Auton Agent Multi-Agent Syst (2011) 22:4–30
DOI 10.1007/s10458-009-9115-8

On the logic of preference and judgment aggregation

Thomas Ågotnes · Wiebe van der Hoek · Michael Wooldridge

Received: 15 February 2009 / Accepted: 1 October 2009 / Published online: 28 October 2009

Abstract Agents that must reach agreements with other agents need to reason about how their preferences, judgments, and beliefs might be aggregated with those of others by the social choice mechanisms that govern their interactions. The emerging field of *judgment aggregation* studies aggregation from a *logical* perspective, and considers how multiple sets of logical formulae can be aggregated to a single consistent set. As a special case, judgment aggregation can be seen to subsume classical preference aggregation. We present a modal logic that is intended to support reasoning about judgment aggregation scenarios (and hence, as a special case, about preference aggregation): the logical language is interpreted directly in judgment aggregation rules. We present a sound and complete axiomatisation. We show that the logic can express aggregation rules such as majority voting; rule properties such as independence; and results such as the discursive paradox, Arrow's theorem and Condorcet's paradox—which are derivable as formal theorems of the logic. The logic is parameterised in such a way that it can be used as a general framework for comparing the logical properties of different types of aggregation—including classical preference aggregation. As a case study we present a logical study of, including a formal proof of, the *neutrality lemma*, the main ingredient in a well-known proof of Arrow's theorem.

Keywords Judgment aggregation · Preference aggregation · Modal logic · Complexity · Completeness

T. Ågotnes (✉)
Department of Information Science and Media Studies, University of Bergen, Bergen, Norway
e-mail: thomas.agotnes@infomedia.uib.no

W. van der Hoek · M. Wooldridge
Department of Computer Science, University of Liverpool, Liverpool, UK
e-mail: wiebe@liv.ac.uk

M. Wooldridge
e-mail: mjw@liv.ac.uk

1 Introduction

In this paper, we are interested in knowledge representation formalisms for systems in which agents need to *aggregate* their preferences, judgments, beliefs, etc. For example, an agent may need to reason about majority voting in a group he is a member of. *Preference aggregation*—combining individuals' preference relations over some set of alternatives into a preference relation which represents the joint preferences of the group by so-called *social welfare functions*—has been extensively studied in social choice theory [4]. The emerging field of *judgment aggregation* studies aggregation from a logical perspective, and investigates how, given a consistent set of logical formulae for each agent, representing the agent's beliefs or judgments, we can aggregate these to a single consistent set of formulae. A variety of *judgment aggregation rules* have been developed to this end. As a special case, judgment aggregation can be seen to subsume preference aggregation [7].

In this paper we present a formal logic, called *Judgment Aggregation Logic* (JAL), for reasoning about judgment aggregation. The formulae of the logic are interpreted as statements about judgment aggregation rules, and we give a sound and complete axiomatisation of a large class of such rules. The axiomatisation is parameterised in such a way that we can instantiate it to get a range of different judgment aggregation logics. For example, one instance is an axiomatisation, in our language, of all social welfare functions—thus we get a logic of classical preference aggregation as well. And this is one of the main contributions of this paper: we identify the logical properties of judgment aggregation, and we can compare the logical properties of different classes of judgment aggregation—and of general judgment aggregation and preference aggregation in particular.

Of course, a logic is only interesting as long as it is *expressive*. One of the goals of this paper is to investigate the representational and logical capabilities an agent needs for judgment and preference aggregation; that is, what kind of logical language might be used to represent and reason about judgment aggregation? An agent's knowledge representation language should be able to express: common aggregation rules such as majority voting; commonly discussed properties of judgment aggregation rules and social welfare functions such as independence; paradoxes commonly used to illustrate judgment aggregation and preference aggregation, viz. the discursive paradox and Condorcet's paradox respectively; and other important properties such as Arrow's theorem. In order to illustrate in more detail what such a language would need to be able to express, take the example of a potential property of social welfare functions (SWFs) called independence of irrelevant alternatives (IIA): given two preference profiles (each consisting of one preference relation for each agent) and two alternatives, if for each agent the two alternatives have the same order in the two preference profiles, then the two alternatives must have the same order in the two preference relations resulting from applying the SWF to the two preference profiles, respectively. From this example it seems that a formal language for SWFs should be able to express:

- Quantification on several levels: over alternatives; over preference profiles, i.e., over relations over alternatives (second-order quantification); and over agents.
- Properties of preference relations for different agents, and properties of several different preference relations for the same agent in the same formula.
- Comparison of different preference relations.
- The preference relation resulting from applying a SWF to other preference relations.

Given these requirements, it might seem that such a language would be rather complex (in particular, these requirements seem to rule out a standard propositional modal logic). Perhaps surprisingly, the language of JAL is syntactically and semantically rather simple; and yet the

language is, nevertheless, expressive enough to give elegant and succinct expressions of, e.g., IIA, majority voting, the discursive dilemma, Condorcet's paradox and Arrow's theorem. This means, for example, that Arrow's theorem is a formal theorem of JAL, i.e., a derivable formula; we thus have a formal proof theory for social choice.

The structure of the rest of the paper is as follows. In the next section we review the basics of judgment aggregation as well as preference aggregation, and mention some commonly discussed properties of judgment aggregation rules and social welfare functions. In Sect. 3 we introduce the syntax and semantics of JAL. Formulae of JAL are interpreted directly by, and thus represent properties of, judgment aggregation rules. In Sect. 4 we demonstrate that the logic can express commonly discussed properties of judgment aggregation rules, such as the discursive paradox. In Sect. 5 we study the complexity of the model checking problem, as well as a normal form of the formulae of the logic. We give a sound and complete axiomatisation of the logic in Sect. 6, under the assumption that the agenda the agents make judgments over is finite. As mentioned above, preference aggregation can be seen as a special case of judgment aggregation, and in Sect. 7 we introduce an alternative interpretation of JAL formulae directly in social welfare functions, and we show that Condorcet's paradox and Arrow's theorem can be expressed as formulae which are valid in our logic. We obtain a sound and complete axiomatisation of the logic for preference aggregation as well. In Sect. 8 we go a little deeper into preference aggregation, and look at the *neutrality lemma* as a logical case study. The neutrality lemma is the main ingredient in Geanakoplos' proof of Arrow's theorem [11], and we present a formal proof of the lemma in JAL. Section 9 discusses related work and concludes.

2 Judgment and preference aggregation

Judgment aggregation is concerned with judgment aggregation rules aggregating sets of logical formulae; preference aggregation is concerned with social welfare functions aggregating preferences over some set of alternatives. Let n be a number of *agents*; we write N for the set $\{1, \ldots, n\}$.

2.1 Judgment aggregation rules

Let $\mathbf{L}$ be a logic with language $\mathcal{L}(\mathbf{L})$. We require that the language has negation and material implication, with the usual semantics. We will sometimes refer to $\mathbf{L}$ as "the underlying logic". An *agenda* over $\mathbf{L}$ is a non-empty set $\mathcal{A} \subseteq \mathcal{L}(\mathbf{L})$, where for every formula ϕ that does not have the form of a negation, $\phi \in \mathcal{A}$ iff $\neg\phi \in \mathcal{A}$. We sometimes call a member of $\mathcal{A}$ an *agenda item*. A subset $A' \subseteq \mathcal{A}$ is *consistent* unless A' entails both $\neg\phi$ and ϕ in $\mathbf{L}$ for some $\phi \in \mathcal{L}(\mathbf{L})$; A' is *complete* if either $\phi \in A'$ or $\neg\phi \in A'$ for every $\phi \in \mathcal{A}$ which does not have the form of a negation. An (admissible) *individual judgment set* is a complete and consistent subset $A_i \subseteq \mathcal{A}$ of the agenda. The idea here is that a judgment set A_i represents the choices from $\mathcal{A}$ made by agent i. Two rationality criteria demand that an agents' choices at least be internally consistent, and that each agent makes a decision between every item and its negation. An (admissible) *judgment profile* is an n-tuple $\langle A_1, \ldots, A_n \rangle$, where A_i is the individual judgment set of agent i. $J(\mathcal{A}, \mathbf{L})$ denotes the set of all individual (complete and $\mathbf{L}$-consistent) judgment sets over $\mathcal{A}$, and $J(\mathcal{A}, \mathbf{L})^n$ the set of all judgment profiles over $\mathcal{A}$. When $\gamma \in J(\mathcal{A}, \mathbf{L})^n$, we use γ_i to denote the ith element of γ, i.e., agent i's individual judgment set in judgment profile γ.

A *judgment aggregation rule (JAR)* is a function f that maps each judgment profile $\langle A_1, \ldots, A_n \rangle$ to a complete and consistent *collective judgment set* $f(A_1, \ldots, A_n) \in J(\mathcal{A}, \mathbf{L})$. Such a rule is thus a recipe to enforce a rational group decision, given a tuple of rational choices by the individual agents. Of course, such a rule should to a certain extent be 'fair'. Some possible properties of a judgment aggregation rule f over an agenda $\mathcal{A}$:

Non-dictatorship (**ND1**): There is no agent i such that for every judgment profile $\langle A_1, \ldots, A_n \rangle$, we have $f(A_1, \ldots, A_n) = A_i$

Independence (**IND**): For any $p \in \mathcal{A}$ and judgment profiles $\langle A_1, \ldots, A_n \rangle$ and $\langle B_1, \ldots, B_n \rangle$, if for all agents i ($p \in A_i$ iff $p \in B_i$), then $p \in f(A_1, \ldots, A_n)$ iff $p \in f(B_1, \ldots, B_n)$

Unanimity (**UNA**): For any judgment profile $\langle A_1, \ldots, A_n \rangle$ and any $p \in \mathcal{A}$, if $p \in A_i$ for all agents i, then $p \in f(A_1, \ldots, A_n)$

2.2 Social welfare functions

Social welfare functions (SWFs) are usually defined in terms of ordinal preference structures, rather than cardinal structures such as utility functions. An SWF takes a preference relation, a binary relation over some set of alternatives, for each agent, and outputs another preference relation representing the aggregated preferences.

The best known result about SWFs is Arrow's theorem [3]. Many variants of the theorem appear in the literature, differing in assumptions about the preference relations. In this paper, we make the assumption that all preference relations are linear orders, i.e., that neither agents nor the aggregated preference can be indifferent between distinct alternatives. This gives one of the simplest formulations of Arrow's theorem (Theorem 1 below). See, for example, [4] for a discussion and more general formulations.

Formally, let K be a set of *alternatives*. We henceforth implicitly assume that there are at least two alternatives. A *preference relation* (over K) is, here, a total (linear) order on K, i.e., a relation R over K which is antisymmetric (i.e., $(a, b) \in R$ and $(b, a) \in R$ implies that $a = b$), transitive (i.e., $(a, b) \in R$ and $(b, c) \in R$ imply that $(a, c) \in R$), and total (i.e., either $(a, b) \in R$ or $(b, a) \in R$). We sometimes use the infix notation aRb for $(a, b) \in R$. The set of preference relations over alternatives K is denoted $L(K)$. Alternatively, we can view $L(K)$ as the set of all permutations of K. Thus, we shall sometimes use a permutation of K to denote a member of $L(K)$. For example, when $K = \{a, b, c\}$, we will sometimes use the expression acb to denote the relation $\{(a, c), (a, b), (c, b), (a, a), (b, b), (c, c)\}$. aRb means that b is preferred over a if a and b are different. R^s denotes the *irreflexive*, or, in our case of linear orders, *strict* version of R, i.e., $R^s = R \setminus \{(a, a) : a \in K\}$. aR^sb means that b is strictly preferred over a, i.e., b is preferred over a and $a \neq b$.

A *preference profile* for N over alternatives K is a tuple $(R_1, \ldots, R_n) \in L(K)^n$, consisting of one preference relation R_i for each agent i. A *social welfare function* (SWF) is a function

$$F : L(K)^n \to L(K)$$

mapping each preference profile to an aggregated preference relation. The class of all SWFs over alternatives K is denoted $\mathcal{F}(K)$.

Properties of SWFs F corresponding to the judgment aggregation rule properties discussed in Sect. 2.1 are:

Non-dictatorship (**ND2**) (corresponds to **ND1**): There is no agent i such that for all preference profiles $(R_1, \ldots, R_n)$, we have $F(R_1, \ldots, R_n) = R_i$

Independence of irrelevant alternatives (**IIA**) (corresponds to **IND**): For all alternatives a and b and preference profiles $(R_1, \ldots, R_n)$ and $(S_1, \ldots, S_n)$, if for all agents i (aR_ib iff aS_ib), then ($aF(R_1, \ldots, R_n)b$ iff $aF(S_1, \ldots, S_n)b$)

Pareto Optimality (**PO**) (corresponds to **UNA**): For any preference profile $(R_1, \ldots, R_n)$ and any alternatives a and b, if aR_i^sb for all agents i, then $aF(R_1, \ldots, R_n)^sb$

Arrow's theorem says that the three properties above are inconsistent if there are more than two alternatives.

