Reasoning about Judgment and Preference Aggregation

Thomas Ågotnes
Department of Computer Engineering, Bergen University College
PB. 7030, N-5020 Bergen, Norway
tag@hib.no

Wiebe van der Hoek
Department of Computer Science, University of Liverpool
Liverpool L69 7ZF, UK
wiebe@csc.liv.ac.uk

Michael Wooldridge
Department of Computer Science, University of Liverpool
Liverpool L69 7ZF, UK
mjw@csc.liv.ac.uk

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 recently 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 of such rules. 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.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods—*Modal logic*

General Terms

Theory

Keywords

Judgment aggregation, preference aggregation, modal logic

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 [2]. The recently emerging field of *judgment aggregation* studies aggregation from a logical perspective, and discusses 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 [5].

In this paper we present a 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 all 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:

AAMAS'07 May 14–18 2007, Honolulu, Hawai'i, USA.

978-81-904262-7-5 (RPS) ©2007 IFAAMAS

- 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.

From these points 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 Section 3 we introduce the syntax and semantics of JAL, and study the complexity of the model checking problem. Formulae of JAL are interpreted directly by, and thus represent properties of, judgment aggregation rules. In Section 4 we demonstrate that the logic can express commonly discussed properties of judgment aggregation rules, such as the discursive paradox. We give a sound and complete axiomatisation of the logic in Section 5, 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 Section 6 we introduce an alternative interpretation of JAL formulae directly in social welfare functions. We obtain a sound and complete axiomatisation of the logic for preference aggregation as well. Sections 7 and 8 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 Σ 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 start with 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 start with 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 hence is a recipe to enforce a rational group decision, given an 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 $\langle A_1, \ldots, A_n \rangle, f(A_1, \ldots, A_n) = A_i$

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 most well known result about SWFs is Arrow's theorem [1]. Many variants of the theorem appear in the literature, differing in assumptions about the preference relations. In this paper, we take 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). Cf., e.g., [2] for a discussion and more general formulations.

Formally, let K be a set of *alternatives*. We henceforth implicitly assume that there are always 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 $(b, c) \in R$ implies 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 version of R, i.e., $R^s = R \setminus \{(a, a) : a \in K\}$. aR^sb means that b is preferred over a and that $a \neq b$.

A *preference profile* for Σ 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 Section 2.1 are:

Non-dictatorship (ND2) $\neg\exists_{i\in\Sigma}\forall(R_1,\ldots,R_n)\in L(K)^n$ $F(R_1,\ldots,R_n)=R_i$ (corresponds to **ND1**)

Independence of irrelevant alternatives (IIA) $\forall(R_1,\ldots,R_n)\in L(K)^n\forall(S_1,\ldots,S_n)\in L(K)^n\forall a\in K\forall b\in K((\forall i\in\Sigma(aR_ib\Leftrightarrow aS_ib))\Rightarrow(aF(R_1,\ldots,R_n)b\Leftrightarrow aF(S_1,\ldots,S_n)b))$ (corresponds to **IND**)

Pareto Optimality (PO) $\forall(R_1,\ldots,R_n)\in L(K)^n\forall a\in K\forall b\in K((\forall i\in\Sigma aR_i^sb)\Rightarrow aF(R_1,\ldots,R_n)^sb)$ (corresponds to **UNA**)

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**.*

3. JUDGMENT AGGREGATION LOGIC: SYNTAX AND SEMANTICS

The language of *Judgment Aggregation Logic (JAL)* is parameterised by a set of agents $\Sigma=\{1,2,\ldots,n\}$ (we will assume that there are at least two agents) and an agenda $\mathcal{A}$. The following atomic propositions are used:

$$\Pi=\{i,\sigma,\mathbf{h}_p\mid p\in\mathcal{A},i\in\Sigma\}$$

The language $\mathcal{L}(\Sigma,\mathcal{A})$ 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 structures consisting of an agenda item, a judgment profile and a judgment aggregation function; informally, i means that the agenda item is in agent i's judgment set in the current judgment profile; σ means that the agenda item is in the aggregated judgment set of the current judgment profile; $\mathbf{h}_p$ means that the 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 $\mathcal{L}(\Sigma,\mathcal{A})$. A *model* wrt. $\mathcal{L}(\Sigma,\mathcal{A})$ 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.

