, a quota rule f_q is induced by a function $q : \Phi \rightarrow \{0, 1, \dots, n+1\}$, mapping formulas to thresholds:

$$f_q(\mathbf{J}) = \{\varphi \in \Phi \mid |N_\varphi^\mathbf{J}| \geq q(\varphi)\}$$

Note that if the quota of a given formula is 0, then this means that that formula will always get accepted, while quota $n+1$ means that the formula in question will never get accepted. The rule f_q is called a *uniform* quota rule if q maps all formulas to the same number λ (in such a case, we simply write f_λ). For example, the (strict) majority rule is the uniform quota rule $f_{\lceil \frac{n+1}{2} \rceil}$, while f_n is the *intersection rule*, mapping any given profile \mathbf{J} to the judgment set $J_1 \cap \dots \cap J_n$.

It is intuitively clear that, if we increase the quota, then it should be less likely that we obtain a collective judgment set that is inconsistent. For example, take once more the agenda $\Phi = \{p, \neg p, q, \neg q, p \wedge q, \neg(p \wedge q)\}$. Then, provided we impose a

uniform quota strictly greater than $\frac{2}{3}n$, the collective judgment set will be consistent for every possible profile.¹⁰ This is a consequence of the following result:

Proposition 3 (Dietrich and List, 2007b) *Let k be the size of the largest minimally inconsistent subset of the agenda Φ . Then every uniform quota rule f_λ with a quota of $\lambda > \frac{k-1}{k} \cdot n$ is consistent.*

Proof. For the sake of contradiction, suppose there exists a profile $\mathbf{J} \in \mathcal{J}(\Phi)^n$ for which $f_\lambda(\mathbf{J})$ is inconsistent. Take an arbitrary minimally inconsistent subset $X \subseteq f_\lambda(\mathbf{J})$. By assumption, we have $|X| \leq k$. For each formula $\varphi \in X$, there must have been at least λ judges accepting it. Hence, summing over all the formulas in X , there must have been at least $|X| \cdot \lambda$ occasions of a formula in X being accepted by some judge. By the pigeon hole principle, at least one of the n judges must have accepted at least $\frac{|X| \cdot \lambda}{n}$ of these formulas. But due to $\lambda > \frac{k-1}{k} \cdot n$, we get $\frac{|X| \cdot \lambda}{n} > |X| - \frac{|X|}{k}$, i.e., (as $\frac{|X|}{k} \leq 1$) that judge must have accepted *all* of the formulas in X , contradicting our assumption of individual consistency. \square

Thus, if we are willing to give up completeness (which is violated by quota rules with high quotas), then we can circumvent the impossibility of Theorem 2.

Proposition 3 can be strengthened to say that for any quota *not* satisfying the constraint $\lambda > \frac{k-1}{k} \cdot n$ the corresponding rule is *not* consistent. In our earlier example, $\{p, q, \neg(p \wedge q)\}$ with size 3 is the largest minimally inconsistent subset of the agenda $\Phi = \{p, \neg p, q, \neg q, p \wedge q, \neg(p \wedge q)\}$. So if, for example, $n = 30$, then any quota of 21 or above will guarantee consistency, while quota 20 does not.

An important question in social choice theory is whether a given set of axiomatic requirements fully characterizes a particular aggregator (or family of aggregators). If this is the case and if we have strong normative support for the axioms in question, then this forces us to adopt the corresponding aggregator. Let us briefly review a number of such characterization results for quota rules.

Proposition 4 (Dietrich and List, 2007b) *An aggregator is anonymous, independent, and monotonic if and only if it is a quota rule.*

Proof. For the right-to-left direction, observe that every quota rule clearly has those three properties. For the left-to-right direction, the claim follows from Lemma 1: By independence, we can decide formula by formula. By monotonicity, the set of winning coalitions is upward closed. By anonymity, only the size of coalitions matters. Taken together, this means that for every formula φ there exists a number λ_φ such that φ is collectively accepted if and only if at least λ_φ judges accept φ . In other words, λ_φ is the quota associated with φ , i.e., $q(\varphi) = \lambda_\varphi$. \square

¹⁰ On the downside, the collective judgment set may not always be complete (namely when both φ and $\neg\varphi$ are accepted by fewer than two thirds of the judges).

A quota rule is neutral if and only if it is a uniform quota rule.¹¹ Thus, as an immediate consequence of Proposition 4, we obtain the following characterization:

Corollary 5 *An aggregator is anonymous, neutral, independent, and monotonic if and only if it is a uniform quota rule.*

We now want to add completeness and complement-freeness to our requirements (but not consistency, which we already know to be impossible from Theorem 2). Intuitively speaking, low quotas should favor completeness (as generally more formulas will get accepted), while high quotas should favor complement-freeness (as generally fewer formulas will get accepted). Let λ be the uniformly imposed quota. If k individuals accept φ , then $n-k$ individuals accept $\sim\varphi$. Thus, to guarantee completeness, we require $\max\{k, n-k\} \geq \lambda$ for all $k \leq n$, as that way at least one of φ and $\sim\varphi$ will get accepted under all circumstances. Similarly, to guarantee complement-freeness, we require $\lambda > \min\{k, n-k\}$ for all $k \leq n$. The closer k is to $\frac{n}{2}$, the harder it gets to satisfy *both* of these constraints. Recall that we assume n to be odd. For $k = \frac{n+1}{2}$, there is only a single value for λ that will satisfy both constraints, namely $\lambda = \frac{n+1}{2}$. That is, only the majority rule satisfies all of our requirements and we obtain the following corollary to Corollary 5:

Corollary 6 *An aggregator is anonymous, neutral, independent, monotonic, complete, and complement-free if and only if it is the (strict) majority rule.*

We stress that Corollary 6 is true only when the number n of individual judges is odd. As pointed out in Footnote (9) already, when n is even, there exists no aggregator that satisfies all of these requirements. For the special case of an agenda with just a single pair of complementary formulas, Corollary 6 reduces to May's Theorem on the characterization of the majority rule in voting theory (May, 1952).¹²

17.3.2 Distance-based Aggregation / Winner Determination

Another approach to aggregation is to select as outcome a judgment set that, in some sense, minimizes the distance to the profile. If we measure distances between judgment sets using the *Hamming distance* and if we interpret the distance from an outcome to a profile as the *sum* of the distances between that outcome and each of the individual judgment sets, then we obtain the following aggregator:

$$f_{\text{kem}}(\mathbf{J}) = \underset{J \in \mathcal{J}(\Phi)}{\operatorname{argmin}} \sum_{i \in N} H(J, J_i)$$

That is, we go through all complete and consistent judgment sets J and return the one that minimizes the sum of the Hamming distances to the individual judgment

¹¹ This characterization holds, provided the agenda is nontrivial. See also Footnote (6).

¹² May allows for ties and uses a more sophisticated monotonicity axiom, known as *positive responsiveness*, so as to be able to also cover the case of an even number of individuals. Refer to Chapter 2 (Zwicker, 2016) for full details.

sets. This rule, discussed under a variety of different names in the literature, is the most widely used distance-based aggregator in JA. If we apply it to the preference agenda (see Section 17.2.2), then we obtain the *Kemeny rule* familiar from preference aggregation (Kemeny, 1959) and discussed in depth in Chapter 4 (Fischer et al., 2016); this is why we shall refer to it as the *generalized Kemeny rule*. To be precise, f_{kem} is an *irresolute* aggregator, as it may return a *set* of several judgment sets (that are all equally close to the profile). To obtain a resolute rule fitting our definition of Section 17.2.1, we have to combine f_{kem} with a *tie-breaking* rule.

The generalized Kemeny rule is clearly consistent: only consistent judgment sets are considered as possible outcomes during the optimization process. It achieves consistency by sacrificing independence.