Theorem 1 (*Arrow*) *If there are more than two alternatives, no SWF has all the properties **PO**, **ND2** and **IIA**.*[1]

3 Judgment aggregation logic: Syntax and semantics

The language of *Judgment Aggregation Logic* (JAL) is parameterised by a set of agents $N = \{1, 2, \ldots, n\}$ (we will assume that there are at least two agents) and an agenda $\mathcal{A}$. The following atomic propositions are used, where σ is a constant denoting that the current agenda item is chosen using the current aggregation rule (see below):

$$\Pi = N \cup \{\mathbf{h}_p \mid p \in \mathcal{A}\} \cup \{\sigma\}$$

The language $\mathcal{L}(N, \mathcal{A})$ of JAL is defined by the following grammar:

$$\phi ::= \alpha \mid \Box\phi \mid \blacksquare\phi \mid \phi \wedge \phi \mid \neg\phi$$

where $\alpha \in \Pi$. This language will be formally interpreted in triplets consisting of an agenda item p, a judgment profile γ and a judgment aggregation function f; informally, the proposition i (where $i \in N$) means that the current agenda item p is in agent i's judgment set in the current judgment profile; σ means that the current agenda item is in the aggregated judgment set $f(\gamma)$ of the current judgment profile γ; $\mathbf{h}_p$ means that the current agenda item is p; $\Box\phi$ means that ϕ is true in every judgment profile; $\blacksquare\phi$ means that ϕ is true in every agenda item.

We define $\Diamond\psi = \neg\Box\neg\psi$, intuitively meaning "$\psi$ is true for some judgment profile", and $\blacklozenge\psi = \neg\blacksquare\neg\psi$, intuitively meaning "$\psi$ is true for some agenda item", as usual, in addition to the usual derived propositional connectives.

We now define the formal semantics of $\mathcal{L}(N, \mathcal{A})$. A *model* wrt. $\mathcal{L}(N, \mathcal{A})$ and underlying logic $\mathbf{L}$ is a judgment aggregation rule f over $\mathcal{A}$. Recall that $J(\mathcal{A}, \mathbf{L})^n$ denotes the set of complete and $\mathbf{L}$-consistent judgment profiles over $\mathcal{A}$. A *table* is a tuple $T = \langle f, \gamma, p\rangle$ such that f is a model, $\gamma \in J(\mathcal{A}, \mathbf{L})^n$ and $p \in \mathcal{A}$. A formula is interpreted on a table as follows.

$$\begin{array}{ll}
f, \gamma, p \models_{\mathbf{L}} \mathbf{h}_q & \Leftrightarrow p = q \\
f, \gamma, p \models_{\mathbf{L}} i & \Leftrightarrow p \in \gamma_i \\
f, \gamma, p \models_{\mathbf{L}} \sigma & \Leftrightarrow p \in f(\gamma) \\
f, \gamma, p \models_{\mathbf{L}} \Box\psi & \Leftrightarrow \forall\gamma' \in J(\mathcal{A}, \mathbf{L})^n \ f, \gamma', p \models_{\mathbf{L}} \psi \\
f, \gamma, p \models_{\mathbf{L}} \blacksquare\psi & \Leftrightarrow \forall p' \in \mathcal{A} \ f, \gamma, p' \models_{\mathbf{L}} \psi \\
f, \gamma, p \models_{\mathbf{L}} \phi \wedge \psi & \Leftrightarrow f, \gamma, p \models_{\mathbf{L}} \phi \text{ and } f, \gamma, p \models_{\mathbf{L}} \psi \\
f, \gamma, p \models_{\mathbf{L}} \neg\phi & \Leftrightarrow f, \gamma, p \not\models_{\mathbf{L}} \phi
\end{array}$$

So, e.g., we have that $f, \gamma, p \models_{\mathbf{L}} \bigwedge_{i \in N} i$ if everybody chooses p in γ.

[1] Note that the property referred to as *unrestricted domain* is implicit in our definition of the preference aggregation framework.

Table 1 Example of aggregation in a voting scenario

	p	$p \to q$	q
1	Yes	Yes	Yes
2	No	Yes	Yes
3	Yes	No	No
f_{maj}	Yes	Yes	Yes

Example 1 A committee of three agents are voting on the following three propositions: "the candidate is qualified" (p), "if the candidate is qualified he will get an offer" ($p \to q$), and "the candidate will get an offer" (q). One possible voting scenario is illustrated in Table 1. In the table, the results of *proposition-wise majority voting*, i.e., the JAR f_{maj} accepting a proposition iff it is accepted by a majority of the agents, are also shown. This example can be modelled by taking the agenda to be $\mathcal{A} = \{p, p \to q, q, \neg p, \neg(p \to q), \neg q\}$ (recall that agendas are closed under single negation) and $\mathbf{L}$ to be propositional logic. The agents' votes can be modelled by the following judgment profile: $\gamma = \langle \gamma_1, \gamma_2, \gamma_3 \rangle$, where $\gamma_1 = \{p, p \to q, q\}, \gamma_2 = \{\neg p, p \to q, q\}, \gamma_3 = \{p, \neg(p \to q), \neg q\}$. We then have that:

- $f_{\text{maj}}, \gamma, p \models_{\mathbf{L}} 1 \wedge \neg 2 \wedge 3$
  agents 1 and 3 judges p to be true in the profile γ, while agent 2 does not;
- $f_{\text{maj}}, \gamma, p \models_{\mathbf{L}} \sigma$
  majority voting on p given the preference profile γ leads to acceptance of p;
- $f_{\text{maj}}, \gamma, p \models_{\mathbf{L}} \blacklozenge(1 \wedge 2)$
  agents 1 and 2 agree on some agenda item, under the judgment profile γ. Note that this formula does not depend on which agenda item is on the table;
- $f_{\text{maj}}, \gamma, p \models_{\mathbf{L}} \Diamond((1 \leftrightarrow 2) \wedge (2 \leftrightarrow 3) \wedge (1 \leftrightarrow 3))$
  there is some judgment profile on which all agents agree on p. Note that this formula does not depend on which judgment profile is on the table;
- $f_{\text{maj}}, \gamma, p \models_{\mathbf{L}} \Diamond\blacksquare((1 \leftrightarrow 2) \wedge (2 \leftrightarrow 3) \wedge (1 \leftrightarrow 3))$
  there is some judgment profile on which all agents agree on all agenda items. Note that this formula does not depend on any of the elements on the table;
- $f_{\text{maj}}, \gamma, p \models_{\mathbf{L}} \Box\blacksquare(\sigma \leftrightarrow \bigvee_{G \subseteq \{1,2,3\}, |G| \geq 2} \bigwedge_{i \in G} i)$
  the JAR f_{maj} implements majority voting.

We write $f \models_{\mathbf{L}} \phi$ iff $f, \gamma, p \models_{\mathbf{L}} \phi$ for every γ over $\mathcal{A}$ and $p \in \mathcal{A}$; $\models_{\mathbf{L}} \phi$ iff $f \models_{\mathbf{L}} \phi$ for all models f. Given a possible property of a JAR, such as, e.g., independence, we say that a formula *expresses the property* if the formula is true in an aggregation rule f iff f has the property.

Note that when we are given a formula $\phi \in \mathcal{L}(N, \mathcal{A})$, validity, i.e., $\models_{\mathbf{L}} \phi$, is defined with respect to models of the particular language $\mathcal{L}(N, \mathcal{A})$ defined over the particular agenda $\mathcal{A}$ (and similar for validity with respect to a JAR, i.e., $f \models_{\mathbf{L}} \phi$). The agenda, like the set of agents N, is given when we define the language, and is thus implicit in the interpretation of the language.[2]

[2] Likewise, in classical modal logic the language is parameterised with a set of primitive propositions, and validity is defined with respect to all models with valuations over that particular set.

Let an *outcome* o be a maximal conjunction of literals $\kappa_1 \wedge \cdots \wedge \kappa_n$, where each κ_i is either i or $\neg i$. The set O is the set of all possible outcomes. Note that the decision of the society is not incorporated here: an outcome only collects votes of agents from N.

3.1 Some properties

We have thus defined a language which can be used to express properties of judgment aggregation rules. An interesting question is then: what are the universal properties of aggregation rules expressible in the language; which formulae are valid? Here, in order to illustrate the logic, we discuss some of these logical properties. In Sect. 6 we give a complete axiomatisation of all of them.

Recall that we defined the set O of outcomes as the set of all conjunctions with exactly one, possibly negated, atom for each $i \in N$. Let $P = \{o \wedge \sigma, o \wedge \neg\sigma : o \in O\}$; $p \in P$ completely describes the decisions of the agents and the aggregation function. Let $\triangledown$ denote "exclusive or".

We have that:

$\models_{\mathbf{L}} \triangledown_{p \in P} p$	any agent and the JAR always have to make up their mind regarding any agenda item, and this is done unambiguously;
$\models_{\mathbf{L}} (i \wedge \neg j) \rightarrow \Diamond\neg i$	if *some* agent can think differently about an item than i does, then also i can change his mind about it. In fact this principle can be strengthened to;
$\models_{\mathbf{L}} (\Diamond i \wedge \Diamond\neg j) \rightarrow \Diamond(\neg i \wedge j)$	(for $i \neq j$) if two agents can give opposing recommendations (possibly for different items), then they can also disagree on one and the same item;
$\models_{\mathbf{L}} \Box\blacklozenge x$	for any $x \in \{i, \neg i, \sigma, \neg\sigma : i \in N\}$ – both the individual agents and the JAR will always judge some agenda item to be true, and conversely, some agenda item to be false;
$\models_{\mathbf{L}} \Diamond\blacklozenge(i \wedge j)$	there exist admissible judgment sets such that agents i and j agree on some judgment;
$\models_{\mathbf{L}} \Diamond\blacksquare(i \leftrightarrow j)$	there exist admissible judgment sets such that agents i and j always agree.

The interpretation of formulae depends on the agenda $\mathcal{A}$ and the underlying logic $\mathbf{L}$, in the quantification over the set $J(\mathcal{A}, \mathbf{L})^n$ of admissible, e.g., complete and $\mathbf{L}$-consistent, judgment profiles. Note that this means that some JAL formula might be valid under one underlying logic, while not under another. For example, if the agenda contains some formula which is inconsistent in the underlying logic (and, by implication, some tautology), then the following holds:

$\models_{\mathbf{L}} \Box\blacklozenge(i \wedge \sigma)$	for every judgment profile, there is some agenda item (take a tautology) which both agent i and the JAR judges to be true

But this property does not hold when every agenda item is consistent with respect to the underlying logic. One such combination of an agenda and an underlying logic will be discussed in Sect. 7.

4 Expressivity examples

The JAR properties discussed in Sect. 2.1 can be expressed as follows:

$$ND = \bigwedge_{i \in N} \Diamond\blacklozenge\neg(\sigma \leftrightarrow i) \tag{1}$$

$$IND = \Box \bigwedge_{o \in O} \blacksquare((o \wedge \sigma) \rightarrow \Box(o \rightarrow \sigma)) \tag{2}$$

$$UNA = \Box\blacksquare((1 \wedge \cdots \wedge n) \rightarrow \sigma) \tag{3}$$

Proposition 1

1. $f \models_{\mathbf{L}} ND$ *iff* f *has the property* ***ND1***.
2. $f \models_{\mathbf{L}} IND$ *iff* f *has the property* ***IND***.
3. $f \models_{\mathbf{L}} UNA$ *iff* f *has the property* ***UNA***.

Proposition 1 shows that commonly discussed properties of judgment aggregation rules can be expressed in the modal language $\mathcal{L}(N, \mathcal{A})$.

4.1 The discursive paradox

As illustrated in Example 1, the following formula expresses proposition-wise majority voting over some proposition p

$$MV = \sigma \leftrightarrow \bigvee_{G \subseteq N, |G| > \frac{n}{2}} \bigwedge_{i \in G} i \tag{4}$$

i.e., the following property of a JAR f and admissible profile $\langle A_1, \ldots, A_n \rangle$:

$$p \in f(A_1, \ldots, A_n) \Leftrightarrow |\{i : p \in A_i\}| > |\{i : p \notin A_i\}|$$

$f \models_{\mathbf{L}} MV$ exactly iff f has the above property for all judgment profiles and propositions.

However, we have the following in our logic. Assume that the agenda contains at least two distinct formulae and their material implication (i.e., $\mathcal{A}$ contains $p, q, p \rightarrow q$ for some $p, q \in \mathcal{L}(\mathbf{L})$).

