Ontology Merging as Social Choice

Daniele Porello and Ulle Endriss

Institute for Logic, Language and Computation (ILLC)
University of Amsterdam

Abstract. The problem of merging several ontologies has important applications in the Semantic Web, medical ontology engineering, and other domains where information from several distinct sources needs to be integrated in a coherent manner. We propose to treat ontology merging as a problem of social choice, i.e., as a problem of aggregating the input of a set of individuals into an adequate collective decision, and we show how to apply the methodology of social choice theory in this new domain. We do this for the case of ontologies that are modelled using description logics. Specifically, we formulate a number of desirable properties for ontology merging procedures, we identify the incompatibility of some of these properties, and we define and analyse several concrete procedures.

1 Introduction

Merging a number of ontologies originating from different sources is a pressing problem in applications ranging from medical informatics to the Semantic Web [13, 6]. We propose to add a new perspective to this problem by treating it as a problem of *social choice*. Social choice theory (SCT) is a branch of economic theory that deals with the design and analysis of mechanisms for aggregating opinions of individual agents to arrive at a basis for a collective decision [7]. A typical example is voting. In the context of ontology merging, we may think of the provider of each ontology as a voter, and these voters try to "elect" a collective ontology that adequately and fairly represents the information provided by each of them.

As an example, imagine a possible Semantic Web scenario. Suppose several sources provide different encyclopedia entries of the same word. Naturally, encyclopedias might differ with respect to the information provided, the degree of exhaustiveness attained, or the aspects chosen as relevant. Of course, there might be conflicts about the views provided by the different sources. We might imagine an agent who is searching the web for a given definition who is interested in knowing an answer that best represents the *class* of encyclopedias he has access to, rather than checking each source by itself. This problem is clearly related to the problem of aggregating several points of view into a collective point of view, where we do not have enough information to discriminate the reliability of the various sources. With respect to such a scenario, the kind of axioms usually discussed in SCT are relevant, because they provide precise definitions of the idea of collective information.

Our aim in this paper is to make the idea of viewing ontology merging as a problem of social choice precise by providing a suitable formal framework for its analysis and to propose a number of simple procedures that fit this framework, together with an initial analysis of some of their most fundamental properties. We concentrate on high-level properties that are broadly related to "fairness" and we restrict attention to what one might want to call "coarse" merging: the ontology to be constructed will be a list of some of the formulas included in the individual ontologies. We do not deal with "fine" merging, where we might also want to construct entirely new formulas from those provided by the individuals. We use ontologies expressed in a simple description logic [1] as an example, although the choice of logic is in fact not critical to our proposal.

In the remainder of this paper we shall use the term *ontology aggregation* to refer to our specific approach based on SCT, to distinguish it from the broader and established research area of *ontology merging*.

What we propose is closely related to *judgment aggregation* (JA), a branch of SCT that deals with the aggregation of individual judgments regarding the truth or falsehood of a set of interrelated propositions modelled as formulas of propositional logic [10]. The main points of interest of our proposal from the viewpoint of the JA literature are the following:

(1) First, the agenda, i.e., the set of formulas which may or may not be accepted by individuals, is not closed under complementation (which is a standard assumption in JA).
(2) Second, we operate under an *open world assumption*, meaning that an agent's failure to explicitly include a formula in her ontology does not necessarily mean that she rejects the truth of that formula.
(3) Third, description logical ontologies make a separation between terminological and assertional knowledge, and this conceptual distinction can guide the aggregation process (cf. the discussion of "premises" and "conclusions" in the JA literature).

The problem of modelling ontology change is of course a very general and protean task, dealing with a vast number of interrelated phenomena such as updating after new information has arrived, revision, or debugging for inconsistencies [6]. Contributions to ontology merging range from sophisticated engineering solutions (see e.g. [13]) to works in mathematical logic. Applications of AGM belief revision to ontology merging and debugging have been discussed, for instance, by [14]. However, even though the connections between SCT and belief merging are clearly recognised in AI [8], this methodology seems not yet to have been applied to ontology merging.

The remainder of the paper is organised as follows. In Section 2, we define a formal framework for ontology aggregation in description logics. In Section 3, we then define a number of axioms (i.e., desirable properties) that a specific aggregation procedure may or may not satisfy. Finally, in Section 4, we present a number of such procedures based on simple principles and discuss to what extent they satisfy the axioms defined earlier. We conclude with a brief discussion of possible directions for future work.

2 A Framework for Ontology Aggregation

We first define our framework for aggregating ontologies expressed in a description logic with a common alphabet. We begin by recalling some basic notation and terminology from description logics.

2.1 Preliminaries: Description Logics

Description logics are languages for knowledge representation with a formal syntax and semantics that balance expressive power as dictated by applications with computational efficiency requirements. The best known and mostly widely used basic description logic is $\mathcal{ALC}$. Our approach is not tied to any particular description logic, but for reasons of clarity of exposition we shall restrict attention to $\mathcal{ALC}$. The following review of the basics of description logics and $\mathcal{ALC}$ is fairly succinct; for full details we refer to the literature [1].