A fundamental question in computational social choice concerns the computational complexity of the *winner determination* problem of a given method of aggregation. For a resolute aggregator f this can be cast as a decision problem by asking, for a given profile \mathbf{J} and a given formula $\varphi \in \Phi$, whether it is the case that $\varphi \in f(\mathbf{J})$. By answering this question for every formula φ in the agenda, we can compute the outcome, the ‘winning’ judgment set. Thus, the complexity of our decision problem is directly relevant to the algorithmic feasibility of the pragmatic problem of applying rule f . For irresolute aggregators, such as the generalized Kemeny rule, an appropriate formulation of the question is whether, for a given profile \mathbf{J} and a given set $L \subseteq \Phi$, there exists a winning set $J^* \in f(\mathbf{J})$ such that $L \subseteq J^*$. For the generalized Kemeny rule, winner determination is highly intractable:

Theorem 7 (Endriss et al., 2012) *The winner determination problem of the generalized Kemeny rule is Θ_2^P -complete.*

The complexity class Θ_2^P is the class of problems that can be solved, in polynomial time, by a (hypothetical) machine that has access to an oracle capable of deciding NP-complete problems in an instant, with the restriction that the number of queries to the oracle must be at most logarithmic in the size of the problem. Equivalently, we may ask a polynomial number of oracle queries, provided we ask them all in parallel. That is, the generalized Kemeny rule is *complete for parallel access to NP*. The proof of Theorem 7 is beyond the scope of this chapter. It is based on a reduction from the KEMENY WINNER problem in voting theory, for which Θ_2^P -completeness was established by Hemaspaandra et al. (2005). For a discussion of that problem we refer to Chapter 4 (Fischer et al., 2016). For comparison, it is easy to see that for any quota rule the winner determination problem is polynomial.

The generalized Kemeny rule is not the only distance-based aggregator in JA. We may use other distance metrics than the Hamming distance and we may vary the way we define the distance to a profile in terms of that basic metric. Here we only give one further example of a generalization of a rule familiar from preference aggregation: Under the *Slater rule* we first compute the *majority graph* and then return the preference order that minimizes the number of edges in the majority

graph we have to invert before we obtain a transitive order (Slater, 1961).¹³ This idea is easily transposed to JA to obtain a *generalized Slater rule*:

$$f_{\text{sla}}(\mathbf{J}) = \underset{J \in \mathcal{J}(\Phi)}{\operatorname{argmin}} H(J, f_{\text{maj}}(\mathbf{J}))$$

That is, we first compute the majority outcome $f_{\text{maj}}(\mathbf{J})$ and then find the complete and consistent judgment set closest to it (again, more than one judgment set may be closest and we may have to break ties if we need a unique winner).

The following example, with φ_1 and φ_2 representing two syntactically distinct formulas that are both semantically equivalent to $p \vee (q_1 \wedge q_2) \vee (r_1 \wedge r_2 \wedge r_3)$, demonstrates that the Kemeny rule and the Slater rule do not always coincide:

| | p | q_1 | q_2 | r_1 | r_2 | r_3 | φ_1 | φ_2 |
|------------|-----|-------|-------|-------|-------|-------|-------------|-------------|
| 1 judge: | Yes | No | No | No | No | No | Yes | Yes |
| 10 judges: | No | Yes | Yes | No | No | No | Yes | Yes |
| 10 judges: | No | No | No | Yes | Yes | Yes | Yes | Yes |
| Kemeny: | No | Yes | Yes | No | No | No | Yes | Yes |
| Slater: | Yes | No | No | No | No | No | Yes | Yes |

Consult Section 17.7 for references to other distance-based aggregators.

17.3.3 Premise-based Aggregation / Strategic Manipulation

For some application scenarios it will be natural to think of certain formulas in the agenda as *premises* and the others as *conclusions*. Suppose we can partition the agenda Φ into a set Φ^p of premises and a set Φ^c of conclusions such that each of them is closed under complementation, $\Phi = \Phi^p \cup \Phi^c$, and $\Phi^p \cap \Phi^c = \emptyset$.

The *premise-based rule* f_{pre} works by first applying the majority rule to the premises and then accepting those conclusions that logically follow from the collectively accepted premises:

$$f_{\text{pre}}(\mathbf{J}) = \Delta \cup \{\varphi \in \Phi^c \mid \Delta \models \varphi\}, \text{ where } \Delta = \{\varphi \in \Phi^p \mid |N_\varphi^\mathbf{J}| > \frac{n}{2}\}$$

Clearly, in the general case f_{pre} inherits all the problems of the majority rule and cannot guarantee consistency (e.g., when *every* formula is declared a premise). In addition, f_{pre} may fail to be complete. For example, when $p \wedge q$ and $\neg(p \wedge q)$ are conclusions and no premise has either p or q as a subformula, then neither $p \wedge q$ nor $\neg(p \wedge q)$ will be logically entailed by whatever (consistent) set of premises we end up accepting. However, if we impose suitable constraints on the definition of premises and conclusions, we can obtain a well-behaved aggregator. The most common set of assumptions, which we shall also adopt here, is the following:

¹³ For an in-depth discussion of the Slater rule we refer to Chapter 3 (Brandt et al., 2016).

- (i) The set of premises is the set of literals in the agenda.
- (ii) The agenda is closed under propositional variables, i.e., if φ is a formula in the agenda, then so is every propositional variable occurring within φ .

Assumption (i) guarantees consistency; a set of collectively accepted premises now directly corresponds to an assignment of truth values to propositional variables. Assumption (ii) then guarantees completeness, because, for every conclusion φ , either φ or $\sim\varphi$ will be satisfied by that truth assignment. That is, we obtain the following result (given our assumption of n being odd):

Fact 8 *For any agenda that is closed under propositional variables, the premise-based rule, with the premises being the literals, is complete and consistent.*

Thus, by sacrificing both neutrality and independence, the premise-based rule allows us to circumvent the impossibility of Theorem 2.

We have seen that distance-based aggregators achieve consistency at the price of high computational complexity. This is not the case for f_{pre} :

Fact 9 *The winner determination problem of the premise-based rule, with the premises being the literals, is polynomial.*

Indeed, applying the majority rule to the premises is certainly polynomial. Deciding whether a given conclusion should be accepted then amounts to a model checking problem for propositional logic, which can also be done in polynomial time.

One of the central phenomena studied in social choice theory is strategic behavior. Under what circumstances do individuals have an incentive to truthfully report their private information to the aggregation mechanism? This is a very natural question in the context of voting or resource allocation, where this private information concerns the individuals' preferences and where we can use these very same preferences to define what it means for an individual to have an incentive to perform a given action. In JA, on the other hand, individuals do not have preferences, so it is not obvious what it may mean for an individual to have an incentive one way or the other. To be able to analyze strategic behavior in JA, we need to endow our judges with preferences. For the purposes of this brief exposition, let us assume that every judge's preferences are induced by her true judgment set and the Hamming distance: judge i with true judgment set J_i prefers J to J' , denoted as $J \succ_i J'$, if and only if $H(J_i, J) < H(J_i, J')$. We say that she has *Hamming preferences*.¹⁴ Now suppose three judges truthfully report their judgment sets:

| | p | q | r | $p \vee q$ | $p \vee r$ |
|----------|-----|-----|-----|------------|------------|
| Judge 1: | No | No | No | No | No |
| Judge 2: | Yes | No | No | Yes | Yes |
| Judge 3: | No | Yes | Yes | Yes | Yes |

¹⁴ Preferences induced by the Hamming distance are also discussed in Chapter 9 on voting in combinatorial domains (Lang and Xia, 2016)

Under the premise-based rule, all three premises will get rejected, which means that also the two conclusions get rejected. For judge 3 the Hamming distance from this outcome to her own judgment set is 4. But if she lies and answers ‘yes’ for p , then p will get accepted and thus also the two conclusions. Her distance to this new outcome is only 3, i.e., she has an incentive to *manipulate* in this manner. We say that an aggregator f is *immune to manipulation* for a given method of deriving preferences from judgment sets, if every judge i (at least weakly) prefers the sincere outcome $f(\mathbf{J})$ to the manipulated outcome $f(\mathbf{J}_{-i}, J'_i)$ for any profile $\mathbf{J} \in \mathcal{J}(\Phi)^n$ and any alternative insincere judgment set $J'_i \in \mathcal{J}(\Phi)$ for judge i .