Proposition 2 (Discursive Paradox) *Let* $\bot = \sigma \wedge \neg\sigma$, *then*

$$\models_{\mathbf{L}} \Diamond((\blacksquare MV) \rightarrow \bot) \text{ or, equivalently, } \models_{\mathbf{L}} \Diamond\blacklozenge\neg MV$$

when there are at least three agents and the agenda contains at least two distinct formulae and their material implication.

Proof Assume the opposite, i.e., that $\mathcal{A} = \{p, p \rightarrow q, q, \neg p, \neg(p \rightarrow q), \neg q, \ldots\}$ and there exists an aggregation rule f over $\mathcal{A}$ such that $f \models_{\mathbf{L}} \Box\blacksquare(\sigma \leftrightarrow \bigvee_{G \subseteq N, |G| > \frac{n}{2}} \bigwedge_{i \in G} i)$. Let γ be the judgment profile $\gamma = \langle A_1, A_2, A_3 \rangle$ where $A_1 = \{p, p \rightarrow q, q, \ldots\}$, $A_2 = \{p, \neg(p \rightarrow q), \neg q, \ldots\}$ and $A_3 = \{\neg p, p \rightarrow q, \neg q, \ldots\}$. We have that $f, \gamma, p' \models_{\mathbf{L}} \blacksquare(\sigma \leftrightarrow \bigvee_{G \subseteq N, |G| > \frac{n}{2}} \bigwedge_{i \in G} i)$ for any p', so $f, \gamma, p \models_{\mathbf{L}} \sigma \leftrightarrow \bigvee_{G \subseteq N, |G| > \frac{n}{2}} \bigwedge_{i \in G} i$. Because $f, \gamma, p \models_{\mathbf{L}} 1 \wedge 2$, it follows that $f, \gamma, p \models_{\mathbf{L}} \sigma$. In a similar manner it follows that $f, \gamma, p \rightarrow q \models_{\mathbf{L}} \sigma$ and $f, \gamma, q \models_{\mathbf{L}} \neg\sigma$. In other words, $p \in f(\gamma)$, $p \rightarrow q \in f(\gamma)$ and $q \notin f(\gamma)$. Since $f(\gamma)$ is complete, $\neg q \in f(\gamma)$. But that contradicts the fact that $f(\gamma)$ is required to be consistent. □

Proposition 2 is a logical statement of a variant of the well-known discursive dilemma: if three agents are voting on propositions p, q and $p \rightarrow q$, proposition-wise majority voting might not yield a consistent result.

5 Model checking and normal form

5.1 Model checking

Model checking is currently one of the most active areas of research with respect to reasoning in modal logics [6], and it is natural to investigate the complexity of this problem for judgment aggregation logic. Intuitively, the model checking problem for judgment aggregation logic is as follows:

Given f, γ, p and formula ϕ of JAL, is it the case that $f, \gamma, p \models_{\mathbf{L}} \phi$ or not?

While this problem is easy to understand mathematically, it presents some difficulties if we want to analyse it from a computational point of view. Specifically, the problem lies in the *representation* of the judgment aggregation rule, f. Recall that this function maps judgment profiles to complete and consistent judgment sets. A JAR must be defined for *all* judgment profiles over some agenda, i.e., it must produce an output for all these possible inputs. But how are we to represent such a rule? The simplest representation of a function $f : X \to Y$ is as the set of ordered pairs $\{(x, y) \mid x \in X \ \& \ y = f(x)\}$. However, this is not a feasible representation for JARs, as there will be exponentially many judgment profiles in the size of the agenda, and so the representation would be unfeasibly large in practice. If we *did* assume this representation for JARs, then it is not hard to see that model checking for our logic would be decidable in polynomial time: the naive algorithm, derivable from semantics, serves this purpose.

However, we emphasise that this result is of no practical significance, since it assumes an unreasonable representation for models—a representation that simply could not be used in practice for examples of anything other than trivial size.

So, what is a more realistic representation for JARs? Let us say a representation R_f of a JAR f is *reasonable* if: (*i*) the size of R_f is polynomial in the size of the agenda; and (*ii*) there is a polynomial time algorithm A, which takes as input a representation R_f and a judgment profile γ, and produces as output $f(\gamma)$. There are, of course, many such representations R_f for JARs f. Here, we will look at a very general one: where the JAR is represented as a polynomially bounded two-tape Turing machine T_f, which takes on its first tape a judgment profile, and writes on its second tape the resulting judgment set. The requirement that the Turing machine should be polynomially bounded roughly corresponds to the requirement that a JAR is "reasonable" to compute; if there is some JAR that cannot be represented by such a machine, then it is arguably of little value, since it could not be used in practice.[3] With such a representation, we can investigate the complexity of our model checking problem.

In modal logics, the usual source of complexity, over and above the classical logic connectives, is the modal operators. With respect to judgment aggregation logic, the operator $\Box$ quantifies over all judgment profiles, and hence over all consistent subsets of the agenda. It follows that this is a rather powerful operator: as we will see, it can be used as an NP oracle [20, p. 339]. In contrast, the operator $\blacksquare$ quantifies over members of the agenda, and is hence much weaker, from a computational perspective (we can think of it as a conjunction over elements of the agenda).

The power of the $\Box$ quantifier suggests that the complexity of model checking judgment aggregation logic over relatively succinct representations of JAR is going to be relatively

[3] Of course, we have no *general* way of checking whether any given Turing machine is guaranteed to terminate in polynomial time; the problem is undecidable. As a consequence, we cannot always check whether a particular Turing machine representation of a JAR meets our requirements. However, this does not prevent specific JARs being so represented, with corresponding proofs that they terminate in polynomial time.

high; we now prove that the complexity of model checking judgment aggregation logic is as hard as solving a polynomial number of NP-hard problems [20, pp. 424–429].

Theorem 2 *The model checking problem for judgment aggregation logic, assuming the representation of JARs described above, is Δ_2^p-hard; it is* NP*-hard even if the formula to be checked is of the form $\Diamond\psi$, where ψ contains no further $\Box$ or $\Diamond$ operators.*

Proof For Δ_2^p-hardness, we reduce SNSAT ("sequentially nested satisfiability"). An instance is given by a series of equations of the form

$$z_1 = \exists X_1.\phi_1(X_1) \;\; z_2 = \exists X_2.\phi_2(X_2, z_1) \;\; z_3 = \exists X_3.\phi_3(X_3, z_1, z_2)$$
$$\ldots$$
$$z_k = \exists X_k.\phi_k(X_k, z_1, \ldots, z_{k-1})$$

where $X_1, \ldots, X_k$ are pairwise disjoint sets of variables, and each $\phi_i(X_i, z_1, \ldots, z_{i-1})$ is a propositional logic formula over the variables X_i which also uses the values of z_j $(j < i)$; the idea is that we first check whether $\phi_1(X_1)$ is satisfiable, and if it is, we assign z_1 the value true, otherwise assign it false; we then check whether ϕ_2 is satisfiable under the assumption that z_1 takes the value just derived, and so on. Thus the result of each equation depends on the value of the previous one. The goal is to determine whether z_k is true.

To reduce this problem to judgment aggregation logic model checking, we first fix the JAR: this rule simply copies whatever agent 1's judgment set is. (Clearly this can be implemented by a polynomially bounded Turing machine.) The agenda is assumed to contain the variables $X_1 \cup \cdots \cup X_k \cup \{z_1, \ldots, z_k\}$ and their negations. We fix the initial judgment profile γ to be $X_1 \cup \cdots \cup X_k \cup \{z_1, \ldots, z_k\}$, and fix $p = x_1$. Given a variable x_i, define x_i^* to be $\blacklozenge(\mathbf{h}_{x_i} \wedge 1)$. If ϕ_i is one of the formulae $\phi_1, \ldots, \phi_k$, define ϕ_i^* to be the formula obtained from ϕ_i by systematically substituting x_i^* for each variable x_i and z_i^* similarly.

Now, we define the function ξ_i for natural numbers $i > 0$ as:

$$\xi_i = \begin{cases} z_1^* \leftrightarrow \Diamond(\phi_1^*) & \text{if } i = 1 \\ z_i^* \leftrightarrow \Diamond(\phi_i^* \wedge \bigwedge_{j=1}^{i-1} \xi_j) & \text{otherwise.} \end{cases}$$

And we define the formula to be model checked as:

$$\Diamond\left(\phi_k^* \wedge \bigwedge_{j=1}^{k-1} \xi_j\right)$$

It is now straightforward from construction that this formula is true under the interpretation iff z_k is true in the SNSAT instance. The proof of the latter half of the theorem is immediate from the special case where $k = 1$. □

5.2 Normal form

We show that in the case of a finite agenda, every formula of the logic is equivalent to one of a particular normal form. This normal form demonstrates, among other things, that we never need to put a modal operator in its own scope, and (hence) the overall maximal modal depth can be restricted to two.

Assume that the agenda is finite. The idea is to explicitly describe every table satisfying a given formula. For any formula ϕ, let

$$d(\phi) = \bigvee_{f,\gamma,p \models_{\mathbf{L}} \phi} d(f, \gamma, p)$$

where the *description* $d(f, \gamma, p)$ of f, γ, p is defined as follows:

$$\begin{aligned} d(f,\gamma,p) = \mathbf{h}_p \wedge & \\ & \textstyle\bigwedge_{p' \in \mathcal{A}} \blacklozenge(\mathbf{h}_{p'} \wedge \kappa(p',\gamma)) \wedge \\ & \textstyle\bigwedge_{\gamma' \in J(\mathcal{A},\mathbf{L})^n} \Diamond \bigwedge_{p' \in \mathcal{A}} \blacklozenge(\mathbf{h}_{p'} \wedge \kappa(p',\gamma) \wedge \chi(p',f,\gamma'))\}) \end{aligned}$$

where

$$\kappa(p,\gamma) = \bigwedge\{i : p \in \gamma_i\} \wedge \bigwedge\{\neg i : p \notin \gamma_i\} \quad \text{and} \quad \chi(p,f,\gamma) = \begin{cases} \sigma & p \in f(\gamma) \\ \neg\sigma & \text{otherwise} \end{cases}$$

The three main conjuncts in $d(f, \gamma, p)$ describe the current agenda item, the judgment profile, and the aggregation function, respectively. Note that this construction relies on the fact that the agenda, and thereby the set of all judgment profiles and the set of all aggregation functions, is finite.

Theorem 3 *For any f, γ, p and formula ϕ,*

$$f, \gamma, p \models_{\mathbf{L}} \phi \quad \Leftrightarrow \quad f, \gamma, p \models_{\mathbf{L}} d(\phi)$$

Proof For the direction to the left, let $f, \gamma, p \models_{\mathbf{L}} d(\phi)$. That is, there are f'', γ'', p'' such that

$$f'', \gamma'', p'' \models_{\mathbf{L}} \phi \tag{5}$$

$$f, \gamma, p \models_{\mathbf{L}} d(f'', \gamma'', p'') \tag{6}$$

In order to show that $f, \gamma, p \models_{\mathbf{L}} \phi$, we show that (1) $p'' = p$, (2) $\gamma'' = \gamma$ and (3) $f'' = f$.

1. From (6) (first conjunct) we have that $f, \gamma, p \models_{\mathbf{L}} \mathbf{h}_{p''}$, so $p = p''$.
2. From (6) (second conjunct) we have that

$$f, \gamma, p \models_{\mathbf{L}} \bigwedge_{p' \in \mathcal{A}} \blacklozenge(\mathbf{h}_{p'} \wedge \bigwedge\{i : p'' \in \gamma_i''\} \wedge \bigwedge\{\neg i : p' \notin \gamma_i''\})$$

   By the semantic definitions, this means that for every $p' \in \mathcal{A}$, $p' \in \gamma_i \Leftrightarrow p' \in \gamma_i''$ for any i, which means that $\gamma = \gamma''$.
3. Let $\overline{\gamma}$ be an arbitrary judgment profile. We will show that $f(\overline{\gamma}) = f''(\overline{\gamma})$. From (6) (third conjunct) we have that

$$f, \gamma, p \models_{\mathbf{L}} \Diamond \bigwedge_{p' \in \mathcal{A}} \blacklozenge(\mathbf{h}_{p'} \wedge \kappa(p',\overline{\gamma}) \wedge \chi(p', f'', \overline{\gamma}))$$

   According to the semantic definitions, this means that there exists a judgment profile $\hat{\gamma}$ such that for any $p' \in \mathcal{A}$:

$$f, \hat{\gamma}, p' \models_{\mathbf{L}} \kappa(p',\overline{\gamma}) \wedge \chi(p', f'', \overline{\gamma})$$

   This again means that there is a $\hat{\gamma}$ such that for any p', $p' \in \hat{\gamma}_i \Leftrightarrow p' \in \overline{\gamma}_i$ and $p' \in f(\hat{\gamma}) \Leftrightarrow p' \in f''(\overline{\gamma})$, i.e, $\hat{\gamma} = \overline{\gamma}$ and $f(\overline{\gamma}) = f''(\overline{\gamma})$.

