



HAL
open science

Introduction to Judgment Aggregation

Davide Grossi, Gabriella Pigozzi

► **To cite this version:**

Davide Grossi, Gabriella Pigozzi. Introduction to Judgment Aggregation. ESSLLI 2011, Aug 2011, Ljubljana, Slovenia. pp.160-209, 10.1007/978-3-642-31485-8_5 . hal-01497075

HAL Id: hal-01497075

<https://hal.science/hal-01497075>

Submitted on 28 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Introduction to Judgment Aggregation

*Lecture Notes for ESSLLI'11*¹

Davide Grossi[†] and Gabriella Pigozzi[‡]

[†] ILLC - University of Amsterdam
d.grossi@uva.nl

[‡] LAMSADE - Université Paris Dauphine
gabriella.pigozzi@lamsade.dauphine.fr

¹introJA.tex, Tuesday 26th July, 2011, 20:33.

Abstract

The present notes serve as material for the course *Introduction to Judgment Aggregation* to be given at the 23rd European Summer School on Logic, Language and Information (ESSLLI'11, Ljubljana).

Comments are very much welcome!

Acknowledgments.

We would like to thank Umberto Grandi for useful comments that helped us improving the present version of these notes. Davide Grossi wishes to acknowledge support by the *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* under the NWO VENI grant 639.021.816.

Contents

1	Logic meets Social Choice	3
1.1	A <i>social</i> view on logic	3
1.1.1	From the doctrinal paradox to the discursive dilemma . . .	3
1.1.2	Preference aggregation and judgment aggregation	5
1.2	Preliminary notions	7
1.2.1	Propositional languages	7
1.2.2	Agendas and individual judgments	7
1.2.3	Aggregation functions	8
2	Impossibility	11
2.1	Preliminaries	11
2.1.1	Agenda conditions	11
2.1.2	Aggregation conditions	15
2.2	An impossibility theorem	16
2.2.1	Winning coalitions	16
2.2.2	Lemmata	17
2.2.3	The theorem	20
2.3	More on impossibility	21
2.3.1	The structure of impossibility theorems in JA	21
2.3.2	Other impossibility results	22
2.3.3	Impossibility via ultrafilters	22
3	Coping with impossibility	24
3.1	Relaxing the input constraints	25
3.1.1	Unidimensional alignment	25
3.1.2	Value-restriction	26
3.2	Relaxing the output constraints	27
3.2.1	Abstention	27
3.2.2	Quota rules	27
3.3	Relaxing independence	28
3.3.1	The premise-based approach	29
3.3.2	The sequential priority approach	30
3.3.3	The distance-based rules	32
4	Manipulability	34
4.1	Agenda manipulation	34
4.2	Vote manipulation	35
4.2.1	A case of manipulation in the $\pm\{p, q, p \wedge q\}$ agenda	35

4.2.2	Manipulability: definition and a characterization	36
4.3	The impossibility of non-manipulability	38
4.3.1	Preliminaries	38
4.3.2	Lemmata	38
4.3.3	The theorem	41
4.4	Strategic issues in JA	41
4.4.1	Preferences over judgment sets	42
4.4.2	Strategy-proofness and manipulability	42
5	The research agenda of JA	43
5.1	JA vs. PA	43
5.2	Aggregation under integrity constraints	44
5.3	Agenda safety	45
5.4	JA and abstract argumentation	45

Chapter 1

Logic meets Social Choice

Outline of the chapter. We start by presenting the paradox that originated the whole field of judgment aggregation (Section 1.1.1). 1.1.2 looks at how judgment aggregation relates to the older theory of preference aggregation, and how the doctrinal paradox relates to the classical Condorcet paradox. Section 1.2 formally introduces the three central notions in the theory of judgment aggregation, namely agendas, judgment sets and aggregation functions.

1.1 A *social* view on logic

1.1.1 From the doctrinal paradox to the discursive dilemma

The idea that groups make better decisions than individuals dates back to 18th century social theorists like Rousseau and Condorcet. However, as we will see in much detail, majority voting—the exemplary democratic aggregation rule—is unable to ensure consistent social positions under all situations—this is the bottom line of the now famous Condorcet paradox, to which we will turn in Section 1.1.2 below.

Whereas voting theory studies the aggregation of individual preferences, the recent theory of judgment aggregation investigates how individual opinions on logically related propositions can be consistently aggregated into a collective position. Judgment aggregation has its roots in jurisprudence, building on the *doctrinal paradox* that Kornhauser and Sager discovered in the decision making procedure of collegial courts [43, 44, 42].

It is instructive to recall the original example that Kornhauser and Sager used to illustrate the doctrinal paradox [44]. A three-member court has to reach a verdict in a breach of contract case between a plaintiff and a defendant. According to the contract law, the defendant is liable (the *conclusion*, here denoted by proposition r) if and only if there was a valid contract and the defendant was in breach of it (the two *premises*, here denoted by propositions p and q respectively). Suppose that the three judges cast their votes as in Table 1.1.

The court can rule on the case either directly, by taking the majority vote on the conclusion r (*conclusion-based procedure*) or indirectly, by taking the judges' recommendations on the premises and inferring the court's decision on r via the rule $(p \wedge q) \leftrightarrow r$ that formalizes the contract law (*premise-based*

	Valid contract p	Breach q	Defendant liable r
Judge 1	1	1	1
Judge 2	1	0	0
Judge 3	0	1	0
Majority	1	1	0

Table 1.1: An illustration of the doctrinal paradox

	Valid contract p	Breach q	Legal doctrine $(p \wedge q) \leftrightarrow r$	Defendant liable r
Judge 1	1	1	1	1
Judge 2	1	0	1	0
Judge 3	0	1	1	0
Majority	1	1	1	0

Table 1.2: The discursive dilemma

procedure). The problem is that the court’s decision depends on the procedure adopted. Under the conclusion-based procedure, the defendant will be declared not liable, whereas under the premise-based procedure, the defendant would be sentenced liable. As Kornhauser and Sager stated:

We have no clear understanding of how a court should proceed in cases where the doctrinal paradox arises. Worse, we have no systematic account of the collective nature of appellate adjudication to turn to in the effort to generate such an understanding. [44, p. 12]

The systematic account to the understanding of situations like the one in Table 1.1 has been provided by judgment aggregation. The first step was made by political philosopher Pettit [62], who recognized that the paradox illustrates a more general problem than a court decision. Pettit introduced the term *discursive dilemma* to indicate a group decision in which propositionwise majority voting on related propositions may yield an inconsistent collective judgment.

Then, List and Pettit [50] reconstructed Kornhauser and Sager’s example as shown in Table 1.2. By adding the legal doctrine to the set of issues on which the judges have to vote, List and Pettit attained a great generality which provided clear analytical advantages. The discursive dilemma is characterized by the fact that the group reaches an inconsistent decision, like $\{p, q, (p \wedge q) \leftrightarrow r, \neg r\}$ in Table 1.2.

Is the move from the doctrinal paradox to the discursive dilemma an innocent one? Recently, Mongin and Dietrich [56, 55] have investigated such shift and observed that:

[T]he discursive dilemma shifts the stress away from the conflict of methods to *the logical contradiction within the total set of propositions that the group accepts*. [...] Trivial as this shift seems, it has far-reaching consequences, because all propositions are now being

treated alike; indeed, the very distinction between premisses and conclusions vanishes. This may be a questionable simplification to make in the legal context, but if one is concerned with developing a general theory, the move has clear analytical advantages. [55, p. 2]

Instead of premisses and conclusions, List and Pettit focused their attention to *judgment sets*, the sets of propositions that individuals accept. The theory of judgment aggregation becomes then a formal investigation of the conditions under which consistent individual judgment sets can collapse into an inconsistent judgment set. Their approach combines a logical formalization of the judgment aggregation with an axiomatic approach in the spirit of Arrow's social choice theory. The first question they can address is how general the judgment aggregation problem is, that is, whether the culprit is majority voting or whether the dilemma arises also with other aggregation rules. They obtained a first general impossibility theorem stating that there exists no aggregation rule that satisfies few desirable conditions and that can ensure a consistent collective outcome (see Chapter 2).

1.1.2 Preference aggregation and judgment aggregation

How individual preferences can be aggregated into a collectively preferred alternative is traditionally studied by social choice theory [1, 67]. The *Condorcet paradox* shows how challenging the problem is: given a set of individual preferences, we compare each of the alternatives in pairs. For each pair, we determine the winner by majority voting, and the final collective ordering is obtained by a combination of all partial results. Unfortunately, this method can lead to cycles in the collective ordering.

Let a set of alternatives X . For all $x, y \in X$, $x < y$ denotes that alternative x is strictly preferred to y . The desired properties of preference relations viewed as strict linear orders are:

- (P1) $\forall x, y((x < y) \rightarrow \neg(y < x))$ (asymmetry)
(P2) $\forall x, y(x \neq y \rightarrow (x < y \vee y < x))$ (connectedness)
(P3) $\forall x, y, z((x < y \wedge y < z) \rightarrow x < z)$ (transitivity)

Example 1 (Condorcet paradox). *Suppose that there are three possible alternatives x, y and z and three voters V_1, V_2 and V_3 , who express their total preferences in the following way: $V_1 = \{x < y, y < z\}$, $V_2 = \{y < z, z < x\}$ and $V_3 = \{z < x, x < y\}$. According to Condorcet's method, $x < y$ has the majority of the voters (V_1 and V_3), so does $y < z$ (V_1 and V_2) and also $z < x$ (V_2 and V_3). This leads us to the collective outcome $x < y, y < z$ and $z < x$, which together with transitivity (P3) violates (P1). This is an instance of the so-called Condorcet paradox.¹*

Arrow [1] showed that this is not a problem specific of pairwise majority comparison. In his famous impossibility theorem, Arrow proved that, when there are three or more alternatives, the only aggregation procedures that satisfy few desirable properties (like the absence of cycles in the collective preference) are dictatorial ones.

¹We will come back later in Chapter 2 to another simple formalization of the Condorcet paradox (Example 6).

The lesson is that, when we combine individual choices into a collective one, we may lose something that held at the individual level, like transitivity or logical consistency.

A natural question is how the theory of judgment aggregation is related to the theory of preference aggregation. We can address this question in two ways: we can consider what are the possible interpretations of aggregating judgments and preferences, and we can investigate the formal relations between the two theories. On the first consideration, Kornhauser and Sager see the possibility of being right or wrong as the discriminating factor:

When an individual expresses a preference, she is advancing a limited and sovereign claim. The claim is limited in the sense that it speaks only to her own values and advantage. The claim is sovereign in the sense that she is the final and authoritative arbiter of her preferences. The limited and sovereign attributes of a preference combine to make it perfectly possible for two individuals to disagree strongly in their preferences without either of them being wrong. [...] In contrast, when an individual renders a judgment, she is advancing a claim that is neither limited nor sovereign. [...] Two persons may disagree in their judgments, but when they do, each acknowledges that (at least) one of them is wrong. [43, p. 85]²

Regarding the formal relations between judgment and preference aggregation, Dietrich and List [13] (extending an earlier work by List and Pettit [51]) show that Arrow's theorem for strict and complete preferences can be derived from an impossibility result in judgment aggregation.

In order to represent preference relations, they consider a first-order language with a binary predicate P representing strict preference for all $x, y \in X$. The inference relation is enriched with the asymmetry, connectedness and transitivity axioms of P . The Condorcet paradox can then be represented as a judgment aggregation problem as Table 1.3 illustrates. Voters of Table 1.3 are perfectly consistent likewise the judges of Table 1.1. The difference is that in the doctrinal paradox individuals are consistent in terms of propositional logic, while in the Condorcet example consistency corresponds to the transitive and complete conditions imposed on preferences. However, it is worth mentioning that Kornhauser and Sager [42] notice that the doctrinal paradox resembles the Condorcet paradox, but that the two paradoxes are not equivalent. Indeed, as stated also by List and Pettit:

[W]hen transcribed into the framework of preferences instances of the discursive dilemma do not always constitute instances of the Condorcet paradox; and equally instances of the Condorcet paradox do not always constitute instances of the discursive dilemma. [51, pp. 216-217]

Given the analogy between the Condorcet paradox and the doctrinal paradox, List and Pettit' first question was whether an analogous of Arrow's theorem could be found for the judgment aggregation problem. Arrow showed that the

²Different procedures for judgment aggregation have been assessed with respect to their truth-tracking capabilities, see [3, 35].

	xPy	yPz	xPz	yPx	zPy	zPx
$V_1 = \{x < y, y < z\}$	1	1	1	0	0	0
$V_2 = \{y < z, z < x\}$	0	1	0	1	0	1
$V_3 = \{z < x, x < y\}$	1	0	0	0	1	1
Majority	1	1	0	0	0	1

Table 1.3: The Condorcet paradox as a doctrinal paradox

Condorcet paradox hides a much deeper problem that does not affect only majority rule. The same question could be posed in judgment aggregation: is the doctrinal paradox only the surface of a more troublesome problem arising when individuals cast judgments on a given set of propositions? The answer to this question is positive and that was the starting point of the new theory of judgment aggregation. We will mention some recent work investigating the formal similarities and differences between preference and judgement aggregation in Chapter 5. It is now time to introduce some formal definitions.

1.2 Preliminary notions

In this section we introduce the three central notions underlying the formal theory of judgment aggregation: agendas, judgment sets and aggregation functions.

1.2.1 Propositional languages

We will work with the aggregation of judgments formulated in a standard propositional language:

$$\varphi := p \in X \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \varphi \rightarrow \varphi$$

where X is a set of atoms.

1.2.2 Agendas and individual judgments

The following is the first key definition of the framework of JA:

Definition 1 (Judgment aggregation structure). *Let $\mathcal{L}(X)$ be a propositional language on a given set of atoms X .³ A judgment aggregation structure for \mathcal{L} is a tuple $\mathcal{J} = \langle N, A \rangle$ where:*

- ▶ N is a non-empty set of agents;
- ▶ $A \subseteq \mathcal{L}$ (the agenda) such that $A = \{\varphi \mid \varphi \in I\} \cup \{\neg\varphi \mid \varphi \in I\}$ for some $I \subseteq \mathcal{L}$ (the set of issues) which contains only positive (i.e., non-negated) contingent⁴ formulae. An agenda based on a set of issues I will often be denoted $\pm I$.

³We will often drop the reference to X when clear from the context.

⁴I.e., which are neither a tautology nor a contradiction.

In other words, the agenda is a set of formulae which is closed under complementation, i.e., $\forall\varphi: \varphi \in A$ iff $\neg\varphi \in A$, and which contains only formulae with at most one negation.

Definition 2 (Judgment sets and profiles). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure. A judgment set for \mathcal{J} is a set of formulae $J \subseteq X$ such that:*

- ▶ *J is consistent, i.e., it has a model;*
- ▶ *J is maximal (or complete) w.r.t. A , i.e., $\forall\varphi \in A$, either $\varphi \in J$ or $\neg\varphi \in J$.*

The set of all judgment sets is denoted $\mathbf{J} \subseteq \wp(X)$. A judgment profile $P = \langle J_i \rangle_{i \in N}$ is a $|N|$ -tuple of judgment sets. We denote with \mathbf{P} the set of all judgment profiles.

That a formula φ follows from a judgment set J will be denoted $J \models \varphi$. For a φ in the agenda, the same notation will be often used interchangeably with $\varphi \in J$ to indicate that φ belongs to J . Slightly abusing notation, we will often indicate that a judgment set J_i belongs to a profile P by writing $J_i \in P$.