The language of $\mathcal{ALC}$ is based on an alphabet consisting of *atomic concepts*, *role names*, and *object names*. The set of *concept descriptions* is generated by the following grammar (where A represents atomic concepts and R role names):

$$C ::= A \mid \neg C \mid C \sqcap C \mid C \sqcup C \mid \forall R.C \mid \exists R.C$$

A *TBox* is a finite set of formulas of the form $A \sqsubseteq C$ and $A \equiv C$ (where A is an atomic concept and C a concept description). It is used to store terminological knowledge regarding the relationships between concepts. An *ABox* is a finite set of formulas of the form $A(a)$ ("object a is an instance of concept A") and $R(a, b)$ ("objects a and b stand to each other in the R-relation").[1] It is used to store assertional knowledge regarding specific objects. The semantics of $\mathcal{ALC}$ is defined in terms of *interpretations* that map each object name to an element of its domain, each atomic concept to a subset of the domain, and each role name to a binary relation on the domain. The truth of a formula in such an interpretation is defined in the usual manner [1]. For instance, $\forall R.C$ is true in a given interpretation at point a if all elements related to a via (the interpretation of) R belong to the (interpretation of) C. A set of (TBox and ABox) formulas is *satisfiable* if there exists an interpretation in which they are all true. A consequence relation $\models$ is defined on top of this semantics in the standard way.

2.2 Ontology Aggregators

Let us now fix a particular alphabet. This induces a fixed finite set of ABox formulas (but the set of TBox formulas is infinite). Let us fix a finite set $\mathcal{L}$ of $\mathcal{ALC}$ formulas over this alphabet that includes all ABox formulas that can be

[1] Note that limiting the ABox to "atomic" formulas is not a restriction, as A may be given a complex definition in the TBox.

expressed.[2] We call $\mathcal{L}$ the *agenda* and any set $O \subseteq \mathcal{L}$ an *ontology*.[3] Any such ontology O can be divided into a TBox O^T and an ABox O^A. We denote the set of all those ontologies that are *satisfiable* by $\mathrm{On}(\mathcal{L})$. Also recall that the *closure* of a set of formulas $\Phi \subseteq \mathcal{L}$ is the set of all formulas that logically follow from those in Φ. It is denoted by $\mathrm{Cl}(\Phi) := \{\varphi \in \mathcal{L} \mid \Phi \models \varphi\}$.

Let $\mathcal{N} = \{1, \ldots, n\}$ be a finite set of *agents* (or *individuals*, or *experts*). Each agent $i \in \mathcal{N}$ provides a satisfiable ontology $O_i \in \mathrm{On}(\mathcal{L})$. An *ontology profile* $\boldsymbol{O} = (O_1, \ldots, O_n) \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ is a vector of such ontologies, one for each agent. We write $\mathcal{N}_\varphi^{\boldsymbol{O}} := \{i \in \mathcal{N} \mid \varphi \in O_i\}$ for the set of agents including φ in their ontology under profile $\boldsymbol{O}$.

The question we shall address in this paper is how to best aggregate an ontology profile into a single collective ontology. That is, our object of study are *ontology aggregators*.

Definition 1 (Ontology aggregators). *An ontology aggregator is a function* $F : \mathrm{On}(\mathcal{L})^{\mathcal{N}} \to 2^{\mathcal{L}}$ *mapping any profile of satisfiable ontologies to an ontology.*

Observe that, according to this definition, the ontology we obtain as the outcome of an aggregation process need not be satisfiable. Of course, we will be particularly interested in ontology aggregators that are *satisfiable*, i.e., aggregators F for which $F(O_1, \ldots, O_n)$ is satisfiable whenever all O_i are.

2.3 Example

A simplistic example for an ontology aggregator is F with $F(\boldsymbol{O}) := O_1 \cup \cdots \cup O_n$, which simply returns the union of the individuals ontologies. Of course, this will usually not be a good choice, as F clearly is not a satisfiable aggregator. Another simple natural choice is the *majority rule:* accept a formula if and only if a majority of the agents do. This can also lead to unsatisfiable outcomes, as we can easily simulate the *doctrinal paradox* familiar from JA [10]. Suppose three agents share a common TBox with two formulas:

$$C_3 \sqsubseteq C_1 \sqcap C_2 \qquad C_4 \sqsubseteq \neg C_3$$

Furthermore, suppose the three ABoxes are as follows:

	$C_1(a)$	$C_2(a)$	$C_3(a)$	$C_4(a)$
Agent 1	yes	yes	yes	no
Agent 2	yes	no	no	yes
Agent 3	no	yes	no	yes
Majority	yes	yes	no	yes

Even though all individual ontologies are satisfiable, the ontology obtained by applying the majority rule is not.

[2] The finite set of TBox formulas in $\mathcal{L}$ might be all TBox formulas of a certain maximum length or the union of all TBox formulas that a given population of agents choose to include in their TBoxes.

