Introduction to Judgment Aggregation

Davide Grossi[1] and Gabriella Pigozzi[2]

[1] Department of Computer Science, University of Liverpool, UK
d.grossi@liverpool.ac.uk
[2] LAMSADE, Université Paris Dauphine, France
gabriella.pigozzi@lamsade.dauphine.fr

Abstract. The present notes are an improved version of the notes that served as teaching materials for the course *Introduction to Judgment Aggregation* given at the 23rd European Summer School on Logic, Language and Information (ESSLLI'11, Ljubljana). The notes are structured as follows: Section 1 introduces the field of judgment aggregation, its relations to preference aggregation and some formal preliminaries. Section 2 shows that the paradox that originated judgment aggregation is not a problem limited to propositionwise majority voting but a more general issue, illustrated by an impossibility theorem of judgment aggregation that is here proven. The relaxation of some conditions used in impossibility results in judgment aggregation may lead to escape routes from the impossibility theorems. These escape routes are explored in Section 3. Section 4 presents the issue of manipulation that arises when voters strategically misrepresent their true vote in order to force a different outcome in the aggregation process. Finally, we conclude by sketching a list of on-going research in the field of judgment aggregation (Section 5).

1 Logic Meets Social Choice Theory

Outline. We start by presenting the paradox that originated the whole field of judgment aggregation and by looking at how judgment aggregation relates to the older theory of preference aggregation. 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

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 [13]. 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 later in this section.

Whereas voting theory studies the aggregation of individual preferences, the recent theory of judgment aggregation investigates how individual opinions on

N. Bezhanishvili et al. (Eds.): ESSLLI 2010/2011, Lectures, LNCS 7388, pp. 160–209, 2012.

logically related propositions can be consistently aggregated into a collective position.[1] 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 [47,48,46].

It is instructive to recall the original example that Kornhauser and Sager used to illustrate the doctrinal paradox [48]. 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.

Table 1. An illustration of the doctrinal paradox

	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

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 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. [48, p. 12]

The systematic account to the understanding of situations like the one in Table 1 has been provided by judgment aggregation. The first step was made by political philosopher Pettit [67], 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.

[1] Among the existing surveys and tutorial papers on judgment aggregation, we recall [56,53].

Then, List and Pettit [55] reconstructed Kornhauser and Sager' example as shown in Table 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 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 2.

Table 2. The discursive dilemma

	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

Is the move from the doctrinal paradox to the discursive dilemma an innocent one? Recently, Mongin and Dietrich [60,59] 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. [59, p. 2]

Instead of premises and conclusions, List and Pettit focused their attention on *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. Impossibility results will be the topic of Section 2.

Preference Aggregation and Judgment Aggregation. How individual preferences can be aggregated into a collectively preferred alternative is traditionally studied by social choice theory [2,72], whose origins can be traced back to the works by Borda and Condorcet [5,13]. In particular, Condorcet aimed at an aggregation procedure that would maximize the probability that a group of people take the right decision. His result, known as the *Condorcet Jury Theorem*, showed that - under certain conditions - majority voting was a good truth-tracking method. However, he also discovered a disturbing problem of majority voting. Given a set of individual preferences, the method suggested by Condorcet consisted in the comparison of 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: the *Condorcet paradox*.

Let a set of alternatives X. Let P be a binary predicate interpreted on a binary relation over X, that we denote by $\succ$ (or $\succ_i$ when we need to make the agent explicit). xPy means "x is strictly preferable to y". The desired properties of preference relations viewed as strict linear orders are:

(P1) $\forall x, y((xPy) \rightarrow \neg(yPx))$ (asymmetry)
(P2) $\forall x, y(x \neq y \rightarrow (xPy \vee yPx))$ (completeness)
(P3) $\forall x, y, z((xPy \wedge yPz) \rightarrow xPz)$ (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. Let $\succ_i$ denote agent i's preference over X (without any index, $\succ$ denotes the collective preference relation). The three voters' total preferences are the following: $V_1 = \{x \succ_1 y, y \succ_1 z\}$, $V_2 = \{y \succ_2 z, z \succ_2 x\}$ and $V_3 = \{z \succ_3 x, x \succ_3 y\}$. According to Condorcet's method, a majority of the voters (V_1 and V_3) prefers x to y, a majority (V_1 and V_2) prefers y to z, and another majority (V_2 and V_3) prefers z to x. This leads us to the collective outcome $x \succ y$, $y \succ z$ and $z \succ x$, which together with transitivity (P3) violates (P1). Each voter's preference is transitive, but transitivity fails to be mirrored at the collective level. This is an instance of the so-called Condorcet paradox.[2] The profile of preferences here considered is known as 'latin square', that is, each alternative is ranked top, second and last in someone's preference.

Condorcet's 1785 *Essai* was read and understood by a few until Edward J. Nanson (1882) [61] and Duncan Black (1958) [4]. In 1951 a young economist, the future Nobel prize winner Kenneth Arrow, showed that the Condorcet paradox is not a problem specific of pairwise majority comparison [2]. 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. That is, the collective preference coincides with the preference of one and the same individual of the group.

Thus, when we combine individual choices into a collective one, we may lose something that held at the individual level, like transitivity (in the case of preference aggregation) or logical consistency (in the case of judgment aggregation).

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

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. [47, p. 85][3]

Regarding the formal relations between judgment and preference aggregation, Dietrich and List [17] (extending earlier work by List and Pettit [55]) 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, completeness and transitivity axioms of P. The Condorcet paradox can then be represented as a judgment aggregation problem as Table 3 illustrates. Voters of Table 3 are perfectly consistent just like the judges of Table 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 [46] 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. [55, 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

[3] Different procedures for judgment aggregation have been assessed with respect to their truth-tracking capabilities, see [6,39].

Table 3. The Condorcet paradox as a doctrinal paradox

	xPy	yPz	xPz	yPx	zPy	zPx
$V_1 = \{x \succ_1 y, y \succ_1 z\}$	1	1	1	0	0	0
$V_2 = \{y \succ_2 z, z \succ_2 x\}$	0	1	0	1	0	1
$V_3 = \{z \succ_3 x, x \succ_3 y\}$	1	0	0	0	1	1
Majority	1	1	0	0	0	1

Condorcet paradox hides a much deeper problem that does not affect only the 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 Section 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.

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.

Agendas and Individual Judgments. The following is the first key definition of the framework of judgment aggregation:

Definition 1 (Judgment aggregation structure). *Let $\mathcal{L}(X)$ be a propositional language on a given set of atoms X.*[4] *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) contingent*[5] *formulae. An agenda based on a set of issues I will often be denoted $\pm I$.*

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 where double negations are eliminated so that each formula contains at most one negation.

[4] We will often drop the reference to X when clear from the context.

[5] I.e., which are neither a tautology nor a contradiction.

Definition 2 (Judgment sets and profiles). *Let $\mathcal{J} = \langle N, A \rangle$ be a judgment aggregation structure. A judgment set for $\mathcal{J}$ is a set of formulae $J \subseteq A$ 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 an $|N|$-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 (Judgment sets as valuation functions). Judgment sets are consistent and maximal subsets of the agenda. As such, they can equivalently be viewed as functions $J : A \longrightarrow \{1, 0\}$ preserving the meaning of propositional connectives. 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 $J : X \longrightarrow \{1, 0\}$.