For the direction to the right, let $f, \gamma, p \models_{\mathbf{L}} \phi$. It is easy to see that $f, \gamma, p \models_{\mathbf{L}} d(f, \gamma, p)$, and it follows that $f, \gamma, p \models_{\mathbf{L}} d(\phi)$. □

It follows from the theorem that every expressible property can be expressed with no modality in the scope of more than one other modality, and without "boxes" in the sense that no modal diamond is in the scope of a negation. Again, the construction is possible because the agenda is finite. However, note that the formula $d(\phi)$ can be extremely long even for very simple arguments ϕ, and will often be exponential in the number of elements in the agenda. Thus, for practical purposes, more succinct formulae are needed.

Table 2 The logic JAL($\mathbf{L}$) for the language $\mathcal{L}(N, \mathcal{A})$. p, p_i, q range over the agenda $\mathcal{A}$; ϕ, ψ, ψ_i over $\mathcal{L}(N, \mathcal{A})$; x over $\{\sigma, i : i \in N\}$; $\boxdot$ over $\{\Box, \blacksquare\}$; i, j over N; o over the set of outcomes O. $\mathbf{h}'_p$ means $\mathbf{h}_q$ when $p = \neg q$ for some q, otherwise it means $\mathbf{h}_{\neg p}$. $\mathbf{L}$ is the underlying logic

$\neg(\mathbf{h}_p \wedge \mathbf{h}_q)$	$p \neq q$	*Atmost*
$\bigvee_{p \in \mathcal{A}} \mathbf{h}_p$		*Atleast*
$\blacklozenge \mathbf{h}_p$		*Agenda*
$\blacklozenge(\mathbf{h}_p \wedge \varphi) \rightarrow \blacksquare(\mathbf{h}_p \rightarrow \varphi)$		*Once*
$\blacklozenge(\mathbf{h}_p \wedge x) \vee \blacklozenge(\mathbf{h}'_p \wedge x)$		*CpJS*
All instantiations of propositional tautologies		taut
$\boxdot(\psi_1 \rightarrow \psi_2) \rightarrow (\boxdot\psi_1 \rightarrow \boxdot\psi_2)$		*K*
$\boxdot\psi \rightarrow \psi$		*T*
$\boxdot\psi \rightarrow \boxdot\boxdot\psi$		4
$\neg\boxdot\psi \rightarrow \boxdot\neg\boxdot\psi$		5
$(\Diamond i \wedge \Diamond\neg j) \rightarrow \bigwedge_{o \in O} \Diamond o$		*C*
$\Box\blacksquare\psi \leftrightarrow \blacksquare\Box\psi$		*COMM*
$\Diamond\blacksquare\psi \rightarrow \blacksquare\Diamond\psi$		*CR*
$\mathbf{h}_p \rightarrow \Box\mathbf{h}_p$		*CA*
$\Diamond(\blacklozenge(\mathbf{h}_{p_1} \wedge o_1) \wedge \cdots \wedge \blacklozenge(\mathbf{h}_{p_k} \wedge o_k))$	For some judgment profile $\gamma = \langle\gamma_1, \ldots, \gamma_n\rangle$, where $\{p_1, \ldots, p_k\} = \mathcal{A}$ is the agenda, and $o_1, \ldots, o_k \in O$ are outcomes such that the ith conjunct in o_j is positive (does not start with a negation) iff $i \in \gamma_i$.	*UD*
From $p_1, \ldots p_n \vdash_{\mathbf{L}} q$ infer $\vdash_{JAL(\mathbf{L})} (\blacklozenge(\mathbf{h}_{p_1} \wedge x) \wedge \cdots \wedge \blacklozenge(\mathbf{h}_{p_n} \wedge x)) \rightarrow (\blacksquare(\mathbf{h}_q \rightarrow x) \wedge \blacksquare(\mathbf{h}'_q \rightarrow \neg x))$		*Closure*
From $\vdash_{JAL(\mathbf{L})} \varphi \rightarrow \psi$ and $\vdash_{JAL(\mathbf{L})} \varphi$ infer $\vdash_{JAL(\mathbf{L})} \psi$		*MP*
From $\vdash_{JAL(\mathbf{L})} \psi$ infer $\vdash_{JAL(\mathbf{L})} \boxdot\psi$		*Nec*

6 Axiomatisation

Given an underlying logic $\mathbf{L}$, a finite agenda $\mathcal{A}$ over $\mathbf{L}$, and a set of agents N, *Judgment Aggregation Logic* (JAL($\mathbf{L}$), or just JAL when $\mathbf{L}$ is understood) for the language $\mathcal{L}(N, \mathcal{A})$, is defined in Table 2. This is a Hilbert style presentation with axioms *Atmost*—*UD* and inference rules *Closure*, *MP* and *Nec*. We write $\vdash_{JAL(\mathbf{L})} \psi$ to denote that there exists a derivation of ψ using the axioms and rules from JAL. Likewise, for a formula schema Φ, $\vdash_{JAL(\mathbf{L})+\Phi} \psi$ denotes the existence of a derivation from the axioms of JAL together with Φ, using the rules of JAL, of ψ.

The first 5 axioms represent properties of a table and of judgment sets. Axiom *Atmost* says that there is at most one item on the table at a time, and *Atleast* says that we always have an item on the table. Axiom *Agenda* says that every agenda item will appear on the table, whereas *Once* says that every item of the agenda only appears on the table once. Note that a

conjunction $\mathbf{h}_p \wedge x$ reads: item p is on the agenda, and x is in favour of it, or x judges it true. Also, note the difference between $\mathbf{h}'_p$ and $\neg\mathbf{h}_p$: the first implies the latter, but not vice versa: from the fact that p is not the current agenda item on the table it does not follow that $\neg p$ must be it! Axiom $CpJS$ corresponds to the requirement that judgment sets are complete. Note that from $Agenda$, $CsJS$ and $CpJS$ we derive the scheme $\blacklozenge x \wedge \blacklozenge\neg x$, which says that everybody should at least express one opinion in favour of something, and against something.

The axioms $taut$ through 5 are well-known from modal logic: they directly reflect the unrestricted quantification in the truth definition of $\Box$ and $\blacksquare$. Axiom C says that for any agenda item for which it is possible to have opposing opinions, every possible outcome for that item should be achievable. $COMM$ says that everything that is true for an arbitrary profile and item, is also true for an arbitrary item and profile. Axiom CR (Church-Rosser) says that if there is a profile such that for all agenda items something is the case, it follows that for every agenda item there is a profile such that that something is the case. CA says that the agenda item on the table does not change when the preference profile is changed. To understand UD (Universal Domain), note that $\blacklozenge(\mathbf{h}_p \wedge o)$, where o is an outcome, says whether or not the agenda item p is in the judgment set of each of the agents. UD then says that there exists a judgment profile for any combination of such description of all items on the agenda; in other words that agents can choose any combinations of judgment sets. $Closure$ guarantees that agents behave consistently with respect to consequence in the logic $\mathbf{L}$. MP and Nec are standard.

Theorem 4 *If the agenda is finite, we have that for any formula* $\psi \in \mathcal{L}(N, \mathcal{A})$, $\vdash_{JAL(\mathbf{L})} \psi$ *iff* $\models_{\mathbf{L}} \psi$.

We point out that JAL has all the axioms $taut$, K, T, 4, 5 and the rules MP and Nec of the modal logic S5. However, *uniform substitution*, a principle of all normal modal logics (cf., e.g., [5]), does *not* hold. A counter example is the fact that the following is valid:

$$\Box\blacklozenge\sigma \tag{7}$$

– no matter what preferences the agents have, the JAR will always make some judgment – while this is not valid:

$$\Box\blacklozenge(\sigma \wedge i) \tag{8}$$

– the JAR will not necessarily make the same judgments as agent i.

As an example, we have that the discursive paradox is provable in JAL($\mathbf{L}$): $\vdash_{JAL(\mathbf{L})} \Diamond((\blacksquare MV) \rightarrow \bot)$. An example of a derivation of the less complicated (valid) property $\blacklozenge\Diamond(i \wedge j)$ is shown in Table 3.

6.1 Soundness completeness

We give a proof of Theorem 4. The structure of the completeness proof is rather standard for modal logics [5]: we only need to take care that the canonical model we build can be conceived as an aggregation function. We build a JAL table for a consistent formula ψ as follows.

Let MCS denote the set of all maximal and consistent sets of formulae. By viewing the set $\Pi = \{i, \ldots, n, \sigma, \mathbf{h}_p : p \in \mathcal{A}\}$ as primitive propositions, we can now build a canonical

Table 3 Summary of a JAL derivation of $\blacklozenge\Diamond(i \wedge j)$. The overall structure is that on line 4 we have $\vdash_{JAL(\mathbf{L})}$ $(A \wedge C) \vee (A \wedge D) \vee (B \wedge C) \vee (B \wedge D)$ and then, for each disjunct $X \in \{(A \wedge C), (A \wedge D), (B \wedge C), B \wedge D)\}$, we have $\vdash_{JAL(\mathbf{L})}$ $X \rightarrow \blacklozenge\Diamond(i \wedge j)$

1	$\blacklozenge(\mathbf{h}_p \wedge i) \vee \blacklozenge(\mathbf{h}'_p \wedge i)$	$CpJS(i)$
2	$\blacklozenge(\mathbf{h}_p \wedge j) \vee \blacklozenge(\mathbf{h}'_p \wedge j)$	$CpJS(j)$
3	Call 1 $A \vee B$ and 2 $C \vee D$	abbreviation, 1, 2
4	$(A \wedge C) \vee (A \wedge D) \vee (B \wedge C) \vee (B \wedge D)$	taut, 3
5	derive $\Diamond\blacklozenge(i \wedge j)$ from every disjunct of 4	strategy is $\vee$ elim
6	$\blacklozenge(\mathbf{h}_p \wedge i) \wedge \blacklozenge(\mathbf{h}_p \wedge j)$	assume $A \wedge C$
7	$\blacksquare(\mathbf{h}_p \rightarrow (i \wedge j))$	$Once$, 6, $K(\blacksquare)$
8	$\blacklozenge(i \wedge j)$	7, $Agenda$
9	$\Diamond\blacklozenge(i \wedge j)$	8, $T(\Box)$
10	$\blacklozenge(\mathbf{h}_p \wedge i) \wedge \blacklozenge(\mathbf{h}'_p \wedge j)$	assume $A \wedge D$
11	$\blacklozenge(\mathbf{h}_p \wedge x) \leftrightarrow \blacklozenge(\mathbf{h}'_p \wedge \neg x)$	$Agenda$, $Closure$
12	$\blacklozenge(\mathbf{h}_p \wedge i) \wedge \blacklozenge(\mathbf{h}_p \wedge \neg j)$	10, 11
13	$\blacklozenge(\mathbf{h}_p \wedge i \wedge \neg j)$	12, $Once$, $K(\blacksquare)$
14	$\blacklozenge(i \wedge \neg j)$	13, taut
15	$\Diamond\blacklozenge(i \wedge \neg j)$	14, $K(\Box)$
16	$\blacklozenge\Diamond(i \wedge \neg j)$	15, $COMM$
17	$\blacklozenge(\Diamond i \wedge D\neg j)$	16, $K(\blacksquare)$
18	$\blacklozenge\Diamond(i \wedge j)$	17, C
19	$\blacklozenge(\mathbf{h}'_p \wedge i) \wedge \blacklozenge(\mathbf{h}'_p \wedge j)$	assume $B \wedge D$
20	goes as 6-9	
21	$\blacklozenge(\mathbf{h}'_p \wedge i) \wedge \blacklozenge(\mathbf{h}_p \wedge j)$	assume $B \wedge C$
22	goes as 10 - 18	
23	$\blacklozenge\Diamond(i \wedge j)$	$\vee$-elim, 1, 2, 9, 18, 20, 22

Kripke structure $M^c = (MCS, R_{\Box}, R_{\blacksquare}, V)$, with MCS as the set of states, in the usual way:

$$(\Delta, \Delta') \in R_{\Box} \text{ iff for all } \Box\phi \in \Delta, \phi \in \Delta'$$
$$(\Delta, \Delta') \in R_{\blacksquare} \text{ iff for all } \blacksquare\phi \in \Delta, \phi \in \Delta'$$
$$V(p) = \{\Delta : p \in \Delta\} \text{ for } p \in \Pi$$

Let us use $M^c, \Delta \models \phi$ to denote the fact that the JAL formula ϕ is true in state Δ of M^c, defined in the usual modal logic sense (treating Π as atomic propositions). By this definition we have a normal modal logic, and the truth lemma follows by the standard result (see, e.g., [5]):

Lemma 1 *For every $\Delta \in MCS$ and every* JAL *formula ϕ,*

$$M^c, \Delta \models \phi \textit{ iff } \phi \in \Delta$$

We now proceed to construct a satisfying table from the canonical model.