Proposition 10 (Dietrich and List, 2007a) *Any independent and monotonic aggregator is immune to manipulation by judges with Hamming preferences.*

Proof. Independence means that the would-be manipulator can consider one formula at a time. Monotonicity then means that it is always in her best interest to drive up the support for formulas in her judgment set and to reduce the support for those not in her judgment set, i.e., it is in her best interest to report truthfully. \square

As discussed in depth in Chapter 6 (Conitzer and Walsh, 2016), for aggregators that are not immune to manipulation a relevant question is whether it may be the case that manipulation is a computationally intractable problem, as this may provide at least some level of protection against unwanted strategic behavior. The *manipulation problem* for a given aggregator f and a given method of deriving preferences from judgment sets is the problem of deciding whether a judge may obtain a preferred outcome by misreporting her judgment set. While we have seen that the premise-based rule can be manipulated, doing so is hard, at least in the worst case and for agendas involving large formulas:¹⁵

Proposition 11 (Endriss et al., 2012) *The manipulation problem for the premise-based rule and judges with Hamming preferences is NP-complete.*

The proof is not difficult, but beyond the scope of this chapter. It involves a reduction from the NP-hard *satisfiability problem* for propositional logic to our manipulation problem. Given a formula φ , the core idea is to build a profile where one judge has an incentive to misreport her judgment set if and only if φ is satisfiable.

It is important to stress that the particular manipulation problem of Proposition 11 is just one of several natural definitions for which complexity results are known (Baumeister et al., 2013). For instance, we may vary the way in which preferences are induced from judgment sets or we may assume that the manipulator is only interested in some of the formulas in the agenda.

¹⁵ Given that for distance-based aggregators the winner determination problem is already intractable, questions regarding the complexity of manipulation are considerably less interesting here.

17.4 Agenda Characterization Results

Suppose we are given an agenda Φ and an aggregator f . Then we may want to ask: can we be certain that the outcome $f(\mathbf{J})$ will be consistent for every possible consistent profile $\mathbf{J} \in \mathcal{J}(\Phi)^n$ over this agenda?

We want to answer such questions not only for one specific agenda Φ , but would like to be able to offer a full *characterization* of *all* agendas for which the answer is positive. In this section, we provide such characterizations in terms of *agenda properties* that tally the structural richness of an agenda. Going beyond individual aggregators f , we also prove agenda characterization results for *families of aggregators* \mathcal{F} , defined in terms of the axioms they satisfy. Such results come in two forms: *existential results* talk about conditions under which at least *some* aggregator in \mathcal{F} is consistent, while *universal results* establish conditions under which *all* aggregators in \mathcal{F} are. For the latter, in particular, a relevant question is how difficult it is to verify whether a given agenda satisfies certain agenda properties. Therefore, we briefly discuss the computational complexity of this problem.

17.4.1 Consistent Aggregation under the Majority Rule

In the introduction we have seen that for one specific aggregator and one specific agenda, namely the majority rule and the agenda $\{p, \neg p, q, \neg q, p \wedge q, \neg(p \wedge q)\}$, we may encounter a discursive dilemma. That is, for this agenda it is possible to construct a consistent profile such that the majority rule will return a judgment set that is inconsistent. In Section 17.2.4 we have seen a generalization of this insight: the problem persists for any aggregator belonging to an entire *class of aggregators* (cf. Theorem 2). We now want to discuss a different generalization and ask: what is the *class of agendas* for which the majority rule has this deficiency?

Intuitively speaking, the discursive dilemma comes about when there is an inconsistent subset $X \subseteq \Phi$, such that each judge accepts *sufficiently few* of the formulas in X to be individually consistent, but at the same time *sufficiently many* to ensure that every formula in X gathers a majority. That is, the *size* of inconsistent sets, or more precisely *minimally inconsistent sets*, that we can build from the formulas in the agenda seems to matter. We say that an agenda Φ satisfies the *median property* if every inconsistent subset of Φ does itself have an inconsistent subset of size at most 2.¹⁶ For example, the agenda $\{p, \neg p, q, \neg q, p \wedge q, \neg(p \wedge q)\}$ does not have the median property, because one of its minimally inconsistent subsets, namely $\{p, q, \neg(p \wedge q)\}$, has three elements. This agenda property allows us to give a full characterization of the agendas on which the majority rule is consistent:

¹⁶ The term *median property* is due to Nehring and Puppe (2007), who coined this term in the context of work on social choice theory over a class of vector spaces known as median spaces. Agendas that satisfy the median property are sometimes also called *simple* agendas.

Lemma 12 (Nehring and Puppe, 2007) *The majority rule is consistent for a given agenda Φ if and only if Φ has the median property.*

Proof. The right-to-left direction is an instance of Proposition 3 for $k = 2$, since Φ having the median property means that the largest minimally inconsistent subset of Φ has at most size 2. For the other direction, let Φ be an agenda that violates the median property, i.e., there exists a minimally inconsistent set $X = \{\varphi_1, \dots, \varphi_k\} \subseteq \Phi$ with $k \geq 3$. Now consider the profile \mathbf{J} , in which judge i accepts all formulas in X except for $\varphi_{1+(i \bmod 3)}$. Note that \mathbf{J} is consistent. But the majority rule will accept all formulas in X , i.e., $f_{\text{maj}}(\mathbf{J})$ is inconsistent. \square

Deciding whether a given agenda has the median property is Π_2^P -complete (Endriss et al., 2012). Π_2^P , a complexity class located at the second level of the polynomial hierarchy, is the class of decision problems for which a certificate for a negative answer can be verified in polynomial time by a machine that has access to an oracle for answering NP-complete problems in an instant. Deciding whether a quantified Boolean formula of the form $\forall x_1 \dots \forall x_n \exists y_1 \dots \exists y_m \varphi$ is true is the canonical example of a problem that is complete for this class. Π_2^P includes the other complexity classes featuring in this chapter, particularly NP and Θ_2^P (and this inclusion is believed, but not known, to be proper).

Thus, we have a doubly negative result: not only do we have to restrict ourselves to highly simplistic agendas if we want to be certain to avoid instances of the discursive dilemma (this is the import of Lemma 12), but on top of this determining whether a given agenda is sufficiently simplistic is very difficult.

17.4.2 Existential Agenda Characterization

Next we want to go beyond specific aggregators (such as the majority rule), and instead prove agenda characterization results for *classes* of aggregators, defined in terms of certain axioms. For the first such result we take the axioms defining the majority rule (cf. Corollary 6), except that we substantially weaken anonymity and instead only require that our aggregator is not a dictatorship. Formally, an aggregator f is a *dictatorship* if there exists an individual $i^* \in N$ (the dictator) such that $f(\mathbf{J}) = J_{i^*}$ for every profile \mathbf{J} ; otherwise f is *nondictatorial*.¹⁷ Lemma 12 shows that only agendas with the median property guarantee consistent outcomes under the majority rule. We will now see that, even if we allow ourselves to pick freely from the much larger class of aggregators we obtain when we replace anonymity by nondictatoriality, we cannot do better than that.

Before we state this result formally, let us say something about the technique we are going to use to prove it. Recall that any independent and neutral aggregator is

¹⁷ To appreciate that anonymity really is much stronger than nondictatoriality, observe that the former rules out all weighted majority rules, while the latter does not.

defined by a family $\mathcal{W} \subseteq 2^N$ of winning coalitions (cf. Lemma 1). In mathematics, a set \mathcal{W} of subsets of a set N is called an *ultrafilter* if it satisfies the following three conditions (see, e.g., Davey and Priestley, 2002):¹⁸

- (i) \mathcal{W} does not include the empty set: $\emptyset \notin \mathcal{W}$.
- (ii) \mathcal{W} is closed under taking intersections: $C \in \mathcal{W}$ and $C' \in \mathcal{W}$ entail $C \cap C' \in \mathcal{W}$ for all $C, C' \subseteq N$.
- (iii) \mathcal{W} is maximal: $C \in \mathcal{W}$ or $\overline{C} \in \mathcal{W}$ for all $C \subseteq N$.