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 judgment aggregation structure. An aggregation function for $\mathcal{J}$ is a function $f : \mathbf{P} \longrightarrow \mathbf{J}$.*

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

$$\varphi \in f_{maj}(P) \text{ IFF } |\{J_i \in P \mid \varphi \in J_i\}| \geq q, \quad (1)$$

with $q = \lceil (|N| + 1)/2 \rceil$, where $\lceil x \rceil$ is the smallest integer $\geq x$.
I.e., φ is collectively accepted iff there is a majority of voters accepting it.

Propositionwise Unanimity

$$\varphi \in f_u(P) \text{ IFF } \forall i \in N, \varphi \in J_i \tag{2}$$

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

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.

$$f_{pbp}(P) = f_{maj}(Prem) \cup \{\varphi \in Conc \mid f_{maj}(Prem) \models \varphi\} \tag{3}$$

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.

$$f_{cbp}(P) = f_{maj}(Conc) \tag{4}$$

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 **P** to **J**, or as a function from **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 section.

We conclude this section 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 [17]:

p: Current CO_2 emissions lead to global warming.
$p \to q$: If current CO_2 emissions lead to global warming, then we should reduce CO_2 emissions.
q: We should reduce CO_2 emissions.

The profile consisting of the three judgment sets $J_1 = \{p, p \to q, q\}$, $J_2 = \{p, \neg(p \to q), \neg q\}$ and $J_3 = \{\neg p, p \to q, \neg q\}$, once aggregated via propositionwise majority, gives rise to an inconsistent collective judgment set $J = \{p, p \to q, \neg q\}$. If we assume that $Prem = \{p, p \to q\}$ 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	$p \to q$	q
J_1	1	1	1
J_2	1	0	0
J_3	0	1	0
f_{maj}	1	1	0
f_{pbp}	1	1	1
f_{cbp}			0

Notice that f_u, propositionwise unanimity, yields the empty judgment set.

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 section shows that the latter is the right answer.

Outline. 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 sections. 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 judgment aggregation literature. The section builds on material and results taken from [17,41,53,65].

2.1 Preliminaries

The discursive dilemma unveils a problematic aspect of propositionwise majority voting. In this section 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.

[6] Notice that the agenda at issue is closed under atoms, hence each judgment set can be viewed as a propositional valuation function (cf. Remark 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 section and in Section 4.

Non-simplicity. The first one is almost self-explanatory, 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) $X \subseteq A$, which is consistent with φ and with $\neg\psi$, $\{\varphi\} \cup X \models \psi$.*

If $\varphi \models_c \psi$ but $\varphi \not\models \psi$ then, by the above definition, there exists $X \neq \emptyset$ such that $X \cup \{\varphi, \neg\psi\}$ is inconsistent. Now, by a well-known property of propositional logic, i.e., its compactness[7], we can conclude that there exists a smallest X' such that $X' \cup \{\varphi, \neg\psi\}$ is inconsistent and, therefore, minimally inconsistent.

Even Negations. The second agenda condition is slightly more involved.[8]

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

- *A contains a minimally inconsistent set $X \subseteq A$ s.t. it contains at least two formulae which, if negated, make the set consistent.*

Notice that this condition requires the existence of a minimally inconsistent set without any cardinality constraint (like in the definition of non-simplicity).

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

---

[7] Roughly, the compactness for propositional logic guarantees that, if $X \models \varphi$ (with X possible infinite), then there exits a smallest (finite) set $X' \subseteq X$ such that $X' \models \varphi$.

[8] In [17] the condition is referred to as *minimal connectedness*, in [53] as even number negatability. It is known to be equivalent to the non-affineness condition introduced in [22].

Example 3 (Non evenly negatable agendas). Consider $A = \pm\{p, q, p \leftrightarrow q\}$. We have the following minimally inconsistent sets: $\{p, \neg q, p \leftrightarrow q\}$, $\{\neg p, q, p \leftrightarrow q\}$, $\{p, q, \neg(p \leftrightarrow q)\}$. No consistent set can be obtained from any of these sets by negating any two formulae at the same time.

Example 4 (Evenly negatable agendas). Consider the agenda (which we have encountered in Section 1) $\pm\{a \prec b, b \prec c, c \prec a\}$ where $a \prec b$, $b \prec c$ and $c \prec a$ are taken to be atomic propositions. Assume now the following constraints, for $x, y \in \{a, b, c\}$ s.t. $x \neq y$:

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

These induce $\succ$ to behave according to a linear order.[9] 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 [19]) or *total-blockedness* (e.g. in [63]).

Definition 7 (Path-connected agendas). *An agenda A is* path connected *(PC) iff for all $\varphi, \psi \in A$ there exists a sequence $\varphi_1, \ldots, \varphi_n$ 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 direct or an indirect 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. Here are two examples:

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 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 $\pm\{p, q, p \vee q\}$.[10]

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 shows. Another agenda, although with 8 elements, satisfying PC which we have encountered in the previous section is the one of the discursive dilemma: $\pm\{p, q, r, r \leftrightarrow (p \wedge q)\}$.

[9] It might be instructive to compare the first-order formulation of these conditions given in Section 1.1.

[10] Recall that $p \rightarrow q$ is $\neg(p \wedge \neg q)$ and $p \vee q$ is $\neg(\neg p \wedge \neg q)$.

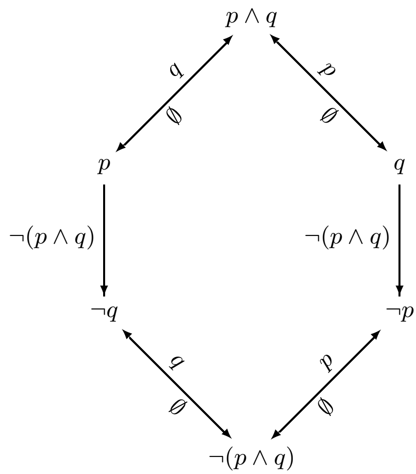

Fig. 1. The conditional entailment graph of agenda $\pm\{p, q, p \wedge q\}$. The lower elements are not connected to the upper elements. The arrows are labeled with (the elements of) the set X which establishes the conditional entailment.

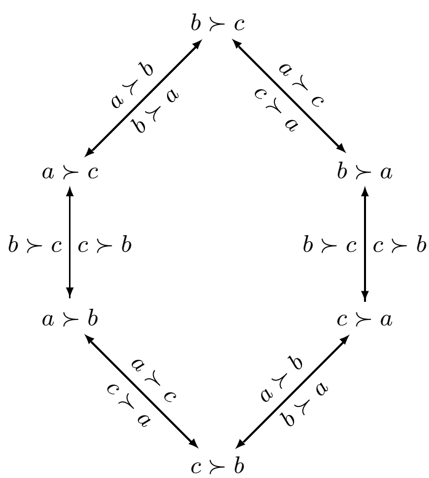

Fig. 2. The conditional entailment graph of agenda $\pm\{a \prec b, b \prec c, c \prec a\}$. Each formula is reachable from any other formula.

