



## UvA-DARE (Digital Academic Repository)

### Complexity of judgment aggregation: safety of the agenda

Endriss, U.; Grandi, U.; Porello, D.

**Publication date**

2010

**Document Version**

Final published version

**Published in**

AAMAS 2010

[Link to publication](#)

**Citation for published version (APA):**

Endriss, U., Grandi, U., & Porello, D. (2010). Complexity of judgment aggregation: safety of the agenda. In W. van der Hoek, G. A. Kaminka, Y. Lespérance, M. Luck, & S. Sen (Eds.), *AAMAS 2010: the 9th International Conference on Autonomous Agents and Multiagent Systems, May 10-14, 2010, Toronto, Canada : conference proceedings* (Vol. 1, pp. 359-366). IFAAMAS. <https://dl.acm.org/citation.cfm?id=1838257>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Complexity of Judgment Aggregation: Safety of the Agenda

Ulle Endriss, Umberto Grandi and Daniele Porello  
Institute for Logic, Language and Computation  
University of Amsterdam

## ABSTRACT

Aggregating the judgments of a group of agents regarding a set of interdependent propositions can lead to inconsistent outcomes. One of the parameters involved is the agenda, the set of propositions on which agents are asked to express an opinion. We introduce the problem of checking the safety of the agenda: for a given agenda, can we guarantee that judgment aggregation will never produce an inconsistent outcome for any aggregation procedure satisfying a given set of axioms? We prove several characterisation results, establishing necessary and sufficient conditions for the safety of the agenda for different combinations of the most important axioms proposed in the literature, and we analyse the computational complexity of checking whether a given agenda satisfies these conditions.

## Categories and Subject Descriptors

I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence—*Multiagent Systems*; J.4 [Social and Behavioral Sciences]: Economics

## General Terms

Theory, Economics

## Keywords

Judgment Aggregation, Computational Social Choice

## 1. INTRODUCTION

*Judgment aggregation* (JA) is a branch of social choice theory that studies the properties of procedures for amalgamating individual judgments on a set of related propositions of the members of a group into a collective judgment reflecting the views of that group as a whole [9]. For example, the propositions to be decided upon might be “trading patterns are unusual” ( $p$ ), “if trading patterns are unusual, then raise the alarm” ( $p \rightarrow q$ ), and “raise the alarm” ( $q$ ). A possible aggregation procedure is the *majority rule*, which accepts a proposition for the collective judgment set if a majority of the individual agents do. Unfortunately, this can lead to a *paradox* (also known as the *discursive dilemma* [8]), as first

**Cite as:** Complexity of Judgment Aggregation: Safety of the Agenda, U. Endriss, U. Grandi and D. Porello, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. 359–366  
Copyright © 2010, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

observed in the literature on legal theory [7]. Suppose there are three agents making judgments:

	$p$	$p \rightarrow q$	$q$
Agent 1:	Yes	Yes	Yes
Agent 2:	No	Yes	No
Agent 3:	Yes	No	No
Majority:	Yes	Yes	No

Each individual agent’s judgment set is *consistent* (e.g., agent 3 does accept  $p$ , but disagrees with the rule  $p \rightarrow q$ , so can consistently reject  $q$ ), but the collective judgment set derived using the majority rule is not.

The literature on JA has largely developed in outlets associated with Philosophy, Economic Theory, Political Science, and Logic, but recently JA has also come to be recognised as being highly relevant to Artificial Intelligence (AI) and Multiagent Systems (MAS). The reasons are clear: in a multiagent system, different autonomous software agents may have different “opinions” on the same issues (maybe due to a difference in access to the relevant information, or due to different reasoning capabilities), and some joint course of action needs to be extracted from these diverse views. Indeed, in AI, the related problem of *belief merging* [6] has been studied for some time, and there are interesting parallels between that literature and JA [13].

Given the relevance of JA to MAS, it is important to understand its computational aspects. However, to date, these have only received little attention in the literature. This can of course be explained by the origins of the field in Law, Economics, and Philosophy. As a first step towards bridging this gap, in this paper, we set out to analyse the *computational complexity* of an important problem arising in JA (to the best of our knowledge, this is the first such attempt). This approach is inspired by the very successful research programme of applying tools from complexity theory to the domain of *voting*, pursued in the area of Computational Social Choice (see Chevaleyre et al. [2] and Faliszewski et al. [5] for partial surveys of this line of work).

The problem we analyse is what we call the problem of the *safety of the agenda* (SoA). The *agenda* is the set of propositions on which agents are asked to make a judgment. Whether or not a paradox arises depends on the aggregation procedure used, the agenda, and the actual judgments made by the individual agents. In its simplest form, the SoA problem asks: for a given aggregation procedure  $F$  and a given agenda  $\Phi$ , can we guarantee consistency of the collective judgment set, independently from the individual (consistent) judgments made by the agents? Because JA is plagued with paradoxes and impossibility results, there is no

single preferred aggregation procedure. Instead, a number of *axioms* have been formulated in the literature, expressing desiderata for attractive procedures, such as *anonymity*, *neutrality*, *independence*, and *monotonicity* [8, 9]. In its general form, the SoA problem asks: for a given agenda  $\Phi$ , can we guarantee consistency of the collective judgment set, for any aggregation procedure  $F$  satisfying a given set of axioms? As we shall discuss (in Section 3.1), this is related to, but subtly different from, questions previously analysed in the literature. SoA is a critical issue in MAS, because it tells us whether it is feasible to expect that a group of autonomous agents will be able to come to a consistent agreement, if we want the procedures they use to satisfy certain desirable axioms and if the set of issues they have to decide upon comes with a certain level of structural richness. The computational complexity of SoA matters in scenarios where agendas can be large and safety needs to be checked often.

The remainder of this paper is organised as follows. In Section 2, we review the framework of JA, including in particular the axioms we shall be working with. We also state a number of *representation results* (relating axioms to types of aggregation procedures), most of which are implicit in the existing literature, but have rarely been stated formally. In Section 3, we define the SoA problem and prove a number of *characterisation results* that establish necessary and sufficient conditions for an agenda to be safe for any procedure satisfying various combinations of axioms. In Section 4, we establish the *complexity* of checking these conditions: they turn out to be  $\Pi_2^P$ -complete. Section 5 concludes.

## 2. JUDGMENT AGGREGATION

In this section, we define the model of JA we shall be working with, and we review various concepts from the literature. We introduce some new terminology to shed light on the difference between the “syntactic” and “logical” properties of a judgment set, a difference that we believe is worth stressing. All our definitions are closely related to existing ones, resulting in a framework that is essentially equivalent to the version given by List and Puppe [9]. We also introduce a list of axioms specifying desirable properties for an aggregation procedure, and we define various classes of aggregators combining different sets of axioms. In the last part of the section we study these classes in more details, finding for each of them a uniform mathematical representation.