Remark 1. *If the agenda A is closed under atoms, i.e., it contains all the atomic variables occurring in their formulae, then each judgment set J corresponds to a propositional valuation. More precisely, let X be the set of atoms occurring in the formulae in A . First of all, note that $A \subseteq \mathcal{L}(X)$ and each judgment set J corresponds to a function $J : X \rightarrow \{1, 0\}$.*

1.2.3 Aggregation functions

The aggregation of individual judgments into a collective one is viewed as a function:

Definition 3 (Aggregation function). *Let \mathcal{J} be a JA structure. An aggregation function for \mathcal{J} is a function $f : \mathbf{P} \rightarrow \mathbf{J}$.*

Notice that the function takes as domain the set of all possible judgment profiles—the so-called *universal domain* condition. Often, the collective judgment set $f(P)$ resulting from the aggregation of a profile $P = \langle J_i \rangle_{i \in N}$ via f is simply referred to as J .

To substantiate our presentation, let us now give some concrete examples of ways of aggregating judgment profiles, which are arguably of common use, and which we call here *aggregation procedures*:

Propositionwise majority:

$$\varphi \in f_{maj}(P) \quad \text{IFF} \quad |\{J_i \in P \mid \varphi \in J_i\}| > q, \quad (1.1)$$

with $q = \frac{|N|+1}{2}$ if $|N|$ is even, and $q = \frac{|N|}{2}$ if $|N|$ is odd.⁵

I.e., φ is collectively accepted iff there is a majority of voters accepting it.

Propositionwise unanimity:

$$\varphi \in f_u(P) \quad \text{IFF} \quad |\{J_i \in P \mid \varphi \in J_i\}| = |N| \quad (1.2)$$

I.e., φ is collectively accepted iff all voters accept it.

⁵By varying q between 1 and $|N|$, we can also define specific quota rules [14].

Premise-based procedure: In the premise-based procedure only the individual judgments on the premises are aggregated. The collective judgment on the conclusion is determined by entailment from the group decisions on the premises. Let us denote by $Prem \subseteq A$ the subagenda containing the propositions that are premises and their complements, and by $Conc \subseteq A$ the subagenda containing the conclusions and their complements, such that $A = Prem \cup Conc$. Following the literature, we assume that the aggregation rule is the above propositionwise majority.

$$f_{pbp}(P) = f_{maj}(Prem) \cup \{\varphi \in Conc \mid f_{maj}(Prem) \models \varphi\} \quad (1.3)$$

I.e., φ is collectively accepted iff it is a premise and it has been voted by the majority of the individuals or it is a conclusion entailed by the collectively accepted premises.

Conclusion-based procedure: In the conclusion-based procedure only the individual judgments on the conclusions are aggregated. This implies that there will be no group position on the premises. Again, we assume the aggregation rule is propositionwise majority.

$$f_{cbp}(P) = f_{maj}(Conc) \quad (1.4)$$

I.e., φ is collectively accepted iff it is a conclusion and it has been voted by the majority of the individuals.

Remark 2. *Are the above aggregation procedures aggregation functions in the precise sense of Definition 3? A simple inspection of the definitions of the procedures will show that the answer is no. In the light of the doctrinal paradox and the discursive dilemma, it must have already been clear that propositionwise majority cannot be an aggregation function—and this is yet another way of ‘phrasing’ those paradoxes. It is, however, almost an aggregation function: it can either be viewed as a partial function from \mathbf{P} to \mathbf{J} , or as a function from \mathbf{P} to the set of all possibly inconsistent judgment sets. Similar considerations can be made for the other procedures mentioned above. The discrepancy between these procedures and the ‘idealized’ notion of aggregation function can well be viewed as the symptom of some deep difficulty involved with the aggregation of individual opinions. What such a deep difficulty is will be investigated in detail in the next chapter.*

We conclude this chapter with one more variant of the doctrinal paradox:

Example 2. *Let $A = \pm\{p, p \rightarrow q, q\}$. In the literature this agenda is often associated with the following propositions [13]:*

- p : *Current CO₂ emissions lead to global warming.*
- $p \rightarrow q$: *If current CO₂ emissions lead to global warming, then we should reduce CO₂ emissions.*
- q : *We should reduce CO₂ emissions.*

The profile consisting of the three judgment sets $J_1 = \{p, p \rightarrow q, q\}$, $J_2 = \{p, \neg(p \rightarrow q), \neg q\}$ and $J_3 = \{\neg p, p \rightarrow q, \neg q\}$, once aggregated via propositionwise majority, gives rise to an inconsistent collective judgment set $J = \{p, p \rightarrow$

$q, \neg q$. If we assume that $Prem = \{p, p \rightarrow q\}$ and $Conc = \{q\}$, we can also see the outcomes of the premise and of the conclusion-based procedures. Summarizing this into a table:⁶

	p	$p \rightarrow q$	q
J_1	1	1	1
J_2	1	0	0
J_3	0	1	0
f_{maj}	1	1	0
f_{ppp}	1	1	1
f_{cbp}			0

Notice that f_u , propositionwise unanimity, would yield no judgment set.

⁶Notice that the agenda at issue is closed under atoms, hence each judgment set can be viewed as a propositional valuation function (cf. Remark 1).

Chapter 2

Impossibility

Is the discursive dilemma just a problem of propositionwise majority voting? Or is it the symptom of a widespread feature that characterize all seemingly ‘reasonable’ methods of aggregation? The present chapter shows that the latter is the right answer.

Outline of the chapter. Section 2.1 provides some preliminaries and introduces a number of properties of aggregation functions and agendas—which will be of use in this, and later chapters. Section 2.2 presents and proves a theorem—a so-called *impossibility theorem*—showing that rather undemanding conditions on the agenda and the aggregation function force the aggregation to be dictatorial. Section 2.3 discusses the result and its proof further, and provides pointers to other similar results in the JA literature. The chapter builds on material and results taken from [13, 37, 48, 60].

2.1 Preliminaries

The discursive dilemma unveils a problematic aspect of propositionwise majority voting. In this chapter we will show that analogous problems are bound to arise whenever the agenda and the aggregation functions are taken to satisfy some apparently reasonable and appealing conditions.

2.1.1 Agenda conditions

We state three special conditions on agendas, which capture, abstractly, the sort of logical interdependence possibly existing between elements of the agenda. Such interdependences lie at the core of the impossibility theorems we will discuss in this chapter and in Chapter 4.

Non-simplicity

The first one is almost self-explicated, and is sometimes called *non-simplicity*.

Definition 4 (Non-simple agendas). *An agenda A is non-simple (NS) iff it contains at least one set X s.t.:*

- ▶ $3 \leq |X|$;

► X is minimally inconsistent, i.e.:

- X is inconsistent;
- $\forall Y \subset X: Y$ is consistent.

It is easy to see that agenda $\pm\{p, q, p \wedge q\}$ is non simple as the set $\{p, q, \neg(p \wedge q)\}$ is clearly minimally inconsistent. Non-simplicity is the minimal level of complexity for an agenda to run into problems when attempting aggregation. If an agenda is simple, then the propositionwise majority offers a viable procedure.

If X is minimally inconsistent then, for some $\varphi \in X$ it is not only the case that $X - \{\varphi\}$ is consistent, but, clearly, also that $X - \{\varphi\} \models \neg\varphi$. In fact, non-simplicity is related to logical consequence in the following way:

Definition 5 (Conditional entailment). *Let $\varphi, \psi \in A$ s.t. $\varphi \neq \neg\psi$. We say that φ conditionally entails ψ (notation: $\varphi \models_c \psi$) if for some (possibly empty) $X \subseteq A - \{\varphi, \psi\}$, which is consistent with both φ and ψ , $\{\varphi\} \cup X \models \psi$.*

Clearly, if $X \neq \emptyset$ and $\{\varphi\} \cup X \models \psi$ then $X \cup \{\varphi, \neg\psi\}$ is minimally inconsistent and vice versa. It follows that, if an agenda is non simple then it contains two formulae φ, ψ such that $\varphi \models_c \psi$.

Even negations

The second agenda condition is slightly more involved.¹

Definition 6 (Evenly negatable agendas). *An agenda A satisfies the even negations condition (EN) iff:*

- A is non-simple
- A contains a minimally inconsistent set $X \subseteq A$ of even cardinality s.t. it contains at least two formulae which, if negated, make the set consistent.

Notice that this condition requires the existence of two (possibly distinct!) sets: one that witnesses non-simplicity, and one that is minimally inconsistent and can be made consistent by negating at least two of its elements.

Again, it is easy to see that the agenda $\pm\{p, q, p \wedge q\}$ satisfies this property, as well as $A = \pm\{p, q, p \rightarrow q\}$, but not all agendas do:

Example 3 (Non evenly negatable agendas). *Consider $A = \pm\{p, q, p \leftrightarrow q\}$. We have a minimally inconsistent set $X = \{p, \neg q, p \leftrightarrow q\}$. This is clearly such that no consistent set can be obtained by negating two formulae in the set at the same time.*

Example 4 (Evenly negatable agendas). *Consider the agenda (which we have encountered in Chapter 1) $\pm\{a < b, b < c, c < a\}$ under the following assumptions for $x, y \in \{a, b, c\}$ s.t. $x \neq y$:*

- (P1) $x \succ y \rightarrow \neg(y \succ x)$
- (P2) $(x \succ y \vee y \succ x) \wedge \neg(x \succ y \wedge y \succ x)$
- (P3) $(x \succ y \wedge y \succ z) \rightarrow x \succ z$

¹In [13] the condition is referred to as *minimal connectedness*, in [48] as even number negatability.

which simulate the linear order conditions on a finite propositional language.² This interesting agenda satisfies EN as set $\{a \succ b, b \succ c, c \succ a\}$ shows, which is minimally inconsistent and can be made consistent by swapping the first and third elements.

Path connectedness

The third agenda condition is known as *path connectedness* (e.g. in [15]) or *total-blockedness* (e.g. in [58]).

Definition 7 (Path-connected agendas). *An agenda A is path connected (PC) iff for all $\varphi, \psi \in A$ there exist a sequence $\varphi_1, \dots, \varphi_n$ of elements of A s.t.: $\varphi = \varphi_1$, $\psi = \varphi_n$ and $\varphi_i \models_c \varphi_{i+1}$ for $1 \leq i < n$.*

In other words, we call an agenda path connected whenever any two formulae in the agenda are logically connected in either a directed or an undirected way by fixing the truth value of some other formula in the agenda. It might be instructive to notice that PC is equivalent to the requirement that the transitive closure of the conditional entailment relation (Definition 5) covers the cartesian square of the agenda, i.e.: $(\models_c)^* = A \times A$.

Path connectedness is a strong agenda condition. As it turns out, all the agendas we have considered above are not path connected.

Example 5 (Path dis-connected agendas). *The agenda $\pm\{p, q, p \wedge q\}$ is not path-connected. This can be appreciated by noticing that no negative proposition conditionally entails a positive proposition. Figure 2.1 displays the directed graph of conditional entailment for this agenda. Similar considerations can be made for agendas $\pm\{p, q, p \rightarrow q\}$ and $\pm\{p, q, p \vee q\}$.*³

Example 6 (Path connected agendas). *Consider the agenda we have introduced in Example 4. This agenda satisfies PC, as its conditional entailment graph in Figure 2.2 shows. Another agenda, although with 8 elements, satisfying PC which we have encountered in the previous chapter is the one of the discursive dilemma: $\pm\{p, q, r, r \leftrightarrow (p \wedge q)\}$.*

Comparing agenda conditions

The above agenda conditions are logically related as follows:

- EN implies NS;
- PC implies NS.

Conditions PC and EN have a non-empty intersection (e.g. the discursive dilemma agenda $\pm\{p, q, r, r \leftrightarrow (p \wedge q)\}$) but are logically unrelated. For instance, agenda $\pm\{p, q, p \rightarrow q\}$ satisfies EN but not PC. Vice versa, $\pm\{p, q, p \leftrightarrow q\}$ satisfies PC but not EN.

This concludes our presentation of the most common structural conditions on agendas considered in the JA literature, and we now move further to a discussion of the conditions that can be imposed on the aggregation function.

²It might be instructive to compare the first-order formulation of these conditions given in Chapter 1.

³Recall that $p \rightarrow q$ is $\neg(p \wedge \neg q)$ and $p \vee q$ is $\neg(\neg p \wedge \neg q)$.

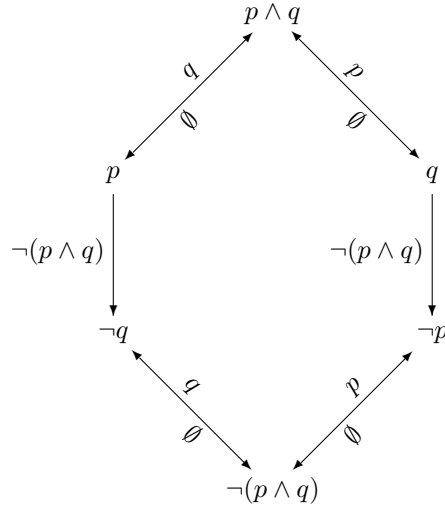


Figure 2.1: The conditional entailment graph of agenda $\pm\{p, q, p \wedge q\}$. The lower elements are not connected to the upper elements. The arrows are labeled with (the elements of) the set X which establishes the conditional entailment.

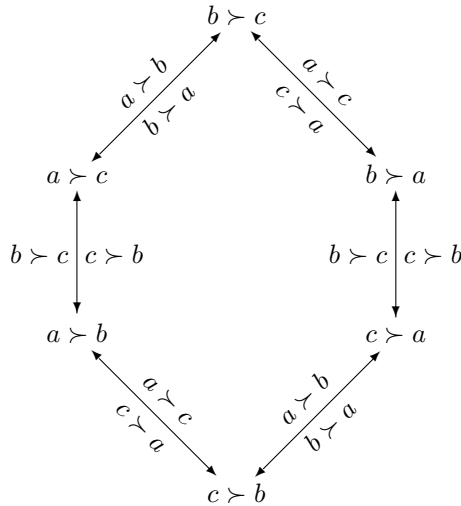


Figure 2.2: The conditional entailment graph of agenda $\pm\{a < b, b < c, c < a\}$. Each formula is reachable from any other formula.

2.1.2 Aggregation conditions

Let us fix first some auxiliary terminology, which will help us to streamline our notation in the coming sections.

Definition 8. We say that, for $J, J' \in \mathbf{J}$ and $P, P' \in \mathbf{P}$:

- ▶ J agrees with J' on formula φ (notation: $J =_{\varphi} J'$) iff $[J \models \varphi \text{ iff } J' \models \varphi]$;
- ▶ P is an i -variant of P' (notation: $P =_{-i} P'$) iff $\forall j \neq i : P_j = P'_j$.

By $J \neq_{\varphi} J'$ we indicate that J does not agree with J' on φ , and by $P \neq_{-i} P'$ that P and P' are not i variants of one another.

These are among the most common conditions on aggregation functions dealt with in the literature on JA:

Definition 9 (Aggregation conditions). Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure and $X \subseteq A$. An aggregation function f is:

Unanimous (U) iff $\forall \varphi \in A, \forall P \in \mathbf{P} : \text{IF } [\forall i \in N : P_i \models \varphi] \text{ THEN } f(P) \models \varphi$.
I.e., if all voters agree on accepting φ , so does also the aggregated judgment set.

Independent (IND) iff $\forall \varphi \in A, \forall P, P' \in \mathbf{P} : \text{IF } [\forall i \in N : P_i \models \varphi \text{ IFF } P'_i \models \varphi] \text{ THEN } f(P) =_{\varphi} f(P')$.
I.e., if all voters in two different profiles agree on the acceptance or rejection of some formula, the aggregated judgments of the two profiles also agree on the acceptance or rejection of the formula.