An ultrafilter is called *principal* if it is of the form $\mathcal{W} = \{C \subseteq N \mid i^* \in C\}$ for some $i^* \in N$. By a well-known fact, every ultrafilter \mathcal{W} on a *finite* set N is principal (Davey and Priestley, 2002).¹⁹ Translating back into the world of winning coalitions, this means that if we can show that the set \mathcal{W} of winning coalitions corresponding to a given (independent and neutral) aggregator f meets the three conditions above, then, given that the set N of judges is finite, f must be dictatorial (with the principal element i^* of the ultrafilter being the dictator).

We are now ready to prove a first important existential agenda characterization theorem, for a class of aggregators defined by a natural combination of axioms:

Theorem 13 (Nehring and Puppe, 2007) *There exists a neutral, independent, monotonic, and nondictatorial aggregator that is complete and consistent for a given agenda Φ if and only if Φ has the median property.*

Proof. The right-to-left direction follows from Lemma 12: if Φ has the median property, then the majority rule is consistent (and the majority rule is also neutral, independent, monotonic, nondictatorial, and complete).

The left-to-right direction in effect establishes an impossibility result. It is equivalent to the following claim: if a given neutral, independent, and monotonic aggregator f is complete and consistent (and thus also complement-free) for a given agenda Φ that violates the median property, then f must be a dictatorship. By Lemma 1, neutrality and independence mean that f is determined by a single family \mathcal{W} of winning coalitions. Let us show that \mathcal{W} must be an ultrafilter on N :

- (i) $\emptyset \notin \mathcal{W}$: If we had $\emptyset \in \mathcal{W}$, then $N \in \mathcal{W}$ by monotonicity (because $\emptyset \subseteq N$), which is in direct conflict with the requirement of being complement-free (as $\overline{\emptyset} = N$).
- (ii) \mathcal{W} is closed under taking intersections: This will be the (only) part in the proof where we make use of the assumption that Φ violates the median property. The median property being violated means that there exists a minimally inconsistent

¹⁸ Sometimes being upward closed is stated as a fourth defining condition of an ultrafilter. But note that we do not need to do so, as this already follows from the other three conditions.

¹⁹ The proof is immediate: Let $C^* := \bigcap_{C \in \mathcal{W}} C$. As N is finite, we can apply (ii) inductively and obtain $C^* \in \mathcal{W}$. By construction, C^* must be the smallest element of \mathcal{W} . If we can show that C^* is a singleton, then we are done. First, by (i), $C^* \neq \emptyset$. Second, for the sake of contradiction, suppose $|C^*| \geq 2$. Take any nonempty proper subset C of C^* . By (iii), either $C \in \mathcal{W}$ or $\overline{C} \in \mathcal{W}$. If it is the latter, then from (ii) and $C^* \in \mathcal{W}$ we get $C^* \cap \overline{C} = C^* \setminus C \in \mathcal{W}$. Hence, in either case a proper subset of C^* must be an element of \mathcal{W} . This is the required contradiction and concludes the proof.

subset $X = \{\varphi_1, \dots, \varphi_k\} \subseteq \Phi$ with $k \geq 3$. Take any two winning coalitions $C, C' \in \mathcal{W}$. (We need to show that $C \cap C' \in \mathcal{W}$.) Observe that it is possible to construct a complete and consistent profile \mathbf{J} with the following properties:

- $N_{\varphi_1}^{\mathbf{J}} = C$
- $N_{\varphi_2}^{\mathbf{J}} = C' \cup (N \setminus C)$
- $N_{\varphi_3}^{\mathbf{J}} = N \setminus (C \cap C')$
- $N_{\varphi_\ell}^{\mathbf{J}} = N$ for all ℓ with $4 \leq \ell \leq k$

That is, each judge accepts exactly $k-1$ of the formulas in X (which means that every judge is consistent). Note that $N_{\sim\varphi_3}^{\mathbf{J}} = C \cap C'$. Now, C being winning implies $\varphi_1 \in f(\mathbf{J})$. By monotonicity, any superset of C' is winning, including $C' \cup (N \setminus C)$. The latter implies $\varphi_2 \in f(\mathbf{J})$. Finally, from $\emptyset \notin \mathcal{W}$ we get $N \in \mathcal{W}$ (i.e., f is unanimous) and therefore all judges accepting all of $X \setminus \{\varphi_1, \varphi_2, \varphi_3\}$ entails $X \setminus \{\varphi_1, \varphi_2, \varphi_3\} \subseteq f(\mathbf{J})$. Taken together, this means $X \setminus \{\varphi_3\} \subseteq f(\mathbf{J})$. Thus, to ensure consistency of the outcome, we must have $\varphi_3 \notin f(\mathbf{J})$. But then, to ensure completeness, we are forced to accept $\sim\varphi_3 \in f(\mathbf{J})$. That is, we are forced to accept a formula that was accepted by precisely the judges in the intersection $C \cap C'$. In other words, $C \cap C'$ is winning, i.e., $C \cap C' \in \mathcal{W}$.

(iii) \mathcal{W} is maximal: This follows immediately from completeness (cf. Lemma 1).

As N is finite, \mathcal{W} must in fact be a principal ultrafilter. But we have already seen that this is equivalent to f being dictatorial, so we are done. \square

The ultrafilter method is very powerful. We will see one more example of its use. Our next theorem concerns independent and unanimous aggregators. The most important difference is that this time we do not want to assume neutrality, i.e., we cannot work with a single family \mathcal{W} of winning coalitions from the outset. To still make the ultrafilter method applicable, we have to *derive* neutrality. As we shall see, this is possible for agendas that are sufficiently rich. Merely violating the median property is not sufficient: for example, if the positive formulas in the agenda are $\{p, q, p \wedge q, r\}$, then the aggregator under which judge 1 dictates the outcome for $\{p, q, p \wedge q\}$ and r is decided upon by majority is nondictatorial, independent, and unanimous. That is, we require an agenda with sufficiently strong connections between its formulas to rule out such ‘localised’ aggregators.

Write $\varphi \xrightarrow{\Phi} \psi$ in case there exists a minimally inconsistent set $X \subseteq \Phi$ with $\varphi, \sim\psi \in X$. This notation indicates that, in the context of the remaining formulas in X , accepting φ forces us to also accept ψ . Let $\xrightarrow{\Phi}$ be the transitive closure of $\xrightarrow{\Phi}$. An agenda Φ is called *totally blocked* if $\varphi \xrightarrow{\Phi} \psi$ for every two formulas $\varphi, \psi \in \Phi$.

Lemma 14 (Neutrality) *Every unanimous and independent aggregator f that is complete and consistent for a totally blocked agenda Φ is neutral.*

Proof. Let Φ be a totally blocked agenda and let f be a unanimous and independent

aggregator that is complete and consistent for Φ . By independence, for each $\varphi \in \Phi$ there exists a $\mathcal{W}_\varphi \subseteq N$ such that $\varphi \in f(\mathbf{J}) \Leftrightarrow N_\varphi^{\mathbf{J}} \in \mathcal{W}_\varphi$ for all profiles \mathbf{J} .

Consider two formulas $\varphi, \psi \in \Phi$ such that $\varphi \xrightarrow{\Phi} \psi$. We will show that this implies $\mathcal{W}_\varphi \subseteq \mathcal{W}_\psi$: Let X be the minimally inconsistent set establishing $\varphi \xrightarrow{\Phi} \psi$. Take any winning coalition $C \in \mathcal{W}_\varphi$. Construct a complete and consistent profile \mathbf{J} in which all judges accept all of $X \setminus \{\varphi, \sim\psi\}$, those in C also accept φ and ψ , and the rest also accept $\sim\varphi$ and $\sim\psi$. By unanimity, $X \setminus \{\varphi, \sim\psi\} \subseteq f(\mathbf{J})$. Due to $C \in \mathcal{W}_\varphi$, furthermore $\varphi \in f(\mathbf{J})$. Hence, consistency forces $\sim\psi \notin f(\mathbf{J})$ and thus completeness forces $\psi \in f(\mathbf{J})$. As it was exactly the judges in C who accepted ψ , this means that C is also a winning coalition for ψ , i.e., $C \in \mathcal{W}_\psi$.