Comparing Agenda Conditions. The above agenda conditions are logically related as follows:

1. EN and NS are logically independent;
2. PC and EN are logically independent;
3. PC implies NS.

As to the first item: there exist agendas which satisfy NS but not EN (e.g., $\pm\{p, q, p \leftrightarrow q\}$) and agendas which satisfy EN but not NS (e.g., $\pm\{p, r, p \wedge q\}$). As to the second item: there exist agendas which satisfy PC but not EN (e.g., $\pm\{p, q, p \leftrightarrow q\}$ again), and agendas which satisfy EN but not PC (e.g., $\pm\{p, q, p \rightarrow q\}$). Notice also that the three conditions have a non-empty intersection (e.g. the discursive dilemma agenda $\pm\{p, q, r, r \leftrightarrow (p \wedge q)\}$). That PC implies NS is a direct consequence of Definition 7.

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

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 $[J \models \varphi$ 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 judgment aggregation:

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

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

Independent *(**IND**) iff $\forall \varphi \in A, \forall P, P' \in \mathbf{P}$: IF $[\forall i \in N : P_i \models \varphi$ IFF $P'_i \models \varphi]$ 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 $\forall \varphi, \psi \in A, \forall P, P' \in \mathbf{P}$: IF $[\forall i \in N : P_i \models \varphi$ IFF $P'_i \models \psi]$ THEN $[f(P) \models \varphi$ 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* $\exists i \in N$ S.T. $\forall \varphi \in A, \forall P \in \mathbf{P} : f(P) \models \varphi$ IFF $P_i \models \varphi$.
I.e., there exists a voter whose judgment set is always identical to the aggregated set.

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

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

Neutral *(***NEU***) iff* $\forall \varphi, \psi \in A, \forall P, P' \in \mathbf{P}$: IF $[\forall i \in N : P_i \models \varphi$ IFF $P_i \models \psi]$ THEN $[f(P) \models \varphi$ IFF $f(P) \models \psi]$.
I.e., if all voters in one same profile accept a formula φ if and only if they accept a formula ψ, then in the aggregated profile φ is accepted if and only ψ is.

These conditions state some very diverse constraints on how the aggregation relates the input—a judgment 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 judgment aggregation 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***,* **IND** *and* **NEU***).* Recall that each judgment set can be seen as a valuation $J : A \longrightarrow \{1, 0\}$ accepting or rejecting each agenda item (Remark 1). 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):

$$\begin{array}{ccc} P_1(\varphi_1) \ldots P_1(\varphi_m) & & P'_1(\varphi_1) \ldots P'_1(\varphi_m) \\ \ldots \quad \ldots \quad \ldots & & \ldots \quad \ldots \quad \ldots \\ P_n(\varphi_1) \ldots P_n(\varphi_m) & & P'_n(\varphi_1) \ldots P'_n(\varphi_m) \end{array}$$

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

What **IND** states is that, for any j with $1 \leq j \leq m$, if the j^{th} column in P consists of the same sequence of zeros and ones as j-th column in P', then the j-th element in the sequence of the aggregated judgment is the same in the two profiles.

Property **NEU** states something similar about two columns in one profile: for any two columns $1 \leq j \neq k \leq m$ in a given profile P, if the j-th column consists of the same sequence of zeros and ones as k-th column, then the two columns are the same also in the aggregated profile.

Finally, property **SYS** pulls **IND** and **NEU** together. It states that, for any two columns $1 \leq j \neq k \leq m$ in, respectively, profile P and profile P', if the

j-th column in P consists of the same sequence of zeros and ones as the k-th column in P', then the j-th and k-th values in the aggregated judgments of the two profiles are the same. So **SYS** is equivalent to the conjunction of **IND** and **NEU**.

Remark 4 (Aggregation of $1-0$ *matrices).* Along the line of Remark 3, if f is systematic then the only information from each profile 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, $\forall \varphi \in A$:[11]

$$f(P_1, \ldots, P_n)(\varphi) = g(P_1(\varphi), \ldots, P_n(\varphi)) \tag{5}$$

This yet more abstract view on aggregation has been systematically pursued in, among others, [22].

2.2 An Impossibility Theorem

The present section will provide an extensive discussion of the following impossibility result, due to [17]:

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

Put differently, 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.

Winning Coalitions. Given any judgment aggregation 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 formula, so does 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:

$$P_\varphi := \{i \in N \mid P_i \models \varphi\} \tag{6}$$

which denotes the set of all agents that accept φ. We can now define the notion of winning coalition as follows:

[11] Cf. [66].

Definition 10 (Winning coalitions for φ). *Let $\mathcal{J} = \langle N, A\rangle$ be a judgment aggregation 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 $\mathcal{W}_\varphi(\mathcal{J}, f)$.[12]

Remark 5. It is important to observe 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.

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 and Characterization by Winning Coalitions.

Lemma 1 (Contagion lemma). *Let $\mathcal{J} = \langle N, A\rangle$ be a judgment aggregation 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**.

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:

$$\mathcal{W} := \{C \subseteq N \mid \forall \varphi \in A, \forall P \in \mathbf{P} : \text{ IF } C = P_\varphi \text{ THEN } f(P) \models \varphi\} \quad (7)$$

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

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

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

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

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 7.

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 judgment aggregation structure and f an aggregation function such that A satisfies NS and EN and f satisfies* **SYS**. *The set $\mathcal{W}$ is an ultrafilter, i.e.:*[13]

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* $C \cap C' \in \mathcal{W}$.[14]

Proof. 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. $C = \{i \mid P_i \models \psi\}$ 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 (Definition 6). 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, the set $(X - \{\varphi, \psi\}) \cup \{\neg\varphi, \neg\psi\}$. Consider now these three coalitions which, notice, form a partition of N:[15]

$$C_1 := C$$
$$C_2 := C' - C$$
$$C_3 := N - C'$$

[13] Ultrafilters 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.

[14] Technically, condition iii) could be dispensed with, as it follows from the other three: if $C' \notin \mathcal{W}$ then—by ii)—$-C' \in \mathcal{W}$ and—by iv)—$C \cap -C' = \emptyset \in \mathcal{W}$, which is impossible given i) and ii). However, standard presentations of ultrafilters include it and the direct proof we will be giving of iii) is an interesting illustration of the sort of arguments underpinning the use of ultrafilters in social choice and judgment aggregation.

[15] See Figure 3.

and consider the judgment profile P thus defined:[16]

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

We can now conclude the following about P. By **U**, we have that $N \in \mathcal{W}$ and hence $X - \{\varphi, \psi\} \in f(P)$. 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 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:[17]

$$C_1 := C \cap C'$$
$$C_2 := C' - C$$
$$C_3 := N - C'$$

and consider the judgment profile P defined as follows:[18]

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

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

[16] This 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:

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

[17] See Figure 4.

[18] This 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, $\varphi := p$, $\psi := q$, $\xi := \neg(p \wedge q)$.

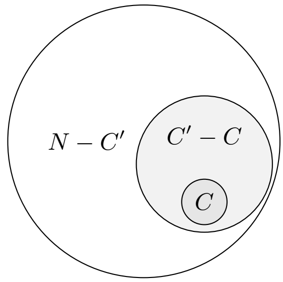

Fig. 3. Voters tripartition in claim iii) Lemma 3

$X \subseteq f(X)$, obtaining an inconsistent set, which is impossible. This proves claim iv) and concludes the proof of the lemma.

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.

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, $\bigcap \mathcal{W} = \{i\}$.

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 judgment aggregation property or construct, but singles out a general property of finite ultrafilters.