[3] In the literature, the term "ontology" is sometimes restricted to terminological knowledge; here we use it in this broader sense.

3 Properties of Ontology Aggregators

We now define a number of properties that a given ontology aggregator may or may not satisfy. Most of these properties relate, in one way or another, to the "fairness" of the aggregation process and are directly inspired by properties of voting rules, JA rules, and other types of aggregators commonly defined in SCT [7, 10]. As in SCT, we refer to these properties as *axioms*.

3.1 Syntactic Axioms

We first define a number of axioms that are "syntactic" in the sense that they relate to the formulas that occur explicitly in the ontologies of individual agents or in the collective ontology. We will later contrast this with "semantic" axioms that also make reference to the formulas that can be *inferred* from those ontologies.

The axiom of *unanimity* postulates then when all individual ontologies include φ, then so should the collective ontology. This clearly is a desirable property in any kind of domain. An aggregator F is *anonymous* if it is symmetric wrt. individual ontologies. This is appropriate if we have reasons to treat all agents equally. In the social choice literature the axiom of anonymity is usually motivated in terms of fairness considerations, which may or may not be relevant in the context of ontology aggregation, depending on the application at hand. But treating all agents equally is also justified, for instance, if we simply do not have any information regarding the reliability of individual agents. F is *independent* if inclusion of φ in the collective ontology only depends on the pattern of its inclusion in the individual ontologies and is independent from which other formulas may or may not have been included. Independence is a more demanding axiom that the previous two; whether or not it should be imposed certainly is debatable. Finally, F is *monotonic* if additional support for a collectively accepted formula will never lead to it being rejected. This, again, is a property that we would usually (though maybe not always) like to see satisfied, certainly in cases where it is reasonable to assume that every agent has at least some degree of reliability. The four axioms introduced so far are formalised as follows:

- **Unanimity:** F is called *unanimous* if $O_1 \cap \cdots \cap O_n \subseteq F(\boldsymbol{O})$ for every profile $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.
- **Anonymity:** F is called *anonymous* if for any profile $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ and any permutation $\pi : \mathcal{N} \to \mathcal{N}$ we have that $F(O_1, \ldots, O_n) = F(O_{\pi(1)}, \ldots, O_{\pi(n)})$.
- **Independence:** F is called *independent* if for any $\varphi \in \mathcal{L}$ and profiles $\boldsymbol{O}, \boldsymbol{O}' \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$, we have that $\varphi \in O_i \Leftrightarrow \varphi \in O'_i$ for all $i \in \mathcal{N}$ implies $\varphi \in F(\boldsymbol{O}) \Leftrightarrow \varphi \in F(\boldsymbol{O}')$.
- **Monotonicity:** F is called *monotonic* if for any $i \in \mathcal{N}$, $\varphi \in \mathcal{L}$, and $\boldsymbol{O}, \boldsymbol{O}' \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ with $O_j = O'_j$ for all $j \neq i$, we have that $\varphi \in O'_i \setminus O_i$ and $\varphi \in F(\boldsymbol{O})$ imply $\varphi \in F(\boldsymbol{O}')$.

A further important axiom from the literature is *neutrality*, which, intuitively, requires all formulas to be treated symmetrically. In fact, there are a number of possible interpretations of this notion, including these:

– **Neutrality:** F is called *neutral* if for any $\varphi, \psi \in \mathcal{L}$ and $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ we have that $\varphi \in O_i \Leftrightarrow \psi \in O_i$ for all $i \in \mathcal{N}$ implies $\varphi \in F(\boldsymbol{O}) \Leftrightarrow \psi \in F(\boldsymbol{O})$.
– **Acceptance-Rejection Neutrality:** F is called *acceptance-rejection neutral* if for any $\varphi \in \mathcal{L}$ and $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ we have that $\varphi \in O_i \Leftrightarrow \psi \notin O_i$ for all $i \in \mathcal{N}$ implies $\varphi \in F(\boldsymbol{O}) \Leftrightarrow \psi \notin F(\boldsymbol{O})$.

The first notion of neutrality is the one that we shall adopt here. It says that if two formulas enjoy the same pattern of acceptance—in the same profile—then either both should be accepted or both should be rejected. The second axiom is closer to the original neutrality axiom in voting theory proposed by [11]. It says that if those patterns of acceptance are complementary, then exactly one of the two formulas should be accepted. The reason we do not believe that acceptance-rejection neutrality is appropriate for ontology aggregation is that it makes the implicit assumption that not explicitly including a formula into one's knowledge base amounts to actively rejecting the validity of that formula. This is an appropriate assumption in JA, but not here.[4]