### 2.1 Basic Definitions

Let  $PS$  be a set of propositional variables, and  $\mathcal{L}_{PS}$  the set of propositional formulas built from  $PS$  (using the usual connectives  $\neg$ ,  $\wedge$ ,  $\vee$ ,  $\rightarrow$ ,  $\leftrightarrow$ , and the constants  $\top$  and  $\perp$ ). If  $\alpha$  is a propositional formula, define  $\sim\alpha$ , the *complement* of  $\alpha$ , as  $\neg\alpha$  if  $\alpha$  is not negated, and as  $\beta$  if  $\alpha = \neg\beta$ .

DEFINITION 1. An **agenda** is a finite nonempty set  $\Phi \subseteq \mathcal{L}_{PS}$  not containing any doubly-negated formulas that is closed under complementation (i.e., if  $\alpha \in \Phi$  then  $\sim\alpha \in \Phi$ ).

In a slight departure from the common definition in the literature [9], note that we *do* allow for tautologies and contradictions in the agenda. Our reason for relaxing the framework in this manner is that one of our interests here is in the complexity of JA, and recognising a tautology or a contradiction is itself a computationally intractable problem.

DEFINITION 2. A **judgment set**  $J$  on an agenda  $\Phi$  is a subset of the agenda  $J \subseteq \Phi$ .

We call a judgment set  $J$ :

- **complete** if  $\alpha \in J$  or  $\sim\alpha \in J$  for all  $\alpha \in \Phi$ ;
- **complement-free**<sup>1</sup> if for all  $\alpha \in \Phi$  it is not the case that both  $\alpha$  and its complement are in  $J$ ;
- **consistent** if there exists an assignment that makes all formulas in  $J$  true.

Denote with  $J(\Phi)$  the set of all complete consistent subsets of  $\Phi$ . Given a set  $N = \{1, \dots, n\}$  of  $n \geq 3$  *individuals* (or *agents*), denote with  $\mathbf{J} = (J_1, \dots, J_n)$  a *profile* of judgment sets, one for each individual.

DEFINITION 3. An **aggregation procedure** for agenda  $\Phi$  and a set of  $n$  individuals is a function  $F : J(\Phi)^n \rightarrow \mathcal{P}(\Phi)$ .

That is, an aggregation procedure maps any profile of individual judgment sets to a single collective judgment set (an element of the powerset of  $\Phi$ ). Since  $F$  is defined on the set of all profiles of consistent and complete judgment sets, we are already assuming a *universal domain*, which is sometimes stated as a separate property [8]. The definition also includes a condition of *individual rationality*: all individual judgment sets are complete and consistent.<sup>2</sup>

### 2.2 Desiderata for Aggregation Procedures

We did not yet put any constraints on the collective judgment set, the outcome of aggregation. This is the role of the following properties. An aggregation procedure  $F$ , defined on an agenda  $\Phi$ , is said to be:

- **complete** if  $F(\mathbf{J})$  is complete for every  $\mathbf{J} \in J(\Phi)$ ;
- **complement-free** if  $F(\mathbf{J})$  is complement-free for every  $\mathbf{J} \in J(\Phi)$ ;
- **consistent** if  $F(\mathbf{J})$  is consistent for every  $\mathbf{J} \in J(\Phi)$ ;
- **null** if  $\Phi$  includes a contradiction  $\varphi^\perp$  and  $\varphi^\perp \in F(\mathbf{J})$  for every profile  $\mathbf{J} \in J(\Phi)$ .

We now present several *axioms* to provide a normative framework in which to state what the desirable (or essential) properties of an acceptable aggregation procedure should be. The first axiom is a very basic requirement, restricting possible aggregators  $F$  in terms of fundamental properties of the outcomes they produce.

**Weak Rationality (WR):**  $F$  is non-null, it is complete, and it is complement-free.

This condition differs from the notion of *collective rationality* often used in the literature [9], because we do not require the collective judgment set to be consistent. The first reason to separate the notion of consistency from the other conditions is that the requirements of (WR) are purely syntactic notions that can be checked automatically in an easy way. The second is that the notion of consistency is not intrinsic to the aggregation function, but depends more on the properties of the agenda. This will be made more precise in Section 3, where we will study the consistency of a class of aggregators depending on the agenda.

The following are the most important axioms for JA discussed in the literature [8, 9]:

<sup>1</sup>This property is called *weak consistency* by Dietrich [3], and *consistency* by List and Pettit [8]. Our choice of terminology is intended to stress the fact that it is a purely syntactic notion, not involving any model-theoretic concept.

<sup>2</sup>Following List and Puppe [9], we will not postulate that individual judgment sets be *deductively closed*, a property already entailed by the two assumptions we do make.

**Unanimity (U):** If  $\varphi \in J_i$  for all  $i$  then  $\varphi \in F(\mathbf{J})$ .

**Anonymity (A):** For any profile  $\mathbf{J}$  and any permutation  $\sigma : N \rightarrow N$  we have  $F(J_1, \dots, J_n) = F(J_{\sigma(1)}, \dots, J_{\sigma(n)})$ .

**Neutrality (N):** For any  $\varphi, \psi$  in the agenda  $\Phi$  and profile  $\mathbf{J} \in J(\Phi)$ , if for all  $i$  we have that  $\varphi \in J_i \Leftrightarrow \psi \in J_i$ , then  $\varphi \in F(\mathbf{J}) \Leftrightarrow \psi \in F(\mathbf{J})$ .

**Independence (I):** For any  $\varphi$  in the agenda  $\Phi$  and profiles  $\mathbf{J}$  and  $\mathbf{J}'$  in  $J(\Phi)$ , if  $\varphi \in J_i \Leftrightarrow \varphi \in J'_i$  for all  $i$ , then  $\varphi \in F(\mathbf{J}) \Leftrightarrow \varphi \in F(\mathbf{J}')$ .

**Systematicity (S):** For any  $\varphi, \psi$  in the agenda  $\Phi$  and profiles  $\mathbf{J}$  and  $\mathbf{J}'$  in  $J(\Phi)$ , if  $\varphi \in J_i \Leftrightarrow \psi \in J'_i$  for all  $i$ , then  $\varphi \in F(\mathbf{J}) \Leftrightarrow \psi \in F(\mathbf{J}')$ .

Unanimity expresses the idea that if all individuals accept a given judgment, then so should the collective. Anonymity states that aggregation should be symmetric with respect to individuals, i.e., all individuals should be treated the same. Neutrality is a symmetry requirement for propositions: if the same subgroup accepts two propositions, then either both or neither should be collectively accepted. Independence says that if a proposition is accepted by the same subgroup under two otherwise distinct profiles, then that proposition should be accepted either under both or under neither profile. Systematicity is satisfied if and only if both neutrality and independence are. While all of these axioms are intuitively appealing, they are stronger than they may seem at first, and several *impossibility theorems*, establishing inconsistencies between certain combinations of axioms with other desiderata, have been proved in the literature. The original impossibility theorem of List and Pettit [8], for instance, shows that there can be no collectively rational aggregation procedure satisfying (A) and (S).

A further important property is monotonicity. We introduce two different axioms for monotonicity. The first is the one commonly used in the literature [4, 9]. It implicitly relies on the independence axiom. The second, which to the best of our knowledge has not been formulated before, is designed to be applied to neutral procedures. For systematic procedures the two formulations are equivalent.

