Ontology Merging as Social Choice:

Judgment Aggregation under the Open World Assumption[*]

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 view 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. That is, we propose to view ontology merging as ontology aggregation. As a first step in this direction, we formulate several desirable properties for ontology aggregators, we identify the incompatibility of some of these properties, and we define and analyse several simple aggregation procedures. Our approach is closely related to work in judgment aggregation, but with the crucial difference that we adopt an open world assumption, by distinguishing between facts not included in an agent's ontology and facts explicitly negated in an agent's ontology.

1 Introduction

Merging a number of ontologies originating from different sources is a challenging problem in applications ranging from medical informatics to the Semantic Web (Noy and Musen, 2000; Flouris et al., 2008). We propose to add a new perspective to this challenge 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 (Gaertner, 2006). 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 the following scenario. Suppose several sources on the Semantic Web 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 also be conflicts amongst 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 obtaining an answer that best represents the *class* of encyclopedias she has access to, rather than checking each source by itself. This problem is thus 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 types of "axioms" (i.e., desirable properties of aggregators) usually discussed in SCT are relevant, because they allow us to approach the elusive notion of *collective information* in a precise and well-defined manner.

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 also do not deal with the problem of aligning the concept names used by different agents, but rather assume that all agents share a common vocabulary. We use ontologies expressed in a simple description logic (Baader and Nutt, 2003) as an example, although the choice of logic is in fact not central to our proposal.

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 (List and Puppe, 2009). Our choice of methodology is motivated as follows. First, it allows us to import axioms from SCT and JA that, besides expressing natural fairness desiderata, can be viewed as conditions modelling constraints on information. For example, the standard axiom of *anonymity* (which, in the context of social choice, encodes the fairness requirement that all agents should carry the same weight) states that all sources are equally reliable. Second, results in JA clarify the connection between the axioms we can expect to be satisfied for certain aggregators and the logical properties of the language that the agents use to express their knowledge. For example, the size of minimally inconsistent sets of formulas expressible in a given logical fragment has been shown to be intimately linked to the quotas we need to impose if we want to guarantee logically consistent outcomes in aggregation using simple quota rules (Nehring

and Puppe, 2007; Dietrich and List, 2007b). Third, the axiomatic analysis inspired by JA can also be applied to ontology merging operators originating in different areas, particularly belief merging and belief revision (Konieczny and Pino Pérez, 2002).

We stress that what we propose is not intended as an alternative approach to ontology merging. Rather, we use the methodology of JA to introduce and discuss several new desiderata that also existing merging operators may or may not satisfy. To take a very simple example, our approach can be applied to check whether an ontology merging operator is anonymous or whether it is weighting some piece of information more than others. The main point of interest of our approach from the viewpoint of ontology merging is that we take an agent-oriented rather than a data-oriented perspective: if we do not have any additional information available besides the ontologies provided by our agents, then the best we can do is to aggregate the information inherent in those ontologies in a fair manner, respecting the contribution of each individual agent. In contrast to this approach, existing methods in ontology merging have focussed on engineering the best possible consensus ontology based on the specific nature of the data present in the different ontologies (e.g., its logical structure or the reliability regarding specific attributes of its provider).

One difference between the standard framework of JA and our approach is that we work with description logics rather than propositional logic. However, as we shall see, this is not the most important difference: at the abstract level at which we present our framework in this paper the precise choice of the underlying logic is not crucial and none of our results heavily relies on it (although, of course, future work might establish results that are more intrinsically related to the expressive power of a specific description logic). Instead, 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* (thus violating a standard assumption in JA). That is, if φ is a formula expressible in our logical language that is available to an agent to be included in her ontology, then this does not necessarily mean that $\neg\varphi$ is also available. Indeed, in most standard description logics it is not possible to express the negation of a concept definition such as $Mother \equiv Parent \sqcap Female$. Importantly, such restrictions are often not due to technical limitations, but rather conceptually motivated (it simply would not be natural to allow an agent to explicitly declare that she does *not* believe that a mother is a female parent).

(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. In standard JA, on the other hand, there is no distinction between the acceptance of $\neg\varphi$ and the rejection of φ. Adopting this open world assumption for ontology merging is crucial: the set of formulas that can be

expressed in principle will be huge and we do not want agents to explicitly provide a judgment on each and every one of them. Also, it is natural to assume that agents may be experts on different domains, i.e., an agent who does not include φ into her ontology is merely expressing that she is agnostic about φ, not that she believes it to be false.

(3) Third, many description logical ontologies make a distinction between *terminological* and *assertional knowledge*, and as we shall see this conceptual distinction can guide the aggregation process. Examples used in the literature on JA often invoke a distinction between "premises" and "conclusions". While this distinction can be meaningful in the context of a specific scenario to be modelled, it has also been criticised for being arbitrary: there is no natural definition of what constitutes a premise and what constitutes a conclusion. In description logics, on the other hand, the distinction of terminological and assertional knowledge is well-defined, and it can thus be exploited also when we are interested in aggregation.