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

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

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" ($p\rightarrow q$), and "the candidate will get an offer" (q). One possible voting scenario is illustrated in the left part of 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 $\mathcal{A}=\{p,p\rightarrow q,q,\neg p,\neg(p\rightarrow q),\neg q\}$ (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 $\gamma_1=\{p,p\rightarrow q,q\}$, $\gamma_2=\{\neg p,p\rightarrow q,q\}$, $\gamma_3=\{p,\neg(p\rightarrow q),\neg q\}$. We then have that:*

	p	$p\rightarrow q$	q
1	yes	yes	yes
2	no	yes	yes
3	yes	no	no
f_{maj}	yes	yes	yes

1	mdc
2	mcd
3	cmd
F_{maj}	mcd

Table 1: Examples

- $f_{maj},\gamma,p\models_{\mathbf{L}}1\wedge\neg2\wedge3$ *(agents 1 and 3 judges p to be true in the profile γ, while agent 2 does not)*
- $f_{maj},\gamma,p\models_{\mathbf{L}}\sigma$ *(majority voting on p given the preference profile γ leads to acceptance of p)*
- $f_{maj},\gamma,p\models_{\mathbf{L}}\blacklozenge(1\wedge2)$ *(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.)*
- $f_{maj},\gamma,p\models_{\mathbf{L}}\Diamond((1\leftrightarrow2)\wedge(2\leftrightarrow3)\wedge(1\leftrightarrow3))$ *(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.)*
- $f_{maj},\gamma,p\models_{\mathbf{L}}\Diamond\blacksquare((1\leftrightarrow2)\wedge(2\leftrightarrow3)\wedge(1\leftrightarrow3))$ *(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.)*
- $f_{maj},\gamma,p\models_{\mathbf{L}}\sigma\leftrightarrow\bigvee_{G\subseteq\{1,2,3\},|G|\geq2}\bigwedge_{i\in G} i$ *(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 $\phi\in\mathcal{L}(\Sigma,\mathcal{A})$, validity, i.e., $\models_{\mathbf{L}}\phi$, is defined with respect to models of the particular language $\mathcal{L}(\Sigma,\mathcal{A})$ 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 Σ, is given when we define the language, and is thus implicit in the interpretation of the language[1].

Let an *outcome* o be a maximal conjunction of literals $(\neg)1,\ldots,(\neg)n$. 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 Σ.

3.1 Model Checking

Model checking is currently one of the most active areas of research with respect to reasoning in modal logics [4], 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 $f,\gamma,p\models\phi$ or not?

[1]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.

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 one 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[2]. 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 [9, 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 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 [9, 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

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

where $X_1, \dots, X_k$ are disjoint sets of variables, and each $\phi_i(Y)$ is a propositional logic formula over the variables Y; the idea is 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 $X_1 \cup \dots \cup X_k \cup \{z_1, \dots, z_k\}$ and their negations. We fix the initial judgment profile γ to be $X_1 \cup \dots \cup X_k \cup \{z_1, \dots, z_k\}$, 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 $\phi_1, \dots, \phi_k$, 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:

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

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 $k = 1$. $\square$

3.2 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 Section 5 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 from Σ. 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 a decision

$\models_{\mathbf{L}} (i \wedge \neg j) \to \Diamond\neg i$ – 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

$\models_{\mathbf{L}} (\Diamond i \wedge \Diamond\neg j) \to \Diamond(\neg i \wedge j)$

$\models_{\mathbf{L}} \Box\blacklozenge x$ – for any $x \in \{i, \neg i, \sigma, \neg\sigma : i \in \Sigma\}$ – 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.

[2]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.

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 hold:

$\models_{\mathbf{L}} \Box\blacklozenge(i \wedge \sigma)$ – 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 agenda and underlying logic will be discussed in Section 6.