**I-Monotonicity (M<sup>I</sup>):** For any  $\varphi$  in the agenda  $\Phi$  and profiles  $\mathbf{J} = (J_1, \dots, J_i, \dots, J_n)$  and  $\mathbf{J}' = (J_1, \dots, J'_i, \dots, J_n)$  in  $J(\Phi)$ , if  $\varphi \notin J_i$  and  $\varphi \in J'_i$ , then  $\varphi \in F(\mathbf{J}) \Rightarrow \varphi \in F(\mathbf{J}')$ .

**N-Monotonicity (M<sup>N</sup>):** For any  $\varphi, \psi$  in the agenda  $\Phi$  and profile  $\mathbf{J}$  in  $J(\Phi)$ , if  $\varphi \in J_i \Rightarrow \psi \in J_i$  for all  $i$  and  $\varphi \notin J_k$  and  $\psi \in J_k$  for some  $k$ , then  $\varphi \in F(\mathbf{J}) \Rightarrow \psi \in F(\mathbf{J})$ .

That is, (M<sup>I</sup>) expresses that if  $\varphi$  is collectively accepted and receives additional support (from  $i$ ), then it should continue to be collectively accepted. Axiom (M<sup>N</sup>) says that if  $\varphi$  is collectively accepted and  $\psi$  is accepted by a strict superset of the individuals accepting  $\varphi$ , then  $\psi$  should also be collectively accepted.

The axioms we have introduced can be used to define different *classes* of aggregation procedures: Given an agenda  $\Phi$  and a list of desirable properties AX provided in the form of axioms, we define  $\mathcal{F}_\Phi[\text{AX}]$  to be the set of all procedures  $F : J(\Phi)^n \rightarrow \mathcal{P}(\Phi)$  that satisfy the axioms in AX.

## 2.3 Representation Results

We now introduce and study various classes of procedures, specified in terms of the set of axioms they satisfy. Some combinations of axioms will be proved to be axiomatisations of the same procedures, and for all natural combinations of axioms we will find a uniform representation. A first interesting result concerning systematicity is the following:

LEMMA 1. *If an agenda  $\Phi$  contains a tautology, then every aggregation procedure for  $\Phi$  that satisfies (WR) and (S) is unanimous (U).*

PROOF. Let  $\varphi^\top$  be the tautology contained in  $\Phi$ , and let  $\psi$  be a formula unanimously accepted in some profile  $\mathbf{J}$ , i.e.,  $\psi \in J_i$  for all  $i$ . As the procedure is non-null, there exists a profile  $\mathbf{J}'$  where  $\varphi^\top \in F(\mathbf{J}')$ . Now  $\psi \in J_i \Leftrightarrow \varphi^\top \in J'_i$ , since every individual judgment set must contain  $\varphi^\top$ , because they are all complete and consistent. But then, by systematicity, we have that  $\psi \in F(\mathbf{J}) \Leftrightarrow \varphi^\top \in F(\mathbf{J}')$ , and since  $\varphi^\top \in F(\mathbf{J}')$  this proves that  $\psi \in F(\mathbf{J})$ .  $\square$

Lemma 1 suggests that unanimity is not that powerful an axiom when the agenda can contain tautologies. This does not render the axiom redundant, but arguably less interesting than the other axioms, and we therefore shall not focus on unanimity in the remainder of this paper.

The next lemma shows that for anonymous and neutral procedures acceptance of a formula depends solely on the number of individuals accepting it. This is a known result; List and Pettit [8], for instance, use this insight in the proof of their impossibility theorem.

We require some further notation: let  $N_\varphi^{\mathbf{J}} = \{i \mid \varphi \in J_i\}$  be the set of individuals accepting  $\varphi$  in a given profile  $\mathbf{J}$ .

LEMMA 2. *If an aggregation procedure  $F$  satisfies (A) and (N), then  $|N_\varphi^{\mathbf{J}}| = |N_\psi^{\mathbf{J}}|$  entails  $\varphi \in F(\mathbf{J}) \Leftrightarrow \psi \in F(\mathbf{J})$ .*

The converse also holds. We immediately get the following, somewhat surprising, result as a corollary:

PROPOSITION 3. *If the number of individuals is even, then there exists no aggregation procedure that satisfies (WR), (A) and (N).<sup>3</sup>*

PROOF. Let  $\varphi$  be a formula and let  $\mathbf{J}$  be a profile such that exactly half of the individuals accept  $\varphi$  and the other half reject it. By completeness, either  $\varphi$  or  $\sim\varphi$  must be in  $F(\mathbf{J})$ , and, by Lemma 2,  $\varphi \in F(\mathbf{J}) \Leftrightarrow \sim\varphi \in F(\mathbf{J})$ , in contradiction with the property of being complement-free.  $\square$

As a second consequence, we can easily prove a representation result for the class  $\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{S}]$ :

PROPOSITION 4. *If the number of individuals is odd, then an aggregation procedure  $F$  satisfies (WR), (A) and (S) if and only if there exists a function  $h : \{0, \dots, |N|\} \rightarrow \{0, 1\}$ , with the property that  $h(i) = 1 - h(|N| - i)$  for all  $i \leq |N|$ , such that  $\varphi \in F(\mathbf{J}) \Leftrightarrow h(|N_\varphi^{\mathbf{J}}|) = 1$ .*

PROOF. By Lemma 2 we know that the acceptance or rejection of a formula depends solely of the number of individuals accepting it. Since systematicity includes independence, we know that these values of acceptance do not depend

<sup>3</sup>It is important to stress that we are not requiring any notion of logical consistency in this statement.

on the profile we are considering. It is therefore sufficient to specify these values with a function  $h : \{0, \dots, |N|\} \rightarrow \{0, 1\}$ . Using the hypothesis of individual rationality and (WR), we get that one of  $h(i)$  and  $h(|N| - i)$  must be equal to 1, so as to guarantee completeness of the collective judgment set; and whenever  $h(i) = 1$  then  $h(|N| - i) = 0$  by the requirement to be complement-free.  $\square$

If we add monotonicity to this set of axioms we get an axiomatisation of the *majority rule* (accepting a proposition if and only if a strict majority of the individuals do):

PROPOSITION 5. *If the number of individuals is odd, an aggregation procedure  $F$  satisfies (WR), (A), (S) and  $(M^I)$  if and only if  $F$  is the majority rule.*

PROOF. By Proposition 4, every  $F$  satisfying (WR), (A) and (S) is determined by a function  $h : \{0, \dots, |N|\} \rightarrow \{0, 1\}$  with the constraint that  $h(i) = h(|N| - i)$  for all  $i \leq |N|$ . If we require also monotonicity, it is easy to see that whenever  $h(i) = 1$ , then for every  $j \geq i$  we have  $h(j) = 1$ . Call  $k$  the minimum  $i$  such that  $h(i) = 1$ . Since for (WR) the function  $F$  must be complete, we get that  $k \leq \frac{|N|+1}{2}$ , given that otherwise there are profiles that lead to incomplete judgment sets. Since  $F$  has also to be complement-free, we get  $k \geq \frac{|N|+1}{2}$ , to avoid acceptance of a formula and its negation. Thus,  $k = \frac{|N|+1}{2}$  and  $F$  is the majority rule.  $\square$

Proposition 5 continues to hold if we weaken systematicity (S) to neutrality (N), replacing  $(M^I)$  with  $(M^N)$ .