Now, by induction, not only $\varphi \xrightarrow{\Phi} \psi$ implies $\mathcal{W}_\varphi \subseteq \mathcal{W}_\psi$, but also $\varphi \xrightarrow{\Phi} \psi$ implies $\mathcal{W}_\varphi \subseteq \mathcal{W}_\psi$. Total blockedness means that $\varphi \xrightarrow{\Phi} \psi$ for all $\varphi, \psi \in \Phi$. Hence, $\mathcal{W}_\varphi = \mathcal{W}_\psi$ for all $\varphi, \psi \in \Phi$, i.e., by Lemma 1, f is indeed neutral. \square

We require one further agenda property: Φ is *even-number-negatable* if it has a minimally inconsistent subset X for which $(X \setminus Y) \cup \{\sim\varphi \mid \varphi \in Y\}$ is consistent for some set $Y \subset X$ of even cardinality. Any minimally inconsistent set X can be made consistent by flipping exactly one of its elements; even-number-negatability requires that for at least one X we can also get consistency by flipping an even number of formulas in X . The agenda property used in the following theorem is intuitively weaker than the median property used in Theorem 13. This is exactly what we would expect, as the axioms we impose are also weaker.

Theorem 15 (Dokow and Holzman, 2010) *There exists a unanimous, independent, and nondictatorial aggregator that is complete and consistent for a given agenda Φ if and only if Φ is not both totally blocked and even-number-negatable.*

Proof. For the right-to-left direction, for any agenda Φ that is either not totally blocked or not even-number-negatable, we show how to construct an aggregator that meets our requirements.

First, suppose Φ is *not totally blocked*. Then we can partition Φ into two disjoint sets Φ_1 and Φ_2 (i.e., $\Phi_1 \cup \Phi_2 = \Phi$ and $\Phi_1 \cap \Phi_2 = \emptyset$) such that there exist no $\varphi_1 \in \Phi_1$ and $\varphi_2 \in \Phi_2$ with $\varphi_1 \xrightarrow{\Phi} \varphi_2$. We use this partition to define an aggregator f : judge 1 dictates acceptance for formulas φ with $\{\varphi, \sim\varphi\} \subseteq \Phi_1$; judge 2 dictates acceptance for formulas φ with $\{\varphi, \sim\varphi\} \subseteq \Phi_2$; and for formulas φ with $\varphi \in \Phi_1$ and $\sim\varphi \in \Phi_2$ we accept φ unless there is unanimous support for $\sim\varphi$. This rule is easily seen to be unanimous, independent, nondictatorial, and complete. It also is consistent: For if not, there would be a profile for which f accepts some minimally inconsistent set X . But X cannot be fully included in Φ_1 , as then judge 1 would be inconsistent. It cannot be fully included in Φ_2 either, as then judge 2 would be inconsistent. Finally, X cannot be split across Φ_1 and Φ_2 : if $\sim\varphi_2 \in \Phi_1$ for all $\varphi_2 \in X \cap \Phi_2$, then all of $X \cap \Phi_2$ would have had to be accepted unanimously, meaning that judge 1 would

have had to accept all of X (contradicting consistency of judge 1); and if $\sim\varphi_2 \in \Phi_2$ for some $\varphi_2 \in X \cap \Phi_2$, then $\varphi_1 \xrightarrow{\Phi} \sim\varphi_2$ for all $\varphi_1 \in X \cap \Phi_1$ (in contradiction to how the partition into Φ_1 and Φ_2 has been defined).

Second, suppose Φ is *not even-number-negatable*. Consider the *parity rule* f_{par} that accepts a formula if and only if an odd number of the judges does.²⁰ This rule clearly is unanimous, independent, nondictatorial, and complete. To show that it is also consistent we will make use of the fact that Φ is not even-number-negatable. Assume, for the sake of contradiction, that there exists a profile $\mathbf{J} \in \mathcal{J}(\Phi)$ for which $f_{\text{par}}(\mathbf{J})$ is inconsistent. Let X be a minimally inconsistent subset of $f_{\text{par}}(\mathbf{J})$. Suppose $|X|$ is odd (the case where $|X|$ is even works analogously). As Φ is not even-number-negatable, every set of the form $(X \setminus Y) \cup \{\sim\varphi \mid \varphi \in Y\}$ with $Y \subset X$ and $|Y|$ even must be inconsistent. Hence, every judge must accept a subset of X of even cardinality. So, if we sum over all judges, we obtain an *even* number of instances of an acceptance of an X -formula by a judge. On the other hand, by definition of f_{par} , each of the odd number of formulas in X must have been accepted by an odd number of judges, which implies that the overall number of acceptances of X -formulas must be *odd*. Thus, we obtain the desired contradiction.

For the left-to-right direction, we have to prove that for any agenda Φ that is totally blocked and even-number-negatable, any aggregator f that is unanimous, independent, complete, and consistent must be dictatorial. As Φ is totally blocked we can apply Lemma 14 and infer that f must be neutral. As Φ is even-number-negatable, it is also *pair-negatable* in the sense that there must exist a minimally inconsistent set X such that $(X \setminus Y) \cup \{\sim\varphi \mid \varphi \in Y\}$ is consistent for some $Y \subset X$ of cardinality 2.²¹ In other words, Φ has a minimally inconsistent subset X including at least three formulas $\varphi_1, \varphi_2, \varphi_3$ such that all of $(X \setminus \{\varphi_1\}) \cup \{\sim\varphi_1\}$, $(X \setminus \{\varphi_2\}) \cup \{\sim\varphi_2\}$, and $(X \setminus \{\varphi_1, \varphi_2\}) \cup \{\sim\varphi_1, \sim\varphi_2\}$ are consistent. In summary, we now have all the conditions of the left-to-right direction of Theorem 13 in place, except for monotonicity; and in addition we can make use of slightly stronger assumptions on Φ . Recall that the only points in the proof of Theorem 13 where we required monotonicity were the derivations of the first two ultrafilter conditions. The first of them, $\emptyset \notin \mathcal{W}$, can instead be inferred from maximality, the third ultrafilter condition, together with $N \in \mathcal{W}$ (which follows from unanimity). We are left with establishing the second ultrafilter condition, closure under taking intersections. Take any two winning coalitions $C, C' \in \mathcal{W}$. We can construct a complete and consistent profile \mathbf{J} with $N_{\varphi_1}^{\mathbf{J}} = C$, $N_{\varphi_2}^{\mathbf{J}} = C'$, $N_{\varphi_3}^{\mathbf{J}} = N \setminus (C \cap C')$, and $N_{\varphi}^{\mathbf{J}} = N$ for all $\varphi \in X \setminus \{\varphi_1, \varphi_2, \varphi_3\}$. Observe that this slightly simplified construction was not available to us in the proof of

²⁰ Here we will make use of our assumption that n , the total number of judges, is odd.

²¹ This may be proved as follows: Suppose Φ is even-number-negatable by virtue of X and Y , with $|Y| = k > 2$. If Φ is not also pair-negatable by virtue of X and some Y' with $|Y'| = 2$, then $(X \setminus \{\varphi_1, \varphi_2\}) \cup \{\sim\varphi_1, \sim\varphi_2\}$ for some $\varphi_1, \varphi_2 \in Y$ is also minimally inconsistent, and we get even-number-negatability by virtue of that set together with $Y \setminus \{\varphi_1, \varphi_2\}$ of cardinality $k - 2$. The claim then follows by induction on $|Y|$.

Theorem 13, because in the absence of pair-negatability we would have had no guarantee that the judgment set of judges in $N \setminus (C \cap C')$, missing two of the formulas in X , is consistent. Now φ_1 gets accepted by virtue of C being winning, φ_2 gets accepted by virtue of C' being winning, and all of $X \setminus \{\varphi_1, \varphi_2, \varphi_3\}$ get accepted by virtue of unanimity. So we must reject φ_3 and thus accept $\sim\varphi_3$. But it is exactly the judges in $C \cap C'$ who support $\sim\varphi_3$, so $C \cap C'$ must be winning. \square

Interestingly, independence and unanimity are also the core axioms in *Arrow's Theorem* in preference aggregation (Arrow, 1963). Indeed, using the embedding of preference aggregation into JA sketched in Section 17.2.2, we can obtain a proof of Arrow's Theorem in the same way as we have shown the left-to-right direction of Theorem 15 (this involves showing that the preference agenda is both totally blocked and even-number-negatable). To be precise, the JA encoding given here allows individuals to, for instance, not accept the formulas for transitivity, while Arrow's Theorem applies even under the assumption that they always do. This can be accounted for by altering the underlying logic and adding the formulas expressing properties of preference orders as logical axioms on top of those of the propositional calculus, rather than including them in the agenda (Dietrich and List, 2007c).