We also propose three axioms that are specific to ontology aggregation and that do not have a counterpart in standard SCT or JA. The first is *groundedness:* a formula should only occur in the collective ontology if it is included in at least one of the individual ontologies, i.e., if it is an element of $O_1 \cup \cdots \cup O_n$, the *support* of a given profile $(O_1, \ldots, O_n)$. In standard JA, due to the assumption that agendas are closed under complementation (and that each agent will accept either φ or its complement), groundedness is implied by unanimity (with consistency) and does not require a separate axiom. The second axiom we propose is *exhaustiveness:* it should not be possible to add any formula from the support to the collective ontology without rendering the latter unsatisfiable. In other words, we should "exhaust" the supply of formulas in the support when building the collective ontology—as long as we do not create any inconsistencies this way. This axiom is desirable if we assume that all information supplied by individuals is (potentially) useful information and if we do not take an agent's omission of a particular formula in their ontology as a vote *against* that formula. That is, exhaustiveness is closely related to the open world assumption. Our third axiom is *group closure*, a weaker version of exhaustiveness: any formula in the support that is logically entailed by the collective ontology should in fact be part of that ontology. We now state these additional axioms formally:

– **Groundedness:** F is called *grounded* if $F(\boldsymbol{O}) \subseteq O_1 \cup \cdots \cup O_n$ for every profile $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.

[4] Dietrich and List [3] use the name "acceptance-rejection neutrality" for a slightly different axiom: for any $\varphi \in \mathcal{L}$ and $\boldsymbol{O}, \boldsymbol{O}' \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$, we have that $\varphi \in O_i \Leftrightarrow \psi \notin O'_i$ for all $i \in \mathcal{N}$ implies $\varphi \in F(\boldsymbol{O}) \Leftrightarrow \psi \notin F(\boldsymbol{O}')$. Arguably, this is closer to an (in)dependence axiom, as it makes reference to *two* profiles.

- **Exhaustiveness:** F is called *exhaustive* if there exists no satisfiable set $\Phi \subseteq O_1 \cup \cdots \cup O_n$ with $F(\boldsymbol{O}) \subset \Phi$ for any profile $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.
- **Group Closure:** F is called *group-closed* if there exists no set $\Phi \subseteq O_1 \cup \cdots \cup O_n$ with $F(\boldsymbol{O}) \models \Phi$ and $F(\boldsymbol{O}) \subset \Phi$ for any profile $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.

All of the above axioms are natural requirements, but we stress that we do *not* impose them in general. Some may be more desirable than others for any given problem domain (but all should certainly be considered).

We are now in a position to make our objection to the axiom of acceptance-rejection neutrality more precise:

Proposition 1. *Any ontology aggregator that satisfies acceptance-rejection neutrality violates exhaustiveness.*

Proof. Any acceptance-rejection neutral aggregator cannot accept both φ and ψ when $\varphi \in O_i \Leftrightarrow \psi \notin O_i$ for all $i \in \mathcal{N}$. But if each is accepted by at least one agent, and if each is logically independent from all other formulas in the support, then an exhaustive aggregator must accept them both. $\square$

3.2 Semantic Axioms

For many applications, the agents providing individual ontologies will not only be worried about the formulas included in the collective ontology but also about the formulas that can be *inferred* from that ontology. This distinction has also been discussed by Flouris et al. [5] in terms of implicitly and explicitly represented knowledge. We therefore formulate semantic variants of the axioms above in which we refer to the closures of ontologies rather than the ontologies themselves. Note that the existing literature on JA only deals with what we have called syntactic axioms above.

Here we only spell out the precise formulation of the semantic variants of the aforementioned axioms for some of them. The remaining ones can be adapted following the same pattern.

- **Semantic Unanimity:** F is called *semantically unanimous* if $\mathrm{Cl}(O_1) \cap \cdots \cap \mathrm{Cl}(O_n) \subseteq \mathrm{Cl}(F(\boldsymbol{O}))$ for every profile $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.
- **Semantic Groundedness:** F is called *semantically grounded* if $\mathrm{Cl}(F(\boldsymbol{O})) \subseteq \mathrm{Cl}(O_1) \cup \cdots \cup \mathrm{Cl}(O_n)$ for every $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.
- **Semantic Exhaustiveness:** F is called *semantically exhaustive* if there exists no satisfiable set $\Phi \subseteq \mathrm{Cl}(O_1) \cup \cdots \cup \mathrm{Cl}(O_n)$ with $\mathrm{Cl}(F(\boldsymbol{O})) \subset \Phi$ for any $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.

That is, semantic unanimity, for instance, is satisfied if whenever each individual ontology suffices to infer some formula φ, then φ should also be derivable from the collective ontology. We believe that all of our semantic properties are generally desirable properties and system designers should be interested in satisfying these axioms—with one exception: semantic groundedness. This axiom postulates that only formulas derivable from at least one individual ontology should be derivable.

This will rarely be a reasonable requirement. On the contrary, we would hope that by combining the information provided by several agents we are able to make new inferences that were not possible before aggregation. For comparison, note that *syntactic* groundedness is perfectly reasonable, at least for what we have called coarse merging above (for fine merging, we *do* want to be able to construct new formulas).