Systematic (SYS) iff $\forall \varphi, \psi \in A, \forall P, P' \in \mathbf{P} : \text{IF } [\forall i \in N : P_i \models \varphi \text{ IFF } P'_i \models \psi] \text{ THEN } [f(P) \models \varphi \text{ IFF } f(P') \models \psi]$.
I.e., if all voters in two different profiles agree on the acceptance or rejection pattern of two formulae (φ is accepted iff ψ is accepted), the aggregated judgments of the two profiles also do.

Dictatorial (D) iff $\exists i \in N \text{ s.t. } \forall \varphi \in A, \forall P \in \mathbf{P} : f(P) \models \varphi \text{ IFF } P_i \models \varphi$.
I.e., there exist a voter whose judgment set is always identical to the aggregated set.

Responsive (RES) iff $\forall \varphi \in A, \exists P, P' \in \mathbf{P} \text{ s.t. } f(P) \models \neg \varphi \text{ AND } f(P') \models \varphi$.
I.e., any formula can possibly be collectively accepted.

Monotonic (MON) iff $\forall \varphi \in A, \forall i \in N, \forall P, P' \in \mathbf{P} : \text{IF } [P =_{-i} P' \text{ AND } P_i \not\models \varphi \text{ AND } P_i \models \varphi] \text{ THEN } [\text{IF } f(P) \models \varphi \text{ THEN } f(P') \models \varphi]$.
I.e., if the collective judgment accepts a formula, then letting one of the voters that rejects that formula switch to acceptance does not modify the collective judgment.

These conditions state some very diverse constraints on how the aggregation relates the input—a strategy profile—to the output—a judgment set. Note that all these conditions have some appeal to our intuitions of what counts as a ‘fair’ or ‘reasonable’ aggregation process. The gist of JA impossibility theorems—one of which will be discussed in detail in the next section—consists in showing that seemingly innocuous combinations of these conditions lead to unacceptable consequences.

Remark 3 (SYS vs IND). Recall that each judgment set can be seen as a valuation $J : A \rightarrow \{1, 0\}$ accepting or rejecting each agenda item. A judgment profile can therefore be viewed as a tuple of such valuations. Consider now two such profiles P and P' , for n voters and an agenda of m elements. Each of them generates a matrix of 1 (acceptance) or 0 (rejection):

$$\begin{array}{cccccc} P_1(\varphi_1) & \dots & P_1(\varphi_m) & & P'_1(\varphi_1) & \dots & P'_1(\varphi_m) \\ \dots & \dots & \dots & & \dots & \dots & \dots \\ P_n(\varphi_1) & \dots & P_n(\varphi_m) & & P'_n(\varphi_1) & \dots & P'_n(\varphi_m) \end{array}$$

where $P_i(\varphi_j)$ is either value 1 (acceptance) or 0 (rejection).

What **IND** states is that, for any column i , if P_i consists of the same sequence of zeros and ones as P'_i then the i -th element in the sequence of the aggregated judgment is the same in the two profiles. Property **SYS** states something stronger. Namely, for any two columns, if P_i consists of the same sequence of zeros and ones as P'_j , then the i -th and j -th values in the aggregated judgments of the two profiles are the same.

Remark 4 (Aggregation of 1–0 matrices). Along the line of Remark 3, if f is systematic then the only information from each profile that it is used by f is its associated matrix of acceptance-rejection. More precisely, there exists a function g which, for each 1–0 matrix generated by a profile, associates a sequence of 1–0 values such that, $\forall \varphi \in A$:⁴

$$f(P_1, \dots, P_n)(\varphi) = g(P_1(\varphi), \dots, P_n(\varphi)) \quad (2.1)$$

This yet more abstract view on aggregation has been systematically pursued in, among others, [18].

2.2 An impossibility theorem

The present section will provide an extensive discussion of the following impossibility result, due to [13]:

Let the agenda be even number negatable: an aggregation function satisfies unanimity and systematicity if and only if it is a dictatorship by some individual.

Put it otherwise, it is impossible to aggregate in a non-trivial way—like dictatorship does—individual judgments into a collective one by respecting unanimity and systematicity.

The section is devoted to provide a proof of this result. To this end, we will proceed by first introducing an important auxiliary notion—the one of winning coalition—and then proving two key lemmata.

2.2.1 Winning coalitions

Given any JA structure and aggregation function we can always ask ourselves for which agents it always holds that, if they all at the same time accept a given

⁴Cf. [61].

formula, so does also the collective judgment. In other words, we can always define for any element φ of the agenda, what is the coalition of agents that can always force φ to be collectively accepted. Such coalitions are called *decisive* or *winning*.

Let, for any profile P and formula φ the set

$$P_\varphi := \{i \in N \mid P_i \models \varphi\} \quad (2.2)$$

which denotes the set of all judgment sets in P that accept φ —the part of P accepting φ . We can now define the notion of winning coalition as follows:

Definition 10 (Winning coalitions for φ). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure, f an aggregation function and $\varphi \in A$. A coalition $C \subseteq N$ is winning for φ iff*

$$\forall P \in \mathbf{P} : \text{IF } C = P_\varphi \text{ THEN } f(P) \models \varphi.$$

The set of winning coalitions for φ in \mathcal{J} under f is denoted $\mathcal{W}_\varphi(\mathcal{J}, f)$.⁵

Remark 5. *It is worth observing the influence of conditions such as **IND** and **SYS** on Definition 10. By both these conditions, if there exists a profile P such that $C = \{i \in N \mid P_i \models \varphi\}$ and $f(P) \models \varphi$, then for all P' such that $C = \{i \in N \mid P'_i \models \varphi\}$ it holds that $f(P') \models \varphi$. In other words, if C is winning for φ in one profile, then it is winning for φ in all profiles.*

2.2.2 Lemmata

In order to prove the result, we will need four lemmata. The first two relate the property of systematicity to the possibility of defining a set of coalitions of voters which can ‘force’ the whole collective judgment. The third one makes explicit the specific structure of such set of coalitions. Finally the fourth one establishes the existence of a dictator.

Contagion & characterization by winning coalitions

Lemma 1 (Contagion lemma). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure and f an aggregation function satisfying **SYS**:*

$$\forall \varphi, \psi \in A : \mathcal{W}_\varphi = \mathcal{W}_\psi$$

Proof. Now suppose, towards a contradiction, that $\mathcal{W}_\varphi \neq \mathcal{W}_\psi$. WLOG, $\exists C \in \mathcal{W}_\varphi$ s.t. $C \notin \mathcal{W}_\psi$. There exist two profiles P, P' s.t. $C = \{i \in N \mid P_i \models \varphi\} = \{i \in N \mid P'_i \models \psi\}$ such that $f(P) \models \varphi$ and $f(P') \not\models \psi$. This contradicts **SYS**. \square

Intuitively, the lemma states that if f is systematic, then the winning coalitions of all elements of the agenda coincide.

Now, since the winning coalitions for all elements of the agenda are the same set, we can define the set of winning coalitions (for a given \mathcal{J}) as follows:

$$\mathcal{W} := \{C \subseteq N \mid \forall \varphi \in A, \forall P \in \mathbf{P} : \text{IF } C = P_\varphi \text{ THEN } f(P) \models \varphi\} \quad (2.3)$$

And we can prove that any systematic function can be characterized in terms of this set of winning coalitions.

⁵We will almost always drop the reference to \mathcal{J} and f as they will usually be clear from the context.

Lemma 2 (Winning coalitions). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure and f an aggregation function satisfying **SYS**. For all $P \in \mathbf{P}$ and $\varphi \in A$:*

$$f(P) \models \varphi \quad \text{IFF} \quad P_\varphi \in \mathcal{W}.$$

Proof. [RIGHT TO LEFT] It holds directly by the above definition of \mathcal{W} . [LEFT TO RIGHT] Assume $f(P) \models \varphi$ and consider the set of voters P_φ . For any $P' \in \mathbf{P}$, by **SYS** we have that if $P_\varphi = P'_\varphi$ then $f(P') \models \varphi$. Hence $P_\varphi \in \mathcal{W}$ according to the above definition in Formula 2.3. \square

Ultrafilters of winning coalitions

We now move to the central lemma, which shows that the set of winning coalitions \mathcal{W} enjoys a set of remarkable structural properties.

Lemma 3 (Ultrafilter lemma). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure and f an aggregation function such that A satisfies EN and f satisfies **SYS**. The set \mathcal{W} is an ultrafilter, i.e.:⁶*

- i) $N \in \mathcal{W}$;
- ii) $C \in \mathcal{W}$ iff $-C \notin \mathcal{W}$;
- iii) \mathcal{W} is upward closed: if $C \in \mathcal{W}$ and $C \subseteq C'$ then $C' \in \mathcal{W}$;
- iv) \mathcal{W} is closed under finite meets: if $C, C' \in \mathcal{W}$ then $C \cap C' \in \mathcal{W}$.

Proof. The proof is given claim by claim.

i) The claim follows from the assumption that f is unanimous.

ii) [LEFT TO RIGHT] By reductio ad absurdum, suppose both $C, -C \in \mathcal{W}$. Consider now a profile where the judgment sets of the agents in C contain φ and those in $-C$ contain $\neg\varphi$. This profile must exist by the definition of aggregation function (Definition 3), and it would be inconsistent, which is impossible according to the same definition. [RIGHT TO LEFT] By contraposition, suppose $C \notin \mathcal{W}$. By Lemma 1, $\nexists \varphi \in A$ s.t. $C \in \mathcal{W}_\varphi$, so $\forall \varphi \in A, \exists P$ s.t. $C = \{i \mid P_i \models \varphi\}$ and $f(P) \not\models \varphi$. Hence, $\forall \varphi \in A, \exists P$ s.t. $-C = P_\varphi$ and $f(P) \not\models \varphi$. By **SYS** (recall Remark 5) it follows that $\forall \varphi \in A, \forall P$ s.t. $-C = P_{\neg\varphi}$ and $f(P) \models \neg\varphi$, hence $-C \in \mathcal{W}$.

iii) We proceed by reductio ad absurdum: assume $C \in \mathcal{W}$, $C \subseteq C'$ and $C' \notin \mathcal{W}$. Take a minimally inconsistent set $X \subseteq A$ s.t. $\exists Y \subset X$ with $Y = \{\varphi, \psi\}$ for $\varphi, \psi \in A$ and s.t. $(X - Y) \cup \neg Y$ is consistent. This set exists by EN. By Definition 4, it follows that $(X - \{\varphi\}) \cup \{\neg\varphi\}$ and $(X - \{\psi\}) \cup \{\neg\psi\}$ are consistent. Consistent is also, by the definition of EN (Definition 6), the set $(X - \{\varphi, \psi\}) \cup \{\neg\varphi, \neg\psi\}$. Consider now these three coalitions which, notice, form a partition of N :⁷

$$\begin{aligned} C_1 &:= C \\ C_2 &:= C' - C \\ C_3 &:= N - C' \end{aligned}$$

⁶Ultrafilters formalize the intuition of what a set of ‘large’ sets is: i) the largest set is a large set; ii) a set is large iff its complement is not large; iii) if a set is large its supersets are also large; iv) the intersection of two large sets is large.

⁷See Figure 2.3.

and consider the judgment profile P thus defined:⁸

$$P_i = \begin{cases} (X - \{\varphi\}) \cup \{\neg\varphi\} & \text{if } i \in C_1 \\ (X - \{\varphi, \psi\}) \cup \{\neg\varphi, \neg\psi\} & \text{if } i \in C_2 \\ (X - \{\psi\}) \cup \{\neg\psi\} & \text{if } i \in C_3 \end{cases}$$

We can now conclude the following about P . By **U**, we have that $N \in \mathcal{W}$ and hence $X - Y \in f(P)$. Since $C \in \mathcal{W}$ by assumption, we also have that $\psi \in f(P)$. Furthermore, by i) and the assumption that $C' \notin \mathcal{W}$ we conclude that $C_3 \in \mathcal{W}$ and consequently that $\varphi \in f(P)$. So, to sum up, we get that $\{\varphi, \psi\} \in f(P)$ and $X - Y \in f(P)$, from which we conclude that $X \in f(P)$, which is impossible. This completes the proof of claim iii).

iv) We proceed by reductio ad absurdum. Assume $C, C' \in \mathcal{W}$ and $C \cap C' \notin \mathcal{W}$. By **EN**, which implies **NS**, there exists a minimally inconsistent set $X \subseteq A$ s.t. $3 \leq |X|$ (Definition 4). Take three elements of X : φ, ψ, ξ . By the same definition we have that for $x \in \{\varphi, \psi, \xi\}$: $(X - x) \cup \{\neg x\}$ is consistent. Consider now these three coalitions which, notice, form a partition of N :⁹

$$\begin{aligned} C_1 &:= C \cap C' \\ C_2 &:= C' - C \\ C_3 &:= N - C' \end{aligned}$$

and consider the judgment profile P defined as follows:¹⁰

$$P_i = \begin{cases} (X - \{\varphi\}) \cup \{\neg\varphi\} & \text{if } i \in C_1 \\ (X - \{\xi\}) \cup \{\neg\xi\} & \text{if } i \in C_2 \\ (X - \{\psi\}) \cup \{\neg\psi\} & \text{if } i \in C_3 \end{cases}$$

By **U** we have that $X - \{\varphi, \psi, \xi\} \subseteq f(P)$. Since $C' = C_1 \cup C_2 \in \mathcal{W}$, it follows that $\psi \in f(P)$. Also, since $C \subseteq C_1 \cup C_3$, by claim iii) above we have that $C_1 \cup C_3 \in \mathcal{W}$. Hence $\xi \in f(P)$. Finally, since by assumption $C \cap C' = C_1 \notin \mathcal{W}$, by claim ii) we have that $C_2 \cup C_3 \in \mathcal{W}$ and hence that $\varphi \in f(P)$. From this we conclude that $X \subseteq f(X)$, obtaining an inconsistent set, which is impossible. This proves claim iv) and concludes the proof of the lemma. \square

Dictators

This lemma consists, actually, of a general fact concerning finite ultrafilters, i.e., ultrafilters which are defined, like in our case, on a finite domain.

⁸This is a concrete example for the discursive dilemma agenda $A = \pm\{p, q, p \wedge q\}$. The minimally inconsistent set is $X = \{p, q, \neg(p \wedge q)\}$ and the profile is:

$$P_i = \begin{cases} \{q, \neg(p \wedge q), \neg p\} & \text{if } i \in C_1 \\ \{\neg p, \neg(p \wedge q), \neg q\} & \text{if } i \in C_2 \\ \{p, \neg(p \wedge q), \neg q\} & \text{if } i \in C_3 \end{cases}$$

⁹See Figure 2.4

¹⁰This is a concrete example for the discursive dilemma agenda $A = \pm\{p, q, p \wedge q\}$. The minimally inconsistent set is, again, $X = \{p, q, \neg(p \wedge q)\}$ and the profile is:

$$P_i = \begin{cases} \{q, \neg(p \wedge q), \neg p\} & \text{if } i \in C_1 \\ \{p, p \wedge q, q\} & \text{if } i \in C_2 \\ \{p, \neg(p \wedge q), \neg q\} & \text{if } i \in C_3 \end{cases}$$

So, $\varphi := p$, $\psi := q$, $\xi := \neg(p \wedge q)$.