Representation results along the lines of Proposition 4 are easy to obtain for various classes of procedures, and we summarise them in Table 1. We have chosen to focus on a number of different weakenings of the axiomatisation of the majority rule. We never drop the anonymity axiom (A), because we find it very appealing for JA. On the other hand, we do consider the case of only neutral or independent aggregation procedures, as a form of weakening systematicity. We usually also consider the axiom (WR) indispensable, but for one class of procedures we do drop (WR) to study the interesting class of *uniform quota rules*, introduced and axiomatised by Dietrich and List [4]:

DEFINITION 4. *Given some  $m \in \{0, \dots, |N|+1\}$  and an agenda  $\Phi$ , the **uniform quota rule with quota  $m$**  is the aggregation procedure  $F_m$  such that  $\varphi \in F_m(\mathbf{J}) \Leftrightarrow |N_\varphi^J| \geq m$ .*

### 3. CHARACTERISATION RESULTS

In this section, we introduce the concept of safety of the agenda (SoA): an agenda is safe for a class of aggregation procedures, if consistency is guaranteed for every procedure in that class. To the best of our knowledge, this is a new concept in the literature on JA, even though it captures a central problem in the application-driven study of the subject. We will characterise safe agendas for all the classes of procedures identified in Table 1, paving the way for a study of the computational complexity of the problem.

#### 3.1 Safety of the Agenda: Problem Definition

An important set of results in the literature on JA are possibility theorems, sometimes called “characterisation results” [11, 9]. Given some axioms as desiderata for the aggregation procedure (always including consistency), such a

Axioms	Representation
$\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{S}, \text{M}^I]$ $\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{N}, \text{M}^N]$	Majority rule
$\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{S}]$	$h : \{0, \dots,  N \} \rightarrow \{0, 1\}$ with the property $h(i) = 1 - h( N  - i)$ s.t. $\varphi \in F(\mathbf{J}) \Leftrightarrow h( N_\varphi^J ) = 1$
$\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{N}]$	$g : J(\Phi)^n \rightarrow \{h : \{0, \dots,  N \} \rightarrow \{0, 1\}\}$ with the property $h(i) = 1 - h( N  - i)$ s.t. $\varphi \in F(\mathbf{J}) \Leftrightarrow g(\mathbf{J})( N_\varphi^J ) = 1$
$\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{I}]$	$g : \Phi \rightarrow \{h : \{0, \dots,  N \} \rightarrow \{0, 1\}\}$ with $g(\varphi)(i) = 1 - g(\neg\varphi)( N  - i)$ s.t. $\varphi \in F(\mathbf{J}) \Leftrightarrow g(\varphi)( N_\varphi^J ) = 1$
$\mathcal{F}_\Phi[\text{A}, \text{S}, \text{M}^I]$	Uniform quota rules

Table 1: Representation Results

possibility theorem characterises agendas where these conditions are satisfiable. Despite their theoretical interest, results of this form are somewhat less relevant for applications. The reason is that actual users are more likely to want an assurance that aggregation *will* be safe (provided certain axioms are satisfied and the agenda has certain properties) rather than learn that there *exists* a safe form of aggregation (satisfying certain axioms). Moreover, in view of the stress we have put on the distinction between “logical” and “syntactic” properties of an aggregation procedure and the collective judgment set it produces, a thorough study of the consistency of a class of procedures depending on the agenda is of immediate relevance. We therefore introduce the following concept:

DEFINITION 5. *An agenda  $\Phi$  is **safe with respect to a class of aggregation procedures  $\mathcal{F}$** , if every procedure in  $\mathcal{F}$  is consistent when applied to judgment sets over  $\Phi$ .*

The example for a *discursive dilemma* presented in the introductory section demonstrates the *unsafety* of the agenda  $\{p, \neg p, q, \neg q, p \rightarrow q, \neg(p \rightarrow q)\}$  with respect to the majority rule. We can give a general formulation of a discursive dilemma for an aggregation procedure  $F$ , defined on an agenda  $\Phi$ , as a tuple  $(J_1, \dots, J_n, J_{n+1})$  of judgment sets, such that  $J_1, \dots, J_n$  are individual sets of judgments,  $J_{n+1} = F(J_1, \dots, J_n)$ , and  $J_{n+1}$  is inconsistent. Definition 5 says that  $\Phi$  is safe for the class of procedures  $\mathcal{F}$  if no procedure in that class generates a discursive dilemma.

#### 3.2 Agenda Properties

While the characterisation results (i.e., possibility theorems) in the literature address a different issue than the one we are interested in here, some of the properties of agendas defined in that context are still potentially useful for our purposes. One of these is the so-called *median property*, which we shall define next. Later, we will use this property, and some of its variants, to characterise agendas that are safe for certain classes of aggregation procedures.<sup>4</sup>

We call an inconsistent set  $\Delta$  *nontrivially inconsistent* if there is no single proposition  $\varphi^\perp \in \Delta$  that is a contradiction.

<sup>4</sup>Other agenda properties defined in the literature, such as *total blockedness* or the *even-number-negation property* [9] turn out not to be relevant for our purposes.

DEFINITION 6. We say that an agenda  $\Phi$  satisfies the **median property** (MP), if every nontrivially inconsistent subset of  $\Phi$  has itself an inconsistent subset of size 2.

The name of this property, introduced by Nehring and Puppe [11], derives from a property of the set of all judgment sets  $J(\Phi)$  viewed as a subset of a particular vector space. The typical phrasing of this property in the literature is that an agenda satisfies the median property if all minimally inconsistent subsets of  $\Phi$  have size 2. For agendas without tautologies the two formulations are equivalent. In our case, we have to include an additional check of nontriviality in case there is a contradictory formula in the agenda. We can generalise the median property as follows:

DEFINITION 7. An agenda  $\Phi$  satisfies the **k-median property** (kMP) for  $k \geq 2$ , if every inconsistent subset of  $\Phi$  has itself an inconsistent subset of size at most  $k$ .

Observe that we have dropped the restriction to nontrivially inconsistent sets in Definition 7, because for trivially inconsistent sets it is always the case that there is an inconsistent subset of size at most  $k$  (namely one of size 1). The MP of Definition 6 and the 2MP are the same property.

Agendas satisfying the MP are already quite simple, but the restriction can be made tighter by requiring all inconsistent subsets to have a particular form:

DEFINITION 8. An agenda  $\Phi$  satisfies the **simplified median property** (SMP), if every nontrivially inconsistent subset of  $\Phi$  has itself an inconsistent subset of the form  $\{\varphi, \psi\}$  with  $\models \varphi \leftrightarrow \neg\psi$ .