An interesting feature of our model is that it allows for stating precisely the relationship between implicitly and explicitly represented knowledge, namely by investigating relationship between syntactic and the semantic axioms. So, what is the relative strength of a syntactic axiom and its semantic variant? For unanimity, for instance, we can show that the syntactic version does not entail the semantic version, nor *vice versa*. First, consider this example, showing that there are syntactically unanimous aggregators that are not semantically unanimous: Suppose three agents share a common TBox including the formulas $C \equiv D$ and $D \equiv E$, and suppose the ABox of the first agent includes only $C(a)$, the second only $D(a)$, and the third only $E(a)$. Now the majority rule will produce an empty ABox. This violates semantic unanimity, as $C(a)$ can be inferred from all three individual ABoxes, but not from the collective ABox. However, the majority rule clearly is (syntactically) unanimous. Second, a trivial counterexample shows that semantically unanimous aggregators need not be syntactically unanimous: Consider the aggregator F mapping any input to a fixed unsatisfiable ontology, such as $\{C \equiv D \sqcap \neg D, C(a)\}$. F is not syntactically unanimous, but it is semantically unanimous (as we can infer anything from a contradictory ontology). Still, intuitively, semantic unanimity is the (much) stronger property. This intuition can be confirmed for "well-behaved" aggregators:

Proposition 2. *Any satisfiable and exhaustive ontology aggregator that is semantically unanimous is unanimous.*

Proof. Take any F that is satisfiable, exhaustive, and semantically unanimous. Now pick any formula φ and any profile $\boldsymbol{O}$ such that $\varphi \in O_1 \cap \cdots \cap O_n$. By satisfiability of F, the outcome $F(\boldsymbol{O})$ is satisfiable and so is its deductive closure. For the sake of contradiction, assume $\varphi \notin F(\boldsymbol{O})$. $\varphi \in O_1 \cap \cdots \cap O_n$ implies $\varphi \in \mathrm{Cl}(O_1) \cap \cdots \cap \mathrm{Cl}(O_n)$. Thus, by semantic unanimity, $\varphi \in \mathrm{Cl}(F(\boldsymbol{O}))$. That is, there exists a formula in the support (namely φ) that could be added to $F(\boldsymbol{O})$ without rendering the set unsatisfiable. But this violates exhaustiveness, and we are done. $\square$

Similar connections between syntactic and semantic variants can be established for the other axioms.

4 Procedures for Ontology Aggregation

We now define a number of simple procedures for ontology aggregation and discuss some of their properties, including both the extent to which they can guarantee that collective ontologies will be satisfiable and the extent to which they satisfy some of the axioms introduced earlier. We stress that these procedures are not sophisticated enough to be employed for real-world ontology aggregation. Rather, our intent is to provide a catalogue of basic procedures that can serve as building blocks for constructing more sophisticated procedures in the future. Fully understanding the properties of these basic procedures is a necessary step towards designing more advanced procedures.

4.1 The Majority Rule

We have already introduced the majority rule informally. Formally, it is defined as follows:

Definition 2 (Majority rule). *The majority rule is the ontology aggregator* M *with* $\mathrm{M}(\boldsymbol{O}) = \{\varphi \in \mathcal{L} \mid |\mathcal{N}_{\varphi}^{\boldsymbol{O}}| > \frac{n}{2}\}$ *for all* $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.

We have seen that the majority rule can produce unsatisfiable collective ontologies. Following Endriss et al. [4], we call $\mathcal{L}$ *safe* for a given aggregator F if $F(\boldsymbol{O})$ is satisfiable for any profile $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$. We will now identify necessary and sufficient conditions for the safety of $\mathcal{L}$ under the majority rule.

Adapting the terminology from JA [10], we recall that an agenda $\mathcal{L}$ satisfies the *median property* if and only if every unsatisfiable set $X \subseteq \mathcal{L}$ contains itself an unsatisfiable set Y with cardinality at most 2. Now a simple reformulation of a known result due to Nehring and Puppe shows that an agenda $\mathcal{L}$ is safe for the majority rule if and only if it satisfies the median property [12, 10, 4]. This result can be refined if we put restrictions on the range of profiles on $\mathcal{L}$ that we consider. Description logical ontologies suggest a natural restriction of this kind due to the division of knowledge into the TBox and the ABox. Suppose we restrict attention to *profiles with a common TBox:* all agents agree on the TBox but still need to aggregate their ABoxes. Fix such a TBox $\mathcal{T}$. We say that $\mathcal{L}$ satisfies the $\mathcal{T}$*-median property* if and only if for every set of ABox formulas $X \subseteq \mathcal{L}^A$ such that $\mathcal{T} \cup X$ is unsatisfiable there exists a set $Y \subseteq X$ with cardinality at most 2 such such $\mathcal{T} \cup Y$ is also unsatisfiable. We obtain the following characterisation:

Proposition 3. *The majority rule will return a satisfiable ontology for any profile with a common TBox* $\mathcal{T}$ *if and only if the agenda* $\mathcal{L}$ *satisfies the* $\mathcal{T}$*-median property.*