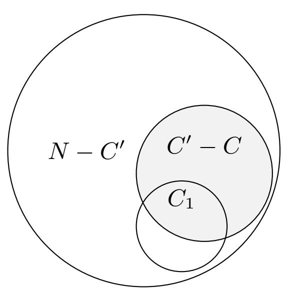

Fig. 4. Voters tripartition in claim iv) Lemma 3

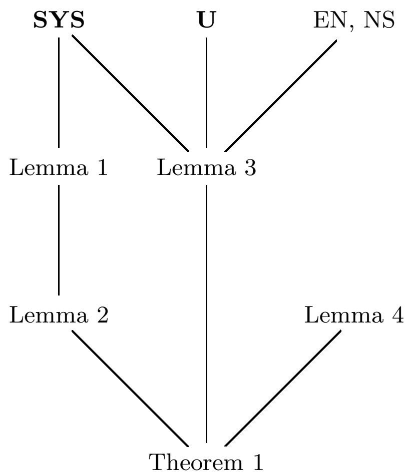

Fig. 5. Structure of the proof of Theorem 1

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

Theorem 1 ([17]). *Let $\mathcal{J} = \langle N, A \rangle$ be a judgment aggregation structure such that A satisfies NS and 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 $\varphi \in A$:

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

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

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

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

The first impossibility theorem of judgment aggregation, proven in the paper that initiated the field [54], is a direct consequence of Theorem 1. Agendas such as $\pm\{p, q, p \wedge q\}$ or $\pm\{p, q, p \rightarrow q\}$, which satisify NS and 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 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.

The Structure of Impossibility Theorems in Judgment Aggregation. Three types of constraints lie at the core of the aggregation problem:

- constraints on the level of connectedness of the agenda (*agenda constraints*);
- 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 4, 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 judgment aggregation 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.

To wrap up, impossibility theorems typically exhibit this structure:

> *If* the agenda satisfies constraints $C_1, \ldots, C_n$, *and* the domain of the aggregation function satisfies constraints $C'_1, \ldots, C'_n$, *then* the aggregation function satisfies constraints $C''_1, \ldots, C''_n$ *if and only if* the aggregation function is a dictatorship.

This wording makes also clear that these impossibility results, as well as their social choice theory counterparts, are actually also *possibility* results in the sense that dropping any of the aggregation constraints, under the same agenda and rationality constraints, guarantees that the aggregation function is not a dictatorship.[19]

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 4, which we adapted from [53], recapitulates in a compact way some of the better-known impossibility results that have been established in the judgment aggregation literature. One more impossibility result of special relevance for the notion of manipulability of an aggregation process will be studied in detail in Section 4.

[19] In fact, Arrow himself refers to his theorem in [1,2] as a "General Possibility Theorem".

Table 4. 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).

Agenda conditions	Aggregation conditions	Proof
NS, EN	$\mathbf{SYS}, \mathbf{U}$	[17]
NS	$\mathbf{SYS}, \mathbf{MON}$	[64]
PC, EN	$\mathbf{IND}, \mathbf{U}$	[17,22]
PC	$\mathbf{IND}, \mathbf{MON}, \mathbf{U}$	[64]

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 [11]. 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 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 is due to [30], which offered an alternative proof of Arrow's theorem. In judgment aggregation, several proofs resort, directly or indirectly, to this technique (e.g. [32,41] but also [17] itself).

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 Section 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-negatability) are not demanding and yet play a central role in the impossibility results.[20] Thus, relaxing the agenda constraints would imply to restrict judgment aggregation to trivial decision problems.

Therefore, following the structure of [53], escape routes must be found in relaxing the *input conditions*, relaxing the *output conditions*, or relaxing the aggregation constraints, more specifically **IND**. In this section we present some of the approaches that have been investigated in the literature and that guarantee the existence of a judgment aggregation function.

Outline. 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 section builds on material and results taken from [49,32,18,21,44,68].

3.1 Relaxing the Input Constraints

Unidimensional Alignment. In 1948 Duncan Black [3] 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-peakedeness capture 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. [2, pp.75-76]

An example of single-peaked preferences is given in Figure 6. The three candidates (a, b and c) are ordered on the x-axis in such a way that the preferences of the three voters have a peak.

Voters of the Condorcet paradox seen in Section 1.1 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 [3]. This theorem says that if all voters' preferences are single-peaked, the peak of the median voter is the *Condorcet winner*. The median voter in Figure 6 is the voter represented by the dotted line. In an election,

[20] Nehring 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 [63]. See also Table 4.

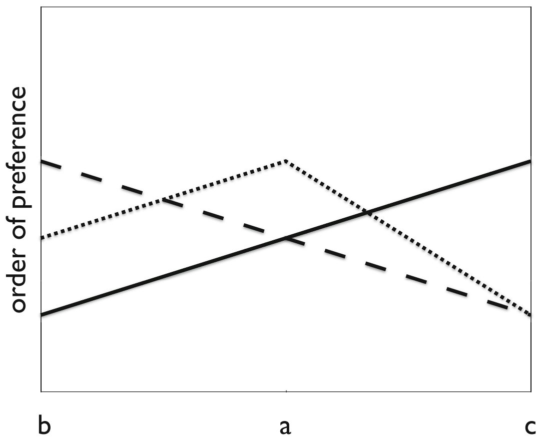

Fig. 6. Single-peaked preferences

Table 5. An example of a unidimensionally aligned profile

	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

the *Condorcet winner* is the candidate that receives the highest number of votes in all pairwise comparison. The Condorcet paradox shows that a Condorcet winner does not always exists.

Researchers explored whether it was possible to transpose the idea of single-peaked preferences to judgment aggregation. 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-peakedeness, List [49,51] 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 is not unidimensionally aligned. Instead, an example of a unidimensionally aligned profile is given in Table 5 [49].

List showed that, under such a 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 have equal weight in the aggregation process.

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 5). Since we assume that individuals are logically consistent, the collective judgment will also be logically 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).

Value-Restriction. The exploration of domain restriction conditions in judgment aggregation continued in [20]. Here, Dietrich and List introduced other sufficient conditions for majority consistency. In particular, they generalized another well-known condition in the theory of preference aggregation: the value-restricted preferences, introduced by Sen in 1966 [71], 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." [31, p. 44][21]

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 $i \in N$.

Value-restriction can be seen as an agreement among individuals that, for every minimal inconsistent set of the agenda, there are two propositions that nobody in the group supports. Clearly, value restriction is sufficient to avoid that an inconsistent judgment set is selected as the group outcome.

Example 7. To illustrate value-restriction, let us consider the agenda $\pm\{p, q, p \vee q\}$. A minimal inconsistent subset is $\{p \vee q, \neg p, \neg q\}$. The value-restriction condition says that (for this minimal inconsistent subset) $\{p \vee q, \neg p\}$ and $\{p \vee q, \neg q\}$ and $\{\neg p, \neg q\}$ should not be in the profile. Thus, for example, the following profile is value-restricted and does not lead to paradoxical outcomes.

[21] As we have seen in Example 1, value restriction is violated in the Condorcet paradox.

	p	q	$p \vee q$
V_1	0	0	0
V_2	0	1	1
V_3	0	1	1
f_{maj}	0	1	1

Dietrich and List also 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 [53], different groups display different levels of pluralism. If there is empirical evidence showing that in a specific group, confronted with a particular decision problem, the conditions above are met, then individual judgments can be safely aggregated into a collective judgment set.