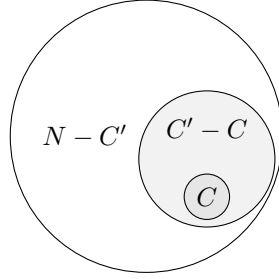


Figure 2.3: Voters tripartition in claim iii) Lemma 3

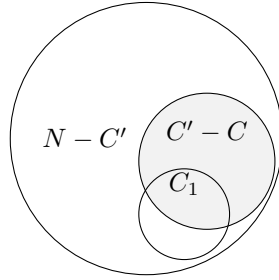


Figure 2.4: Voters tripartition in claim iv) Lemma 3

Lemma 4 (Existence of a dictator). *Let \mathcal{W} be an ultrafilter on a finite domain N . Then \mathcal{W} is principal, i.e.:*

$$\exists i \in N \text{ s.t. } \{i\} \in \mathcal{W}.$$

Proof. Consider $\bigcap \mathcal{W}$, which is well-defined as N is finite. We have that $\bigcap \mathcal{W} \neq \emptyset$. For suppose not, then there must be $C, C' \in \mathcal{W}$ s.t. $C \cap C' = \emptyset$ and hence, WLOG there must be C'' s.t. $C' \subseteq C''$ and $C = -C''$, which is impossible by properties ii) and iii) of Lemma 3. So, WLOG, assume $i \in \bigcap \mathcal{W}$ for $i \in N$.

By property ii), for some $i \in N$ either $\{i\} \in \mathcal{W}$ or $-\{i\} \in \mathcal{W}$, but the second option is impossible as $i \in \bigcap \mathcal{W}$. Hence, $\bigcap \mathcal{W} = \{i\}$. \square

In other words, there exists a voter who is a winning coalition. Such voter is the (unique) dictator. It is worth stressing that this lemma does not hinge on any specific JA property or construct, but singles out a general property of finite ultrafilters.

2.2.3 The theorem

We are now in the position of giving a precise formulation of the theorem and prove it.

Theorem 1 ([13]). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure such that A satisfies EN, and let f be an aggregation function: f satisfies **U** and **SYS** iff f satisfies **D**.*

Proof. [RIGHT TO LEFT] It is easy to verify that if f satisfies **D** then it trivially satisfies **U** and **SYS**. [LEFT TO RIGHT] By Corollary 2, for any $P \in \mathbf{P}$ and

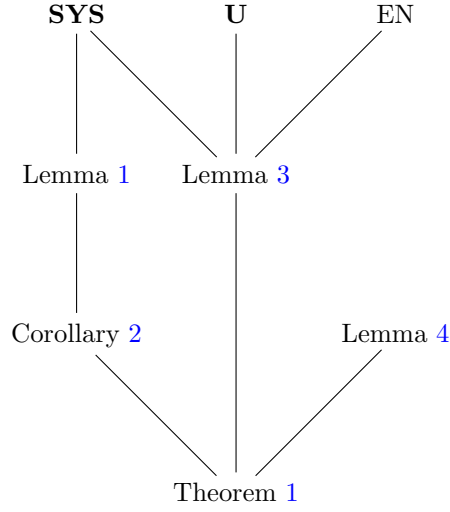


Figure 2.5: Structure of the proof of Theorem 1

$\varphi \in A$:

$$f(P) \models \varphi \quad \text{IFF} \quad P_\varphi \in \mathcal{W}.$$

Then, by Lemma 3 and 4 we have that $\{i\} \in \mathcal{W}$ for some $i \in N$ and hence:

$$P_\varphi \in \mathcal{W} \quad \text{IFF} \quad i \in P_\varphi$$

which concludes the proof: $f(P) \models \varphi$ iff $P_i \models \varphi$. \square

The first impossibility theorem of JA, proven in the paper that initiated the field [49], is a direct consequence of Theorem 1. Agendas such as $\pm\{p, q, p \wedge q\}$ or $\pm\{p, q, p \rightarrow q\}$, which satisfy EN, can be aggregated only in a trivial way, via a dictatorship, if we are to guarantee that the aggregation is unanimous and systematic.

The diagram in Figure 2.5 recapitulates the structure of the proof highlighting the dependences between assumptions, lemmata, and the final statement.

2.3 More on impossibility

In this section we wrap up pointing the reader to further impossibility results, discussing their common features and highlighting some interesting aspects of the proof method we used in the previous section.

2.3.1 The structure of impossibility theorems in JA

Three types of constraints lie at the core of the aggregation problem:

- constraints on the level of connectedness of the agenda (*agenda constraints*);

Agenda conditions	Aggregation conditions	Proof
EN	SYS, U	[13]
NS	SYS, MON	[59]
PC, EN	IND, U	[13]
PC	IND, MON, U	[59]

Table 2.1: Combinations of agenda and aggregation conditions. If the agenda has the property on the left, then the property of the aggregation (middle column) is equivalent to dictatorship. The first row corresponds to Theorem 1. For all rows except the first, the converse also holds (for the first the converse holds for the aggregation being dictatorial *or* inversely dictatorial).

- ▶ constraints on the aggregation function (*aggregation constraints*)
- ▶ constraints on the admissible domain and codomain of the aggregation function (*rationality constraints*)

While the first two are clear from the statement of Theorem 1 and Table 2.1, the third one has been somehow hidden in our set up of the aggregation problem and more precisely in Definition 3. There, the aggregation function is taken to operate from the set of all judgment profiles (domain) to the set of all judgment sets (codomain). Since judgment sets are assumed to be consistent, we have somehow built in our representation of the JA problem the assumption sometimes called *collective rationality*, i.e., that the collective judgment set be consistent and complete. A simple inspection of the proof of Theorem 1 will show that this assumption plays a crucial role for obtaining the desired result.

2.3.2 Other impossibility results

Other theorems analogous to Theorem 1 can be obtained by varying the logical strength of agenda and aggregation conditions: e.g. by strengthening EN with PC and weakening at the same time **SYS** to **IND**. Table 2.1, which we adapted from [48], recapitulates in a compact way some of the better-known impossibility results that have been established in the JA literature. One more impossibility result of special relevance for the notion of manipulability of an aggregation process will be studied in detail in Chapter 4.

2.3.3 Impossibility via ultrafilters

The proof we have provided of Theorem 1 has relied on a well-established technique, based on the use of ultrafilters, structures first introduced in [8]. This technique can be summarized as follows:

To establish impossibility results one shows that the conditions imposed on the agenda and the aggregation function force the set of winning coalitions to be an ultrafilter on the power-set of the set of voters. If the set of voters is finite, one can then conclude that the ultrafilter is principled, i.e., it is generated by one single element that belongs to all winning coalitions, hence establishing the existence of one dictator.

So, we might say, the technique happens to be based on a happy mathematical coincidence.

The first application of this technique to social choice theory was proposed in [26], providing an alternative proof of Arrow's theorem. In JA, several proofs resort, directly or indirectly, to this technique (e.g. [28, 37] but also [13] itself).

Chapter 3

Coping with impossibility

One may view the impossibility results in the literature of judgment aggregation as theorems stating that aggregation functions satisfying a certain number of desirable properties do not exist. There is also, however, a more positive interpretation of such results, namely, they indicate which conditions to relax in order to ensure possibility results.

In Chapter 2 we have seen one impossibility theorem in detail and we have observed that, at the core of the aggregation problems, lie three types of constraints: agenda constraints, aggregation constraints, and rationality constraints (i.e. constraints on the input and on the output of the aggregation function). This indicates that possibility results can be obtained by relaxing any of these types of constraints. However, relaxing the agenda constraints does not appear as a good escape route. The reason is that the two agenda conditions we have considered (non-simplicity and even-number-negativity) are not demanding and yet play a central role in the impossibility results.¹ Thus, relaxing the agenda constraints would imply to restrict judgment aggregation to trivial decision problems.

Therefore, escape routes must be found in relaxing the *input conditions*, relaxing the *output conditions*, or relaxing the aggregation constraints, more specifically the independence condition (**IND**). In this chapter we present some of the approaches that have been investigated in the literature and that guarantee the existence of a judgment aggregation function.

Outline of the chapter. Section 3.1 presents the results obtained when we restrict the domain of the aggregation function, while Section 3.2 reviews what happens when we relax collective rationality. Finally, in Section 3.3 we present the third investigated escape route in judgment aggregation, consisting in dropping independence.

The chapter builds on material and results taken from [45, 28, 14, 17, 40, 63].

¹Nehring and Puppe, for example, showed that, if the agenda is non-simple, every aggregation function satisfying universal domain, collective rationality, **SYS** and **MON** is a dictatorship [58]. See also Table 2.1.

	Judge 3	Judge 2	Judge 5	Judge 4	Judge 1
p	0	0	0	1	1
q	1	1	0	0	0
r	1	0	0	0	0

Table 3.1: An example of a unidimensionally aligned profile

3.1 Relaxing the input constraints

3.1.1 Unidimensional alignment

In 1948 Duncan Black [2] introduced single-peaked preferences in the theory of preference aggregation. These are individual preferences where there is a peak, which represents the most preferred alternative. On either side of the peak lie the less preferred alternatives (unless the peak is at the extreme left or right). The alternatives are ordered in such a way that their desirability declines the farther they are from the peak. Single-peakedness captures the structure of several real world decisions over a single-dimension of choice. As Arrow illustrates:

An example in which this assumption is satisfied is the party structure of prewar European parliaments, where there was a universally recognized Left-Right ordering of the parties. Individuals might have belonged to any of those parties; but each recognized the same arrangement, in the sense that, of two parties to the left of his own, the individual would prefer the program of the one less to the left, and similarly with parties on the right. [1, pp.75-76]

Voters of the Condorcet paradox seen in Section 1.1.2 do not hold single-peaked preferences. Single-peaked preferences are important because they ensure a possibility theorem for preference aggregation under majority voting, the so-called *median voter* theorem [2].

The impossibility theorems of judgment aggregation assume universal domain, which requires that all profiles of consistent and complete judgment sets on a given agenda are admissible inputs for the aggregation function. Inspired by single-peakedness, List [45, 47] introduced a similar condition for judgment aggregation, called *unidimensional alignment*:

Unidimensional alignment A profile is *unidimensionally aligned* if the individuals in N can be ordered from left to right such that, $\forall \varphi \in A$, the individuals accepting φ are either to the left or to the right of all the individuals rejecting φ .

The profile in Table 1.1 is not unidimensionally aligned. An example of a unidimensionally aligned profile is in Table 3.1 [45].

List showed that, under such domain restriction, propositionwise majority voting is the *only* aggregation procedure that guarantees complete and consistent collective judgment sets and that satisfies **SYS** and *anonymity*:

Anonymity (An) $\forall P, P' \in \mathbf{P}$ which are permutations of each other, $f(P) = f(P')$.

I.e., all voters count the same in the decision-making.

The reason why unidimensional alignment is *sufficient* for reaching consistent majority judgment sets is that individuals are ordered in such a way that those accepting a propositions are opposite those rejecting the same propositions. Thus, if the number of individuals is odd, the majority must coincide with the median voter's judgment set (Judge 5 in Table 1.2). Since we assume that individuals are logically consistent, this ensures the collective judgment to be equally consistent. If there is an even number of individuals, the majority will be the intersection of the judgment sets of the two median pair of voters (which will still be a consistent judgment set).

3.1.2 Value-restriction

The exploration of domain restriction conditions in judgment aggregation continued in [16]. Here, Dietrich and List introduce other sufficient conditions for majority consistency. In particular, they generalize another well-known condition in the theory of preference aggregation: the value-restricted preferences, introduced by Sen in 1966 [66], from which he proved a possibility result. The condition of value restriction states that:

In a triple (x, y, z) there is some alternative, say x , such that all concerned individuals agree that it is not worst, or agree that it is not best, or agree that it is not in the middle." [27, p. 44]

The translation of the above condition in the context of judgment aggregation led Dietrich and List to formulate the *value-restricted* condition below:

Value-restriction A profile P is value-restricted if every minimal inconsistent set $X \subseteq A$ has a two-element subset $Y \subseteq X$ that is not accepted by *any* individual $i \in N$.

Clearly, value restriction is sufficient to avoid that an inconsistent judgment set is selected as the group outcome. Moreover, Dietrich and List introduce a *necessary and sufficient* domain-restriction condition:

Majority-consistency A profile P is majority-consistent if every minimal inconsistent set $X \subseteq A$ contains a proposition that is not accepted by a majority.

Domain-restriction conditions can represent plausible escape-routes to the impossibility results in *some* decision-making contexts. As observed in [48], different groups display different levels of pluralism. If there is empirical evidence showing that in a specific group, confronted with a specific decision problem, the conditions above are met, then aggregation can be considered viable in that restricted context.

3.2 Relaxing the output constraints

Besides the input condition of universal domain, impossibility theorems of judgment aggregation assume the output condition of collective rationality (see 2.3). While consistency seems an indispensable requirement, completeness can be dismissed, at least in some contexts (though, from a pragmatic perspective, we may see the agenda as the set of issues on which the group needs to take a decision). Yet, domain restriction conditions cannot offer a general solution to the problem of judgment aggregation. In those contexts in which completeness is a fair price to pay to avoid paradoxical results, this can be achieved in several ways, as we are going to illustrate.

3.2.1 Abstention

Gärdenfors [28] has criticized the completeness requirement as being too strong and unrealistic. Hence, he investigated what happens if we allow voters to abstain from expressing judgments on some propositions in the agenda. He proved that, if the judgment sets need not be complete (but are deductively closed² and consistent), then every aggregation function that is **IND** and Paretian must be oligarchic.

Paretian An aggregation function f is *Paretian* if, $\forall \varphi \in A$, f assigns a collective position to φ whenever $\forall i \in N$ adopted the same position on φ .

Oligarchy An aggregation function is *oligarchic* if $\forall \varphi \in A$, the group adopts a position 0 (resp. 1) if and only if all the members of a subset of the group (the oligarchy) $D \subseteq N$ adopt position 0 (resp. 1) on that issue.

Clearly, when D contains only one member, oligarchy reduces to dictatorship. At the other end of the spectrum, when $D = N$, the decision procedure is anonymous but only unanimous issues are taken by the group.³

Gärdenfors' framework requires the agenda to have a very rich logical structure (with an infinite number of issues). Later, Dokow and Holzman [19] extended Gärdenfors' result and consider finite agendas. Again, the findings are that dictatorial rules are replaced by oligarchic ones, which are very restricted classes of aggregation rules. Hence, relaxing the completeness assumption does not provide a completely satisfactory escape to the impossibility results.

3.2.2 Quota rules

In [14] Dietrich and List explore *quota rules*, where a proposition is accepted if and only if that proposition is accepted by a number of individuals greater than a prefixed threshold. The appeal of quota rules comes from the intuition that different problems may require different social support to be declared collective outcomes. For example, a decision that has a high impact on a group may require to be supported by 2/3 of the individuals. Similarly, in a given agenda, one issue may be more important than another and, hence, different propositions

²According to *deductive closure*, any $\varphi \in A$ that is logically entailed by a given judgment set J , is also contained in it: if $J \models \varphi$, then $\varphi \in J$.

³Note that unanimity rules guarantee deductive closure at the expense of significant incompleteness.

may have different thresholds. Clearly, majority voting is a special kind of quota rules, with the same majority threshold for each proposition.

Dietrich and List consider four rationality conditions on the individual and collective judgment sets (complete, weakly consistent, consistent and deductively closed), and they show that a given quota rule satisfies a rationality condition if a certain inequality is verified. However, whether such inequalities are satisfied depends on the logical structure of the agenda (in particular, on the minimal inconsistent subsets of the agenda and on their size). For rich agendas, these can be demanding conditions. Nevertheless, it is worth mentioning that if we weaken collective rationality to consistency alone (so, dropping completeness), supermajority rules produce consistent collective judgments. This happens when the supermajority threshold q is greater than $(k - 1)/k$, where k is the size of the largest minimally inconsistent subset of the agenda.

3.3 Relaxing independence

In Section 2.3 we have recalled some impossibility results obtained by strengthening or weakening the conditions of Theorem 1. In particular, Table 2.1 listed some impossibility results obtained by weakening **SYS** to **IND**.⁴ The more recent impossibility theorems in judgment aggregation assume **IND**.