Proof. One direction is proved by the doctrinal paradox we have seen earlier. For the other direction, assume the $\mathcal{T}$-median property holds but $\mathrm{M}(\boldsymbol{O})$ is unsatisfiable. By definition of the majority rule, the TBox of $\mathrm{M}(\boldsymbol{O})$ is exactly the common TBox $\mathcal{T}$. Thus, by the $\mathcal{T}$-median property, there must be a set Y of ABox formulas in $\mathrm{M}(\boldsymbol{O})$ with $|Y| \leqslant 2$ such that $\mathcal{T} \cup Y$ is unsatisfiable. First, Y cannot be empty as that would mean that $\mathcal{T}$ is unsatisfiable, contradicting our assumption that individual ontologies are satisfiable. Second, $|Y| = 1$ is also not possible, as that would mean that at least one individual ontology must have included that one formula in Y (together with $\mathcal{T}$), which would again contradict our assumption that individual ontologies are satisfiable. So suppose that

$|Y| = 2$ with $Y = \{\varphi, \psi\}$. These formulas could only have been accepted by M if $|\mathcal{N}^{\boldsymbol{O}}_{\varphi}| > \frac{n}{2}$ and $|\mathcal{N}^{\boldsymbol{O}}_{\psi}| > \frac{n}{2}$. But this means that at least one agent must have accepted both φ and ψ (and $\mathcal{T}$). This again contradicts the assumption that individual ontologies are satisfiable. □

In fact, from a purely technical point of view, we can prove the same kind of result for any division of the agenda into two disjoint sets: those formulas on which there is certain agreement (here the TBox) on those on which there is not (here the ABox). For any such division we can formulate a weakened version of the median property (relative to the first) and prove a corresponding (strengthened) characterisation theorem. In the context of ontology aggregation, we argue, such a division is particularly natural.

4.2 Quota Rules

We can generalise the idea underlying the majority rule and accept a formula for the collective ontology whenever the number of agents who do so meet a certain quota. This gives rise to the family of quota rules:

Definition 3 (Quota rules). *Let $q \in [0, 1]$. The quota rule with quota q is the ontology aggregator F_q with $F_q(\boldsymbol{O}) = \{\varphi \in \mathcal{L} \mid |\mathcal{N}^{\boldsymbol{O}}_{\varphi}| \geqslant q \cdot n\}$ for all $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.*

We could also generalise further and allow different quotas for different formulas; Dietrich and List [2] make a distinction between general and *uniform* quota rules. Observe that we obtain the majority procedure for $q = \frac{1}{2} + \epsilon$ for any positive $\epsilon < \frac{1}{n}$. Also observe that for $q \leqslant \frac{1}{n}$ the aggregator F_q simply returns the union of all individual ontologies.

We have seen earlier that the majority rule violates semantic unanimity. In fact, any quota rule does, unless we lower the quota so far as to obtain the trivial union aggregator:

Proposition 4. *A quota rule with quota q for n agents is semantically unanimous if and only if $q \leqslant \frac{1}{n}$.*

Proof (sketch). First, it is easy to check that if the quota is at most $\frac{1}{n}$, then semantic unanimity holds. To see that the axiom does not hold as soon as $q > \frac{1}{n}$, consider the following example. All agents agree on the same TBox $\{C_1 \equiv C_2, C_2 \equiv C_3, \ldots, C_{n-1} \equiv C_n\}$ and, for each $i \in \mathcal{N}$, the ABox of agent i consists of the single formula $C_i(a)$. Then $C_1(a)$ can be inferred from each agent's ontology, but it will not be accepted if $q > \frac{1}{n}$. □

Quota-based rules are (syntactically) anonymous, neutral, independent and monotonic [2]. We can strengthen Proposition 4 and show that anonymity and independence together with semantic unanimity are sufficient to single out the trivial union aggregator:

Proposition 5. *If F is anonymous, independent and semantically unanimous, then $F(\boldsymbol{O}) = O_1 \cup \cdots \cup O_n$ for any $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.*

Proof (sketch). By a standard argument [9, 4], if F is anonymous and independent, then there exists a family of functions $\{g_\varphi : \mathbb{N} \to \{0,1\}\}_{\varphi\in\mathcal{L}}$ such that $\varphi \in F(\boldsymbol{O})$ if and only if $g_\varphi(|\{i \in \mathcal{N} \mid \varphi \in O_i\}|) = 1$. That is, whether of not φ is accepted only depends on the number of agents accepting φ. Now, using a similar construction as in the proof of Proposition 4, we can show that semantic unanimity forces us to accept a formula as soon as any positive number of individual agents do. □

4.3 A Support-Based Procedure

The next aggregation procedure we introduce works as follows: we order the formulas in terms of the number of agents supporting them; we then accept formulas in decreasing order, but drop any formula that would render the ontology constructed thus far unsatisfiable. To decide which of two formulas with the same number of agents supporting it to try first, we introduce a *priority rule* $\gg$ mapping each profile $\boldsymbol{O}$ to a strict linear order $\gg_{\boldsymbol{O}}$ on $\mathcal{L}$ such that $\varphi \gg_{\boldsymbol{O}} \psi$ implies $|\mathcal{N}^{\boldsymbol{O}}_\varphi| \geqslant |\mathcal{N}^{\boldsymbol{O}}_\psi|$ for all $\varphi, \psi \in \mathcal{L}$.