3.2 Relaxing the Output Constraints

We have seen that domain restriction conditions cannot offer a general solution to the problem of judgment aggregation. 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). In those situations 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.

Abstention. One of the first to have proposed to relax the completeness of collective judgment sets was Gärdenfors [32], who 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 to be complete (but are deductively closed[22] and consistent), then every aggregation function that is **IND** and **U** must be oligarchic.

[22] According 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 $\varphi \in J$.

Oligarchic. iff $\exists O \subseteq N, \forall \varphi \in A, \forall P \in \mathbf{P}$: $[\forall i \in O : P_i \models \varphi$ IFF $f(P) \models \varphi]$.
I.e., 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) $O \subseteq N$ adopt position 0 (resp. 1) on that issue.

Clearly, when O contains only one member, oligarchy reduces to dictatorship. At the other end of the spectrum, when $O = N$, we have a unanimity rule, which can thus be seen as the oligarchy of the whole set of individuals N. It is worth noticing that, when $O = N$, the decision procedure is anonymous but only unanimous issues are taken by the group.[23]

Gärdenfors' framework requires the agenda to have a very rich logical structure (with an infinite number of issues). Later, Dokow and Holzman [23] extended Gärdenfors' result and consider finite agendas. Again, the findings are that dictatorial rules are replaced by oligarchic ones.

Quota Rules. In [18] Dietrich and List explored *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 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 considered four rationality conditions on the individual and collective judgment sets (complete, weakly consistent, consistent and deductively closed), and they showed 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 4 listed some impossibility results obtained by weakening **SYS** to **IND**.[24] The more recent impossibility theorems in judgment aggregation assume **IND**.

[23] Note that unanimity rules guarantee deductive closure at the expense of significant incompleteness.

[24] Recall that **IND** is **SYS** without **NEU** (Remark 3).

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 individual 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 (and more specifically in Theorem 2), **IND** is a key condition to ensure that an aggregation function is non manipulable [16,19], 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, [12,58]). 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. [58, 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**.

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. The premise-based procedure has been introduced in [48] under the name of "issue-by-issue voting" and studied in [21,58]. 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.

Table 6. Paretian dilemma. Premises: p = duty, q = negligence, r = causation. Conclusion: $x = (p \wedge q \wedge r)$ = damages.

	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

In Section 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 [6] 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 the probability that majority voting yields the right collective judgement on that proposition increases with the 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 [6,52] 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 [62] 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. [62, p. 1]

Suppose that the judges vote as in Table 6. The Paretian dilemma is disturbing because, if the judges follow the premise-based procedure, they 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. [62, 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 [58] 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.

The Sequential Priority Approach. Another possibility to relax **IND** is the sequential priority approach. Sequential procedures [50,18] proceed in 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 [18] 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 equal to 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.

Table 7. An example of sequential majority rule

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

In order to illustrate how a sequential priority rule works and the problem of path-dependence, we recall here the example used in [18].

Example 8 (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 in Table 7. 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.

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 [68,57] have been originally derived from distance-based merging operators for belief bases introduced in computer science [43,42,44]. 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 sets[25] 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 $f(d(J_1, J^*), \ldots, d(J_n, J^*))$, where

[25] Uniqueness of the collective outcome is not guaranteed.

$J_1, \ldots, J_n$ are the individual judgment sets. There are many possible variations on this definition (in [57] 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 $J_i, J_j \subseteq X$.[26] 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. However, it should be stressed that the Hamming distance is only one among many possible distance functions that we may use [44,45].

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 $\Delta^{d,\sum}$ 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 9 (Distance-based aggregation). Let us consider the doctrinal paradox. The three judgment sets corresponding to the three judges are:

$$J_1 = \{p, q, r\}$$
$$J_2 = \{p, \neg q, \neg r\}$$
$$J_3 = \{\neg p, q, \neg r\}$$

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 $d(., J_1)$, $d(., J_2)$ and $d(., J_3)$ are the Hamming distances of each J_i from the correspondent collective judgment candidate. 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

[26] We 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.)

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 pointed out in [7]. One can of course define other distance minimization rules [57,43,44]. 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 [25] have shown that, for a given profile, minisum and minimax may select two opposite collective outcomes.

We conclude this overview over distance-based aggregation rules with a word on the manipulation issue, thereby also introducing the next section. In general, distance-based merging operators are not strategy-proof [29]. In particular, the family of merging operator using the Hamming distance and where the function f varies, are not strategy-proof (unless we 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 [26] 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 [40]. The fact [74] that Kemeny's rule is the only preference aggregation rule that is neutral, consistent and satisfies the Condorcet property[27], might in particular be adduced as a justification for the use of the belief merging operator $\Delta^{d,\Sigma}$ [27].

4 Manipulability

We address here 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. Section 4.1 studies manipulation as it can be exercised by voters, providing one first characterization of the property of non-manipulability of an aggregation function. Section 4.2 presents an impossibility result connecting non-manipulability to dictatorship in the spirit of Theorem 1. Finally, Section 4.3 concludes by pointing to a strategic dimension in the issue of manipulability relating judgment aggregation to the theory of games [73]. The section builds on material and results which we have mainly taken from [16,19]. Some notions and methods that had been central in Section 2 play an important role here too.

[27] A 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.

4.1 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 Section 2 whose focus will be a property of aggregation functions called, indeed, non-manipulability.

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

$\longmapsto$

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

$\longmapsto$

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

The perspective we have assumed in this example introduces a whole new dimension into judgment aggregation, which has to make with the strategic behavior of voters. While strategic behavior is the realm of the theory of games [73]—and we will briefly touch upon it in Section 4.3—in the next sections we will assume a typically social-choice theory perspective: the example has shown that premise-based aggregation is manipulable. The question is then: do non-trivial non-manipulable aggregation rules exist?

Manipulability: Definition and a Characterization. Recall first the terminology we have introduced in Definition 8 in Section 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 judgment aggregation structure. An aggregation function f is:*

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

*A function is said to be non-manipulable (non-**MAN**) otherwise.*

In words, an aggregation function is manipulable whenever there exists some profile where, for some agent i, the aggregation yields for some formula φ an outcome which is different from i's opinion over φ, and which can be modified if i were to input a different judgment set in the aggregation function.

It is worth observing that, literally, this condition 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 the conditions of independence and monotonicity introduced in Definition 9.

Theorem 2 ([19]). *Let $\mathcal{J} = \langle N, A \rangle$ be a judgment aggregation structure and f an aggregation function. The following assertions are equivalent:*

i) *f does not satisify* **MAN***,*
ii) *f satisfies* **IND** *and* **MON***.*

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

[(i) implies (ii)] Assume non-**MAN**. 1) We prove that **MON** 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 $P_i \neq_\varphi f(P)$ and by non-**MAN** $f(P) =_\varphi f(P')$. 2) We prove that **IND** 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 **IND**). WLOG suppose $f(P) \models \varphi$ and assume towards a contradiction that $f(P) \not\models \varphi$. It follows that there exists a profile, namely P', such that $P' \models \varphi$ but $f(P') \not\models \varphi$ and that there exits a profile, namely P, such that $P_i =_\varphi P'_i$ and $f(P) \models \varphi$, thereby implying **MAN**, against the assumption.

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. All in all, Theorem 2 establishes a strict link between independence and non-manipulability.