IND rephrases in the context of judgment aggregation the *independence of the irrelevant alternatives* condition in Arrow's theorem. Independence of irrelevant alternatives warrants that the group ranking over any pair of alternatives depends solely on the personal rankings over the same pair of alternatives. The intuition is that the social ranking over, for example, x and y should be determined exclusively on how the individuals rank x compared to y and not on other (irrelevant) alternatives, like z . This requirement has been introduced in judgment aggregation as **IND**. This ensures that the collective judgment on each proposition depends exclusively on the individual judgments on that proposition (and not on other – assumed to be independent — propositions).

As we will see in Section 4.1, **IND** is a key condition to ensure that an aggregation function is non manipulable [12, 15], i.e., robust against strategic voting. This makes **IND** an (instrumentally) attractive condition, as it happens for the independence of the irrelevant alternatives condition in preference aggregation. However, **IND** has also been severely criticized in the literature (see, for example, [9, 54]). Several authors deem **IND** incompatible with a framework whose aim is to aggregate logically *interrelated* propositions. Mongin, for example, writes:

[...] the condition remains open to a charge of irrationality. One would expect society to pay attention not only to the individuals' judgments on φ , but also to their *reasons* for accepting or rejecting this formula, and these reasons may be represented by other formulas that φ in the individual sets. [54, p. 105]

These criticisms make **IND** among the first aggregation constraints to be relaxed in order to achieve possibility results. In this section we will consider three main options to relax **IND**.

⁴We recall that **IND** is **SYS** without neutrality.

3.3.1 The premise-based approach

The first possibility to relax **IND** is to resort to the premise-based procedure, which we encountered already in Section 1.1.1. The premise-based procedure has been introduced in [44] under the name of “issue-by-issue voting” and studied in [17, 54]. The agenda is assumed to be partitioned into two subsets: *premises* and *conclusions*. The premises have to be logically independent. In the premise-based procedure the individuals express their judgments on the premises only. The collective judgment set is the propositionwise aggregation rule (for example, majority rule) on the premises. From the collective outcome on the premises, the collective conclusions are derived using either the logical relationships between, or some external constraints regarding, the agenda issues.

On the one hand, the premise-based approach avoids the doctrinal paradox by ensuring a collective consistent position. Furthermore, it escapes the charges of irrationality of **IND**, being **IND** applied only to logically independent propositions. On the other hand, it is not always clear how to partition an agenda into premises and conclusions.

In Section 1.1.1 we have also considered the conclusion-based procedure and we have noticed that this approach may give an opposite result than the premise-based. Hence, one natural question is how to choose between premise and conclusion-based procedures. One answer has been given by Bovens and Rabinowicz [3] and by List [52]. The idea is to evaluate and compare the two aggregation procedures in their *truth-tracking* reliabilities. It is assumed that a group judgement is factually right or wrong and, thus, the question is how reliable various approaches are at selecting the right judgment set.

If the individuals are better than randomizers at judging the truth or falsity of a proposition (in other words, if the probability of each agent at getting the right judgment on a proposition is greater than 0.5), and if they form their opinions independently, then majority voting eventually yields the right collective judgement on that proposition with increasing size of the group. This general fact, known as the *Condorcet Jury Theorem*, links the competence of the agents to the reliability of majority voting. It also motivates the use of majority-based decision making in the judgement aggregation problem. The general finding of [3, 52] is that premise-based procedure is a better truth-tracking approach than majority rule on the conclusion.

Despite all these good news, the premise-based procedure can lead to unwelcome results. Because in the premise-based procedure the collective judgment on the conclusion is derived from the individual judgments on the premises, it can happen that the premise-based procedure violates a unanimous vote on the conclusion! In [57] Nehring presents a variation of the discursive dilemma, which he calls the *Paretian dilemma*. In his example, a three-judges court has

to decide whether a defendant has to pay damages to the plaintiff. Legal doctrine requires that damages are due if and only if the following three premises are established: 1) the defendant had a duty to take care, 2) the defendant behaved negligently, 3) his negligence caused damage to the plaintiff. [57, p. 1]

Suppose that the judges vote as in Table 3.2 below:

	p	q	r	$x = (p \wedge q \wedge r)$
Judge 1	1	1	0	0
Judge 2	0	1	1	0
Judge 3	1	0	1	0
Majority	1	1	1	0

Table 3.2: Paretian dilemma. Premises: p = duty, q = negligence, r = causation. Conclusion: $x = (p \wedge q \wedge r) =$ damages.

The Paretian dilemma is disturbing because, if the judges would follow the premise-based procedure, they would condemn the defendant to pay damages contradicting the *unanimous* belief of the court that the defendant is *not* liable. Nehring proves a general result according to which the only aggregation functions satisfying independence and monotonicity on the premises and unanimity on the conclusion are dictatorial.

If all well-behaved (i.e. anonymous or non-dictatorial) aggregation rules are prone to the Paretian dilemma, then no reason-based group decision can be guaranteed. How negative is this result? Nehring argues that when the reasons are epistemically independent

all relevant information about the outcome decision is contained in the agents' premise judgments. [...] Indeed, under epistemic independence of premises it is easy to understand how a group aggregation rule can *rightly* override a unanimous outcome judgment. [57, p.36].

Furthermore, the normative force of the Pareto criterion depends on the type of social decision. The Pareto criterion should be ensured when the individuals have a *shared self-interest* in the final outcome, whereas it can be relaxed when they *share responsibility* for the decision. Judicial decisions are clear instances of shared responsibility situations, while other group decisions may be self-interest driven. Nehring's analysis concludes that the Pareto criterion and reason-based group decisions are two principles that may come into conflict. However, such conflict does not mean that one of these two principles is ill-founded.

Independently, Mongin [54] proved that—for sufficiently rich agendas—the only aggregation rule that satisfies universal domain, an independence condition restricted to the atomic propositions (which may be viewed as the premises) and a unanimity condition, is dictatorship.

3.3.2 The sequential priority approach

Another possibility to relax IND is the sequential priority approach. Sequential procedures [46, 14] proceed this way: the elements of the agenda are considered sequentially, following a fixed linear order over the agenda (corresponding, for instance, to temporal precedence or to priority), and earlier decisions constrain later ones. Thus, individuals vote on each proposition p in the agenda, one by one, following the fixed order. If the collective judgment on a proposition p is consistent with the collective judgments obtained on the previous issues of the agenda, the collective judgment on p becomes the group position on p . However,

in case the collective position on p conflicts with the group judgments on the propositions aggregated earlier, the collective judgment on p will be derived from the earlier group judgments.

Collective consistency is guaranteed by definition. Of course, in the general case, the result depends on the choice of the order, *i.e.* it is *path-dependent*. Path-dependence is tightly linked to manipulability: the agenda-setter can manipulate the social outcome by setting a specific order in which the items in the agenda are considered. It is also the case that individuals can strategically vote if the rule is path-dependent. Dietrich and List [14] proved that the absence of path-dependence is equivalent to strategy-proofness in a sequential priority approach. In particular, they show that a sequential majority rule is strategy-proof only when the size of the largest minimal inconsistent set is less or ≤ 2 . On the other hand, a sequential unanimity rule is clearly always strategy-proof but this come with a price, namely incompleteness of the group judgment set.

In order to illustrate how a sequential priority rule works and the problem of path-dependence, we recall here the example used in [14].

Example 7 (Sequential priority rules). *Suppose that the Presidents of three governments have to decide on the following propositions:*

- p : Country X has weapons of mass destruction.
- q : Action Y should be taken against country X .
- $p \leftrightarrow q$: Action Y should be taken against country X if and only if country X has weapons of mass destruction.

Suppose that the individual judgments on the issues in the agenda are as follows:

	p	$p \leftrightarrow q$	q
President 1	1	1	1
President 2	0	0	1
President 3	0	1	0
Majority	0	1	1

Table 3.3: An example of sequential majority rule.

Let us suppose that simple majority is used. We can now consider two different sequential paths. In the first, the items of the agenda are aggregated according to the following order: $p, p \leftrightarrow q, q$. In the second path, agents are asked to vote in the following order: $q, p \leftrightarrow q, p$. We obtain two different collective judgments: $\{\neg p, p \leftrightarrow q, \neg q\}$ when the first path is followed and $\{p, p \leftrightarrow q, q\}$ when the second path is followed. In both cases, the three Presidents agree that action Y should be taken against country X if and only if country X has weapons of mass destruction. However, whilst they will take action against country X if the first path is followed, they will take no action against country X if the second path is used.

Finally, note that premise-based procedure is a specific instance of sequential priority procedures.

3.3.3 The distance-based rules

The third approach that relaxes **IND** and that we consider here is the distance-based approach. Distance-based judgment aggregation rules [63, 53] have been originally derived from distance-based merging operators for belief bases introduced in computer science [39, 38, 40]. Unlike the premise-based procedure and the sequential priority approach, the distance-based approach considers *all* the elements of a judgment set.

Distance-based rules assume a predefined distance between judgment sets and between judgment sets and profiles, and choose as collective outcomes the *consistent* judgment sets⁵ which are closest (for some notion of closeness) to the individual judgments. More precisely, given a distance d and an aggregation function f , the collective judgment J^* minimizes $f(d(J_1, J^*), \dots, d(J_n, J^*))$, where J_1, \dots, J_n are the individual judgment sets. There are many possible variations on this definition (in [53] four general methods are introduced and compared).

Let $d : \mathbf{J} \times \mathbf{J} \mapsto \mathbb{R}^+$ be a distance function between any two judgment sets $J_i, J_j \subseteq X$.⁶ Well-known is the *Hamming distance*, which counts the number of propositions on which two judgment sets disagree. For example, if $J_i = \{a, \neg b, c\}$ and $J_j = \{\neg a, \neg b, c\}$, the Hamming distance d_H between the two judgment sets is 1 as they differ only on the evaluation of proposition a : $d_H(J_i, J_j) = 1$. In the following we use the Hamming distance because of its intuitiveness and wide applicability. But the Hamming distance is only one among many possible distance functions that we may use [40, 41].

The function d assigns a distance to each judgment set of a given profile P and any judgment set that can be selected to be the collective judgment set. Once all these distances are calculated, we need to calculate the distance between the profile and each possible collective judgment. This is done with the help of an aggregation function f that, for example, *sums* the distances obtained between the individual judgment sets in P and each possible collective judgment. The idea is that a distance-based rule $\Delta^{d, \Sigma}$ will select those collective judgment sets which are at minimal distance from P .

The best way to illustrate how a particular distance-based rule works is with an example.

Example 8 (Distance-based aggregation). *Let us consider the doctrinal paradox. The three judgment sets corresponding to the three judges are:*

$$\begin{aligned} J_1 &= \{p, q, r\} \\ J_2 &= \{p, \neg q, \neg r\} \\ J_3 &= \{\neg p, q, \neg r\} \end{aligned}$$

The table below shows the result of a distance-based aggregation rule where d is the Hamming distance and f is $\sum_{i \in N} d_H(J, J_i)$, where J is a possible consistent collective judgment set. The first column lists all the consistent judgment sets. The numbers in the columns of $d(\cdot, J_1)$, $d(\cdot, J_2)$ and $d(\cdot, J_3)$ are the Hamming distances of each J_i from the correspondent collective judgment candidate.

⁵Uniqueness of the collective outcome is not guaranteed.

⁶We recall that d is a distance function if and only if for all $J_i, J_j \subseteq X$ we have that (i) $d(J_i, J_j) = d(J_j, J_i)$ and (ii) $d(J_i, J_j) = 0$ if and only if $J_i = J_j$. Also, we slightly abuse language here, since d is only a pseudo-distance (triangular inequality is not required.)

Finally, in the last column is the sum of the distances over all the individual judgment sets in the profile.

	$d_H(., J_1)$	$d_H(., J_2)$	$d_H(., J_3)$	$\sum(d_H(., P))$
(1, 1, 1)	0	2	2	4
(1, 0, 0)	2	0	2	4
(0, 1, 0)	2	2	0	4
(0, 0, 0)	3	1	1	5

Thus, in this example, the consistent judgment sets that are closest to the profile P correspond exactly to the individual judgment sets in P (they are at distance 4 rather than 5). Thus, by considering only the consistent judgment sets as candidates for the collective position, we avoid the paradox. However, the procedure does not necessarily output a unique solution, as the example above shows. One can construct examples in which a distance-based rule like the one we considered here selects a unique collective judgment.

The equivalence between propositionwise majority voting and the distance minimization rule (called *minisum*) has been put forward in [4]. One can of course define other distance minimization rules [53, 39, 40]. For example, a widely used distance-based aggregation rule is the *minimax*, which selects the collective judgment set that minimizes the maximal distance to the individual judgment sets. Eckert and Klamler [21] have shown that, for a given profile, minisum and minimax may select two opposite collective outcomes.

We can conclude this overview over distance-based aggregation rules with a word on the manipulation issue, thereby also introducing the next chapter. In general, distance-based merging operators are not strategy-proof [25]. In particular, the family of merging operator using the Hamming distance and where the function f varies, are not strategy-proof (unless we to assume a profile with only two judgment sets and a restrictive satisfaction index, which defines the individual preference over the possible outcomes).

Remark 6 (Distance-based rules in voting theory). *It has been shown [22] that, for a preference agenda, the distance-based rule seen in the example above is equivalent to the Kemeny rule, a well known preference aggregation rule [36]. The fact [69] that Kemeny's rule is the only preference aggregation rule that is neutral, consistent and satisfies the Condorcet property⁷, might in particular be adduced as a justification for the use of the belief merging operator $\Delta^{d, \Sigma}$ [23].*

⁷A preference aggregation rule satisfies the Condorcet property if, whenever an alternative x defeats another alternative y in pairwise majority voting, it can never be the case that y is ranked immediately above x in the social preference.

Chapter 4

Manipulability

This chapter addresses the issue of *manipulability* of an aggregation problem. That is, how agenda setters or voters can strategically influence the aggregation in order to induce specific outcomes.

Outline of the chapter. Section 4.1 discusses the problem of agenda manipulation, i.e., the manipulation of the aggregation problem by an agenda setter. Section 4.2 addresses the issue of manipulation as it can be exercised by voters, providing one first characterization of the property of non-manipulability of an aggregation function. Section 4.3 presents an impossibility result connecting non-manipulability to dictatorship in the spirit of Theorem 1. Finally, Section 4.4 concludes by pointing to a strategic dimension in the issue of manipulability relating JA to the theory of games [68]. The chapter builds on material and results which we have mainly taken from [12, 15]. Some notions and methods that had been central in Chapter 2 play an important role here too.

4.1 Agenda manipulation

An agenda setter, possessing enough information about individual opinions on the issues at hands, is able, in given profiles, to determine the collective judgment on some of the issues.

Example 9 (Agenda manipulation). *Consider again Example 2 and assume the same three individual judgment sets over agenda $A = \pm\{p, p \rightarrow q, q\}$. Suppose now the voters are asked to vote over the sub-agenda $A_1 = \pm\{p, q\}$. Under propositionwise majority, the collective judgment is $\{p, \neg q\}$. An agenda manipulator could now swap p for $p \rightarrow q$ in A_1 and obtain an agenda $A_2 = \pm\{p, p \rightarrow q\}$. The collective judgment would so, under propositionwise majority, become $\{p, p \rightarrow q\}$ and entail q .*

	p	q			p	$p \rightarrow q$
J_1	1	1		J_1	1	1
J_2	1	0	\mapsto	J_2	1	0
J_3	0	0		J_3	0	1
J	1	0		J	1	1

In this example, in order to change the collective judgment over q , the manipulator had to remove q itself from the agenda. However, under propositionwise majority, it turns out to be impossible to modify the collective judgment on a given formula without removing that same formula from the agenda (to appreciate this, try to devise a few examples!).