4. EXPRESSIVITY EXAMPLES

Non-dictatorship can be expressed as follows:

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

LEMMA 1. *$f \models_{\mathbf{L}}$ ND iff f has the property **ND1**.*

Independence can be expressed as follows:

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

LEMMA 2. *$f \models_{\mathbf{L}}$ IND iff f has the property **IND**.*

Unanimity can be expressed as follows:

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

LEMMA 3. *$f \models_{\mathbf{L}}$ UNA iff f has the property **UNA**.*

4.1 The Discursive Paradox

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

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

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

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

$f \models MV$ 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 1 (DISCURSIVE PARADOX).

$$\models_{\mathbf{L}} \Diamond((\blacksquare MV) \rightarrow \bot)$$

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

PROOF. Assume the opposite, e.g., 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 $f \models_{\mathbf{L}} \Box\blacksquare(\sigma \leftrightarrow \bigvee_{G\subseteq\Sigma,|G|>\frac{n}{2}} \bigwedge_{i\in G} i)$. 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 $f, \gamma, p' \models_{\mathbf{L}} \blacksquare(\sigma \leftrightarrow \bigvee_{G\subseteq\Sigma,|G|>\frac{n}{2}} \bigwedge_{i\in G} i)$ for any p', so $f, \gamma, p \models_{\mathbf{L}} \sigma \leftrightarrow \bigvee_{G\subseteq\Sigma,|G|>\frac{n}{2}} \bigwedge_{i\in G} i$. 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 1 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. AXIOMATISATION

Given an underlying logic $\mathbf{L}$, a finite agenda $\mathcal{A}$ over $\mathbf{L}$, and a set of agents Σ, *Judgment Aggregation Logic* (JAL($\mathbf{L}$), or just JAL when $\mathbf{L}$ is understood) for the language $\mathcal{L}(\Sigma,\mathcal{A})$, is defined in Table 2.

$\neg(\mathbf{h}_p \wedge \mathbf{h}_q)$ if $p \neq q$	*Atmost*
$\bigvee_{p\in\mathcal{A}} \mathbf{h}_p$	*Atleast*
$\blacklozenge\mathbf{h}_p \quad p \in \mathcal{A}$	*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*)
From $p_1,\ldots p_n \vdash_{\mathbf{L}} q$ infer $\blacklozenge(\mathbf{h}_{p_1} \wedge x) \wedge \cdots \wedge \blacklozenge(\mathbf{h}_{p_n} \wedge x) \rightarrow \blacksquare(\mathbf{h}_q \rightarrow x) \wedge \blacksquare(\mathbf{h}'_q \rightarrow \neg x)$	*Closure*
From $\varphi \rightarrow \psi$ and φ infer ψ	*MP*
From ψ infer $\boxdot\psi$	*Nec*

Table 2: The logic JAL($\mathbf{L}$) for the language $\mathcal{L}(\Sigma,\mathcal{A})$. p, p_i, q range over the agenda $\mathcal{A}$; ϕ,ψ,ψ_i over $\mathcal{L}(\Sigma,\mathcal{A})$; x over $\{\sigma, i : i \in \Sigma\}$; $\boxdot$ over $\{\Box,\blacksquare\}$; i, j over Σ; 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.

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. 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* – 5 are well familiar 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. *Closure* guarantees that agents behave consistently with respect to consequence in the logic $\mathbf{L}$. *MP* and *Nec* are standard. We use $\vdash_{JAL(\mathbf{L})}$ to denote derivability in JAL($\mathbf{L}$).

THEOREM 3. *If the agenda is finite, we have that for any formula $\psi \in \mathcal{L}(\Sigma,\mathcal{A})$, $\vdash_{JAL(\mathbf{L})} \psi$ iff $\models_{\mathbf{L}} \psi$.*

PROOF. Soundness is straightforward. For completeness (we focus on the main idea here and leave out trivial details), we build a