Lemma 2 *For every $\Delta \in MCS$ and agenda item q, there is a unique outcome o^q_Δ such that $\blacklozenge(\mathbf{h}_q \wedge o^q_\Delta) \in \Delta$.*

Proof For existence, suppose that no such outcome exists, i.e., that for all outcomes o, $\neg\blacklozenge(\mathbf{h}_q \wedge o) \in \Delta$. In other words, $\blacksquare(\mathbf{h}_q \rightarrow \neg o) \in \Delta$ for any $o \in O$. By standard

modal reasoning it follows that $\blacksquare(\mathbf{h}_q \rightarrow \neg \bigvee_{o \in O} o) \in \Delta$. But $\bigvee_{o \in O} \in \Delta$ (a propositional tautology), so $\blacksquare \bigvee_{o \in O} \in \Delta$ by *Nec* and thus $\blacksquare(\mathbf{h}_q \rightarrow \bigvee_{o \in O}) \in \Delta$. It follows that $\blacksquare\neg\mathbf{h}_q \in \Delta$, which contradicts *Agenda*.

For uniqueness, suppose both $\blacklozenge(\mathbf{h}_q \wedge o) \in \Delta$ and $\blacklozenge(\mathbf{h}_q \wedge o') \in \Delta$. By *Once*, $\blacksquare(\mathbf{h}_q \rightarrow o) \in \Delta$, and thus $\blacklozenge(\mathbf{h}_q \wedge o' \wedge o) \in \Delta$. If $o \neq o'$, then $o' \wedge o$ is a a propositional contradiction, which would imply that $\blacklozenge(o' \wedge o) \notin \Delta$ by *Nec*. Thus, $o \neq o'$. □

Given a $\Delta \in MCS$, we extract a table f^Δ, γ^Δ, p^Δ as follows:

- p^Δ is the unique agenda item such that $\mathbf{h}_{p^\Delta} \in \Delta$ (existence and uniqueness guaranteed by *Atmost* and *Atleast*)
- For any agenda item q, $q \in \gamma_i^\Delta$ iff the literal i in o_Δ^q is positive (i.e., the ith conjunct in o does not start with a negation).
- Let $\gamma = \langle \gamma_1, \ldots, \gamma_n \rangle$ be an arbitrary judgment profile. $f^\Delta(\gamma)$ is defined as follows. Let $\beta = \Diamond(\blacklozenge(\mathbf{h}_{p_1} \wedge o_1) \wedge \cdots \wedge \blacklozenge(\mathbf{h}_{p_k} \wedge o_k))$ be the instance of *UD* defined by γ. We have that $\beta \in \Delta$. By the truth lemma, there is a Δ_1 such that $(\Delta, \Delta_1) \in R_\Box$ and $\blacklozenge(\mathbf{h}_{p_j} \wedge o_j) \in \Delta_1$ for each $j \in \{1, \ldots, k\}$. Thus, for each j, there is a Δ_2^j such that $(\Delta_1, \Delta_2^j) \in R_\blacksquare$ and $\mathbf{h}_{p_j} \wedge o_j \in \Delta_2^j$. Let, for each $p_j \in \mathcal{A} = \{p_1, \ldots, p_k\}$,

$$\begin{array}{ll} p_j \in f^\Delta(\gamma) & \text{if } \sigma \in \Delta_2^j \\ \neg p_j \in f^\Delta(\gamma) & \text{otherwise} \end{array}$$

Lemma 3 *For any* Δ, γ^Δ *is a judgment profile.*

Proof We must show that for each i, γ_i^Δ is consistent and complete.

Assume that $\gamma_i^\Delta = \{q_1, \ldots, q_k\}$ is not consistent. Let $r \in \mathcal{A}$ be some arbitrary agenda item not starting with negation. Because of inconsistency, we have that $\{q_1, \ldots, q_k\} \vdash_{\mathbf{L}} r$. By construction of γ_i^Δ, we have that $\blacklozenge(\mathbf{h}_{q_1} \wedge i) \wedge \cdots \wedge \blacklozenge(\mathbf{h}_{q_k} \wedge i) \in \Delta$. It follows by *Closure* that $\blacksquare(\mathbf{h}_r \rightarrow i) \in \Delta$. But, again because of inconsistency, we also have that $\{q_1, \ldots, q_k\} \vdash_{\mathbf{L}} \neg r$, so by *Closure* again we have that $\blacksquare(\mathbf{h}_r \rightarrow \neg i) \in \Delta$. It follows that $\blacksquare\neg\mathbf{h}_r \in \Delta$, which contradicts *Agenda*. Thus, γ_i^Δ is consistent.

For completeness, let $r \in Agenda$ be some arbitrary agenda item not being in the form of a negation. By *CpJS* either $\blacklozenge(\mathbf{h}_r \wedge i) \in \Delta$, or $\blacklozenge(\mathbf{h}_{\neg r} \wedge i) \in \Delta$. Wlog. assume the former. Then the ith literal in o_Δ^r is positive, and $r \in \gamma_i^\Delta$. □

The following lemma can be shown in a similar way to Lemma 3.

Lemma 4 *For any* Δ, f^Δ *is a judgment aggregation rule.*

The following can now be shown by induction over the length of the formula.

Lemma 5 *For all* JAL *formulae* ϕ *and all* $\Delta \in MCS$,

$$M^c, \Delta \models \phi \text{ iff } f^\Delta, \gamma^\Delta, p^\Delta \models_{\mathbf{L}} \phi$$

Proof (of Theorem 4) Let a finite $\mathcal{A}$ be given. Soundness (for all $\psi \in \mathcal{L}(N, \mathcal{A})$, $\vdash_{JAL(\mathbf{L})} \psi \Rightarrow \models_{\mathbf{L}} \psi$) is easy and left to the reader, which leaves us to prove that for all $\psi \in \mathcal{L}(N, \mathcal{A})$, $\models_{\mathbf{L}} \psi \Rightarrow \vdash_{JAL(\mathbf{L})} \psi$. This is the same as saying that for all ψ, $\nvdash_{JAL(\mathbf{L})} \psi \Rightarrow \not\models_{\mathbf{L}} \psi$. So suppose $\nvdash_{JAL(\mathbf{L})} \psi$. This means that $\phi = \neg\psi$ is consistent. By Lindenbaum's Lemma ([5]), there must be $\Delta \in MCS$ with $\phi \in \Delta$. By Lemma 1, we have $M^c, \Delta \models \phi$. By Lemma 5, we find a model f^Δ, a profile γ^Δ and an agenda item p such that $f^\Delta, \gamma^\Delta, p^\Delta \models_{\mathbf{L}} \phi$. In other words, ϕ is satisfiable, so $\neg\phi$ is not valid, so that $\not\models_{\mathbf{L}} \psi$. □

7 Preference aggregation

Recently, Dietrich and List [7] showed that preference aggregation can be embedded in judgment aggregation. In this section we show that judgment aggregation logic also can be used to reason about preference aggregation.

Given a finite set K of alternatives, [7] defines a simple predicate logic $\mathbf{L}^K$ with language $\mathcal{L}(\mathbf{L}^K)$ as follows:

- $\mathcal{L}(\mathbf{L}^K)$ has one constant a for each alternative $a \in K$, variables $v_1, v_2, \ldots$, a binary identity predicate $=$, a binary predicate P for strict preference, and the usual propositional and first order connectives
- Z is the collection of the following axioms:
  $\forall v_1\, \forall v_2\, (v_1 P v_2 \rightarrow \neg v_2 P v_1)$;
  $\forall v_1\, \forall v_2\, \forall v_3\, ((v_1 P v_2 \wedge v_2 P v_3) \rightarrow v_1 P v_3)$;
  $\forall v_1\, \forall v_2\, (\neg v_1 = v_2 \rightarrow (v_1 P v_2 \vee v_2 P v_1))$
- When $\Gamma \subseteq \mathcal{L}(\mathbf{L}^K)$ and ϕ is a formula, $\Gamma \models \phi$ is defined to hold iff $\Gamma \cup Z$ entails ϕ in the standard sense of predicate logic

It is easy to see that there is a one-to-one correspondence between the set of preference relations (total linear orders) over K and the set of $\mathbf{L}^K$-consistent and complete judgment sets over the *preference agenda*

$$\mathcal{A}^K = \{aPb, \neg aPb : a, b \in K, a \neq b\}$$

Given a SWF F over K, the corresponding JAR f^F over the preference agenda $\mathcal{A}^K$ is defined as follows: $f^F(A_1, \ldots, A_n) = A$, where A is the consistent and complete judgment set corresponding to $F(L_1, \ldots, L_n)$ where L_i is the preference relation corresponding to the consistent and complete judgment set A_i.

Thus we can use JAL to reason about preference aggregation as follows. Take the logical language $\mathcal{L}(N, \mathcal{A}^K)$, for some set of agents N, and take the underlying logic to be $\mathbf{L}^K$. We interpret our formulae in an SWF F over K, a preference profile $L \in L(K)$ and a pair $(a, b) \in K \times K, a \neq b$, as follows:

$$F, L, (a, b) \models^{swf} \phi \Leftrightarrow f^F, \gamma^L, aPb \models_{\mathbf{L}^K} \phi$$

where γ^L is the judgment profile corresponding to the preference profile L.

While in the general judgment aggregation case a formula is interpreted in the context of an agenda item, in the preference aggregation case a formula is thus interpreted in the context of a *pair of alternatives*.

Example 2 Three agents must decide between going to dinner (d), a movie (m) or a concert (c). Their individual preferences are illustrated in Table 4 in Sect. 3, along with the result of a SWF F_{maj} implementing *pair-wise majority voting*.

Let $L = \langle mdc, mcd, cmd \rangle$ be the preference profile corresponding to the preferences in the example. We have the following:

Table 4 Example of preference aggregation

1	mdc
2	mcd
3	cmd
F_{maj}	mcd

- $F_{\text{maj}}, L, (m, d) \models^{swf} 1 \wedge 2 \wedge 3$
  all agents agree, under the individual rankings L, on the relative ranking of m and d – they agree that d is better than m;
- $F_{\text{maj}}, L, (m, d) \models^{swf} \blacklozenge\neg(1 \leftrightarrow 2)$
  under the individual rankings L, there is some pair of alternatives on which agents 1 and 2 disagree;
- $F_{\text{maj}}, L, (m, d) \models^{swf} \Diamond\blacklozenge(1 \wedge 2)$
  agents 1 and 2 can choose their preferences such that they will agree on some pair of alternatives;
- $F_{\text{maj}}, L, (m, d) \models^{swf} \sigma \leftrightarrow \bigvee_{G\subseteq\{1,2,3\},|G|\geq 2} \bigwedge_{i\in G} i$
  the SWF F_{maj} implements pair-wise majority voting.

As usual, we write $F \models^{swf} \phi$ when $F, L, (a, b) \models^{swf} \phi$ for any L and (a, b), and so on. Thus, our formulae can be seen as expressing properties of social welfare functions.

Example 3 Take the formula $\Diamond\blacksquare(i \leftrightarrow \sigma)$. When this formula is interpreted as a statement about a social welfare function, it says that there exists a preference profile such that for all pairs (a, b) of alternatives, b is preferred over a in the aggregation (by the SWF) of the preference profile if and only if agent i prefers b over a.

7.1 Expressivity examples

We make precise the claim in Sect. 2.2 that the three mentioned SWF properties correspond to the three mentioned JAR properties, respectively. Recall the formulae defined in Sect. 4.

Proposition 3

1. $F \models^{swf} ND$ iff F has the property **ND2**
2. $F \models^{swf} IND$ iff F has the property **IIA**
3. $F \models^{swf} UNA$ iff F has the property **PO**

Proposition 3 expresses properties of SWFs in terms of $\mathcal{L}(N, \mathcal{A})$ formulas. Let us now look at properties of the set of *alternatives* K we can express. Properties involving *cardinality* are often of interest, for example in Arrow's theorem. Let:

$$MT2 = \Diamond\,(\blacklozenge(1 \wedge 2) \wedge \blacklozenge(1 \wedge \neg 2))$$

Intuitively, $MT2$ ('more than 2 (alternatives)') says that there is a profile such that for one agenda item both the agents 1 and 2 are in favour, and for one agenda item, agent 1 is in favour, but 2 is not.

Proposition 4 *Let* $F \in \mathcal{F}(K).|K| > 2$ *iff* $F \models^{swf} MT2$.