A further simplification yields:

DEFINITION 9. An agenda  $\Phi$  satisfies the **syntactic simplified median property** (SSMP), if every nontrivially inconsistent subset of  $\Phi$  has itself an inconsistent subset of the form  $\{\varphi, \neg\varphi\}$ .

Agendas satisfying the SSMP are composed of uncorrelated formulas, i.e., they are essentially equivalent to agendas composed of atoms alone. The SMP is less restrictive, allowing for logically equivalent but syntactically different formulas.

Observe that every agenda that satisfies the SMP also satisfies the MP. The converse is not true:  $\Phi = \{p, \neg p, p \wedge q, \neg(p \wedge q)\}$  satisfies the MP, but not the SMP. Similarly, every agenda that satisfies the SSMP also satisfies the SMP. Again, the converse is not true:  $\Phi = \{p, \neg p, p \wedge p, \neg(p \wedge p)\}$  satisfies the SMP, but not the SSMP.

### 3.3 Linking Agenda Properties and Axioms

We now prove several characterisation results for the safe aggregation of judgments. For all classes of aggregation procedures introduced in Section 2.3 we will give necessary and sufficient conditions for an agenda to be safe on that class. The first theorem is familiar from the literature [11], although it is presented there in a different formulation.

THEOREM 6. An agenda  $\Phi$  is safe for the majority rule<sup>5</sup> if and only if  $\Phi$  satisfies the MP.

<sup>5</sup>Recall that we have seen two alternative axiomatisations of this “class” consisting of just one procedure (see Table 1).

The theorem is a direct consequence of a result proved by Nehring and Puppe [11] (see Theorem 3 in the survey by List and Puppe [9] for a formulation in the framework of JA). In that work, the authors show that if the number of individuals is odd, under the assumption of collective rationality (WR plus consistency), monotonicity, unanimity, systematicity, and anonymity, there exists an aggregation procedure on agenda  $\Phi$  if and only if  $\Phi$  satisfies the median property. The witness they give as a consistent aggregator is nothing other than the majority rule. Since Theorem 6 speaks of a “class” consisting only of a single procedure, namely the majority rule, the concept of safety of the agenda and the kind of concept inherent in a possibility theorem coincide and our result is a direct consequence of theirs. Unfortunately, the same kind of approach cannot be used to adapt other possibility theorems available in the literature, because the classes of procedures we consider in the sequel each contain more than just a single procedure.

THEOREM 7. An agenda  $\Phi$  is safe for  $\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{S}]$  if and only if  $\Phi$  satisfies the SMP.

PROOF. ( $\Leftarrow$ ) Suppose that  $\Phi$  satisfies the SMP and suppose, for the sake of contradiction, that there exists a profile  $\mathbf{J}$  such that  $F(\mathbf{J})$  is inconsistent. Note first that  $F(\mathbf{J})$  cannot be trivially inconsistent: if there is a contradiction  $\varphi^\perp$  in the agenda, then Lemma 1 implies that every function in  $\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{S}]$  is unanimous, thus accepting  $\sim\varphi^\perp$  in every profile, which would contradict the assumption that  $F$  is complement-free (part of WR). Therefore,  $F(\mathbf{J})$  contains a minimally inconsistent subset of size 2 of the form  $\{\varphi, \neg\psi\}$  with  $\models \varphi \leftrightarrow \psi$ . Now, since every individual judgment set is consistent, we have that  $\varphi \in J_i \Leftrightarrow \psi \in J_i$  for all  $i \in N$ , which implies, by systematicity, that  $\varphi \in F(\mathbf{J}) \Leftrightarrow \psi \in F(\mathbf{J})$ . As we have  $\{\varphi, \neg\psi\} \subseteq F(\mathbf{J})$ , this entails  $\psi \in F(\mathbf{J})$ , which is a contradiction, since the outcome must be complement-free.

( $\Rightarrow$ ) For the other direction, suppose that  $\Phi$  violates the SMP, i.e., there exists a nontrivially inconsistent subset that does not contain two formulas one of which is equivalent to the negation of the other. This set must contain a minimally inconsistent subset, which we shall call  $X$ . In case  $X$  has size  $\geq 3$ , also the MP will be violated and, by Theorem 6, the majority rule will generate a discursive dilemma.<sup>6</sup> In case  $X$  has size 2, it must be of the form  $\{\varphi, \psi\}$  with  $\varphi \models \neg\psi$  but  $\neg\psi \not\models \varphi$ . Consider then the following aggregation procedure for 3 individuals, defined with the notation used in Proposition 4:  $h(0) = h(1) = 1$  and  $h(2) = h(3) = 0$ .  $F_h$  accepts a proposition only if it is accepted by 0 or 1 individual. Consider the following profile, restricted to  $\varphi$  and  $\psi$  and their complements:  $J_1 = \{\sim\varphi, \sim\psi\}$ ,  $J_2 = \{\varphi, \sim\psi\}$ ,  $J_3 = \{\sim\varphi, \psi\}$ . (Note that each of these sets is consistent.) This profile (opportunistically extended to a profile on the whole agenda) will generate an inconsistent outcome, since both  $\varphi$  and  $\psi$  are accepted by only one of the individuals. This proves that when the SMP is violated there always exists a function satisfying (WR), (S) and (A) that generates an inconsistent outcome.  $\square$

With similar arguments we can prove the following:

THEOREM 8. An agenda  $\Phi$  is safe for  $\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{N}]$  if and only if  $\Phi$  satisfies the SMP and does not contain a contradictory formula.

<sup>6</sup>This, in turn, implies that the agenda is not safe on the class  $\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{S}]$ , which includes the majority rule.

Observe that if  $\Phi$  does contain a contradiction  $\varphi^\perp$ , then there always exists a non-null neutral procedure that rejects  $\varphi^\perp$  in some profile and accepts it elsewhere, making every such agenda trivially unsafe. For the rest of the proof it is sufficient to note that the assumption of systematicity in the first part of the previous proof can be relaxed to neutrality, and that the left-to-right direction of Theorem 7 entails the analogous direction for Theorem 8.

A more peculiar, though even more restricting, characterisation result is the following:

**THEOREM 9.** *An agenda  $\Phi$  is safe for  $\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{I}]$  if and only if  $\Phi$  satisfies the SSMP.*

**PROOF.** ( $\Leftarrow$ ) Suppose  $\Phi$  satisfies the SSMP. If there exists a profile  $\mathbf{J}$  such that  $F(\mathbf{J})$  is inconsistent, we can assume that  $F(\mathbf{J})$  is nontrivially inconsistent, because the independence axiom implies that every non-null procedure rejects contradictory formulas in every profile. The SSMP now tells us that there must exist a formula  $\varphi \in \Phi$  such that  $\{\varphi, \neg\varphi\} \subseteq F(\mathbf{J})$ , in contradiction with the property of being complement-free. (Note that if an agenda does not contain contradictions and satisfies the SSMP, then *any* weakly-rational procedure is consistent.)