JAL table for a consistent formula ψ as follows. In fact, our axiomatisation completely determines a table, except for the behaviour of f. To be more precise, let a *table description* be a conjunction of the form $\mathbf{h}_p \wedge o \wedge (\neg)\sigma$. It is easy to see that table descriptions are mutually exclusive, and, moreover, we can derive $\bigvee_{\tau \in T} \tau$, where T is the set of all table descriptions. Let $\mathcal{D}$ be the set of all maximal consistent sets Δ. We don't want all of those: it might well be that ψ requires σ to be in a certain way, which is incompatible with some Δ's. We define two accessibility relations in the standard way: $R_\Box \Delta_1 \Delta_2$ iff for all $\Box\psi$: $\Box\psi \in \Delta_1 \Rightarrow \psi \in \Delta_2$. Similarly for $R_\blacksquare$ with respect to $\blacksquare$. Both relations are equivalences (due to *taut*-5), and moreover, when $R_\Box \Delta_1 \Delta_2$ and $R_\blacksquare \Delta_2 \Delta_3$ then for some Δ_2', also $R_\blacksquare \Delta_1 \Delta_2'$ and $R_\Box \Delta_2' \Delta_3$ (because of axiom *COMM*).

Let Δ_0 be a MCS containing ψ. We now define the set *Tables* =

$$\{\Delta_0\} \cup \{\Delta_1, \Delta_2 \mid (R_\Box \Delta_0 \Delta_1 \text{ and } R_\blacksquare \Delta_1 \Delta_2) \text{ or } (R_\blacksquare \Delta_0 \Delta_1 \text{ and } R_\Box \Delta_1 \Delta_2)\}$$

Every $\Delta \in$ *Tables* can be conceived as a pair γ, p, since every Δ contains a unique $\blacklozenge(\mathbf{h}_q \wedge o \wedge (\neg)\sigma)$ for every $\mathbf{h}_q$ and a unique $\mathbf{h}_p$.

It is then easy to verify that, for every $\Delta \in$ *Tables*, and every formula φ, $\Delta \models \varphi$ iff $\varphi \in \Delta$, where $\models$ here means truth in the ordinary modal logic sense when the set of states is taken to be *Tables*. Now, we extract an aggregation function f and pairs γ, p as follows:

For every $\Delta \in$ *Tables*, find a conjunction $\mathbf{h}_p \wedge o \wedge (\neg)\sigma$. There will be exactly one such p. This defines the p we are looking for. Furthermore, the γ is obtained, for every agent i, by finding all q for which $\blacklozenge(\mathbf{h}_q \wedge i)$ is currently true. Finally, the function f is a table of all tuples $\mathbf{h}_p, o(p), \sigma$ for which $\blacklozenge(\mathbf{h}_p \wedge o(o) \wedge \sigma)$ is contained in some set in *Tables*. □

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., [3]), does *not* hold. A counter example is the fact that the following is valid:

$$\Box\blacklozenge\sigma \tag{5}$$

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

$$\Box\blacklozenge(\sigma \wedge i) \tag{6}$$

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

So, for 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.

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

Table 3: JAR derivation of $\blacklozenge\Diamond(i \wedge j)$

6. PREFERENCE AGGREGATION

Recently, Dietrich and List [5] showed that preference aggregation can be embedded in judgment aggregation. In this section we show that our judgment aggregation logic also can be used to reason about preference aggregation.

Given a set K of alternatives, [5] 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 an 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 $\mathcal{L}(\Sigma, \mathcal{A}^K)$, for some set of agents Σ, and take the underlying logic to be $\mathbf{L}^K$. We can then interpret our formulae in an SWF F over K, a preference profile $L \in L(K)$ and a pair $(a, b) \subseteq K \times K$, $a \neq b$, 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 on the right in Table 1 in Section 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:*

- $F_{maj}, L, (m, d) \models^{swf} 1 \wedge 2 \wedge 3$ *(all agents agree, under the individual rankings L, on the relative ranking of m and d – they agree that d is better than m)*
- $F_{maj}, L, (m, d) \models^{swf} \blacklozenge\neg(1 \leftrightarrow 2)$ *(under the individual rankings L, there is some pair of alternatives on which agents 1 and 2 disagree)*

- $F_{maj}, L, (m, d) \models^{swf} \Diamond\blacklozenge(1 \wedge 2)$ *(agents 1 and 2 can choose their preferences such that they will agree on some pair of alternatives)*
- $F_{maj}, L, (m, d) \models^{swf} \sigma \leftrightarrow \bigvee_{G\subseteq\{1,2,3\},|G|\geq 2} \bigwedge_{i\in G} i$ *(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.*

6.1 Expressivity Examples

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

Proposition 2.