Proof For the direction to the left, let $F \models^{swf} MT2$. Thus, there is a γ such that there exists $(a^1, b^1), (a^2, b^2) \in K \times K$, where $a^1 \neq b^1$, and $a^2 \neq b^2$, such that (i) $a^1 P b^1 \in \gamma_1$, (ii) $a^1 P b^1 \in \gamma_2$, (iii) $a^2 P b^2 \in \gamma_1$ and (iv) $a^2 P b^2 \notin \gamma_2$. From (ii) and (iv) we get that $(a^1, b^1) \neq (a^2, b^2)$, and from that and (i) and (iii) it follows that γ_1 contains two different pairs $a^1 P b^1$ and $a^2 P b^2$ each having two different elements. But that is not possible if $|K| = 2$, because if $K = \{a, b\}$ then $\mathcal{A}^K = \{aPb, \neg aPb, bPa, \neg bPa\}$ and thus it is impossible that $\gamma_1 \subseteq \mathcal{A}^K$ since we cannot have $aPb, bPa \in \gamma_1$.

For the direction to the right, let $|K| > 2$; let a, b, c be three distinct elements of K. Let γ_1 be the judgment set corresponding to the ranking abc and γ_2 the judgment set corresponding to acb. Now, for any aggregation rule f, $f, \gamma, aPb \models_{\mathbf{L}^K} 1 \wedge 2$ and $f, \gamma, bPc \models_{\mathbf{L}^K} 1 \wedge \neg 2$. Thus, $F \models^{swf} MT2$, for any SWF F. □

We now have everything we need to express Arrow's statement as a formula. It follows from his theorem that the formula is valid on the class of all social welfare functions.

Theorem 5 $\models^{swf} MT2 \rightarrow \neg(UNA \wedge ND \wedge IND)$

Proof Note that MT2, UNA, ND and IND are true SWF properties, their truth value wrt. a table is determined solely by the SWF. For example, $F, L, (a, b) \models^{swf} MT2$ iff $F \models^{swf} MT2$, for any F, L, a, b. Let $F \in \mathcal{F}(K)$, and $F, L, (a, b) \models^{swf} MT2$ for some L and a, b. By Proposition 4, K has more than two alternatives. By Arrow's theorem, F cannot have all the properties **PO**, **ND2** and **IIA**. W.l.o.g assume that F does not have the **PO** property. By Proposition 3, $F \not\models^{swf} UNA$. Since UNA is a SWF property, this means that $F, L, (a, b) \not\models^{swf} UNA$ (satisfaction of UNA is independent of L, a, b), and thus that $F, L, (a, b) \models^{swf} \neg UNA \vee \neg ND \vee \neg IND$. □

Note that the formula in Theorem 5 does not mention any agenda items (i.e., pairs of alternatives) such as $\mathbf{h}_{aPb}$ directly in an expression. This means that the formula is a member of $\mathcal{L}(N, \mathcal{A}^K)$ for *any* set of alternatives K, and is valid no matter which set of alternatives we assume.

The formula MV which in the general judgment aggregation case expresses proposition-wise majority voting, expresses in the preference aggregation case pair-wise majority voting, as illustrated in Example 2. The preference aggregation correspondent to the discursive paradox of judgment aggregation is the well known Condorcet's voting paradox, stating that pair-wise majority voting can lead to aggregated preferences which are cyclic (even if the individual preferences are not). We can express Condorcet's paradox as follows, again as a universally valid logical property of SWFs.

Proposition 5 $\models^{swf} MT2 \rightarrow \Diamond\blacklozenge\neg MV$, *when there are at least three agents.*

Proof The proof is similar to the proof of the discursive paradox. Let $f^F, \gamma, aPb \models_{\mathbf{L}^K} MT2$; there are thus three distinct elements $a, b, c \in K$. Assume that $f^F, \gamma, aPb \models_{\mathbf{L}^K} \Box\blacksquare MV$. Let γ' be the judgment profile corresponding to the preference profile $X = (abc, cab, bca)$. We have that $f^F, \gamma', aPb \models_{\mathbf{L}^K} 1 \wedge 2$ and, since $f^F, \gamma', aPb \models_{\mathbf{L}^K} MV$, we have that $f^F, \gamma', aPb \models_{\mathbf{L}^K} \sigma$ and thus that $aPb \in f^F(\gamma')$ and $(a, b) \in F(X)$. In a similar manner we get that $(c, a) \in F(X)$ and $(b, c) \in F(X)$. But that is impossible, since by transitivity we would also have that $(a, c) \in F(X)$ which contradicts the fact that $F(X)$ is antisymmetric. Thus, it follows that $f^F, \gamma, aPb \not\models_{\mathbf{L}^K} \Box\blacksquare MV$.

7.2 Axiomatisation and logical properties

We immediately get, from Theorem 4, a sound and complete axiomatisation of preference aggregation over a finite set of alternatives.

Corollary 1 *If the set of alternatives K is finite, we have that for any formula $\psi \in \mathcal{L}(N, \mathcal{A}^K)$, $\vdash_{JAL(\mathbf{L}^K)} \psi$ iff $\models^{swf} \psi$.*

Proof Follows immediately from Theorem 4 and the fact that for any JAR f, there is a SWF F such that $f = f^F$. □

So, for example, Arrow's theorem is provable in JAL($\mathbf{L}^K$): $\vdash_{JAL(\mathbf{L}^K)} MT2 \rightarrow \neg(UNA \wedge ND \wedge IND)$. Of course, this argument is existential rather than constructive, and an actual formal JAL proof of the theorem would be of additional interest. While we do not present a

complete proof here, we give a partial one in the next section. The main building block in a well-known proof of Arrow's theorem [10] is a so-called *Neutrality Lemma*. In Sect. 8 we not only give a syntactic derivation of this Lemma, but we also demonstrate how our object language can help to sort out some subtleties in its formulation.

Every formula which is valid with respect to judgment aggregation rules is also valid with respect to social welfare functions, so all general logical properties of JARs are also properties of SWFs.

Depending on the agenda, SWFs may have additional properties, induced by the logic $\mathbf{L}^K$, which are not always shared by JARs with other underlying logics. One such property is $\Diamond i$. While we have

$$\models^{swf} \Diamond i,$$

for other agendas there are underlying logics $\mathbf{L}$ such that

$$\not\models_{\mathbf{L}} \Diamond i$$

To see the latter, take an agenda with a formula p which is inconsistent in the underlying logic $\mathbf{L}$ – p can never be included in a judgment set. To see the former, take an arbitrary pair of alternatives (a, b). There exists some preference profile in which agent i prefers b over a.

Technically speaking, the formula $\Diamond i$ holds in SWFs because the agenda $\mathcal{A}^K$ does not contain a formula which (alone) is inconsistent wrt. the underlying logic $\mathbf{L}^K$. By the same reason, the following properties also hold in SWFs but not in JARs in general.

$$\models^{swf} \bigwedge_{o \in O} \Diamond o$$

– for any pair of alternatives (a, b), any possible combination of the relative ranking of a and b among the agents is possible.

$$\models^{swf} i \rightarrow \Diamond \neg i$$

– given an alternative b which is preferred over some other alternative a by agent i, there is some other pair of alternatives c and d such that d is *not* preferred over c – namely $(c, d) = (b, a)$.

$$\models^{swf} \Box(\blacksquare(i \vee j) \rightarrow \blacklozenge(i \wedge \neg j))$$

– if, given preferences of agents and a SWF, for any two alternatives it is always the case that either agent i or agent j prefers the second alternative over the first, then there must exist a pair of alternatives for which the two agents disagree. A justification is that no single agent can prefer the second alternative over the first for *every* pair of alternatives, so in this case if i prefers b over a then j must prefer a over b. Again, this property does not necessarily hold for other agendas, because the agenda might contain an inconsistency the agents could not possibly disagree upon.

Proof-theoretically, these additional properties of SWFs are derived using the $Closure$ rule. In the next section we discuss preference aggregation further.

8 A logical study of neutrality

We now present a preference aggregation case study for our logic, representing and formally proving Geanakoplos' **Strict Neutrality Lemma**. This lemma featured in the paper *Three*

brief proofs of Arrow's Impossibility Theorem [10,11], in particular it is the main argument in the third proof. Interestingly, Geanakoplos formulates two slightly different versions of this lemma in [10] and [11]. We will call the two formulations SN1 and SN2, respectively.

The property of neutrality is in the literature of judgment aggregation also known as *Systematicity*, see for instance [19]. Like neutrality, systematicity occurs in a number of variants: see [18].

What we do in this section is the following. First of all, our object language gives a neat characterisation of the properties SN1 and SN2. In fact we formulate a slight generalisation of both principles, **SN** in our language, and we give a formal derivation of **SN** in our logic JAL. By doing so, we illustrate how an important step in one proof [10,11] of Arrow's theorem, and a property that is important in the literature on judgment aggregation [18], can be derived in our logic, provided that we add the principles UNA and CA.

Let us start with SN1, [10, p. 4]:

> All binary social rankings are made the same way. Consider two pairs of alternatives ab and $\alpha\beta$. Suppose each voter strictly prefers a to b, or b to a, and suppose each voter has the same relative ranking of $\alpha\beta$ as he does of ab. Then the social preference between ab is identical to the social preference between $\alpha\beta$, and both social preferences are strict.

Our mathematical formulation of it:

Strict Neutrality Lemma (**SN1**) $\forall(R_1, \ldots, R_n) \in L(K)^n \forall a, b, \alpha, \beta \in K$
$((\forall i \in N(aR_ib \Leftrightarrow \alpha R_i\beta)) \Rightarrow (aF(R_1, \ldots, R_n)b \Leftrightarrow \alpha F(R_1, \ldots, R_n)\beta))$

We claim that this is the proper formalisation in the context of the preferences being a linear order. Then, the conditions 'one agent strictly prefers a to b' (or b to a) and '$a \neq b$', 'all agents strictly prefer a to b or b to a' are all equivalent. More formally, reading the main implication in **SN1** as $\Phi \Rightarrow \Psi$, in our setting, this is equivalent to $(\Phi \wedge a \neq b \wedge \alpha \neq \beta) \Rightarrow (\Psi \wedge a \neq b \wedge \alpha \neq \beta)$ (from right to left: note that $a = b$ with Φ implies $\alpha = \beta$ and both Φ and Ψ become vacuously true).

Now, property **SN1** can be expressed as follows. Let $C \subseteq N$. Slightly abusing notation, we will use C in the object language to mean $C \leftrightarrow \bigwedge_{i\in C} i \wedge \bigwedge_{j\notin C} \neg j$. Note that under this notation, C is equal to an outcome o. Let:

$$SN_1 = (C \wedge \sigma) \rightarrow \blacksquare(C \rightarrow \sigma) \tag{9}$$

We make a number of remarks concerning (9) here. First of all from SN_1 it follows that

$$(C \wedge \neg\sigma) \rightarrow \blacksquare(C \rightarrow \neg\sigma) \tag{10}$$

The following is a proof: suppose that SN_1 holds, but (10) does not. That is, we have $(C \wedge \neg\sigma) \wedge \blacklozenge(C \wedge \sigma)$. Applying SN_1 would yield $(C \wedge \neg\sigma) \wedge \blacklozenge\blacksquare(C \rightarrow \sigma)$. Using the $S5$ properties of $\blacksquare$, we get $(C \wedge \neg\sigma) \wedge \blacksquare(C \rightarrow \sigma)$ and, again because of $S5$, $(C \wedge \neg\sigma) \wedge (C \rightarrow \sigma)$, a contradiction. In fact, (10) is *equivalent* to SN_1 (use the same proof, reversing the roles of σ and $\neg\sigma$).

Secondly, note that in SN_1 we can avoid a lot of quantification that is present in **SN1**. This is first of all because, as we will claim, SN_1 is *valid* and hence there is an implicit quantification over all profiles $(R_1, \ldots, R_n)$ and all outcomes a and b. In the same spirit, C is used as an (arbitrary) variable over coalitions. And secondly, the quantification over alternatives α and β is captured by the use of $\blacksquare$, which has of course a universally quantified interpretation.

Thirdly, our language facilitates to note a remarkable similarity between SN_1, $(C \wedge \sigma) \rightarrow \blacksquare(C \rightarrow \sigma)$, and IND, $(C \wedge \sigma) \rightarrow \Box(C \rightarrow \sigma)$. In words, SN_1 says

> Given a profile and two alternatives, if we know how everybody's judgment regarding the two alternatives is and what the aggregation says about them, then, *if we keep the profile fixed but look at two arbitrary alternatives,* if everybody's judgment would stay the same regarding those two alternatives, then the aggregation decision for them should be the same as well.

Compare this with our reading of IND:

> Given a profile and two alternatives, if we know how everybody's judgment regarding the two alternatives is and what the aggregation says about them, then, *if we keep the alternatives fixed but look at an arbitrary profile,* if everybody's judgment would stay the same regarding the two alternatives, then the aggregation decision for them should be the same as well.