Definition 4 (Support-based procedure). *Given a priority rule $\gg$, the support-based procedure with $\gg$ is the ontology aggregator* $\mathrm{SBP}_\gg$ *mapping any profile $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ to $\mathrm{SBP}_\gg(\boldsymbol{O}) := \Phi$ for the unique set $\Phi \subseteq \mathcal{L}$ for which $\varphi \in \Phi$ if and only if*

(i) $\mathcal{N}^{\boldsymbol{O}}_\varphi \neq \emptyset$ *and*
(ii) $\{\psi \in \Phi \mid \psi \gg_{\boldsymbol{O}} \varphi\} \cup \{\varphi\}$ *is satisfiable.*

We can also define an *irresolute* aggregator that returns the set of all ontologies obtained by *some* choice of priority rule: $\mathrm{SBP}(\boldsymbol{O}) := \{O \mid \mathrm{SBP}_\gg(\boldsymbol{O}) = O \text{ for some } \gg\}$.

The SBP clearly satisfies the axioms of anonymity, monotonicity, groundedness (due to condition (*i*)), and exhaustiveness (due to condition (*ii*)). Neutrality is violated by virtue of having to fix a priority rule $\gg$. Independence is also violated (because φ may cease to be accepted if a formula it is contradicting receives additional support).

Several variants and generalisations of the SBP are possible and interesting. For instance, we can replace $\gg$ as defined above with *any* other function mapping each profile $\boldsymbol{O}$ to a linear order $\gg_{\boldsymbol{O}}$ on $\mathcal{L}$. Each choice of $\gg$ corresponds to a different *greedy procedure* that attempts to accept as many formulas as possible without violating satisfiability in order of priority as specified by $\gg_\varphi$. For instance, a priority rule $\gg$ for which $\varphi \gg_{\boldsymbol{O}} \psi$ holds whenever $\mathcal{N}^{\boldsymbol{O}}_\varphi \supseteq \mathcal{N}^{\boldsymbol{O}}_\psi$ does but not necessarily whenever $|\mathcal{N}^{\boldsymbol{O}}_\varphi| \geqslant |\mathcal{N}^{\boldsymbol{O}}_\psi|$ does will be appropriate to aggregate ontologies from sources with different degrees of reliability (i.e., when the violation of anonymity is acceptable). Another attractive variant would be a *semantic* SBP, where we define $\gg$ in terms of $\{i \in \mathcal{N} \mid O_i \models \varphi\}$ instead of $\mathcal{N}^{\boldsymbol{O}}_\varphi$. That is, under this procedure we accept formulas (supported by at least one agent) in order of priority defined in terms of the number of agents who were able to *infer* those formulas from their own ontologies (but not necessarily included them explicitly).

4.4 A Distance-Based Procedure

In voting theory, many voting rules can be defined using a notion of distance. The well-known *Kemeny rule* is a natural example [7]. Similar ideas have also been used in belief merging [8].

We will now define an aggregation procedure that chooses from a class of acceptable ontologies (namely the satisfiable ones) that ontology that minimises the sum of the distances to the individual ontologies. A common choice is the *Hamming distance:* the distance between two ontologies O and O' is the number of formulas that are included in one and only one of O and O'. In fact, the Hamming distance is not appropriate here, because it gives the same weight to a formula φ that an agent has stated but that will not be included in the collective ontology as to a formula ψ that she has omitted but that will be included (when in fact the former should be much worse; indeed, she may be entirely indifferent to the latter). That is, distances *stricto sensu*, which are symmetric, are not suitable for our purposes. With a slight abuse of terminology, we shall still call the function $d : (A, B) \mapsto |\{\varphi \mid \varphi \in A \text{ and } \varphi \notin B\}|$ a distance.

Definition 5 (Distance-based procedure). *The distance-based procedure is the (irresolute) ontology aggregator* DBP *mapping any profile* $\boldsymbol{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ *to the following set of satisfiable ontologies:*

$$\mathrm{DBP}(\boldsymbol{O}) = \mathrm{argmin}_{O \in \mathrm{On}(\mathcal{L})} \sum_{i \in \mathcal{N}} d(O_i, O)$$

To obtain a resolute aggregator, the DBP needs to be combined with a tie-breaking rule, which will violate either anonymity or neutrality. It also violates independence, because O does not range over all possible ontologies. On the other hand, it is satisfiable by construction. Note that if we choose a tie-breaking rule that selects a maximal set (wrt. set-inclusion), then the DBP will always return a maximally satisfiable set and thus satisfy the axiom of exhaustiveness.