Agenda Manipulation. Before closing this section, it is worth pointing at another form of manipulation, concerning the way agendas are set up. 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 10 (Agenda manipulation). Consider again Example 2 and assume the same three individual judgment sets over agenda $A = \pm\{p, p \to q, q\}$. 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 $p \to q$ in A_1 and obtain an agenda $A_2 = \pm\{p, p \to q\}$. The collective judgment would so, under propositionwise majority, become $\{p, p \to q\}$ and entail q.

	p	q
J_1	1	1
J_2	1	0
J_3	0	0
J	1	0

$\longmapsto$

	p	$p \to q$
J_1	1	1
J_2	1	0
J_3	0	1
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**). By this condition, it is possible to modify a collective judgment on a formula only by removing that same formula from the agenda.

4.2 The Impossibility of Non-manipulability

This section can be seen on a par with Section 2.2. 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.

Preliminaries. We will prove the theorem by resorting to the same technique we used in Section 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—and with different 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.[28] The proof in its full generality can be found in [19].

Lemmata

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

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

Proof. Assume there 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 2, 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.

Contagion and Characterization by Winning Coalitions. Recall first Definition 10 which has been introduced in Section 2.

Lemma 6 (Contagion lemma). *Let $\mathcal{J} = \langle N, A \rangle$ be a judgment aggregation structure where A satisfies PC and f be an aggregation function that satisfies non-***MAN** *and* **RES***:*

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

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 $\varphi = \varphi_1 \models_c \ldots \models_c \varphi_k = \psi$ for some $\varphi_1, \ldots, \varphi_k \in A$. 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 **U** (Lemma 5) we have that $X \subseteq f(P)$. 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 $P'_{\varphi_{j+1}} = C$. For, suppose not, i.e., $C \notin \mathcal{W}_{\varphi_{j+1}}$. By **IND** (Theorem 2), it follows that for any P', if $P'_{\varphi_{j+1}} = C$ then $f(P') \not\models \varphi_{j+1}$. Contradiction.

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

[28] The theorem for this agenda will hence be a variant of the Gibbard-Sattherthwaite theorem (cf. [31, Ch. 5]).

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 7, 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 judgment aggregation structure where A satisfies PC and f be an aggregation function that does not satisfy* **MAN**. *For all $P \in \mathbf{P}$ and $\varphi \in A$:*

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

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 2 we have that if $P_\varphi = P'_\varphi$ then $f(P') \models \varphi$. Hence $P_\varphi \in \mathcal{W}$ according to the above definition.

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 judgment aggregation 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 $Y_3 = \{c \succ a\}$:

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

where, recall, $\neg(x \succ y) = y \succ x$.

Ultrafilters of Winning Coalitions.

Lemma 9 (Ultrafilter lemma). *Let $\mathcal{J} = \langle N, A \rangle$ be a judgment aggregation 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 $-C \notin \mathcal{W}$;*
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) [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 $\neg\varphi$. By **IND** (Theorem 2) this profile would yield an inconsistent collective judgment, which is impossible by Definition 3. [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. $-C = \{i \mid P_i \not\models \psi\}$ and $f(P) \not\models \varphi$. By **IND** (Theorem 2) it follows that $\forall \varphi \in A, \forall P$ s.t. $-C = \{i \mid P_i \models \neg\varphi\}$ and $f(P) \models \neg\varphi$, hence $-C \in \mathcal{W}$.[29]

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 $-C_2$ 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 $c \succ a$. Contradiction.

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.

The Theorem

Theorem 3 ([19]). *Let $\mathcal{J} = \langle N, A \rangle$ be a judgment aggregation 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 $\varphi \in A$:

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

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

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

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

[29] Notice 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**.

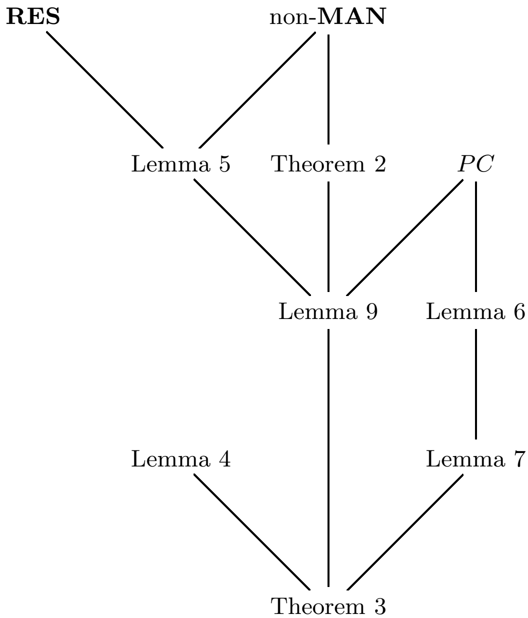

Fig. 7. Structure of the proof of Theorem 3

.

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 7 recapitulates the structure of the proof.

4.3 Strategic Issues in Judgment Aggregation

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 judgment aggregation.

Preferences over Judgment Sets. Talking about incentives of voters means talking about their preferences, i.e., total preorders[30] over the set of all possible judgment sets **J**. Preferences are, however, not included in the representation of aggregation problems as provided by judgment aggregation structures (Definition 1). So two possibilities arise. Either judgment aggregation structures

[30] We recall that a total preorder is a binary relation which is reflexive, transitive and total.

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 $g : \mathbf{J} \longrightarrow \mathbf{J} \times \mathbf{J}$ such a function. The property of strategy-proofness of an aggregation function can be defined as follows:

Definition 12 (Strategy-proofness). *Let $\mathcal{J} = \langle N, A \rangle$ be a judgment aggregation 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.

Strategy-proofness and Manipulability. It turns out that, despite introducing this strategic dimension, strategy-proofness is equivalent to the property of non-manipulability under some rather reasonable assumptions.

Definition 13 (Closeness). *Let $\mathcal{J} = \langle N, A \rangle$ be a judgment aggregation 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 [19].

Theorem 4. *Let $\mathcal{J} = \langle N, A \rangle$ be a judgment aggregation 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.

5 The Research Agenda of Judgment Aggregation

In this section we briefly sketch a somewhat loose list of topics of on-going research in the field of judgment aggregation. 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. 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 sections we have also seen that judgment aggregation commonly studies whether there exist aggregation functions that satisfy a 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* [28], 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 section with an overview on the recent interest of researchers working in abstract argumentation for judgment aggregation problems. This section builds on material taken mainly from [35,36,33,34,28,9]

5.1 Judgment Aggregation vs. Preference Aggregation

Since the early days of judgment aggregation, a natural question that presented itself to judgment aggregation researchers was whether judgment aggregation could be showed to subsume preference aggregation. The question was answered positively by Dietrich and List in [17], which obtained Arrow's theorem—preference aggregation'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.[31]

So Arrow's theorem can be viewed as an instance of a more general judgment aggregation impossibility. But can we go the other way around too? That is, can we view judgment aggregation impossibility results as instances of preference aggregation impossibilities? This question has been partially investigated by Grossi in [35,36] and has been given a first positive answer. That work provides a number of results at the interface of judgment aggregation, preference aggregation and many-valued logics (see, for instance, [38]) 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 [15]; ii) numerical functions can ground logical semantics, like it happens in many-valued logic

[31] It is worth mentioning, in passing, another interesting generalization of Arrow's theorem via lattice theory provided in [14].