( $\Rightarrow$ ) The fact that  $\Phi$  does not satisfy the SSMP is equivalent to the existence of two distinct formulas  $\varphi$  and  $\psi$  in  $\Phi$  such that  $\varphi \models \psi$ . Consider then the constant function that accepts  $\varphi$  and rejects  $\psi$  in every profile: this is clearly a weakly-rational, independent, and anonymous function, and it generates for every profile an inconsistent outcome.  $\square$

The only class of procedures listed in Table 1 not yet covered is  $\mathcal{F}_\Phi[\text{A}, \text{S}, \text{M}^I]$ , corresponding to the uniform quota rules. Here, a characterisation result of the kind we seek is available in the literature for certain subclasses of  $\mathcal{F}_\Phi[\text{A}, \text{S}, \text{M}^I]$ , namely uniform quota rules with a specific bound on the quota [4]. We state this interesting result as follows:

**THEOREM 10.** *Let  $k \geq 2$ . An agenda  $\Phi$  is safe for the class of uniform quota rules  $F_m$  for  $n$  individuals satisfying  $m > n - \frac{n}{k}$  if and only if  $\Phi$  satisfies the  $k\text{MP}$ .*

Theorem 10 is a reformulation of Corollary 2(a) in the work of Dietrich and List [4], and we shall not prove it here.

## 4. COMPLEXITY RESULTS

In this section, we establish the complexity of deciding whether an agenda satisfies the median property (or one of its variants), and we use these results to show that checking the safety of an agenda is  $\Pi_2^P$ -complete for several classes of aggregators, each characterised by a combination of the most important axioms for JA discussed in the literature.

### 4.1 Background: Complexity Theory

We shall assume familiarity with the basics of complexity theory up to the notion of NP-completeness (helpful introductions include the textbooks by Papadimitriou [12] and by Arora and Barak [1]).

We will work with  $\Pi_2^P$  (also known as  $\text{coNP}^{\text{NP}}$  or “coNP with an NP oracle”), a complexity class located at the second level of the polynomial hierarchy [1, 12]. This is the class of decision problems for which a negative answer can be computed in polynomial time by a nondeterministic machine that has access to an oracle for answering queries to SAT

(or any other NP-complete problem). To prove a problem  $\Pi_2^P$ -complete, we have to prove both membership in  $\Pi_2^P$  and  $\Pi_2^P$ -hardness. To prove membership, we need to provide an algorithm that, when provided with a *certificate* intended to establish a negative answer, can verify the correctness of that certificate in polynomial time, if we assume that the algorithm has access to a SAT-oracle.

The main challenge is typically to prove hardness. This can be done by giving a polynomial-time reduction from a problem that is already known to be  $\Pi_2^P$ -hard to the problem under investigation. For this purpose, we will make use of *quantified boolean formulas* (QBFs). While QSAT, the satisfiability problem for general QBFs, is PSPACE-complete, by imposing suitable syntactic restrictions we can generate complete problems for any level of the polynomial hierarchy. Consider a QBF of the following form:

$$\forall x_1 \cdots x_r \exists y_1 \cdots y_s. \varphi(x_1, \dots, x_r, y_1, \dots, y_s)$$

Here  $\varphi$  is an arbitrary propositional formula and  $\{x_1, \dots, x_r\} \cup \{y_1, \dots, y_s\}$  is the set of all propositional variables occurring in  $\varphi$  (that is, above could be any QBF for which any existential quantifiers occur inside the scope of all universal quantifiers). The problem of checking whether a formula of this form is *satisfiable* (i.e., *true*), which we shall denote  $\text{coQSAT}_2$ , is known to be  $\Pi_2^P$ -complete [1, 12].

In the sequel, we shall abbreviate formulas of the above type by writing  $\forall x \exists y. \varphi(x, y)$ .

### 4.2 Membership

We shall write MP for the problem of deciding whether a given agenda  $\Phi$  satisfies the MP, and similarly for the other properties defined in Section 3.2.

**LEMMA 11.** *MP, SMP, SSMP, and  $k\text{MP}$  are all in  $\Pi_2^P$ .*

**PROOF.** We shall sketch the proof for  $k\text{MP}$ , which is intuitively the most difficult of the four problems. The proofs for the other three problems are very similar and omitted for lack of space.

We need to give an algorithm that decides the correctness of a certificate for the *violation* of the  $k\text{MP}$  in polynomial time, assuming it has access to a SAT-oracle. For a given agenda  $\Phi$  (with  $n = |\Phi|$ ), such a certificate is a set  $\Delta \subseteq \Phi$  that (a) needs to be inconsistent and that (b) must *not* have any inconsistent subsets of size  $\leq k$ . Inconsistency of  $\Delta$  can be checked with a single query to the SAT-oracle. If  $n' = |\Delta|$ , then there are  $\sum_{i=1}^k \binom{n'}{i}$  nonempty subsets of  $\Delta$ , which is polynomial in  $n'$  (and thus also in  $n$ ).<sup>7</sup> Hence, the second condition can be checked by a further polynomial number of queries to the oracle.  $\square$

### 4.3 Hardness

To help intuition, observe that, similarly to  $\text{coQSAT}_2$ , the median property and its variants ask questions beginning with a universal and ending in an existential quantification (roughly: “for *all* subsets ... there *exists* a subset ...”). To formally prove  $\Pi_2^P$ -hardness, we need to show that, although  $\text{coQSAT}_2$  may seem a more general problem, it can be reduced to our seemingly more specific problems.

We first prove a technical lemma. Let  $\text{coQSAT}_2^2$  be the problem of checking whether a QBF of the following form is

<sup>7</sup>This figure is not polynomial in  $k$ , but note that this does not affect the argument, because  $k$  is a constant.

true, given that we already know that (i)  $\varphi$  is not a tautology, (ii)  $\varphi$  is not a contradiction, and (iii)  $\varphi$  is not logically equivalent to a literal:

$$\forall \underline{x} \exists \underline{y}. \varphi(\underline{x}, \underline{y}) \wedge \forall \underline{x} \exists \underline{y}. \neg \varphi(\underline{x}, \underline{y})$$

LEMMA 12.  $\text{coQSAT}_2^2$  is  $\Pi_2^P$ -hard.

PROOF. By reduction from  $\text{coQSAT}_2$ : Given any QBF of the form  $\forall \underline{x} \exists \underline{y}. \varphi(\underline{x}, \underline{y})$ , checking satisfiability is equivalent to running  $\text{coQSAT}_2^2$  on  $(\varphi \vee a) \wedge b$ , for two new propositional variables  $a$  and  $b$  not occurring in  $\varphi$ , i.e., to checking the satisfiability of the formula

$$\forall \underline{x} \forall a \exists \underline{y} \exists b. [(\varphi(\underline{x}, \underline{y}) \vee a) \wedge b] \wedge \forall \underline{x} \forall a \exists \underline{y} \exists b. \neg [(\varphi(\underline{x}, \underline{y}) \vee a) \wedge b].$$

This is so, because the second of the above conjuncts is always satisfiable (by making  $b$  false), while the first is satisfiable exactly when the original formula  $\forall \underline{x} \exists \underline{y}. \varphi(\underline{x}, \underline{y})$  is true. (Note that  $(\varphi \vee a) \wedge b$  cannot be a tautology, a contradiction, or equivalent to a literal, so the side constraints specified in the definition of  $\text{coQSAT}_2^2$  are satisfied.)  $\square$

We first prove hardness for the SSMP:

LEMMA 13. SSMP is  $\Pi_2^P$ -hard.