This sort of robustness rests on the fact that propositionwise majority satisfies the condition of independence (**IND**). As the following results show, such condition is equivalent to the impossibility of agenda manipulation by simple addition or removal of formulae on the agenda other than the to-be-manipulated formula.

Let us first fix some notation. Given a JA structure $\mathcal{J} = \langle N, A \rangle$, an aggregation function f and $A_i \subseteq A$, $f_i \subseteq f$ denotes the reduction of f to the domain yielded by agenda A_i , i.e., profiles P_i obtained by appropriately restricting each P .

Theorem 2. *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure, let f be an aggregation function for \mathcal{J} . Consider now a sub-agenda $A_1 \subset A$. For any profile P (for \mathcal{J}), formula $\varphi \in A_1$ and any $A_2 \subseteq A$ such that $\varphi \in A_1 \cap A_2$, it holds that:*

$$f_1(P_1) \models \varphi \quad \text{IFF} \quad f_2(P_2) \models \varphi$$

*if and only if f_1 and f_2 satisfy **IND**.*

Proof. [Right to left] The claim follows from the fact that P_1 and P_2 are identical w.r.t. φ and f_1 and f_2 also coincide w.r.t. φ . Hence, by **IND**, the collective judgment over φ is the same. [Left to right] By contraposition, suppose WLOG that f_1 violates **IND**. Then there exists two identical profiles w.r.t. some $\varphi \in A_1$ such that their collective judgments over φ are inconsistent. Assume WLOG that $\varphi \in f_1(P_1)$ for one of the two profiles P_1 . Now, for the agenda $\pm\{\varphi\}$ there exists a profile P_2 such that $f_2(P_2) \models \neg\varphi$ \square

Roughly, **IND** is a good antidote against agenda manipulation. If it is settled what the individual judgments are over an agenda A , then the aggregation with respect to a subagenda A_1 cannot be changed with respect to any formula of A other than by removing that formula itself.

4.2 Vote manipulation

In this section we address that form of manipulation that arises when voters strategically misrepresent their true vote in order to force a different outcome in the aggregation process. The central results of the section will consist in an impossibility theorem in the spirit of the one already presented in Chapter 2 whose focus will be a property of aggregation functions called, indeed, non-manipulability.

4.2.1 A case of manipulation in the $\pm\{p, q, p \wedge q\}$ agenda

Consider once more the discursive dilemma and suppose the judges are to apply premise-based voting but are mainly interested in the logical conclusions of the aggregation process, namely whether the defendant is to be found liable or not.

Judge 1 believes the defendant is guilty ($p \wedge q$) and he will vote accordingly hence accepting both premises p and q . Now suppose the other two judges are aware of how Judge 1 will vote. How will they vote? As they are both convinced the defendant is not guilty ($\neg(p \wedge q)$), on the basis of the information they have about Judge 1, they know that if they both reject both assumptions ($\neg p$ and $\neg q$) they will be able to force their view through the aggregation.

	p	q	$p \wedge q$			p	q	$p \wedge q$
J_1	1	1	1	\mapsto	J_1	1	1	1
J_2	1	0	0		J_2	0	0	0
J_3	0	1	0		J_3	0	0	0
J	1	1	1		J	0	0	0

In fact, Judge 3 alone could force outcome $\neg(p \wedge q)$ provided that she knows what Judge 2 would vote and that she would vote truthfully. Same holds, obviously, for Judge 2 with respect to Judge 3.

	p	q	$p \wedge q$			p	q	$p \wedge q$
J_1	1	1	1	\mapsto	J_1	1	1	1
J_2	1	0	0		J_2	1	0	0
J_3	0	1	0		J_3	0	0	0
J	1	1	1		J	1	0	0

The perspective we have assumed in this example introduces a whole new dimension into JA, which has to make with the strategic behavior of voters. While strategic behavior is the realm of the theory of games [68]—and we will briefly touch upon it in Section 4.4—in the next sections we will assume a typically social-choice theory perspective: the example has shown that premise-based aggregation is clearly manipulable. The question is then: do non-trivial non-manipulable aggregation rules exist?

4.2.2 Manipulability: definition and a characterization

Recall first the terminology we have introduced in Definition 8 in Chapter 2. We now state a formal definition of (non-)manipulability as a property of aggregation functions.

Definition 11 (Manipulability). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure and $X \subseteq A$. An aggregation function f is:*

Manipulable on X (MAN_X) iff $\exists P \in \mathbf{P}, i \in N, \varphi \in X$ s.t. $f(P) \neq_\varphi P_i$ and $P_i =_\varphi f(P')$ for some $P' \in \mathbf{P}$ s.t. $P =_{-i} P'$.

*Simple manipulability **MAN** is defined simply by setting $X = A$ in MAN_X . A function is said to be non-manipulable otherwise.*

In words, an aggregation function is manipulable, w.r.t. a set of issues X , whenever there exists some profile where, for some agent i , the aggregation yields, for some formula φ in the given sub-agenda X , an outcome which is different from her individual opinion over φ and which can be modified as she to input a different judgment set in the aggregation function.

It is worth observing that, literally, this conditions states a ‘possibility’ of manipulation. Whether such possibility is attractive or not for the potential manipulator, is a different issue having to do with the manipulator’s incentives.

Manipulability is a strong condition. We will now characterize it in terms of two other aggregation conditions which are simple variants of the conditions of independence and monotonicity introduced in Definition 9.

Definition 12. Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure and $X \subseteq A$. An aggregation function f is:

Independent on X (\mathbf{IND}_X) iff $\forall \varphi \in X, \forall P, P' \in \mathbf{P}$: IF $[\forall i \in N : P_i \models \varphi$ IFF $P'_i \models \varphi]$ THEN $[f(P) \models \varphi$ IFF $f(P') \models \varphi]$;

Monotonic on X (\mathbf{MON}_X) iff $\forall \varphi \in X, \forall i \in N, \forall P, P' \in \mathbf{P}$: IF $[P =_{-i} P'$ AND $P_i \not\models \varphi$ AND $P_i \models \varphi]$ THEN $[$ IF $f(P) \models \varphi$ THEN $f(P') \models \varphi]$.

Note that these conditions are nothing but **IND** and **MON** whose scope has been constrained to a subset X of the agenda. We thus obtain the following characterization of non-manipulability:

Theorem 3 (A characterization of manipulability). Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure, $Y \subseteq A$ and f an aggregation function. The following assertions are equivalent:

- i) f does not satisfy \mathbf{MAN}_X
- ii) f satisfies \mathbf{IND}_X and \mathbf{MON}_X

Proof. [(ii) implies (i)] Assume f satisfies \mathbf{IND}_X and \mathbf{MON}_X and suppose $\exists \varphi \in X, P \in \mathbf{P}$ and $i \in N$ s.t. $P_i \not\models_{\varphi} f(P)$. We will show that $\forall P' \in \mathbf{P}$ s.t. $P' =_{-i} P$ we have that $P_i \not\models_{\varphi} f(P')$, thus proving non- \mathbf{MAN}_X . There are two cases: 1) $P_i =_{\varphi} P'_i$, or 2) $P_i \not\models_{\varphi} P'_i$. As to 1), by \mathbf{IND}_X it follows that $f(P') =_{\varphi} f(P)$, and hence it is still the case that $f(P') \not\models_{\varphi} P_i$. As to 2), since $P_i \not\models_{\varphi} f(P)$, it also follows that $P'_i =_{\varphi} f(P)$ (as there can be only either acceptance or rejection of φ). By \mathbf{MON}_X , it follows that $f(P) =_{\varphi} f(P')$ and hence that $f(P') \not\models_{\varphi} P_i$ as required.

[(i) implies (ii)] Assume non- \mathbf{MAN}_X . 1) We prove that \mathbf{MON}_X follows. Take any $\varphi \in X, i \in N$ and $P, P' \in \mathbf{P}$ s.t. $P =_{-i} P'$. WLOG assume that $P_i \not\models \varphi$ and $P' \models \varphi$. Now, if $f(P) \models \varphi$, then $P_i \not\models_{\varphi} f(P)$ and by non- \mathbf{MAN}_X $f(P) =_{\varphi} f(P')$. 2) We prove that \mathbf{IND}_X follows. Consider any $\varphi \in X$ and $P, P' \in \mathbf{P}$ s.t., $\forall i \in N: P_i =_{\varphi} P'_i$ (the antecedent of \mathbf{IND}_X). Now we transform P into P' by substituting P'_i for P_i for all agents i . We show that these substitutions do not modify the collective outcome $f(P)$. For suppose not, then $f(P) \not\models_{\varphi} f(P')$. From this it follows that $\exists i \in N$ such that either $P_i \not\models_{\varphi} f(P)$ or $P'_i \not\models_{\varphi} f(P')$ and hence that, in both cases, f would be manipulable, against the assumption. It follows that $f(P) =_{\varphi} f(P')$ (consequent of \mathbf{IND}_X). \square

Intuitively, the theorem proves that non-manipulability amounts precisely to the combination of monotonicity and independence. It is also instructive to notice that the result holds independently of any further assumptions on the structure of the agenda. More noticeably, it would still hold by dropping the assumption of collective rationality of f , which is not necessary to derive the result, as a simple inspection of the proof shows.

Before moving to the next section, it is worth observing that theorems 2 and 3 establish a strict link between independence and non-manipulability of both agendas and (together with monotonicity) votes.

4.3 The impossibility of non-manipulability

This section can be seen on a par with Section 2.2 in Chapter 2. We will introduce, prove and discuss an impossibility result relating manipulability with dictatorship:

For path-connected agendas, an aggregation function is responsive and non-manipulable if and only if it is a dictatorship.

In other words, when the agenda is path-connected (Definition 7) it is impossible to aggregate in a non-trivial way individual judgments into a collective one without violating responsiveness or introducing the possibility of manipulation.

4.3.1 Preliminaries

We will prove the theorem by resorting to the same technique we used in Chapter 2, the ultrafilter method. This allows us to reuse the dictatorship lemma (Lemma 4). We will then have to prove a variant of the ultrafilter lemma (Lemma 3), as this time we are working with weaker aggregation conditions—we do not have systematicity—but stronger agenda conditions—we have path-connectedness.

In order to make the argument behind the proof more intelligible, we will prove some of the lemmata establishing the theorem with respect to the agenda $\pm\{a \succ b, b \succ c, c \succ a\}$ which, we have seen in Example 6, is path connected.¹ The proof in its full generality can be found in [15].

4.3.2 Lemmata

Unanimity

We first show that responsiveness and non-manipulability imply unanimity.

Lemma 5 (Unanimity). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure. If an aggregation function f satisfies **RES** and non-**MAN**, then it satisfies **U**.*

Proof. Assume $\exists \varphi \in A$ s.t. $\forall i \in N : P_i \models \varphi$. By **RES**, $\exists P'$ s.t. $f(P') \models \varphi$. Recall that, by Theorem 3, f satisfies **IND** and **MON**. Take now P' and replace, for all i , P'_i with P_i , thus obtaining P . For each replacement we have two cases: 1) $P'_i \models \varphi$, or 2) $P'_i \not\models \varphi$. If 1) is the case then, by **MON**, we have that $f(P) \models \varphi$. If 2) is the case then, by **IND**, we also have that $f(P) \models \varphi$, which concludes the proof. \square

Contagion & characterization by winning coalitions

Recall first Definition 10 which has been introduced in Chapter 2.

Lemma 6 (Contagion lemma). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure where A satisfies **PC** and f be an aggregation function that does not satisfy **MAN**:*

$$\forall \varphi, \psi \in A : \mathcal{W}_\varphi = \mathcal{W}_\psi$$

¹The theorem for this agenda will hence be a variant of the Gibbard-Satterthwaite theorem.

Proof. We split the proof in two directions. $[\mathcal{W}_\psi \supseteq \mathcal{W}_\varphi]$ Assume $C \in \mathcal{W}_\varphi$. By PC (Definition 7), we have that $\varphi = \varphi_1 \models_c \dots \models_c \varphi_k = \psi$ for some $\varphi_1, \dots, \varphi_k \in A$. We show that, $\forall j : 1 \leq j \leq k, C \in \mathcal{W}_{\varphi_j}$. Proceed by induction.

B: Let $j = 1$, then the claim holds by assumption.

S: Let $1 \leq j < k$, and assume (IH) that $C \in \mathcal{W}_{\varphi_j}$. We prove that $C \in \mathcal{W}_{\varphi_{j+1}}$. Since $\varphi_j \models_c \varphi_{j+1}$ by assumption, there exists $X \subseteq A$ s.t. $X \cup \{\varphi_j, \neg\varphi_{j+1}\}$ is inconsistent but $X \cup \{\varphi_j\}$, $X \cup \{\neg\varphi_{j+1}\}$, $X \cup \{\varphi_j, \varphi_{j+1}\}$ and $X \cup \{\neg\varphi_j, \neg\varphi_{j+1}\}$ are all consistent. Define now a profile P as follows:

$$P_i = \begin{cases} X \cup \{\varphi_j, \varphi_{j+1}\} & \text{if } i \in C \\ X \cup \{\neg\varphi_j, \neg\varphi_{j+1}\} & \text{if } i \in N - C \end{cases}$$

We can observe the following. First, by Lemma 5, $X \subseteq f(P)$. Then, by IH, $C \in \mathcal{W}_{\varphi_j}$ and since $P_{\varphi_j} = C$ we have that $\varphi_j \in f(P)$. Now, since $X \cup \{\varphi_j, \neg\varphi_{j+1}\}$, we also have that $\varphi_{j+1} \in f(P)$. For any profile P' we therefore have that if $P'_{\varphi_j} = C$ then $P'_{\varphi_{j+1}} = C$. From which we conclude $C \in \mathcal{W}_{\varphi_{j+1}}$ as desired.

$[\mathcal{W}_\varphi \subseteq \mathcal{W}_\psi]$ The proof of this direction is similar. \square

Intuitively, the lemma states that if f is non-manipulable on a path-connected agenda, then the winning coalitions of all elements of the agenda coincide. It might be instructive to compare this lemma and its proof to Lemma 1, which establishes the same result by using systematicity.

The winning coalitions for all elements of the agenda are the same. So we can work with one unique set of winning coalitions, which has been defined earlier in Formula 2.3, and prove that any non-manipulable function, on path-connected agendas, can be characterized in terms of the set of winning coalitions.

Lemma 7 (Winning coalitions). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure where A satisfies PC and f be an aggregation function that does not satisfy MAN. For all $P \in \mathbf{P}$ and $\varphi \in A$:*

$$f(P) \models \varphi \quad \text{IFF} \quad P_\varphi \in \mathcal{W}.$$

Proof. [RIGHT TO LEFT] It holds directly by the above definition of \mathcal{W} . [LEFT TO RIGHT] Consider the set of voters P_φ . For any $P' \in \mathbf{P}$, by IND, which follows by Theorem 3 we have that if $P_\varphi = P'_\varphi$ then $f(P') \models \varphi$. Hence $P_\varphi \in \mathcal{W}$ according to the above definition. \square

Voters tripartition

We establish now a property of an aggregation problem which follows from path-connectedness. We will prove it for the special case of the path-connected agenda $\pm\{a \succ b, b \succ c, c \succ a\}$.