Even more loosely, one might interpret SN_1 as "the aggregation decision between two alternatives should not depend on the particular alternatives", and IND as "the aggregation decision between two alternatives should not depend on the particular profile".

Let us now consider Geanakoplos' second formulation SN2 of the Strict Neutrality Lemma [11, p. 214]:

> All binary social rankings are made the same way. Consider two pairs of alternatives ab and $\alpha\beta$. Suppose that in some profile π each voter strictly prefers a to b, or b to a, and suppose that in another profile π' each voter has the same relative ranking of $\alpha\beta$ as he does of ab in π. Then the social preference between ab in π is identical to the social preference between $\alpha\beta$ in π' and both preferences are strict.

Our mathematical formulation of it:

> Strict Neutrality Lemma (**SN2**) $\forall(R_1, \ldots, R_n), (S_1, \ldots, S_n) \in L(K)^n \forall a, b, \alpha, \beta \in K$
> $((R_1, \ldots, R_n) \neq (S_1, \ldots, S_n) \Rightarrow$
> $((\forall i \in N(a R_i b \Leftrightarrow \alpha S_i \beta)) \Rightarrow (aF(R_1, \ldots, R_n)b \Leftrightarrow \alpha F(S_1, \ldots, S_n)\beta)))$

Here $(R_1, \ldots, R_n) \neq (S_1, \ldots, S_n)$ means $R_i = S_i$ for all i. This is equality between sets: this equality does not hold iff $\exists i \in N \exists x, y \in K(x R_i y \wedge \neg x S_i y)$.

Again, we make some observations regarding **SN2**. First of all, note that is it not required that the pair $\alpha\beta$ is *different* from the pair ab, and that, for the case that $\alpha = a$ and $\beta = b$, we get exactly Independence (**IND**) back. In other words, **SN2** is stronger than **IND**. Note that also in the case where the antecedent of **SN2** does not hold (making **SN2** vacuously true), i.e. when $(R_1, \ldots, R_n) = (S_1, \ldots, S_n)$, **IND** reduces to the tautology $aF(R_1, \ldots, R_n)b \Leftrightarrow aF(S_1, \ldots, S_n)b$.

Secondly, note that, whereas in **SN1** it was demanded that we stay in the *same* profile, **SN2** requires that we change to a *different* profile. In modal logic, one can use de Rijke's *Difference* operator D to reason about what holds in different states: $D\varphi$ being true in state s means that φ is true in all states that are different from s. In our logic, in principle, we can *define* such an operator when K is finite. This would work as follows. Suppose we want to define $D_\Box$, with $f, \gamma, p \models_{\mathbf{L}} D_\Box\varphi$ meaning: in all profiles γ' different from γ, we have $f, \gamma', p \models_{\mathbf{L}} \varphi$. Let the cardinality of K be k. Let $\pi = \langle x_1, \ldots, x_k \rangle$ be a way to order K and let Π be the set of all such π. For given $\pi = \langle x_1, \ldots, x_k \rangle$, and agent i, we write $P_i\pi$ for

$\blacklozenge x_1 P_i x_2 \wedge \blacklozenge x_2 P_i x_3 \wedge \cdots \wedge \blacklozenge x_{k_1} P_i x_k$. This statement is true in f, γ, p when given γ, agent i's preferences are exactly like the order in π, with x_1 being the most preferred, and x_k the least. Now we can define $D_\Box$ as

$$D_\Box \varphi = \bigwedge_{i \in N} \bigwedge_{\pi_i \in \Pi} (P_i \pi_i \rightarrow \Box(\bigvee_{i \in N} \neg P_i \pi \rightarrow \varphi)) \tag{11}$$

Note that although this is written down relatively compactly, it in fact represents a very long formula.

However, here we can do without the difference operator and instead look at a simple generalisation of **SN1** and **SN2**. Note that **SN1** is a property of one preference profile, and **SN2** is one of two different profiles. A straightforward way to generalise this is to say that the property holds for any two profiles, whether those profiles equal each other (**SN1**) or be different (**SN2**).

Strict Neutrality Lemma (**SN**) $\forall (R_1, \ldots, R_n), (S_1, \ldots, S_n) \in L(K)^n \forall a, b, \alpha, \beta \in K$
$((\forall i \in N(aR_i b \Leftrightarrow \alpha S_i \beta)) \Rightarrow (aF(R_1, \ldots, R_n)b \Leftrightarrow \alpha F(S_1, \ldots, S_n)\beta))$

We claim that **SN** is expressed by SN below:

$$SN \;=\; (C \wedge \sigma) \rightarrow \blacksquare\Box(C \rightarrow \sigma) \tag{12}$$

Since both $\blacksquare$ and $\Box$ satisfy the T axiom, it is clear that SN implies both SN_1 and IND.

Theorem 6 *Let F be an arbitrary social welfare function.*

1. $F \models^{swf} SN$ *iff F has the property* **SN**
2. $\vdash_{JAL(\mathbf{L}^K)+UNA+IND} SN$

Proof 1 First suppose F satisfies **SN**. Let $f = f^F$ and $\gamma = \gamma^L = (R_1, \ldots, R_n)$ be as explained in Sect. 7, and suppose that for some profile γ and agenda item aPb, we have $f, \gamma, aPb \models_{\mathbf{L}^K} (C \wedge \sigma)$. This implies $(aF(R_1, \ldots, R_n)b$. Now take an arbitrary profile $\gamma' = (S_1, \ldots, S_n)$ and agenda item cPd. It is sufficient to show that $f, \gamma', cPd \models_{\mathbf{L}^K} C \rightarrow \sigma$, so suppose $f, \gamma', cPd \models_{\mathbf{L}^K} C$. This means that for all $i \in N$, we have $aR_i b \Leftrightarrow cS_i b$. By **SN**, we have $(aF(R_1, \ldots, R_n)b \Leftrightarrow cF(S_1, \ldots, S_n)d))$, and hence $cF(S_1, \ldots, S_n)d$, i.e., $f, \gamma', cPd \models_{\mathbf{L}^K} \sigma$. For the converse, suppose that F does *not* satisfy **SN**. This means that for some $(R_1, \ldots, R_n)$ and $(S_1, \ldots, S_n)$, for all agents $i \in N$ we have $aP_i b \Leftrightarrow cP_i d$ but still $aF(R_1, \ldots, R_n)b$ but not $cF(S_1, \ldots, S_n)d$. Let coalition C be the agents who favour a over b. This means that in the associated model we have $f, \gamma, aPb \models_{\mathbf{L}^K} C \wedge \sigma$ and $f, \gamma', cPd \models_{\mathbf{L}^K} C \wedge \neg\sigma$, i.e., $f, \sigma, aPb \models_{\mathbf{L}^K} (C \wedge \sigma) \wedge \neg\Box\blacksquare(C \rightarrow \sigma)$.

2 We can use the previous item and call in completeness. However, for a syntactic derivation, see the following Example 4. □

For the syntactic proof of SN, we first need to establish some (modal) properties.

Lemma 6 *Let φ be an arbitrary formula, and x and $y \in K$.*

1. *Let* $\mathbf{D}$ *be an arbitrary sequence of diamonds, including the empty sequence, i.e.,* $\mathbf{D} \in \{\epsilon, \Diamond, \blacklozenge, \Diamond\Diamond, \Diamond\blacklozenge, \blacklozenge\Diamond, \blacklozenge\blacklozenge, \Diamond\Diamond\Diamond, \ldots\}$. *Then*

$$\vdash_{JAL(\mathbf{L}^K)} \mathbf{D}\varphi \leftrightarrow \bigvee_{aPb \in \mathcal{A}} \mathbf{D}(aPb \wedge \varphi)$$

2. $\vdash_{JAL(\mathbf{L}^K)} \Diamond(xPby \wedge \varphi) \leftrightarrow (xPy \wedge \Diamond\varphi)$
3. $\vdash_{JAL(\mathbf{L}^K)+IND} \Diamond(C \wedge \sigma) \wedge \Diamond(C \wedge \neg\sigma) \rightarrow \bot$.

Proof 1 For $\mathbf{D} = \epsilon$, this follows immediately from *Atleast*. For any non-empty sequence,let $\mathbf{B}$ be the sequence of boxes that is associated with $\mathbf{D}$ (i.e., the nth box in the sequence $\mathbf{B}$ is the dual of the nth diamond in $\mathbf{D}$). Since we have necessitation for each box, from *Atleast* we derive $\mathbf{B} \bigvee_{aPb \in \mathcal{A}} aPb$. From this, derive $\mathbf{D}\varphi \leftrightarrow \mathbf{D}(\bigvee_{aPb \in \mathcal{A}} aPb \wedge \varphi)$ the modal principle here is that $\boxdot\psi$ and $\Diamond\varphi$ imply $\Diamond(\psi \wedge \varphi))$. The latter is equivalent to $\mathbf{D}\varphi \leftrightarrow \bigvee_{aPb \in \mathcal{A}} \mathbf{D}(aPb \wedge \varphi)$ (use $\Diamond(\phi \vee \psi) \leftrightarrow (\Diamond\phi \vee \Diamond\psi)$).

2 By CA, we have $(xPy \wedge \Diamond\varphi) \rightarrow (\Box xPy \wedge \Diamond\varphi)$. By the first modal principle alluded to under item 1, we derive $\Diamond(xPy \wedge \varphi)$. For the other direction, again by CA, $\Diamond(xPy \wedge \varphi) \rightarrow \Diamond(\Box xPy \wedge \varphi)$. In modal logic a diamond distributes over conjunction, giving $\Diamond\Box xPy \wedge \Diamond\varphi$. Since the logic for $\Box$ is $S5$, we have that $\Diamond\Box xPy$ is equivalent to $\Box xPy$, which in turn implies xPy, giving $(xPy \wedge \Diamond\varphi)$.

3 From $\Diamond(C \wedge \sigma) \wedge \Diamond(C \wedge \neg\sigma)$ using IND, infer $\Diamond\Box(C \wedge \sigma) \wedge \Diamond\Box(C \wedge \neg\sigma)$ and, since $\Box$ is $S5$, we get $\Box(C \rightarrow \sigma) \wedge \Box(C \rightarrow \neg\sigma)$. But then, from $\Diamond(C \wedge \neg\sigma)$ and $\Box(C \rightarrow \sigma)$, infer $\Diamond(C \wedge \sigma \wedge \neg\sigma)$, which implies $\bot$. □

Example 4 (Proof that $\vdash_{JAL(\mathbf{L}^K)+UNA+IND} SN$*)* The principle of SN is of the form $A \rightarrow B$, which we will prove by deriving $A \rightarrow (\neg B \rightarrow \bot)$. So assume A, i.e., $(C \wedge \sigma)$. From Lemma 6.1 we know that $(C \wedge \sigma) \leftrightarrow \bigvee_{Pab \in \mathcal{A}}(Pab \wedge C \wedge \sigma)$. So SN now becomes of the form $(\bigvee_x A_x) \rightarrow (\neg B \rightarrow \bot)$, and this is proven if we show that for each disjunct A_x, $A_x \rightarrow (\neg B \rightarrow \bot)$. So let us assume such a disjunct A_x, i.e., assume that $aPb \wedge C \wedge \sigma$. Using the T axiom for both modalities gives us

$$\Diamond\blacklozenge(aPb \wedge C \wedge \sigma) \tag{13}$$

Also note that $\neg B$ is of the form $\Diamond\blacklozenge(C \wedge \neg\sigma)$. Using Lemma 6.1, we know that this is equivalent to $\bigvee_{Pcd} \Diamond\blacklozenge(cPd \wedge C \wedge \neg\sigma)$. So also $\neg B$ is a disjunction. If we can show that for each disjunct $\Diamond\blacklozenge(cPd \wedge C \wedge \neg\sigma)$, that (13) $\rightarrow (\Diamond\blacklozenge(cPd \wedge C \wedge \neg\sigma) \rightarrow \bot)$, we are done. So assume (13) and

$$\Diamond\blacklozenge(cPd \wedge C \wedge \neg\sigma) \tag{14}$$

Now distinguish four cases:

1. $a = c$ and $b = d$. By $COMM$, from (13) and (14), we derive $\blacklozenge\Diamond(aPb \wedge C \wedge \sigma) \wedge \blacklozenge\Diamond(aPb \wedge C \wedge \neg\sigma)$. Using Lemma 6.2, we get $\blacklozenge(aPb \wedge \Diamond(C \wedge \sigma)) \wedge \blacklozenge(aPb \wedge \Diamond(C \wedge \neg\sigma))$. Using *Once* and some modal principles to this yields $\blacklozenge(aPb \wedge \Diamond(C \wedge \sigma) \wedge \Diamond(C \wedge \neg\sigma))$. From Lemma 6.3 we conclude $\blacklozenge(aPb \wedge \bot)$ which yields $\bot$.