PROOF. We shall give a reduction from  $\text{coQSAT}_2^2$  to SSMP; the claim then follows from Lemma 12.

Take any instance of  $\text{coQSAT}_2^2$ , i.e., the question whether  $\forall \underline{x} \exists \underline{y}. \varphi(\underline{x}, \underline{y}) \wedge \forall \underline{x} \exists \underline{y}. \neg \varphi(\underline{x}, \underline{y})$  is true for some  $\varphi$  with  $\not\models \varphi$ ,  $\varphi \not\models \perp$ , and  $\not\models \varphi \leftrightarrow \ell$  for literals  $\ell$ . Suppose  $\underline{x} = \langle x_1, \dots, x_r \rangle$ , and define an agenda as follows:<sup>8</sup>

$$\Phi = \{x_1, \neg x_1, x_2, \neg x_2, \dots, x_r, \neg x_r, (\varphi \wedge \top), \neg(\varphi \wedge \top)\}$$

Now,  $\Phi$  satisfies the SSMP if and only if the answer to our  $\text{coQSAT}_2^2$ -question should be YES.

To see this, consider the following facts. First, as  $\varphi$  is neither a tautology nor a contradiction, any inconsistent subset of  $\Phi$  must be nontrivially inconsistent. Second, by construction of  $\Phi$  (consisting largely of literals), any inconsistent subset of  $\Phi$  not including a pair of syntactic complements must include either  $(\varphi \wedge \top)$  or  $\neg(\varphi \wedge \top)$ , as well as a (complement-free) subset of  $\{x_1, \neg x_1, \dots, x_r, \neg x_r\}$ . That is, the only way of violating the SSMP is to find a subset of literals from  $\{x_1, \neg x_1, \dots, x_r, \neg x_r\}$  to make true that forces either  $(\varphi \wedge \top)$  or  $\neg(\varphi \wedge \top)$  to be false. But this is precisely the situation in which our instance of  $\text{coQSAT}_2^2$  requires a negative answer.  $\square$

Proving hardness for the SMP works similarly:

LEMMA 14. SMP is  $\Pi_2^P$ -hard.

PROOF. The construction used is the same as for the proof of Lemma 13. The only additional insight required is the observation that for the special kind of agenda constructed in that proof, the SMP and the SSMP coincide (by excluding formulas  $\varphi$  that are equivalent to literals, we ensure that there are no inconsistent subsets consisting of one literal and one compound formula only).  $\square$

Finally, for the MP and the  $k$ MP we give proofs using reductions from the SSMP:

<sup>8</sup>Using  $(\varphi \wedge \top)$  rather than  $\varphi$  in  $\Phi$  ensures that the agenda defined does not include doubly-negated formulas.

LEMMA 15. MP is  $\Pi_2^P$ -hard.

PROOF. We shall give a polynomial-time reduction from SSMP, a  $\Pi_2^P$ -complete problem by Lemma 13, to MP.

Let  $\Phi$  be an agenda on which we want to test the SSMP, and divide the formulas in  $\Phi$  into a positive and a negative part:  $\Phi = \Phi_+ \cup \{\neg\varphi \mid \varphi \in \Phi_+\}$ . Let  $\Phi_+ = \{\varphi_1, \dots, \varphi_m\}$ . Now build the set  $\Phi'_+$  in the following way: copy all formulas in  $\Phi_+$   $m$  times, every time renaming the variables occurring in  $\varphi_i$ , obtaining the set of formulas  $\varphi_i^j$  for  $1 \leq i, j \leq m$ . For every  $i$  substitute  $\varphi_i^i$  with  $\varphi_i^i \vee p^i$ , where  $p^i$  is a variable not occurring in any of the  $\varphi_i^j$ . Finally, add  $p^1, \dots, p^m$  to the agenda. We obtain the following set:

$$\begin{aligned} \Phi'_+ = & \{p^1, \varphi_1^1 \vee p^1, \dots, \varphi_m^1, \\ & p^2, \varphi_1^2, \varphi_2^2 \vee p^2, \dots, \varphi_m^2, \\ & \vdots \\ & p^m, \varphi_1^m, \dots, \varphi_m^m \vee p^m\} \end{aligned}$$

Define  $\Phi' = \Phi'_+ \cup \{\neg\varphi \mid \varphi \in \Phi'_+\}$ . We claim that  $\Phi$  satisfies the SSMP if and only if  $\Phi'$  satisfies the MP. One direction is easy: if  $\Phi$  does not satisfy the SSMP, then there exists a minimally inconsistent subset of size  $k \geq 2$  not containing both a formula and its complement. If this subset is  $X = \{\varphi_{i_1}, \dots, \varphi_{i_k}\}$ , then there exists a subset of  $\Phi'$ , namely  $X' = \{\neg p^{i_1}, \varphi_{i_1}^{i_1} \vee p^{i_1}, \varphi_{i_2}^{i_2}, \dots, \varphi_{i_k}^{i_k}\}$ , that is a minimally inconsistent subset not containing any inconsistent subset of size  $k+1 \geq 3$ , thereby falsifying the MP.

For the opposite direction, suppose that  $\Phi'$  does not satisfy the MP. That is, there exists a minimally inconsistent subset of size  $\geq 3$ . By construction of  $\Phi'$ , we know that such a subset must only contain formulas with the same superscript or their complements (all other formulas having different variables). If this subset does not contain any  $p^i$  or  $\neg p^i$ , then we can find a copy of it in  $\Phi$ , which then violates the SSMP. If instead either  $p^i$  or  $\neg p^i$  is contained in this set for some  $i$ , then by minimality also  $\varphi_i^i \vee p^i$  or its negation must be included. We can now reason by cases: if both  $p^i$  and  $\varphi_i^i \vee p^i$  are in the set, then by dropping the disjunction we will still get an inconsistent subset, against the assumption of minimality;  $\neg p^i$  and  $\neg(\varphi_i^i \vee p^i)$  cannot be in the set for the same reason; finally,  $p^i$  together with the negation of  $\varphi_i^i \vee p^i$  are already inconsistent. Therefore, we can conclude that all minimally inconsistent subsets that can be built from  $\Phi'$  are of the form  $\{\neg p^i, \varphi_i^i \vee p^i, (-)\varphi_j^i\}$ , where  $\varphi_j^i$  is a vector of formulas with the same superscript and the prefix  $(-)$  is intended to indicate that any number of formulas in that vector can be negated. It is now easy to see that  $\{\varphi_i, (-)\varphi_j^i\}$  is a minimally inconsistent subset of  $\Phi$  that falsifies the SSMP.  $\square$

LEMMA 16.  $k$ MP is  $\Pi_2^P$ -hard for every  $k \geq 2$ .