Lemma 8 (Voters tripartition). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure such that $A = \pm\{a \succ b, b \succ c, c \succ a\}$. There exists an inconsistent set $X \subseteq A$ and three consistent sets $Y_1, Y_2, Y_3 \subseteq X$ s.t. $\forall i \in \{1, 2, 3\}: (X - Y_i) \cup \neg Y_i$ is consistent, where $\neg Y_i = \{\neg\varphi \mid \varphi \in Y_i\}$.*

Proof. It suffices to provide the desired set $X = \{a \succ b, b \succ c, c \succ a\}$, and $Y_1 = \{a \succ b\}$, $Y_2 = \{b \succ c\}$ and $Y_3 = \{c \succ a\}$:

$$\begin{aligned} (X - Y_1) \cup \neg Y_1 &= \{\neg(a \succ b), b \succ c, c \succ a\} \\ (X - Y_2) \cup \neg Y_2 &= \{a \succ b, \neg(b \succ c), c \succ a\} \\ (X - Y_3) \cup \neg Y_3 &= \{a \succ b, b \succ c, \neg(c \succ a)\} \end{aligned}$$

where, recall, $\neg(x \succ y) = y \succ x$. This completes the proof. \square

Ultrafilters of winning coalitions

Lemma 9 (Ultrafilter lemma). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure such that $A = \pm\{p, q, p \wedge q\}$. The set \mathcal{W} is an ultrafilter, i.e.:*

- i) $N \in \mathcal{W}$;
- ii) if $C \in \mathcal{W}$ then $\neg C \notin \mathcal{W}$;
- iii) \mathcal{W} is upward closed: if $C \in \mathcal{W}$ and $C \subseteq C'$ then $C' \in \mathcal{W}$;
- iv) \mathcal{W} is closed under finite meets: if $C_1, C_2 \in \mathcal{W}$ then $C_1 \cap C_2 \in \mathcal{W}$.

Proof. We prove the lemma claim by claim.

i) The claim follows directly from unanimity (Lemma 5).

ii) The claim follows from the assumption of collective rationality built in the definition of aggregation function. [LEFT TO RIGHT] By reductio ad absurdum, suppose both $C, \neg C \in \mathcal{W}$. Consider now a profile where the judgment sets of the agents in C contain φ and those in $\neg C$ contain $\neg\varphi$. This profile must exist by the definition of aggregation function (Definition 3), and it would be inconsistent, which is impossible according to the same definition. [RIGHT TO LEFT] By contraposition, suppose $C \notin \mathcal{W}$. By Lemma 1, $\nexists \varphi \in A$ s.t. $C \in \mathcal{W}_\varphi$, so $\forall \varphi \in A, \exists P$ s.t. $C = \{i \mid P_i \models \psi\}$ and $f(P) \not\models \varphi$. Hence, $\forall \varphi \in A, \exists P$ s.t. $\neg C = \{i \mid P_i \not\models \psi\}$ and $f(P) \not\models \varphi$. By **IND**, which holds by Theorem 3, it follows that $\forall \varphi \in A, \forall P$ s.t. $\neg C = \{i \mid P_i \models \neg\varphi\}$ and $f(P) \models \neg\varphi$, hence $\neg C \in \mathcal{W}$.²

iii) The claim is a direct consequence of monotonicity.

iv) The claim can be proven by reductio ad absurdum as follows. Assume $C_1, C_2 \in \mathcal{W}$ and suppose that $C_1 \cap C_2 \notin \mathcal{W}$. Now put $C' = C_2 - C_1$ and $C'' = N - C_2$. Notice that $C_1 \cap C_2, C_2 - C_1$ and $\neg C_2$ are three disjoint sets covering N . Define now the following profile, which exists by Lemma 8:

$$P_i = \begin{cases} \{\neg(a \succ b), b \succ c, c \succ a\} & \text{if } i \in C_1 \cap C_2 = C \\ \{a \succ b, \neg(b \succ c), c \succ a\} & \text{if } i \in C_2 - C_1 = C' \\ \{a \succ b, b \succ c, \neg(c \succ a)\} & \text{if } i \in N - C_2 = C'' \end{cases}$$

As $C_1 \cap C_2 \notin \mathcal{W}$ by assumption, from ii) it follows that $N - (C_1 \cap C_2) = C' \cup C'' \in \mathcal{W}$. Since $C_1 \in \mathcal{W}$ by assumption, it follows by iii) that $C_1 \subseteq C' \cup C'' \in \mathcal{W}$. Finally $(C_1 \cap C_2) \cup (C_2 - C_1) = C_2 \in \mathcal{W}$ by assumption. It follows that $f(P)$ satisfies: $a \succ b, b \succ c$ and $c \succ a$, which is impossible. \square

As we can now rely for the existence of a dictator on Lemma 4, we are set to state and prove the theorem we are after.

²Notice that the proof of this claim is identical to the proof of the analogous claim in Lemma 3, except for the fact that we resort here to **IND** instead of **SYS**.

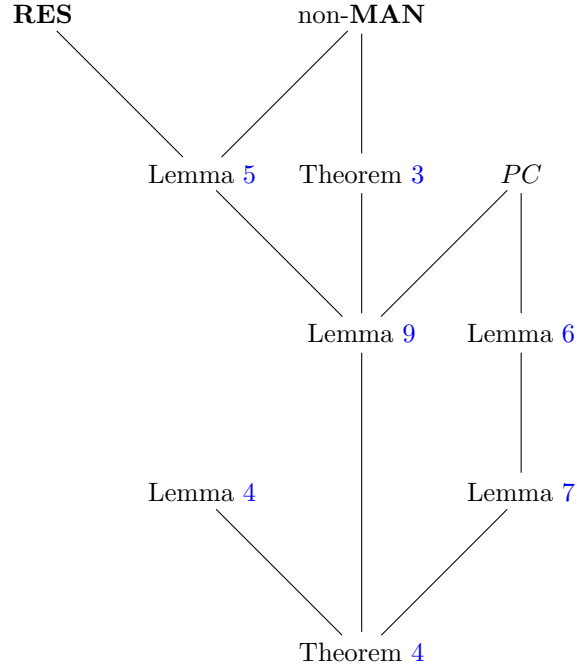


Figure 4.1: Structure of the proof of Theorem 4

4.3.3 The theorem

Theorem 4. *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure such that A satisfies PC . An aggregation function f satisfies **RES** and **non-MAN** iff it satisfies **D**.*

Proof. [RIGHT TO LEFT] If f satisfies **D** then it trivially satisfies **RES** and **non-MAN**. [LEFT TO RIGHT] By Lemma 7, for any $P \in \mathbf{P}$ and $\varphi \in A$:

$$f(P) \models \varphi \quad \text{IFF} \quad P_\varphi \in \mathcal{W}.$$

Then, by Lemmata 9 and 4 we have that $\{i\} \in \mathcal{W}$ for some $i \in N$ and hence:

$$P_\varphi \in \mathcal{W} \quad \text{IFF} \quad i \in P_\varphi$$

which concludes the proof: $f(P) \models \varphi$ iff $P_i \models \varphi$. \square

The theorem provides a characterization of dictatorship in terms of non-manipulability and responsiveness, on path-connected agendas. As such, notice that its statement is fully analogous to Theorem 1, and its proof shares many similarities to the proof of Theorem 1. Figure 4.1 recapitulates the structure of the proof.

4.4 Strategic issues in JA

In the previous sections we have dealt with a notion of manipulability viewed as the mere possibility of manipulation. No considerations about the incentives

that a voter might have for actually realizing the possibility of manipulation have been made. This section obviates to this addressing the notion of *strategy-proofness* in the context of JA.

4.4.1 Preferences over judgment sets

Talking about incentives of voters means talking about their preferences, i.e., total preorders³ over the set of all possible judgment sets \mathbf{J} . Preferences are, however, not included in the representation of aggregation problems as provided by JA structures (Definition 1). So two possibilities arise. Either JA structure are to be extended with an explicit representation of voters preferences, or they can be built, through appropriate stipulations, from each judgment set.

According to this latter option, the preference of a voter will be a function of her judgment set. Let us call $g : \mathbf{J} \rightarrow \mathbf{J} \times \mathbf{J}$ such a function. The property of strategy-proofness of an aggregation function can be defined as follows:

Definition 13 (Strategy-proofness). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure and $X \subseteq A$. An aggregation function f is:*

Strategy-proof (SP) *w.r.t. g iff $\forall i \in N$ and $\forall P, P' \in \mathbf{P}$ s.t. $P =_{-i} P'$: $f(P) \succeq_i f(P')$ where $\succeq_i = g(P_i)$*

I.e., for all voters and all profiles, switching to a different judgment set is never profitable.

Intuitively, f is strategy-proof if no matter what the (g -generated) preferences of voters are, it is never strictly preferable for them to misrepresent their true judgment set. Put it in game-theoretic terms, it is always a weakly dominant strategy for all the voters to be truthful towards the aggregation function.

4.4.2 Strategy-proofness and manipulability

It turns out that, although introducing this strategic dimension, strategy-proofness is equivalent to non-manipulability under some rather reasonable assumptions.

Definition 14 (Closeness). *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure and J, J', J'' be judgment sets. We say that J' is at least as close as J'' to J (notation: $J' \sqsupseteq_J J''$) if and only if $\forall \varphi \in A$ if $J =_{\varphi} J''$ then $J =_{\varphi} J'$.*

Closeness provides a natural way of ordering judgment sets w.r.t. a given judgment set. It is easy to see that closeness generates a total preorder.

The following interesting result can be proven—we omit here the proof that can be found in [15].

Theorem 5. *Let $\mathcal{J} = \langle N, A \rangle$ be a JA structure, f an aggregation function and let g be defined as:*

$$g(J) = \sqsupseteq_J$$

*It holds that f satisfies **SP** w.r.t. g if and only if it does not satisfy **MAN**.*

In other words, if preferences are taken to be dictated by closeness, there is no difference between the possibility of manipulating, and its strategic opportunity.

³We recall that a total preorder is a binary relation which is reflexive, transitive and total.

Chapter 5

The research agenda of JA

In this chapter we briefly sketch a somewhat loose list of topics of on-going research in the field of JA. We by no means claim this list to be exhaustive, as it is definitely biased by the authors' research interests. Furthermore, our exposition will not be comprehensive, but rather geared toward the communication of the main ideas animating those lines of research.

Outline of the chapter. We have seen that Arrow's theorem can be derived as a corollary from one impossibility theorem in judgment aggregation. However, we did not address the other direction: can the theorems of judgment aggregation be derived from results obtained in preference aggregation? In Section 5.1 we review some preliminary answers to this issue.

In the past chapters we have also seen that judgment aggregation commonly studies whether there exists aggregation functions that satisfy few desirable conditions. A different and novel approach investigates aggregation functions with respect to the rationality postulates they can preserve. We present this idea in Section 5.2.

In Section 5.3 we introduce a new property, that of *agenda safety* [24], showing that aggregation problems can be investigated from a different angle, namely whether an agenda is safe with respect to a class of aggregation functions.

We conclude the chapter with an overview on the recent interest of researchers working in abstract argumentation for judgment aggregation problems. This chapter builds on material taken mainly from [31, 32, 29, 30, 24, 6]

5.1 JA vs. PA

Since the early days of JA, a natural question that presented itself to JA researchers was whether JA could be showed to subsume PA. The question was answered positively by Dietrich and List in [13], which obtained Arrow's theorem—PA's central impossibility result—as the corollary of a more general theorem concerning the aggregation of logical formulae. We have very briefly touched upon this theorem in Section 2.3:

If the agenda satisfies EN and PC then the aggregation function satisfies IND and U if and only if it satisfies D.

Arrow’s theorem follows then directly by showing—as we have done in Examples 4 and 6—that its agenda is evenly negatable and path-connected.¹

So Arrow’s theorem can be viewed as an instance of a more general JA impossibility. But can we go the other way around too? That is, can we view JA impossibility results as instances of PA impossibilities? This question has been partially investigated by Grossi in [31, 32] and has been given a first positive answer. That work provides a number of results at the interface of JA, PA and many-valued logics (see, for instance, [34]) and is based on the following simple observation: i) preferences (strict \succ and weak \succeq ones) can be studied in terms of numerical ranking functions u , e.g., on the $[0, 1]$ interval [11]; ii) numerical functions can ground logical semantics, like it happens in many-valued logic [34] where, like in propositional logic, the semantic clause $u(x) \leq u(y)$ typically defines the satisfaction by u of the implication $x \rightarrow y$:

$$u \models x \rightarrow y \text{ iff } u(x) \leq u(y). \quad (5.1)$$

Intuitively, implication $x \rightarrow y$ is true (or accepted, or satisfied) iff the rank of x is at most as high as the rank of y . Preference aggregation (on possibly weak preferences) can then be studied as an instance of JA on many-valued logics. In turn, JA can be studied as an instance of a type of PA defined on dichotomous preferences, thus enriching the picture of the logical relationship between PA and JA.

5.2 Aggregation under integrity constraints

In Chapter 2 we have briefly touched upon the possibility of viewing JA as the problem of aggregating vector of propositional valuations into one valuation:

$$\mathbf{V}^{|N|} \rightarrow \mathbf{V} \quad (5.2)$$

or, to put it otherwise, a matrix of 1-0 (acceptance-rejection) values into one 1-0 vector (recall Remark 4). While this representation of aggregation problems naturally lends itself to the sort of axiomatic analysis which we have carried out in Chapters 2 and 4, recent work presented by Grandi and Endriss in [29, 30] has investigated it from a slightly different angle.

The axiomatic approach to aggregation, as we have presented it, aims at establishing the equivalence (or, more weakly, the respective inclusion) of classes of aggregation functions satisfying certain axioms, under given agenda conditions. E.g., we have seen, the class of aggregation functions satisfying systematicity and unanimity is equivalent to the class of aggregation functions satisfying dictatorship (Theorem 1). Under the standard interpretation, these results establish the *impossibility* of aggregation by showing that the class of aggregation functions which satisfies a given set of axioms—e.g., unanimity, systematicity and non-dictatorship—if intersected with the class of functions which preserves collective rationality—e.g., propositional consistency—yields the empty set.

The work in [29, 30] elaborates upon this latter interpretation of impossibility, by asking what sort of ‘rationality’ each axiom is then able to preserve.

¹It is worth mentioning, in passing, another interesting generalization of Arrow’s theorem via lattice theory provided in [10].

Rationality is then generalized by the notion of *integrity constraint*. For instance, the aggregation w.r.t. agenda $\pm\{p, q, p \wedge q\}$ could be seen, abstractly, as the problem of aggregating vectors of sequences

$$\langle v(p), v(q), v(p \wedge q) \rangle$$

where $v(\varphi) \in \{1, 0\}$. Clearly, some of those $1 - 0$ sequences are ruled out by the assumption we are dealing with propositional valuations, in this case: $\langle 1, 1, 0 \rangle$. Treating $p \wedge q$ as an independent issue r , the integrity constraint corresponding to the propositional consistency assumption in this agenda is, therefore:

$$(p \wedge q \wedge r) \vee (p \wedge \neg q \wedge \neg r) \vee (\neg p \wedge q \wedge \neg r) \vee (\neg p \wedge \neg q \wedge \neg r)$$

In this reader we have worked with constraints formalizable in propositional logic. A typical example of the questions addressed in [29, 30] is: what axiom characterizes the class of aggregation functions that preserve any integrity constraint expressible in propositional logic? The answer is: the class of aggregation functions satisfying *generalized dictatorship*, i.e., those functions which, for any profile, output the judgment set of one of the voters (not necessarily always the same!).

In general, the issue becomes of classifying aggregation functions not so much in terms of aggregation conditions, but in terms of the sort of rationality postulates that they are able to preserve or, to be more precise, in terms of the logical languages whose formulae, viewed as integrity constraints, they are able to preserve.

5.3 Agenda safety

Recent work presented in [24] has introduced a novel property of agendas called *safety*: an agenda A is safe w.r.t. a class C of aggregation functions if every function in C preserves consistency in the A .

The property of safety provides an interesting angle from which to look at aggregation problems. Instead of tackling the question of the existence of aggregation functions satisfying certain properties—this is yet another possible reading of (im)possibility results—the issue becomes of checking whether given aggregation functions satisfy collective rationality w.r.t. the underlying agenda.