   Note that we were able to derive $\neg\Diamond\blacklozenge(aPb \wedge C \wedge \neg\sigma)$ from $(aPb \wedge C \wedge \sigma)$, in other words, we proved

$$(aPb \wedge C \wedge \sigma) \rightarrow \Box\blacksquare((aPb \wedge C) \rightarrow \sigma) \tag{15}$$

2. $a \neq c$ and $b \neq d$. By UD we have:

$$\Diamond(\blacklozenge(cPa \wedge A) \wedge \blacklozenge(aPb \wedge C) \wedge \blacklozenge(bPd \wedge A)) \tag{16}$$

   Applying UNA to the above gives

$$\Diamond(\blacklozenge(cPa \wedge A \wedge \sigma) \wedge \blacklozenge(aPb \wedge C) \wedge \blacklozenge(bPd \wedge A \wedge \sigma)) \tag{17}$$

   Note that $aPb \wedge C \wedge \sigma$ is an assumption, so we can apply (15) to (17), giving

$$\Diamond(\blacklozenge(cPa \wedge A \wedge \sigma) \wedge \blacklozenge(aPb \wedge C \wedge \sigma) \wedge \blacklozenge(bPd \wedge A \wedge \sigma)) \tag{18}$$

Now, by transitivity for P and $Closure$ we get

$$\Diamond\blacklozenge(cPd \wedge C \wedge \sigma) \tag{19}$$

Finally, we show that (14) and (19) are inconsistent. From (19) and (15), we derive $\Diamond\blacklozenge\Box\blacksquare((cPd \wedge C) \rightarrow \sigma)$. Let φ be $((cPd \wedge C) \rightarrow \sigma$. By $COMM$, $\Diamond\blacklozenge\Box\blacksquare\varphi$ is equivalent to $\Diamond\blacklozenge\blacksquare\Box\varphi$, and by $S5$ of $\blacksquare$, this is equivalent to $\Diamond\blacksquare\Box\varphi$. Using $COMM$ again, we get $\Diamond\Box\blacksquare\varphi$, and $S5$ for $\Box$, gives $\Box\blacksquare\varphi$. Indeed, this is inconsistent with (14) which, with our definition of φ is $\Diamond\blacklozenge\neg\varphi$.

3. $a \neq c$ and $b = d$. By UD and UNA we now have:

$$\Diamond(\blacklozenge(cPa \wedge A \wedge \sigma) \wedge \blacklozenge(aPb \wedge C)) \tag{20}$$

Note that $aPb \wedge C \wedge \sigma$ is an assumption, so we can apply (15) to (17), giving (recall that $b = d$)

$$\Diamond(\blacklozenge(Pca \wedge A \wedge \sigma) \wedge \blacklozenge(Pad \wedge C \wedge \sigma)) \tag{21}$$

By transitivity and $Closure$ we get $\Diamond\blacklozenge(cPd \wedge C \wedge \sigma)$ and, as we did from (19), we derive $\perp$.

4. $a = c$ and $b \neq c$. As the previous case.

9 Related work and conclusions

While there has been considerable recent interest [22] in modal logics capturing game theoretic concepts such as Nash equilibrium [13] or the core [1], formal logics related to social choice have focused mostly on the logical representation of preferences when the set of alternatives is large and on the computation properties of computing aggregated preferences for a given representation [14–16].

A notable and recent exception is a logical framework for judgment aggregation developed by Marc Pauly [21], in order to be able to characterise the logical relationships between different judgment aggregation rules. While the motivation is similar to the work in this paper, the approaches are fundamentally different: [21], the possible *results* from applying a rule to some judgment profile are taken as primary and described axiomatically; in our approach the aggregation rule and its possible *inputs*, i.e., judgment profiles, are taken as primary and described axiomatically. The two approaches do not seem to be directly related to each other in the sense that one can be embedded in the other.

The modal logic *arrow logic* [23] is designed to reason about any object that can be graphically represented as an arrow, and has various modal operators for expressing properties of and relationships between these arrows. In the preference aggregation logic JAL($\mathbf{L}^K$) we interpreted formulae in pairs of alternatives – which can be seen as arrows. Thus, (at least) the preference aggregation variant of our logic is related to arrow logic. However, while the modal operators of arrow logic can express properties of preference relations such as transitivity, they cannot directly express most of the properties we have discussed in this paper. Nevertheless, the relationship to arrow logic could be investigated further in future work. In particular, arrow logics are usually proven complete wrt. an algebra. This could mean that it might be possible to use such algebras as the underlying structure to represent individual and collective preferences. Then, changing the preference profile takes us from one algebra to another, and a SWF determines the collective preference, in each of the algebras.

In summary, we have presented a sound and complete logic JAL for representing and reasoning about judgment aggregation. JAL is expressive: it can express judgment aggregation rules such as majority voting; complicated properties such as independence; and important results such as the discursive paradox, Arrow's theorem and Condorcet's paradox. We argue that these results show exactly which logical capabilities an agent needs in order to be able to reason about judgment aggregation. It is perhaps surprising that a relatively simple language provides these capabilities. JAL provides a proof theory, in which results such as those mentioned above can be derived.[4]

The axiomatisation describes the logical principles of judgment aggregation, and can also be instantiated to reason about specific instances of judgment aggregation, such as classical Arrovian preference aggregation. Thus our framework sheds light on the differences between the logical principles behind general judgment aggregation on the one hand and classical preference aggregation on the other. We presented a proof of the Strict Neutrality Lemma, which is a main step in the proof of Arrow's theorem in [11]. A complete proof of the theorem is not possible in the space available here, but we believe that the proof of the lemma provides some insight into the utility of the proof theory for that kind of purpose. Although the object language seems to be good enough to make some subtle points clear, and, as one would expect from a modal language, is free from most of the variables and quantifiers that a first-order formalisation brings with it, we are now in a position to make a critical remark as well. Where our language has been good at *representing* the claim in a clear way, the *reasoning* as presented in our proof relies quite heavily on a 'smart' use of the UD principle: we reason about the profiles that will lead us to a proof, and state using UD that such a profile exists. That in itself does not represent a definitive argument against using an object language like ours, but it *does* raise the question of whether there are languages that do more right to this particular form of reasoning about specific profiles.

Related to this is the following future research question (several of the next suggestions are very thankfully taken from the useful reviews received from the AAMAS journal). We have added Unrestricted Domain as a basic axiom in our logic, but one might wonder whether there is a weaker basic logic that is complete wit respect to models that do not impose this condition of availability of all possible profiles. This would then allow for constraints on preference profiles like *single peakedness* allowing for *possibility* results, rather than *impossibility* results (see, e.g., [8]).

There is currently a renewed interest in formal representations of Arrow's theorem and related impossibility results. In [17], Lin and Tang for instance use induction (over both the number of agents and the number of alternatives) in a first-order setting to prove Arrow's theorem, where the base case is proven using computer programs. Likewise, in [12], Grandi and Endriss presented a First-order Formalisation of Arrow's theorem and on some initial experiments with automated reasoning tools to derive the theorem for a fixed number of agents and alternatives. Although there is an obvious translation from our modal language to that of first-order logic, there is more to be said about the connection between our formalisation and those in [17,12]. Moreover, it would be interesting to see whether modal theorem provers could be employed directly for our modal language. In future work it would also be interesting to relax the completeness and consistency requirements of judgment sets, and try to characterise these in the logical language, as properties of general judgment sets, instead.

[4] Dietrich and List [7] prove a general version of Arrow's theorem for JARs: for a *strongly connected* agenda, a JAR has the **IND** and **UNA** properties iff it does not have the **ND1** property, where strong connectedness is an algebraic and logical condition on agendas. Thus, if we assume that the agenda is strongly connected then $(ND \wedge UNA) \leftrightarrow \neg ND1$ is valid, and derivable in JAR. An interesting possibility for future work is to try to characterise conditions such as strong connectedness directly as a logical formula.

Since modal logic is *the* logic to reason about binary relations, it would be of value to see whether a change from linear orders to total preorders to represent preferences would be easily implementable in our framework. Last but not least, it is interesting to see whether and how our approach can be extended to cater for relaxations on the given number of agents, and how we can capture generalisations of impossibility results to the case with infinitely many agents or voters, as for instance given in [9]. We leave all this for, hopefully near, future work.

Acknowledgements An earlier version of this paper was presented at the AAMAS 2007 conference [2]. The comments of the AAMAS reviewers, the feedback at the AAMAS conference, and the three reviews received from the *Autonomous Agents and Multi Agent Systems* journal gave very helpful directions to the shape of this final version. Discussions with Frank Wolter about product logics were very useful as well.

References

1. Ågotnes, T., van der Hoek, W., & Wooldridge, M. (2009). Reasoning about coalitional games. *Artificial Intelligence, 173*(1), 45–79.
2. Ågotnes, T., Wooldridge, M., & van der Hoek, W. (2007) Reasoning about judgment and preference aggregation. In M. Huhns, O. Shehory, (Eds.), *Proceedings of the sixth international conference on autonomous agents and multiagent systems (AAMAS 2007)* (pp. 554–561). IFAMAAS.
3. Arrow, K. J. (1951). *Social choice and individual values*. London: Wiley.
4. Arrow, K. J., Sen, A. K., & Suzumura, K., (Eds). (2002). Handbook of social choice and welfare, vol 1. North-Holland.
5. Blackburn, P., de Rijke, M., & Venema, Y. (2001). *Modal logic*. Cambridge: Cambridge University Press.
6. Clarke, E. M., Grumberg, O., & Peled, D. A. (2000). *Model checking*. Cambridge, MA: The MIT Press.
7. Dietrich, F., & List, C. (2007). Arrow's theorem in judgment aggregation. *Social Choice and Welfare, 29*(1), 19–33.
8. Ehlers, L., & Storcken, T. (2008). Arrow's possibility theorem for one dimensional single-peaked preferences. *Games and Economic Behavior, 64*(2), 533–547.
9. Fishburn, P. (1970). Arrow's impossibility theorem: Concise proof and infinite voters. *Journal of Economic Theory, 2*(1), 103–106.
10. Geanakoplos, J. (2001). Three brief proofs of Arrow's impossibility theorem. Cowles foundation discussion papers 1123R3, Cowles Foundation, Yale University.
11. Geanakoplos, J. (2005). Three brief proofs of arrow's impossibility theorem. *Economic Theory, 26*(1), 211–215.
12. Grandi, U., & Endriss, U. (2009). First-order formalisation of Arrow's Theorem. Presentation given at a seminar at the University of Amsterdam, http://www.illc.uva.nl/lgc/seminar/docs/Arrow.pdf.
13. Harrenstein, B. P., van der Hoek, W., Meyer, J. -J., & Witteveen, C. (2003). A modal characterization of Nash equilibrium. *Fundamenta Informaticae, 57*(2–4), 281–321.
14. Lafage, C., & Lang, J. (2000). Logical representation of preferences for group decision making. In A. G. Cohn, F. Giunchiglia, & B. Selman (Eds.), *Proceedings of the conference on principles of knowledge representation and reasoning (KR-00)* (pp. 457–470). Morgan Kaufman.
15. Lang, J. (2002). From preference representation to combinatorial vote. In D. Fensel, F. Giunchiglia, D. L. McGuinness, M. -A. Williams (Eds.), *Proceedings of the conference on principles and knowledge representation and reasoning (KR-02), April 22–25, 2002* (pp. 277–290). Morgan Kaufmann.
16. Lang, J. (2004). Logical preference representation and combinatorial vote. *Annals of the Mathematics of Artificial Intellegence, 42*(1-3), 37–71.
17. Lin, F., & Tang, P. (2008). Computer-aided proofs of Arrow's and other impossibility theorems. *Proceedings of the 23rd AAAI conference on artificial intelligence* (pp. 114–119).
18. List, C. (2009). Judgment aggregation. A bibliography on the discursive dilemma, doctrinal paradox and decisions on multiple propositions. Website, see http://personal.lse.ac.uk/LIST/DOCTRINALPARADOX.HTM.
19. List, C., & Pettit, P. (2005). Aggregating sets of judgments: An impossibility result. *Economics and Philosophy, 18*, 89–110.

20. Papadimitriou, C. H. (1994). *Computational complexity*. Reading, MA: Addison-Wesley.
21. Pauly, M. (2006). Axiomatizing collective judgment sets in a minimal logical language. Manuscript.
22. van der Hoek, W., & Pauly, M. (2006). Modal logic for games and information. In P. Blackburn, J. van Benthem, & F. Wolter, (Eds). , *Handbook of modal logic* (pp. 1077–1148). Amsterdam, The Netherlands: Elsevier Science Publishers B.V.
23. Venema, Y. (1996). A crash course in arrow logic. In M. Marx, M. Masuch, & L. Polos, *Arrow logic and multi-modal logic* (pp. 3–34). Stanford: CSLI Publications.