17.4.3 Universal Agenda Characterization

Let us call an agenda Φ *safe* for a given aggregator f if $f(\mathbf{J})$ is consistent for every admissible profile $\mathbf{J} \in \mathcal{J}(\Phi)^n$. That is, Theorems 13 and 15 talk about properties of agendas that guarantee safety for at least one aggregator from a given class. Now we want to establish conditions under which *every* aggregator of a given class guarantees safety. We will see one example of such a universal agenda characterization theorem. As is to be expected, the conditions we have to impose on our agendas will be more restrictive than for existential agenda characterization theorems. We shall work with a restriction of the median property: agenda Φ is said to satisfy the *simplified median property* if every non-singleton minimally inconsistent subset of Φ consists of two formulas that are logical complements, i.e., every such set is of the form $\{\varphi, \psi\}$ with $\models \varphi \leftrightarrow \neg\psi$.²²

Theorem 16 (Endriss et al., 2012) *An agenda Φ is safe for all unanimous, anonymous, neutral, independent, complete and complement-free aggregators if and only if Φ has the simplified median property.*

Proof. First, suppose Φ has the simplified median property. Take any aggregator f that is unanimous, neutral, and complement-free (we will not need the other axioms) and for the sake of contradiction assume there exists a profile $\mathbf{J} \in \mathcal{J}(\Phi)^n$ such that $f(\mathbf{J})$ is inconsistent. By unanimity and complement-freeness, this inconsistency

²² An example of a set that satisfies the median property but not the simplified median property is $\{\neg p, p \wedge q\}$: the two formulas together are inconsistent, but they are not logical complements.

cannot be due to a single inconsistent formula (as no judge would accept that formula), so we must have $\{\varphi, \psi\} \subseteq f(\mathbf{J})$ with $\models \varphi \leftrightarrow \neg\psi$ for some $\varphi, \psi \in \Phi$. By completeness and consistency of the profile, φ and $\neg\psi$ must be accepted by the same coalition. Thus, by neutrality, $\varphi \in f(\mathbf{J})$ implies $\neg\psi \in f(\mathbf{J})$, i.e., $\{\psi, \neg\psi\} \subseteq f(\mathbf{J})$. But this contradicts complement-freeness of f . Hence, $f(\mathbf{J})$ cannot be inconsistent.

For the opposite direction, suppose Φ violates the simplified median property. If Φ also violates the (normal) median property, then we are done, as we already know that the majority rule is inconsistent (cf. Lemma 12) and satisfies all the required axioms (cf. Corollary 6). So, w.l.o.g., suppose Φ has a minimally inconsistent subset consisting of two formulas that are not logical complements, i.e., there exist $\varphi, \psi \in \Phi$ with $\varphi \models \neg\psi$ but $\neg\psi \not\models \varphi$. But now consider the parity rule f_{par} , accepting all those formulas that are accepted by an odd number of judges. It is easy to see that f_{par} is unanimous (as n is odd), anonymous, neutral, independent, complete and complement-free. However, f_{par} is not safe: If one judge accepts (the consistent set) $\{\varphi, \neg\psi\}$, one judge accepts (the consistent set) $\{\neg\varphi, \psi\}$, and all other judges (an odd number) accept (the consistent set) $\{\neg\varphi, \neg\psi\}$, then f will accept the inconsistent set $\{\varphi, \psi\}$. This concludes the proof. \square

Let us consider two examples. First, an agenda that consists solely of literals satisfies the simplified median property. Thus, any such agenda will be safe, not only for the majority rule, but for every aggregator that meets the axioms of Theorem 16. Second, consider the agenda $\{p, \neg p, p \wedge q, \neg(p \wedge q), r, \neg r\}$, which violates the simplified median property, and the following profile:

| | p | $p \wedge q$ | r |
|----------|-----|--------------|-----|
| Judge 1: | Yes | Yes | Yes |
| Judge 2: | No | No | Yes |
| Judge 3: | Yes | No | Yes |

All three judgment sets are consistent. Yet, if we aggregate using the parity rule f_{par} , we have to reject p but accept $p \wedge q$, which corresponds to an inconsistent judgement set. Thus, as predicted by Theorem 16, our agenda is not safe for f_{par} .

Suppose we have some limited information on the method of aggregation a group of agents is going to use (e.g., we might know that they will respect certain axioms). Results such as Theorem 16 allow us to give consistency guarantees in such situations. Unfortunately, however, respecting the simplified median property severely limits the expressive power of the JA framework, and verifying whether a given agenda satisfies the simplified median property is computationally intractable; more specifically, this problem is Π_2^P -complete (Endriss et al., 2012).

To conclude our discussion of agenda characterization results, let us briefly compare existential and universal theorems. For a given class \mathcal{F} of aggregators, the former speak about agenda properties ensuring *some* aggregators in \mathcal{F} are consis-

tent, while the latter speak about agenda properties ensuring that *all* of them are.²³ The former is a natural question from the perspective of economics. For instance, a mechanism designer who is working in a specific application domain (determining the agenda properties) and who wants to respect certain axioms (determining \mathcal{F}), must ask this question to find out whether her desiderata are feasible. The latter problem is more likely to surface in computer science. For instance, a system designer may only have partial knowledge of the decision making methods employed by the users of a platform she is providing (say, enough to determine \mathcal{F} , but not enough to single out a specific aggregator), but still wants to be able to issue guarantees against inconsistencies in the agreements forged over this platform.

17.5 Related Frameworks

In this section, we briefly review three approaches to collective decision making that are closely related to JA, namely belief merging, binary aggregation, and voting in combinatorial domains.

17.5.1 Belief Merging

In computer science, specifically in artificial intelligence and database theory, the problem of *belief merging* had been investigated already some time before the emergence of the modern literature on JA in philosophy and economics (see, e.g., Baral et al., 1992; Liberatore and Schaefer, 1998; Konieczny and Pino Pérez, 2002). Suppose we are given several knowledge bases encoding knowledge provided by experts in a logical language and we want to integrate these individual bases to obtain a large overall knowledge base. The naïve approach of simply taking the union of all individual bases will, in most cases, lead to an inconsistent knowledge base. One possible refinement is to choose a maximally consistent subset of the union (Baral et al., 1992), but this and similar approaches do not track which individual provided which formulas in the knowledge base, i.e., it is not possible to give equal weight to different individuals. The latter is possible using distance-based procedures (Konieczny and Pino Pérez, 2002), which are also briefly discussed in Chapter 9 (Lang and Xia, 2016).

There are important differences between JA and typical work in belief merging. In JA we have an *agenda* and we usually assume that every individual expresses a judgment regarding every formula in the agenda, while in belief merging the knowledge/belief bases coming from distinct individuals need not concern the same set of formulas. Furthermore, in JA we (implicitly) impose the same constraints on different individuals and the collective, while in belief merging we typically impose an *integrity constraint* to be satisfied by the merged base that need not be satisfied

²³ If \mathcal{F} is a singleton, then the two notions coincide, as for Lemma 12.

by every individual base. Like in JA, work in belief merging often evokes the notion of ‘*axiom*’ to define what makes a good procedure for merging, but most of the axioms used in the literature are inspired by work in *belief revision* (and typically are ‘outcome-oriented’, talking about consistency requirements) rather than social choice theory (where axioms tend to have more of an ‘agent-oriented’ flavor).