In this view, the discursive dilemma shows that the agenda $\pm\{p, q, r, p \leftrightarrow (p \wedge q)\}$ is not safe w.r.t. propositionwise majority voting. Natural questions then arise of the following type: what structural conditions should an agenda satisfy in order to be safe for the aggregation functions defined by some given set of aggregation axioms? The work in [24] provides a first set of answers to this questions.

5.4 JA and abstract argumentation

A set of arguments and a defeat relation among them is called *argumentation framework*. Given an argumentation framework, argumentation theory identifies and characterizes the sets of arguments (*extensions*) that can reasonably

survive the conflicts expressed in the argumentation framework, and therefore can collectively be accepted. In general, there are several possible extensions for a set of arguments and a defeat relation on them [20].

For example, in Fig. 5.1, we have that A attacks B and that B attacks A .



Figure 5.1: An argumentation framework.

There are three possible extensions for this argumentation framework, namely those pictured in Fig 5.2. The black color means that the argument is rejected, white means that it is accepted and grey means that it is undecided, i.e. one does not take a position about it.

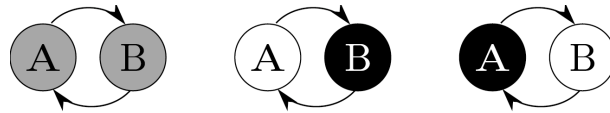


Figure 5.2: The three possible extensions for the argumentation framework of Fig. 5.1.

Hence, given an argumentation framework, different individuals may provide different evaluations regarding what should be accepted and rejected. The question is then the following: how can individual evaluations be mapped into a collective one? Similarly as in judgment aggregation, where the acceptance or rejection of a proposition may yield to the acceptance or rejection of another one, in argumentation the acceptance of one argument may force to reject another one. The aggregation of individual evaluations of a given argumentation framework raises the same problems as the aggregation of individual judgments. Indeed argument-by-argument majority voting may result in an unacceptable extension, as the proposition-wise majority voting may output an inconsistent collective judgment set. Judgment aggregation can be then addressed as the problem of combining different individual evaluations of the situation represented by an argumentation framework.

The first who applied abstract argumentation to judgment aggregation problems were Caminada and Pigozzi in [6]. The reason for using abstract argumentation is twofold: on the one hand, the existence of different argumentation semantics allows us to be flexible when defining which social outcomes are permissible. On the other hand, it allows us to bring judgment aggregation from classical logic to nonmonotonic reasoning.

To mention is also the work by Rahwan and Tohmé [65] who – given an argumentation framework – address the question of how to aggregate individual labellings into a collective position. By drawing on a general impossibility theorem from judgment aggregation, they prove an impossibility result and provide some escape solutions. Moreover, in [64] Rahwan and Larson explore welfare properties of collective argument evaluation.

Whereas the literature on judgment aggregation is concerned with the unpleasant occurrence of irrational collective outcomes, the interest of [6] is not only to guarantee a consistent group outcome, but also that such outcome is

compatible with the individual judgments. Group inconsistency is not the only undesirable outcome. It may happen, for example, that majority rule selects as social outcome a consistent combination of reasons and conclusion that actually no member voted for (a remembrance of another voting paradox, the *multiple election paradox* [5]). Such situation may be not a desirable collective outcome as it may conflict with some of its members' judgments. The research question tackled in [6] was precisely this: when is a group outcome 'compatible' with its members' judgments? The goal is to define a group decision making in which any group member is able to defend the group decision without having to argue against his own private beliefs.

Three operators (sceptical, credulous and super credulous) are defined and investigated in [6], as well as their properties. It is shown that, by iterating the aggregation process, not only a collective consistent decision is guaranteed, but that this is also unique. As a side remark, we should mention that the three operators do not satisfy **IND**. Furthermore, another property that is not satisfied is the preservation of a unanimously supported outcome. This is because the aggregated judgment is not merely the sum of the individual judgments. It is very well possible that the same argument is accepted by different participants for different reasons, but that these reasons cancel each other out when being put together, which is the same effect we saw in the Paretian dilemma in Section 3.2.

It is to be observed also that the existing few papers on abstract argumentation and judgment aggregation have given examples of how to map a judgment aggregation problem into an argumentation framework. However, whether such mapping exists for *all* kinds of judgment aggregation problems is still an open question.

In a follow-up paper [7] the intuition that, although every social outcome that is compatible with one's own labelling is acceptable, some outcomes are more acceptable than others, has been formalised and examined. This observation led to two new research questions:

- (i) Are the social outcomes of the aggregation operators in [6] Pareto optimal if preferences between different outcomes are also taken into account?
- (ii) Do agents have an incentive to misrepresent their own opinion in order to obtain a more favourable outcome? And if so, what are the effects from the perspective of social welfare?

Pareto optimality is a key principle of welfare economics which intuitively stipulates that a social state cannot be further improved. [7] studies whether the compatible social outcomes selected by the aggregation operators of in [6] are Pareto optimal. The results show that the two aggregation operators are Pareto optimal, when a certain distance is used.

The answer to the second question is that, though manipulability is usually considered to be an undesirable property of social choice decision rules and, while the operators considered are manipulable, one of them guarantees that an agent who lies does not only ensure a preferable outcome for himself, but even promotes social welfare, what we call a *benevolent lie*.²

²A similar idea was introduced and studied earlier in [33], where manipulation was seen as a coordinated action of the whole group.

Bibliography

- [1] K. Arrow. *Social choice and individual values*. Cowles Foundation Monograph Series, 1963.
- [2] D. Black. On the rationale of group decision making. *The Journal of Political Economy*, 56:23–34, 1948.
- [3] L. Bovens and W. Rabinowicz. Democratic answers to complex questions. an epistemic perspective. *Synthese*, 150:131–153, 2006.
- [4] S. Brams, M. Kilgour, and R. Sanver. A minimax procedure for electing committees. *Public Choice*, 132(3-4):401–420, 2007.
- [5] S.J. Brams, D.M. Kilgour, and W.S. Zwicker. The paradox of multiple elections. *Social Choice and Welfare*, 15(2):211–236, 1998.
- [6] M. Caminada and G. Pigozzi. On judgment aggregation in abstract argumentation. *Autonomous Agents and Multi-Agent Systems*, 22(1):64–102, 2011.
- [7] M. Caminada, G. Pigozzi, and M. Podlaskowski. Manipulation in group argument evaluation. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011)*, 2011.
- [8] H. Cartan. Filtrés et ultrafiltrés. *Comptes Rendus de l'Académie des Sciences*, pages 777–779, 1937.
- [9] B. Chapman. Rational aggregation. *Politics, Philosophy & Economics*, 1(3):337–354, Sep 2002.
- [10] T. Daniëls and E. Pacuit. A general approach to aggregation problems. *Journal of Logic and Computation*, 19(3):517–536, 2008.
- [11] G. Debreu. Representation of a preference ordering by a numerical function. In R. M. Thrall, C. H. Coombs, and R. L. Davis, editors, *Decision Processes*. John Wiley, 1954.
- [12] F. Dietrich. Judgment aggregation: (im)possibility theorems. *Economic Theory*, 126:286–298, 2006.
- [13] F. Dietrich and C. List. Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, 29(19–33), 2007.
- [14] F. Dietrich and C. List. Judgment aggregation by quota rules: Majority voting generalized. *Journal of Theoretical Politics*, 19:391–424, 2007.

-
- [15] F. Dietrich and C. List. Strategy-proof judgment aggregation. *Economics and Philosophy*, 23:269–300, 2007.
- [16] F. Dietrich and C. List. Majority voting on restricted domains. *Journal of Economic Theory*, 145(2):512–543, Feb 2010.
- [17] F. Dietrich and P. Mongin. The premiss-based approach to judgment aggregation. *Journal of Economic Theory*, 145(2):562 – 582, 2010.
- [18] E. Dokow and R. Holzman. Aggregation of binary evaluations. *Journal of Economic Theory*, 145(2):495–511, 2010.
- [19] Elad Dokow and Ron Holzman. Aggregation of binary evaluations with abstentions. *Journal of Economic Theory*, 145(2):544–561, Mar 2010.
- [20] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n -person games. *Artificial Intelligence*, 77:321–357, 1995.
- [21] D. Eckert and C. Klamler. How puzzling is judgment aggregation? antipolarity in distance-based aggregation rules. *Working paper. University of Graz*, pages 1–7, 2007.
- [22] D. Eckert and J. Mitlöhner. Logical representation and merging of preference information. *Multidisciplinary IJCAI-05 workshop on advances in preference handling*, pages 85–87, 2005.
- [23] D. Eckert and G. Pigozzi. Belief merging, judgment aggregation, and some links with social choice theory. In *Dagstuhl Seminar Proceedings 05321*, 2005.
- [24] U. Endriss, U. Grandi, and D. Porello. Complexity of judgment aggregation: Safety of the agenda. In *Proceedings of AAMAS’10*, 2010.
- [25] P. Everaere, S. Konieczny, and P. Marquis. The strategy-proofness landscape of merging. *Journal of Artificial Intelligence Research*, 28:49–105, 2007.
- [26] P. C. Fishburn. Arrow’s impossibility theorem: Concise proof and infinite voters. *Journal of Economic Theory*, 2(1):103–106, 1970.
- [27] W. Gaertner. *A Primer in Social Choice Theory*. Oxford University Press, 2006.
- [28] P. Gärdenfors. A representation theorem for voting with logical consequences. *Economics and Philosophy*, 22:181–190, 2006.
- [29] U. Grandi and U. Endriss. Lifting rationality assumptions in binary aggregation. In *Proceedings of AAAI’10*, 2010.
- [30] U. Grandi and U. Endriss. Binary aggregation with integrity constraints. In *Proceedings of IJCAI’11*, 2011.
- [31] D. Grossi. Unifying preference and judgment aggregation. In P. Decker and J. Sichman, editors, *Proceedings of the Eighth International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2009)*, pages 217–224. ACM Press, 2009.

- [32] D. Grossi. Correspondences in the theory of aggregation. In G. Bonanno, B. Loewe, and W. van der Hoek, editors, *Logic and the Foundation of Game and Decision Theory—LOFT 08, Revised and Selected Papers*, volume 6006 of *LNAI*, pages 34–60. Springer, 2010.
- [33] D. Grossi, G. Pigozzi, and M. Slavkovik. White manipulation in judgment aggregation. Proceedings of BNAIC 2009 - The 21st Benelux Conference on Artificial Intelligence, 2009.
- [34] R. Hähnle. Advanced many-valued logics. In D. M. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic, 2nd Edition*, volume 2, pages 297–395. Kluwer, 2001.
- [35] S Hartmann, G Pigozzi, and J Sprenger. Reliable methods of judgement aggregation. *Journal of Logic and Computation*, 20(2):603–617, Apr 2010.
- [36] J. Kemeny. Mathematics without numbers. *Daedalus*, 88:577–591, 1959.
- [37] C. Klamler and D. Eckert. A simple ultrafilter proof for an impossibility theorem in judgment aggregation. *Economics Bulletin*, 29(1):319–327, 2009.
- [38] S. Konieczny and E. Grégoire. Logic-based approaches to information fusion. *Information Fusion*, 7:4–18, 2006.
- [39] S. Konieczny, J. Lang, and P. Marquis. DA^2 merging operators. *Artificial Intelligence*, 157:49–79, 2004.
- [40] S. Konieczny and R. Pino-Pérez. Merging with integrity constraints. *Fifth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU’99)*, 7:233–244, 1999.
- [41] S. Konieczny and R. Pino Pérez. On the frontier between arbitration and majority. In *Proceedings of the Eighth International Conference on Principles of Knowledge Representation and Reasoning*, pages 109–118, 2002.
- [42] LA Kornhauser. Modeling collegial courts. ii. legal doctrine. *Journal of Law, Economics, and Organization*, 8:441, 1992.
- [43] L.A. Kornhauser and L.G. Sager. Unpacking the court. *Yale Law Journal*, 96:82–117, 1986.
- [44] L.A. Kornhauser and L.G. Sager. The one and the many: Adjudication in collegial courts. *California Law Review*, 81:1–51, 1993.
- [45] C List. A possibility theorem on aggregation over multiple interconnected propositions. *Mathematical Social Sciences*, 45(1):1–13, Oct 2002.
- [46] C. List. A model of path-dependence in decisions over multiple propositions. *American Political Science Review*, 98(3):495 – 513, 2004.
- [47] C List. Corrigendum to ”a possibility theorem on aggregation over multiple interconnected propositions”. *Mathematical Social Sciences*, 52:109–110, Sep 2005.

- [48] C. List. The theory of judgment aggregation: An introductory review. Lse choice group working papers, London School of Economics, 2010.
- [49] C. List and P. Pettit. Aggregating sets of judgments: An impossibility result. *Economics and Philosophy*, 18:89–110, 2002.
- [50] C. List and P. Pettit. Aggregating sets of judgments: Two impossibility results compared. *Synthese*, 140:207–235, 2004.
- [51] C List and P Pettit. Aggregating sets of judgments: Two impossibility results compared. *Synthese*, 140(1):207–235, 2004.
- [52] Christian List. The probability of inconsistencies in complex collective decisions. *Social Choice and Welfare*, 24(1):3–32, 2005.
- [53] Michael K Miller and Daniel Osherson. Methods for distance-based judgment aggregation. *Social Choice and Welfare*, 32(4):575–601, 2009.
- [54] P Mongin. Factoring out the impossibility of logical aggregation. *Journal of Economic Theory*, 141(1):100–113, Nov 2008.
- [55] P. Mongin. Judgment aggregation. In S.O. Hansson and V.F. Hendricks, editors, *The Handbook of Formal Philosophy*. Springer, 2011.
- [56] P Mongin and F Dietrich. Un bilan interprétatif de la théorie de l’agrégation logique. *Revue d’économie politique*, 120(6):929–972, 2010.
- [57] K. Nehring. The (im)possibility of a Paretian rational. Economics working papers, Institute for Advanced Study, School of Social Science, Nov 2005.
- [58] K. Nehring and C. Puppe. Strategy-proof social choice on single-peaked domains: Possibility, impossibility and the space between. Working Paper, University of California at Davis.
- [59] K. Nehring and C. Puppe. Abstract arrowian aggregation. *Journal of Economic Theory*, 145(2):467–494, 2010.
- [60] P. Odifreddi. Ultrafilters, dictators and gods. In C. Calude and G. Păun, editors, *Finite Versus Infinite. Contributions to an Eternal Dilemma.*, pages 239–246. Springer, 2000.
- [61] M. Pauly and M. van Hees. Logical constraints on judgment aggregation. *Journal of Philosophical Logic*, 35:569–585, 2006.
- [62] P. Pettit. Deliberative democracy and the discursive dilemma. *Philosophical Issues*, 11:268–299, 2001.
- [63] G. Pigozzi. Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. *Synthese*, 152:285–298, 2006.
- [64] I. Rahwan and K. Larson. Welfare properties of argumentation-based semantics. In *Proceedings of the 2nd International Workshop on Computational Social Choice (COMSOC)*, 2008.

- [65] I. Rahwan and F. Tohmé. Collective argument evaluation as judgment aggregation. In *Proc. of 9th AAMAS*, 2010.
- [66] A. Sen. A possibility theorem on majority decisions. *Econometrica*, 34:491–499, Feb 1966.
- [67] A. K. Sen. *Collective Choice and Social Welfare*. Holden Day, 1970.
- [68] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
- [69] H.P. Young and A. Levenglick. A consistent extension of condorcet’s election principle. *SIAM Journal of Applied Mathematics*, 35:285–300, 1978.