[38] where, like in propositional logic, the semantic clause $u(x) \leq u(y)$ typically defines the satisfaction by u of the implication $x \rightarrow y$:

$$u \models x \rightarrow y \text{ iff } u(x) \leq u(y). \tag{8}$$

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 judgment aggregation on many-valued logics. In turn, judgment aggregation can be studied as an instance of a type of preference aggregation defined on dichotomous preferences, thus enriching the picture of the logical relationship between preference aggregation and judgment aggregation.

5.2 Aggregation under Integrity Constraints

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

$$\mathbf{V}^{|N|} \longrightarrow \mathbf{V} \tag{9}$$

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 Sections 2 and 4, recent work presented by Grandi and Endriss in [33,34] 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 [33,34] elaborates upon this latter interpretation of impossibility, by asking what sort of 'rationality' each axiom is then able to preserve. 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 these notes we have worked with constraints formalizable in propositional logic. A typical example of the questions addressed in [33,34] 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 [28] 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 [28] provides a first set of answers to this questions.

5.4 Judgment Aggregation and Abstract Argumentation

The last recent research theme we mention here considers judgment aggregation as an aggregation of individual evaluations of a given argumentation framework. An *argumentation framework* is defined by a set of arguments and a defeat relation among them. 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 [24].

For example, in the argumentation framework in Fig. 8, we have that argument A attacks argument B, and that B attacks A.

There are three possible extensions for this argumentation framework, namely those pictured in Fig 9. 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.

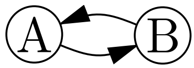

Fig. 8. An argumentation framework

The general idea is that, given an argumentation framework, individuals may provide different evaluations regarding what should be accepted and rejected. The question is then: how can we obtain a collective evaluation from individual ones? 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.

Among the first who applied abstract argumentation to judgment aggregation problems were Caminada and Pigozzi [9]. The reason for using abstract argumentation is twofold: on the one hand, the existence of different argumentation semantics allows to be flexible when defining which social outcomes are permissible. On the other hand, it allows to bring judgment aggregation from classical logic to nonmonotonic reasoning.

To mention is also the work by Rahwan and Tohmé [70] 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 [69] 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 [9] is not only to guarantee a consistent group outcome, but also that such outcome is *compatible* with the individual judgments. Caminada and Pigozzi stress that 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* [8]). Such situation may be not a desirable collective outcome as it may conflict with some of its members' judgments. The research question tackled in [9] was precisely this: when is a group outcome 'compatible' with its members' judgments? By "compatible" it is meant a group decision making in which any group member is able to defend the group decision without having to argue against his own private opinions.

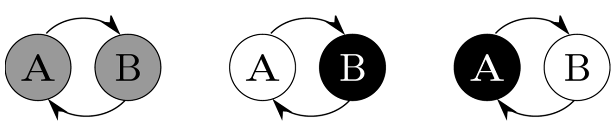

Fig. 9. The three possible extensions for the argumentation framework of Fig. 8

Three operators (sceptical, credulous and super credulous) are defined and investigated in [9], 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. 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 are put together, which is the same effect we saw in the Paretian dilemma in Section 3.2.

It is also to be mentioned that the so far limited literature on abstract argumentation and judgment aggregation has 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 [10] 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 [9] 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. [10] studies whether the compatible social outcomes selected by the aggregation operators in [9]are Pareto optimal. The results show that 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*.[32]

Acknowledgments. We are very grateful to Umberto Grandi, Ulle Endriss, Franz Dietrich and the students of ESSLLI'11 for the many useful comments that helped us improving the present version of these notes. Davide Grossi has been partly supported by the *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* under the NWO VENI grant 639.021.816.

[32] A similar idea was introduced and studied earlier in [37], where manipulation was seen as a coordinated action of the whole group.

References

1. Arrow, K.: A difficulty in the concept of social welfare. Journal of Political Economy 58(4), 328–346 (1950)
2. Arrow, K.: Social Choice and Individual Values, 2nd edn. John Wiley, New York (1963)
3. Black, D.: On the rationale of group decision making. The Journal of Political Economy 56, 23–34 (1948)
4. McLean, I.S., McMillan, A., Monroe, B.L. (eds.): The Theory of Committees and Elections by Duncan Black, and Committee Decisions with Complementary Valuation by Duncan Black and R.A. Newing. Springer (1998)
5. de Borda, J.-C.: Mémoire sur les élections au scrutin par M. de Borda. In: Mémoires de l'Académie Royale des Sciences année 1781, pp. 657–665. l'Imprimerie Royale, Paris (1784)
6. Bovens, L., Rabinowicz, W.: Democratic answers to complex questions. An epistemic perspective. Synthese 150, 131–153 (2006)
7. Brams, S., Kilgour, D., Sanver, R.: A minimax procedure for electing committees. Public Choice 132(3-4), 401–420 (2007)
8. Brams, S., Kilgour, D., Zwicker, W.: The paradox of multiple elections. Social Choice and Welfare 15(2), 211–236 (1998)
9. Caminada, M., Pigozzi, G.: On judgment aggregation in abstract argumentation. Autonomous Agents and Multi-Agent Systems 22(1), 64–102 (2011)
10. Caminada, M., Pigozzi, G., Podlaszewski, M.: Manipulation in group argument evaluation. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011), pp. 121–126 (2011)
11. Cartan, H.: Filtres et ultrafiltres. Comptes Rendus de l'Académie des Sciences, 777–779 (1937)
12. Chapman, B.: Rational aggregation. Politics, Philosophy & Economics 1(3), 337–354 (2002)
13. de Condorcet, M.J.A.N., de Marquis, C.: Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralità des Voix. l'Imprimerie Royale, Paris (1785)
14. Daniëls, T., Pacuit, E.: A general approach to aggregation problems. Journal of Logic and Computation 19(3), 517–536 (2008)
15. Debreu, G.: Representation of a preference ordering by a numerical function. In: Thrall, R.M., Coombs, C.H., Davis, R.L. (eds.) Decision Processes. John Wiley (1954)
16. Dietrich, F.: Judgment aggregation (im)possibility theorems. Economic Theory 126, 286–298 (2006)
17. Dietrich, F., List, C.: Arrow's theorem in judgment aggregation. Social Choice and Welfare 29(19-33) (2007)
18. Dietrich, F., List, C.: Judgment aggregation by quota rules: Majority voting generalized. Journal of Theoretical Politics 19, 391–424 (2007)
19. Dietrich, F., List, C.: Strategy-proof judgment aggregation. Economics and Philosophy 23, 269–300 (2007)
20. Dietrich, F., List, C.: Majority voting on restricted domains. Journal of Economic Theory 145(2), 512–543 (2010)
21. Dietrich, F., Mongin, P.: The premiss-based approach to judgment aggregation. Journal of Economic Theory 145(2), 562–582 (2010)