17.5.2 Binary Aggregation

Binary aggregation (see, e.g., Dokow and Holzman, 2010; Grandi and Endriss, 2013) has its origins in the work of Wilson (1975) and Rubinstein and Fishburn (1986) on abstract (algebraic) aggregation. In binary aggregation, each individual is asked to supply a vector of 0’s and 1’s of some fixed length, and we then have to aggregate this information into a single such vector. The application domain under consideration determines which vectors are *feasible*. This is very closely related to the formula-based framework of JA we have discussed in this chapter. For example, if the set of feasible vectors is $\{000, 010, 100, 111\}$, then this induces the same domain of aggregation as the agenda formulas p , q , and $p \wedge q$. In fact, in the literature the term ‘judgment aggregation’ is sometimes taken to encompass both JA in the narrow sense (i.e., formula-based JA) and binary aggregation.

In the economics literature the set of feasible vectors is usually assumed to be specified explicitly (Dokow and Holzman, 2010), while in the computer science literature the set of feasible vectors is usually defined implicitly by means of an *integrity constraint* expressed in the language of propositional logic (Grandi and Endriss, 2013). For example, the integrity constraint $x_1 \wedge x_2 \wedge x_3 \leftrightarrow x_4$ defines the domain $\{0000, 0010, 0100, 0110, 1000, 1010, 1100, 1111\}$. The advantage of using integrity constraints is that they provide a *compact representation* of the domain of aggregation, which is important when we are interested in algorithmic aspects of aggregation. Integrity constraints also allow for alternative characterizations of aggregators. For instance, an aggregator is unanimous if and only if it maps feasible profiles to feasible outcomes for every integrity constraint expressible as a conjunction of literals; and the only aggregators that ‘lifts’ *all* possible integrity constraints in this sense are the *generalised dictatorships*, i.e., those aggregators that amount to first selecting one of the individuals and then copying that individual’s input vector to produce the output (Grandi and Endriss, 2013).

17.5.3 Voting in Combinatorial Domains

When voting on complex matters, such as the composition of a committee or the outcome of a referendum involving a set of proposals, each of which has to be either accepted or rejected, we face problems not dissimilar to those faced in JA. In such a *multi-attribute election* we have a number of variables, each ranging over a finite domain, and we ask each voter for her preferences regarding the assignment

of values to these variables. For instance, in the committee example the variables are the seats on the committee and the values are the candidates. When each variable is binary (as is the case for referenda), we obtain a similar problem as in binary aggregation. However, a crucial difference is that in binary aggregation we assume that each individual can only communicate one vector (i.e., one set of variable assignments), while in general we may ask each individual to report a complex preference structure over the set of all possible outcomes. This raises interesting questions, not only regarding the aggregation of such information, but also regarding the compact representation of the preferences themselves.

Such questions are studied in the field of *voting in combinatorial domains*. Chapter 9 is devoted to this important topic (Lang and Xia, 2016).

17.6 Applications in Computer Science

In the introduction we have already alluded to the fact that JA has important applications in analytical jurisprudence and political philosophy. We want to conclude this chapter with a few words about possible *applications in computer science*. Maybe the most important of these, and certainly the one most often quoted as an example, is collective decision making in systems of autonomous software agents. In such a *multiagent system* we have several agents, which may have been designed and programmed by different developers and which may act on behalf of different users, and these agents need to interact, both cooperating and competing with each other. Social choice theory is part of the basic repertoire available for modeling the fundamental features of multiagent systems (Wooldridge, 2009; Shoham and Leyton-Brown, 2009). Given that important aspects of the agents themselves, such as their beliefs or goals, are often represented using logic, JA is a useful formalism for modeling agreements made by those agents.

Another important formal tool for modeling the dynamics of multiagent systems is the theory of *abstract argumentation*, widely studied in artificial intelligence, which is concerned with high-level representations of the relationships that hold between arguments for and against a given position. Given a network of arguments together with information on which argument attacks which other argument, different agents may hold competing views on which arguments to accept. JA then is a useful framework for modeling the process of finding a common view in such situations (see, e.g., Caminada and Pigozzi, 2011).

Besides multiagent systems, another potential application for JA is *information merging*, e.g., the merging of knowledge bases provided by different experts briefly discussed in Section 17.5.1. This includes the merging of *ontologies*, which play a central role in the development of the Semantic Web (Porello and Endriss, 2011).

JA and particularly the closely related binary aggregation are also relevant to *crowdsourcing*. Several areas in or related to computer science, e.g., computer vision or computational linguistics, rely on the availability of annotated data, such as

images labelled with object names or text corpora annotated with semantic information. Crowdsourcing platforms provide fast and cheap means for labeling large amounts of data using non-expert annotators. The annotations provided by different workers then need to be aggregated to obtain a collective annotation, a task that shares many similarities with JA (Endriss and Fernández, 2013).

It is important to stress that, at the time of writing, none of the applications of JA sketched here has been developed in depth. Still, JA certainly holds great potential for these, as well as other, areas of computer science.

17.7 Bibliographic Notes and Further Reading

In this section we supply additional bibliographic references and explanations regarding some of the material covered in this chapter. We also provide pointers to further reading on specific topics. Let us begin by pointing out that, while research on JA *proper* started with the work of List and Pettit (2002), the work of Guilbaud (1952) on what he called “the logical problem of aggregation” may be considered an early precursor of JA (Eckert and Monjardet, 2010).

Following the publication of the original impossibility theorem of List and Pettit (our Theorem 2), several other authors derived technically stronger results of a similar nature, replacing in particular anonymity by absence of dictatorships (see, e.g., Pauly and van Hees, 2006; Dietrich, 2007). Gärdenfors (2006) showed that giving up the requirement of completeness amounts to moving from dictatorships to oligarchic aggregators, i.e., aggregators where the outcome is the set of those propositions a fixed set of individuals completely agrees on (dictatorships and the intersection rule are extreme cases of such oligarchic rules). We refer to List (2012) for a systematic discussion on how to cope with impossibility results by relaxing certain desiderata, and to Grossi and Pigazzi (2014) for an in-depth discussion of proofs of impossibility results using the ultrafilter method.

In work integrating the classical axiomatic method with ideas from computer science, Nehama (2013) showed that relaxing the cornerstones of most impossibility theorems, consistency and independence, to *approximate* variants of these desiderata does not allow us to significantly improve on known negative results.

The simulation of preference aggregation in JA sketched in Section 17.2.2 follows Dietrich and List (2007c), who for notational convenience use predicate rather than just propositional logic (which, as we have seen, is not technically required). Also the idea of deriving Arrow’s Theorem from results in JA, as briefly mentioned at the end of the Section 17.4.2, is due to Dietrich and List (2007c).

While initial work in JA had focused on impossibility results and the axiomatic method, work on the design of concrete practically useful methods of aggregation only started several years later. Distance-based methods originally entered JA via the field of belief merging (Revesz, 1997; Konieczny and Pino Pérez, 2002; Pigazzi, 2006). We have only discussed two such methods here; several others are analyzed in

the work of Miller and Osherson (2009), Lang et al. (2011), and Lang and Slavkovik (2013). The premise-based rule has been part of the JA literature from the very beginning, due to its importance in the context of legal decision making in practice (Kornhauser and Sager, 1993). Other approaches for designed JA rules not discussed here include *support-based methods* (Lang et al., 2011; Porello and Endriss, 2011; Everaere et al., 2014), where aggregation is guided by the numbers of judges accepting a given formula, methods inspired by *scoring rules* in voting theory (Dietrich, 2014), and methods that copy the choice of the judge that, in some sense, is the '*most representative*' in the group (Endriss and Grandi, 2014).

The analysis of strategic behavior is a major topic in social choice theory. As we have discussed in Section 17.3.3, how to best model an individual's preferences, and thus her incentives, is debatable in JA. Our definition of strategic manipulation in terms of preferences induced by the Hamming distance is but one possibility (Dietrich and List, 2007a; Baumeister et al., 2013). Manipulation is not the only type of strategic behavior of interest in JA. Baumeister et al. (2011) have initiated a study of the computational complexity of related problems in JA, namely bribery and control problem. For the former, we ask whether a given budget suffices to bribe sufficiently many judges to obtain a particular outcome. An example for a control problem is the question of whether we can obtain a given outcome by removing at most k judges. Alon et al. (2013) have focussed on a specific control problem, where we ask whether a desired outcome can be enforced by conducting separate elections for certain subsets of the agenda.