PROOF. Due to space constraints we shall only give an idea of the proof. As for the previous lemma, we can build a reduction from SSMP to  $k$ MP by building a suitable agenda. This agenda can be built by copying  $m$  times the formulas of the original agenda and replacing  $\varphi_i^i$  with a new propositional symbol  $p_i^i$  and its disjunction with  $\varphi_i^i$ . Instead of adding  $p^1, \dots, p^m$  to the new agenda, one has to add a chain of length  $k-1$  of the form  $\{p_{k-1}^i, p_{k-1}^i \rightarrow p_{k-2}^i, \dots, p_2^i \rightarrow \neg p_1^i\}$ . The remainder of the proof follows the proof of the previous lemma, building this time a minimally inconsistent subset of size bigger than  $k$ .  $\square$



## 4.4 Safety of the Agenda: Complexity

We have shown that deciding whether a given agenda  $\Phi$  satisfies the MP, the SMP, the SSMP, or the  $k$ MP is both in  $\Pi_2^P$  and  $\Pi_2^P$ -hard. Furthermore, in Section 3 we have linked these properties to the safety of  $\Phi$  for varying combinations of axioms. As an immediate corollary to all of these results, we obtain our theorem concerning the complexity of SoA:

**THEOREM 17.** *Checking the safety of an agenda is  $\Pi_2^P$ -complete for any of these classes of aggregation procedures:*

- (i) *the majority rule, corresponding to  $\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{S}, \text{M}^1]$  and  $\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{N}, \text{M}^N]$ ;*
- (ii) *systematic rules:  $\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{S}]$ ;*
- (iii) *neutral rules:  $\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{N}]$ ;*
- (iv) *independent rules:  $\mathcal{F}_\Phi[\text{WR}, \text{A}, \text{I}]$ ;*
- (v) *for any  $k \geq 2$ , the class of uniform quota rules  $F_m$  with  $m > n - \frac{n}{k}$ , where  $n$  is the number of individuals.*

**PROOF.** (i) is a direct consequence of Theorem 6 and Lemma 15. In the same way (ii) is derived from Theorem 7 and Lemma 14, (iii) from Theorem 8 and Lemma 14, and (iv) from Theorem 9 and Lemma 13. Finally, (v) follows from Theorem 10 together with Lemma 16. (Membership in  $\Pi_2^P$  follows from Lemma 11 in all five cases.)  $\square$

## 5. CONCLUSION

We have introduced the notion of *safety of the agenda* into the study of judgment aggregation and proved results of two types for this new concept. The first are *characterisation results*, which identify “logical” properties of the agenda that need to be satisfied if we want to give a guarantee that aggregating judgements over that agenda will never lead to an inconsistent outcome (i.e., to a discursive dilemma) when an arbitrary aggregation procedure from a given class is used. These classes of procedures are (in most cases) defined by means of standard axioms, fixing certain desirable normative conditions of a procedure. The second type are *complexity results*. We have seen that deciding whether a given agenda is safe is a computationally intractable problem for all of the classes of procedures considered in this paper. To be precise, the problem is  $\Pi_2^P$ -complete, which means that there can be no polynomial algorithm to solve this problem (unless  $P=NP$ ), and furthermore that the problem is harder than familiar NP-complete problems (unless the polynomial hierarchy collapses to the first level).

Our complexity results are negative results in the sense that they show us that a problem that we would like to be able to solve efficiently is very difficult. We should stress that this does not render the problem hopeless. Automated theorem provers for QBFs, for instance, could be deployed to check whether an agenda satisfies a given type of median property, providing us with a decision procedure for the SoA problem. Work on QBF solvers has seen a lot of progress in recent years (see, e.g., the annual QBFEVAL competition [10]). Furthermore, understanding how a naturally arising question in judgment aggregation relates to a difficult but well-studied algorithmic problem such as QSAT<sub>2</sub> is interesting and worthwhile in its own right.

While we do not want to claim outright novelty for the representation results given in Section 2.3, we hope that others will find it useful to see these spelt out formally.

While SoA has been a natural problem to start a complexity-theoretic study of judgment aggregation, it is certainly not the only problem in the field that could and should be subjected to this treatment. Regarding future work, we believe that investigating the complexity of checking the *safety of a profile* may also be of interest: even if the agenda is not safe (and possibly if we know *why* it is not safe, e.g., if we have found minimally inconsistent sets of a certain type or size), then we may ask whether a specific profile will generate an inconsistent outcome. Of course, we can always first run the aggregation procedure and compute the outcome and then check its satisfiability (which would be NP-hard). The (open) question is whether (or under what circumstances) we can do better. Another natural question to address is the complexity of strategic manipulation in judgment aggregation (for a suitable notion of preference over alternative judgment sets). This may yield proposals for protecting aggregation against manipulation by means of high computational costs.

## 6. REFERENCES

- [1] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.
- [2] Y. Chevaleyre, U. Endriss, J. Lang, and N. Maudet. A short introduction to computational social choice. In *Proc. SOFSEM-2007*. Springer-Verlag, 2007.
- [3] F. Dietrich. A generalised model of judgment aggregation. *Social Choice and Welfare*, 28(4):529–565, 2007.
- [4] F. Dietrich and C. List. Judgment aggregation by quota rules: Majority voting generalized. *Journal of Theoretical Politics*, 19(4):391–424, 2007.
- [5] P. Faliszewski, E. Hemaspaandra, L. A. Hemaspaandra, and J. Rothe. A richer understanding of the complexity of election systems. In S. S. Ravi and S. K. Shukla, editors, *Fundamental Problems in Computing (Festschrift D.J. Rosenkrantz)*. Springer-Verlag, 2009.
- [6] S. Konieczny and R. Pino Pérez. Merging information under constraints: A logical framework. *Journal of Logic and Computation*, 12(5):773–808, 2002.
- [7] L. A. Kornhauser and L. G. Sager. Unpacking the court. *Yale Law Journal*, 96(1):82–117, 1986.
- [8] C. List and P. Pettit. Aggregating sets of judgments: An impossibility result. *Economics and Philosophy*, 18(1):89–110, 2002.
- [9] C. List and C. Puppe. Judgment aggregation: A survey. In *Handbook of Rational and Social Choice*. Oxford University Press, 2009.
- [10] M. Narizzano, L. Pulina, and A. Tacchella. The QBFEVAL web portal. In *Proc. 10th European Conference on Logics in Artificial Intelligence (JELIA-2006)*. Springer-Verlag, 2006.
- [11] K. Nehring and C. Puppe. The structure of strategy-proof social choice. Part I: General characterization and possibility results on median spaces. *Journal of Economic Theory*, 135(1):269–305, 2007.
- [12] C. H. Papadimitriou. *Computational Complexity*. Addison Wesley, 1994.
- [13] G. Pigozzi. Belief merging and the discursive dilemma. *Synthese*, 152(2):285–298, 2006.