22. Dokow, E., Holzman, R.: Aggregation of binary evaluations. Journal of Economic Theory 145(2), 495–511 (2010)
23. Dokow, E., Holzman, R.: Aggregation of binary evaluations with abstentions. Journal of Economic Theory 145(2), 544–561 (2010)
24. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence 77, 321–357 (1995)
25. Eckert, D., Klamler, C.: How puzzling is judgment aggregation? Antipodality in distance-based aggregation rules. Working Paper, pp. 1–7. University of Graz (2007)
26. Eckert, D., Mitlöhner, J.: Logical representation and merging of preference information. In: Multidisciplinary IJCAI 2005 Workshop on Advances in Preference Handling, pp. 85–87 (2005)
27. Eckert, D., Pigozzi, G.: Belief merging, judgment aggregation, and some links with social choice theory. In: Dagstuhl Seminar Proceedings 05321 (2005)
28. Endriss, U., Grandi, U., Porello, D.: Complexity of judgment aggregation: Safety of the agenda. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010), pp. 359–366 (2010)
29. Everaere, P., Konieczny, S., Marquis, P.: The strategy-proofness landscape of merging. Journal of Artificial Intelligence Research 28, 49–105 (2007)
30. Fishburn, P.C.: Arrow's impossibility theorem: Concise proof and infinite voters. Journal of Economic Theory 2(1), 103–106 (1970)
31. Gaertner, W.: A Primer in Social Choice Theory. Oxford University Press (2006)
32. Gärdenfors, P.: A representation theorem for voting with logical consequences. Economics and Philosophy 22, 181–190 (2006)
33. Grandi, U., Endriss, U.: Lifting rationality assumptions in binary aggregation. In: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI 2010), pp. 780–785 (2010)
34. Grandi, U., Endriss, U.: Binary aggregation with integrity constraints. In: Proceedings of IJCAI 2011, pp. 204–209 (2011)
35. Grossi, D.: Unifying preference and judgment aggregation. In: Decker, P., Sichman, J. (eds.) Proceedings of the Eighth International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2009), pp. 217–224. ACM Press (2009)
36. Grossi, D.: Correspondences in the Theory of Aggregation. In: Bonanno, G., Löwe, B., van der Hoek, W. (eds.) LOFT 2010. LNCS(LNAI), vol. 6006, pp. 34–60. Springer, Heidelberg (2010)
37. Grossi, D., Pigozzi, G., Slavkovik, M.: White manipulation in judgment aggregation. In: Proceedings of BNAIC 2009 - The 21st Benelux Conference on Artificial Intelligence (2009)
38. Hähnle, R.: Advanced many-valued logics. In: Gabbay, D.M., Guenthner, F. (eds.) Handbook of Philosophical Logic, 2nd edn., vol. 2, pp. 297–395. Kluwer (2001)
39. Hartmann, S., Pigozzi, G., Sprenger, J.: Reliable methods of judgement aggregation. Journal of Logic and Computation 20(2), 603–617 (2010)
40. Kemeny, J.: Mathematics without numbers. Daedalus 88, 577–591 (1959)
41. Klamler, C., Eckert, D.: A simple ultrafilter proof for an impossibility theorem in judgment aggregation. Economics Bulletin 29(1), 319–327 (2009)
42. Konieczny, S., Grégoire, E.: Logic-based approaches to information fusion. Information Fusion 7, 4–18 (2006)
43. Konieczny, S., Lang, J., Marquis, P.: DA^2 merging operators. Artificial Intelligence 157, 49–79 (2004)

44. Konieczny, S., Pino Pérez, R.: Merging with Integrity Constraints. In: Hunter, A., Parsons, S. (eds.) ECSQARU 1999. LNCS (LNAI), vol. 1638, pp. 233–244. Springer, Heidelberg (1999)
45. Konieczny, S., Pino Pérez, R.: On the frontier between arbitration and majority. In: Proceedings of the Eighth International Conference on Principles of Knowledge Representation and Reasoning, pp. 109–118 (2002)
46. Kornhauser, L.A.: Modeling collegial courts. II. Legal doctrine. Journal of Law, Economics, and Organization 8, 441 (1992)
47. Kornhauser, L.A., Sager, L.G.: Unpacking the court. Yale Law Journal 96, 82–117 (1986)
48. Kornhauser, L.A., Sager, L.G.: The one and the many: Adjudication in collegial courts. California Law Review 81, 1–51 (1993)
49. List, C.: A possibility theorem on aggregation over multiple interconnected propositions. Mathematical Social Sciences 45(1), 1–13 (2002)
50. List, C.: A model of path-dependence in decisions over multiple propositions. American Political Science Review 98(3), 495–513 (2004)
51. List, C.: Corrigendum to A possibility theorem on aggregation over multiple interconnected propositions. Mathematical Social Sciences 52, 109–110 (2005)
52. List, C.: The probability of inconsistencies in complex collective decisions. Social Choice and Welfare 24(1), 3–32 (2005)
53. List, C.: The theory of judgment aggregation: An introductory review. Synthese (2011)
54. List, C., Pettit, P.: Aggregating sets of judgments: An impossibility result. Economics and Philosophy 18, 89–110 (2002)
55. List, C., Pettit, P.: Aggregating sets of judgments: Two impossibility results compared. Synthese 140(1), 207–235 (2004)
56. List, C., Puppe, C.: Judgment aggregation: A survey. In: Anand, P., Puppe, C., Pattanaik, P. (eds.) The Handbook of Rational and Social Choice. Oxford University Press (2009)
57. Miller, M.K., Osherson, D.: Methods for distance-based judgment aggregation. Social Choice and Welfare 32(4), 575–601 (2009)
58. Mongin, P.: Factoring out the impossibility of logical aggregation. Journal of Economic Theory 141(1), 100–113 (2008)
59. Mongin, P.: Judgment aggregation. In: Hansson, S.O., Hendricks, V.F. (eds.) The Handbook of Formal Philosophy. Springer (2011)
60. Mongin, P., Dietrich, F.: Un bilan interprétatif de la théorie de l'agrégation logique. Revue d'économie politique 120(6), 929–972 (2010)
61. Nanson, E.J.: Methods of election. Transactions and Proceedings of the Royal Society of Victoria 19, 197–240 (1882)
62. Nehring, K.: The (im)possibility of a Paretian rational. Economics working papers. Institute for Advanced Study, School of Social Science (November 2005)
63. Nehring, K., Puppe, C.: Strategy-proof social choice on single-peaked domains: Possibility, impossibility and the space between. Working Paper, University of California at Davis
64. Nehring, K., Puppe, C.: Abstract arrovian aggregation. Journal of Economic Theory 145(2), 467–494 (2010)
65. Odifreddi, P.: Ultrafilters, dictators and gods. In: Calude, C., Păun, G. (eds.) Finite Versus Infinite. Contributions to an Eternal Dilemma, pp. 239–246. Springer (2000)
66. Pauly, M., van Hees, M.: Logical constraints on judgment aggregation. Journal of Philosophical Logic 35, 569–585 (2006)

67. Pettit, P.: Deliberative democracy and the discursive dilemma. Philosophical Issues 11, 268–299 (2001)
68. Pigozzi, G.: Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. Synthese 152, 285–298 (2006)
69. Rahwan, I., Larson, K.: Welfare properties of argumentation-based semantics. In: Proceedings of the 2nd International Workshop on Computational Social Choice, COMSOC (2008)
70. Rahwan, I., Tohmé, F.: Collective argument evaluation as judgment aggregation. In: Proc. of 9th AAMAS (2010)
71. Sen, A.: A possibility theorem on majority decisions. Econometrica 34, 491–499 (1966)
72. Sen, A.K.: Collective Choice and Social Welfare. Holden Day (1970)
73. von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press (1944)
74. Young, H.P., Levenglick, A.: A consistent extension of condorcet's election principle. SIAM Journal of Applied Mathematics 35, 285–300 (1978)