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

The properties expressed above are properties of SWFs. Let us now look at properties of the set of *alternatives* K we can express. Properties involving *cardinality* is often of interest, for example in Arrow's theorem. Let:

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

Proposition 3. *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) $a^1Pb^1 \in \gamma_1$, (ii) $a^1Pb^1 \in \gamma_2$, (iii) $a^2Pb^2 \in \gamma_1$ and (iv) $a^2Pb^2 \notin \gamma_2$. 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^1Pb^1 and a^2Pb^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 $f, f, \gamma, aPb \models 1 \wedge 2$ and $f, \gamma, bPc \models 1 \wedge \neg 2$. 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 4. $\models^{swf} MT2 \rightarrow \neg(PO \wedge ND \wedge IIA)$

Proof. Note that $MT2, PO, ND$ and IIA 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 $F \models MT2$, 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 3, 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 2, $F \not\models^{swf} PO$. Since PO is a SWF property, this means that $F, L, (a, b) \not\models^{swf} PO$ (satisfaction of PO is independent of L, a, b), and thus that $F, L, (a, b) \models^{swf} \neg PO \vee \neg ND \vee \neg IIA$. □

Note that the formula in Theorem 4 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 $\mathcal{L}(\Sigma, \mathcal{A}^K)$ 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 4. *$\models^{swf} MT2 \rightarrow \Diamond\blacklozenge\neg MV$, 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 $f^F, \gamma, aPb \not\models_{\mathbf{L}^K} \Box\blacksquare MV$. □

6.2 Axiomatisation and Logical Properties

We immediately get, from Theorem 3, 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 $\psi \in \mathcal{L}(\Sigma, \mathcal{A}^K)$, $\vdash_{JAL(\mathbf{L}^K)} \psi$ iff $\models^{swf} \psi$.*

Proof. Follows immediately from Theorem 3 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 $\textsc{jal}(\mathbf{L}^K)$: $\vdash_{JAL(\mathbf{L}^K)} MT2 \rightarrow \neg(PO \wedge ND \wedge IIA)$.

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.

7. RELATED WORK

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 [6, 7, 8].

A notable and recent exception is a logical framework for judgment aggregation developed by Marc Pauly in [10], 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: in [10], 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* [11] 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.

8. DISCUSSION

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[3].

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.

In future work it would 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.

9. ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful remarks. Thomas Ågotnes' work on this paper was supported by grants 166525/V30 and 176853/S10 from the Research Council of Norway.

10. REFERENCES

[1] K. J. Arrow. *Social Choice and Individual Values*. Wiley, 1951.

[2] K. J. Arrow, A. K. Sen, and K. Suzumura, eds. *Handbook of Social Choice and Welfare*, volume 1. North-Holland, 2002.

[3] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.

[4] E. M. Clarke, O. Grumberg, and D. A. Peled. *Model Checking*. The MIT Press: Cambridge, MA, 2000.

[5] F. Dietrich and C. List. Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, 2006. Forthcoming.

[6] C. Lafage and J. Lang. Logical representation of preferences for group decision making. In *Proceedings of the Conference on Principles of Knowledge Representation and Reasoning (KR-00)*, pages 457–470. Morgan Kaufman, 2000.

[7] J. Lang. From preference representation to combinatorial vote. *Proceedings of the Eighth International Conference on Principles and Knowledge Representation and Reasoning (KR-02)*, pages 277–290. Morgan Kaufmann, 2002.

[8] J. Lang. Logical preference representation and combinatorial vote. *Ann. Math. Artif. Intell*, 42(1-3):37–71, 2004.

[9] C. H. Papadimitriou. *Computational Complexity*. Addison-Wesley: Reading, MA, 1994.

[10] M. Pauly. Axiomatizing collective judgment sets in a minimal logical language, 2006. Manuscript.

[11] Y. Venema. A crash course in arrow logic. In M. Marx, M. Masuch, and L. Polos, editors, *Arrow Logic and Multi-Modal Logic*, pages 3–34. CSLI Publications, Stanford, 1996.

[3]Dietrich and List [5] 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.