In Section 17.4, we have seen that the normal and the simplified median property are both Π_2^P -complete (Endriss et al., 2012). For most other commonly used agenda properties, at the time of writing, no complexity results are known.

The agenda characterization results in Section 17.4.2 have been adapted to our framework of formula-based JA from original results in binary aggregation (Dokow and Holzman, 2010) and the 'property space' formulation of abstract aggregation (Nehring and Puppe, 2007). For a comprehensive review of existential agenda characterization results we refer to List and Puppe (2009). The ultrafilter method we employed in Section 17.4.2 is useful also in other areas of social choice theory, where it was pioneered by Kirman and Sondermann (1972).

Finally, JA uses logic to define the structure of the domain of aggregation. A broader review of uses of logic in computational social choice, e.g., to describe properties of aggregators in a logical language or to use techniques from automated reasoning to support the verification or discovery of results in social choice theory, is available elsewhere (Endriss, 2011).

Acknowledgements. I would like to thank Dorothea Baumeister, Franz Dietrich, Roosmarijn Goldbach, Umberto Grandi, Davide Grossi, Jérôme Lang, and Daniele Porello for their helpful feedback on earlier versions of this chapter.

References

- N. Alon, D. Falik, R. Meir, and M. Tennenholtz. Bundling attacks in judgment aggregation. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI-2013)*, 2013.
- S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.
- K. J. Arrow. *Social Choice and Individual Values*. John Wiley and Sons, 2nd edition, 1963. First edition published in 1951.
- C. Baral, S. Kraus, J. Minker, and V. S. Subrahmanian. Combining knowledge bases consisting of first-order theories. *Computational Intelligence*, 8(1):45–71, 1992.
- 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-2011)*. Springer-Verlag, 2011.
- D. Baumeister, G. Erdélyi, O. J. 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-2013)*. Springer-Verlag, 2013.
- D. Black. On the rationale of group decision-making. *The Journal of Political Economy*, 56(1):23–34, 1948.
- F. Brandt, M. Brill, and P. Harrenstein. Tournament solutions. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 3. Cambridge University Press, 2016.
- M. Caminada and G. Pigazzi. On judgment aggregation in abstract argumentation. *Autonomous Agents and Multiagent Systems*, 22(1):64–102, 2011.
- V. Conitzer and T. Walsh. Barriers to manipulation. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 6. Cambridge University Press, 2016.
- D. van Dalen. *Logic and Structure*. Springer-Verlag, 5th edition, 2013.
- B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 2nd edition, 2002.
- 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 and C. List. Strategy-proof judgment aggregation. *Economics and Philosophy*, 23(3):269–300, 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. Arrow’s Theorem in judgment aggregation. *Social Choice and Welfare*, 29(1):19–33, 2007c.
- F. Dietrich and C. List. Majority voting on restricted domains. *Journal of Economic Theory*, 145(2):512–543, 2010.
- E. Dokow and R. Holzman. Aggregation of binary evaluations. *Journal of Economic Theory*, 145(2):495–511, 2010.
- D. Eckert and B. Monjardet. Guilbaud’s 1952 theorem on the logical problem of aggregation. *Mathématiques et Sciences Humaines*, 48(189):19–35, 2010.
- U. Endriss. Logic and social choice theory. In A. Gupta and J. van Benthem, editors, *Logic and Philosophy Today*, volume 2, pages 333–377. College Publications, 2011.
- U. Endriss and R. Fernández. Collective annotation of linguistic resources: Basic principles and a formal model. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL-2013)*, 2013.
- U. Endriss and U. Grandi. Binary aggregation by selection of the most representative voter. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI-2014)*, 2014.
- U. Endriss, U. Grandi, and D. Porello. Complexity of judgment aggregation. *Journal of Artificial Intelligence Research (JAIR)*, 45:481–514, 2012.
- P. Everaere, S. Konieczny, and P. Marquis. Counting votes for aggregating judgments. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2014)*, 2014.
- F. Fischer, O. Hudry, and R. Niedermeier. Weighted tournament solutions. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 4. Cambridge University Press, 2016.
- P. Gärdenfors. A representation theorem for voting with logical consequences. *Economics and Philosophy*, 22(2):181–190, 2006.
- U. Grandi and U. Endriss. Lifting integrity constraints in binary aggregation. *Artificial Intelligence*, 199–200:45–66, 2013.
- D. Grossi and G. Pigozzi. *Judgment Aggregation: A Primer*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2014.
- G.-T. Guilbaud. Les théories de l’intérêt général et le problème logique de l’agrégation. *Économie Appliquée*, 5(4):501–584, 1952. English translation appears in *Journal Électronique d’Histoire des Probabilités et de la Statistique*, 4(1):57, 2008.
- E. Hemaspaandra, H. Spakowski, and J. Vogel. The complexity of Kemeny elections. *Theoretical Computer Science*, 349(3):382–391, 2005.
- J. Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959.
- A. P. Kirman and D. Sondermann. Arrow’s Theorem, many agents, and invisible dictators. *Journal of Economic Theory*, 5(3):267–277, 1972.
- S. Konieczny and R. Pino Pérez. Merging information under constraints: A logical framework. *Journal of Logic and Computation*, 12(5):773–808, 2002.

- L. A. Kornhauser and L. G. Sager. The one and the many: Adjudication in collegial courts. *California Law Review*, 81(1):1–59, 1993.
- J. Lang and M. Slavkovik. Judgment aggregation rules and voting rules. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT-2013)*. Springer-Verlag, 2013.
- J. Lang and L. Xia. Voting in combinatorial domains. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 9. Cambridge University Press, 2016.
- 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-XIII)*, 2011.
- P. Liberatore and M. Schaefer. Arbitration (or how to merge knowledge bases). *IEEE Transactions on Knowledge and Data Engineering*, 10(1):76–90, 1998.
- 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: A survey. In P. Anand, P. Pattanaik, and C. Puppe, editors, *Handbook of Rational and Social Choice*. Oxford University Press, 2009.
- K. O. May. A set of independent necessary and sufficient conditions for simple majority decisions. *Econometrica*, 20(4):680–684, 1952.
- I. McLean and A. B. Urken, editors. *Classics of Social Choice*. University of Michigan Press, 1995.
- M. K. Miller and D. Osherson. Methods for distance-based judgment aggregation. *Social Choice and Welfare*, 32(4):575–601, 2009.
- I. Nehama. Approximately classic judgement aggregation. *Annals of Mathematics and Artificial Intelligence*, 68(1-3):91–134, 2013.
- K. Nehring and C. Puppe. The structure of strategy-proof social choice. Part I: General characterization and possibility results on median spaces. *Journal of Economic Theory*, 135(1):269–305, 2007.
- M. Pauly and M. van Hees. Logical constraints on judgement aggregation. *Journal of Philosophical logic*, 35(6):569–585, 2006.
- 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 of paradoxes of judgment aggregation. *Synthese*, 152(2):285–298, 2006.
- D. Porello and U. Endriss. Ontology merging as social choice. In *Proceedings of the 12th International Workshop on Computational Logic in Multiagent Systems (CLIMA-2011)*. Springer-Verlag, 2011.
- P. Z. Revesz. On the semantics of arbitration. *International Journal of Algebra and Computation*, 7(2):133–160, 1997.
- A. Rubinstein and P. C. Fishburn. Algebraic aggregation theory. *Journal of Economic Theory*, 38(1):63–77, 1986.
- Y. Shoham and K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2009.
- P. Slater. Inconsistencies in a schedule of paired comparisons. *Biometrika*, 48(3–4):303–312, 1961.

- R. Wilson. On the theory of aggregation. *Journal of Economic Theory*, 10(1): 89–99, 1975.
- M. Wooldridge. *Multiagent Systems*. John Wiley and Sons, 2nd edition, 2009.
- W. S. Zwicker. Introduction to voting theory. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 2. Cambridge University Press, 2016.