4.5 Two-Stage Procedures

Finally, we briefly sketch an approach for two-stage procedures. Depending on the application, we may give priority to terminological knowledge over assertional knowledge, or *vice versa*, and define aggregation procedures accordingly. This idea is closely related to two classical procedures in JA, the *premise-based procedure*, where individuals vote on the premises by majority and then draw the conclusions, and the *conclusion-based procedure*, where each individual draws her own conclusions and then votes on them by majority [9]. The problem with these procedures is that we lack a convincing general approach for how to label a given proposition as either a premise or a conclusion. There is a significant difference in our case: when we aggregate ontologies, we have a clear separation between two classes of formulas by definition, namely the TBox and the ABox, so we can avoid the problem of splitting the agenda into premises and conclusions.

Definition 6 (Assertion-based procedures). *An (irresolute) assertion-based procedure maps each profile* $\boldsymbol{O}$ *to the set of ontologies obtained as follows:*

(1) Choose an aggregator F_A *restricted to ABox formulas, and let* $F_A(\boldsymbol{O})$ *be the outcome.*

(2) Then the TBox is defined as follows:

$$F_T(\boldsymbol{O}) = \mathrm{argmin}_{O \in \mathrm{On}(\mathcal{L})} \sum_{i \in \mathcal{N}} d(F_A(\boldsymbol{O}) \cup O_i^T, O)$$

An assertion-based procedure stresses the information coming from the ABox. A natural choice for the procedure used in the first step would be the majority rule. In the second step we then select a TBox that is satisfiable in view of the majority ABox and that minimises the cumulative distance to the individual TBoxes. Observe that it is possible that the collective TBox obtained in this manner is empty. An interesting variant of this approach may be to allow agents to revise their TBoxes themselves after the collective ABox has been fixed.

Similarly, we may want to give priority to TBox information and first aggregate TBoxes, then fix a TBox, and finally aggregate ABoxes.

5 Conclusion and Future Work

We have presented a framework for aggregating individual ontologies, consisting of both a TBox and an ABox, inspired by social choice theory. We have discussed axioms that are closely related to well-known fairness conditions and we have introduced new axioms defining a notion of efficiency for the aggregation of ontologies. We have then presented relevant results concerning those axioms and several ontology aggregation procedures we introduced, discussing how they balance fairness and efficiency. We have concentrated on coarse ontology merging, since we wanted to model the aggregation of the information actually provided by agents, as explicitly reflected by our groundedness axiom.

Concerning future work, we believe that the social choice approach provides useful insights also for fine merging. For example, support-based procedures and distance-based procedures can potentially be adapted to deal with *concept merging* (i.e., the construction of new TBox definitions out of definitions stemming from different individual ontologies), providing further qualitative *desiderata* that can be used to select among several possible ways of building concept definitions. We also believe that our work can provide an interesting starting point for future research in judgment aggregation and social choice theory. Ontologies suggest a very rich notion of agent, since they allow for representing the preferences an agent might have over a given set of alternatives together with her information on such alternatives and her criteria for choosing. In this sense, our approach to ontology aggregation can lead to a richer model of collective information and choices.

Acknowledgments. We would like to thank the participants of the Computational Social Choice Seminar at the ILLC for their feedback on an early incarnation of this work, particularly Umberto Grandi, Szymon Klarman and Eric Pacuit.

References

1. Baader, F., Nutt, W.: Basic description logics. In: The Description Logic Handbook. Cambridge University Press (2003)
2. Dietrich, F., List, C.: Judgment aggregation by quota rules: Majority voting generalized. Journal of Theoretical Politics 19(4), 391–424 (2007)
3. Dietrich, F., List, C.: Judgment aggregation with consistency alone. Working paper, London School of Economics (2009)
4. Endriss, U., Grandi, U., Porello, D.: Complexity of judgment aggregation: Safety of the agenda. In: Proc. AAMAS-2010 (2010)
5. Flouris, G., Huang, Z., Pan, J.Z., Plexousakis, D., Wache, H.: Inconsistencies, negations and changes in ontologies. In: Proc. AAAI-2006 (2006)
6. Flouris, G., Manakanatas, D., Kondylakis, H., Plexousakis, D., Antoniou, G.: Ontology change: Classification and survey. Knowledge Engineering Review 23(2), 117–152 (2008)
7. Gaertner, W.: A Primer in Social Choice Theory. Oxford University Press (2006)
8. Konieczny, S., Lang, J., Marquis, P.: DA^2 merging operators. Artificial Intelligence 157(1–2), 49–79 (2004)
9. List, C., Pettit, P.: Aggregating sets of judgments: An impossibility result. Economics and Philosophy 18(1), 89–110 (2002)
10. List, C., Puppe, C.: Judgment aggregation: A survey. In: Handbook of Rational and Social Choice. Oxford University Press (2009)
11. May, K.O.: A set of independent necessary and sufficient conditions for simple majority decision. Econometrica 20(4), 680–684 (1952)
12. Nehring, K., Puppe, C.: 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)
13. Noy, N.F., Musen, M.A.: Algorithm and tool for automated ontology merging and alignment. In: Proc. AAAI-2000 (2000)
14. Ribeiro, M.M., Wassermann, R.: Base revision for ontology debugging. Journal of Logic and Computation 19(5), 721–743 (2009)