One consequence of not requiring ontologies to be closed under complementation is that the property of *completeness*, i.e., the requirement that any set of formulas involved in an aggregation process should include either φ or its complement, which is a common requirement in JA (List and Puppe, 2009), is not applicable here. Some authors in JA have weakened the completeness requirement to a requirement asking those sets of formulas to be at least *deductively closed* (Gärdenfors, 2006; Dietrich and List, 2008). Because of our open world assumption, we shall also not impose this weaker requirement.

From here on, we shall use the term *ontology aggregation* to refer to our specific approach based on SCT and JA, to distinguish it from the broader and established research area of *ontology merging*. This choice of terminology is intended to stress the focus on the agents providing individual ontologies that is central to our approach. As we shall see, ontology aggregation abstracts away from the particular domain of application of a given ontology and also from the internal structure of an ontology, and instead studies the problem of merging several ontologies by investigating properties of classes of functions that take collections of ontologies—one for each agent—as input and return a single ontology as output.

The remainder of the paper is organised as follows. In Section 2, we define our 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. These properties include adaptions of standard axioms from JA; two kinds of new properties (*groundedness* and *exhaustiveness*), the need for which stems from the lack of closure under complementation and the open world assumption; and "semantic" variants of all of these properties, relating to knowledge that can be *inferred* from a given ontology as opposed to being explicitly present in that ontology. Section 3 concludes with a number of simple *impossibility results* that show that certain combinations of axioms are either not

4

possible at all or only permit very simplistic aggregators that are obviously unattractive in practice. In Section 4 we present a number of concrete aggregation 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 in Section 6.

2 A Framework for Ontology Aggregation

In this section we define our framework for aggregating ontologies. We begin by recalling some basic concepts from modelling ontologies using description logics.

2.1 Ontologies

We take an ontology to be a set of formulas in an appropriate logic, describing our domain of interest. Which logic we use precisely will not turn out to be central to our work here, but as much formal work on ontologies makes use of description logics, we will use description logics for all of our examples. 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 $\mathcal{ALC}$, which is the logic we shall be working with here as well. The following review of the basics of description logics and $\mathcal{ALC}$ is fairly succinct; for full details we refer to the literature (Baader and Nutt, 2003).

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):

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

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 (Baader and Nutt, 2003). 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

---

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

the (interpretation of) C. A set of (TBox and ABox) formulas is *consistent* 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.

Also recall that the *closure* of a set of formulas $\Delta \subseteq \Phi$ (with respect to the full set Φ) is the set of all formulas (in Φ) that logically follow from those in Δ. It is denoted by $\mathrm{Cl}(\Delta) := \{\varphi \in \Phi \mid \Delta \models \varphi\}$.

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 Φ of $\mathcal{ALC}$ formulas over this alphabet that includes all ABox formulas that can be expressed.[2] We call Φ the *agenda* and any set $O \subseteq \Phi$ 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 *consistent* by $\mathrm{On}(\Phi)$.

Let $\mathcal{N} = \{1, \ldots, n\}$ be a finite set of *agents* (or *individuals*, or *experts*). Each agent $i \in \mathcal{N}$ provides a consistent ontology $O_i \in \mathrm{On}(\Phi)$. An *ontology profile* $\boldsymbol{O} = (O_1, \ldots, O_n) \in \mathrm{On}(\Phi)^{\mathcal{N}}$ is a vector of such ontologies, one for each agent. We write $N_\varphi^{\boldsymbol{O}} := \{i \in \mathcal{N} \mid \varphi \in O_i\}$ for the set of agents that include 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 $F : \mathrm{On}(\Phi)^{\mathcal{N}} \to 2^{\Phi}$ mapping any profile of consistent ontologies to an ontology.*

Here 2^{Φ} denotes the powerset of Φ. Observe that, according to this definition, the ontology we obtain as the outcome of an aggregation process need not be consistent. Of course, we will be particularly interested in ontology aggregators that are *consistent*, i.e., aggregators F for which $F(O_1, \ldots, O_n)$ is consistent whenever all O_i are.

2.3 Examples

Two very simple examples for ontology aggregators are the *absolute majority rule* and the *union aggregator* (we shall discuss several more sophisticated aggregators in Section 4).

Definition 2 (Absolute majority rule). *The absolute majority rule is the ontology aggregator F_m mapping any given profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ to the following ontology:*

$$F_m(\boldsymbol{O}) \quad := \quad \{\varphi \in \Phi \mid |N_\varphi^{\boldsymbol{O}}| > \frac{n}{2}\}$$

---

[2]The finite set of TBox formulas in Φ might be all TBox formulas of a certain maximum length or the union of all TBox formulas that a given population of agents chose 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.

Definition 3 (Union aggregator). *The union aggregator is the ontology aggregator F_u mapping any given profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ to the following ontology:*

$$F_u(\boldsymbol{O}) \quad := \quad O_1 \cup \cdots \cup O_n$$

That is, under the absolute majority rule, a formula gets accepted if and only if more than 50% of the individual agents accept it. Under the union aggregator, on the other hand, a formula gets accepted as soon as at least one of the individual agents accepts it. Clearly, under most circumstances the union aggregator will not be a good choice. In particular, it is not a consistent aggregator.

The following example will show that also the absolute majority rule, which looks a lot more attractive at first sight, is not a consistent aggregator. Our example is a simple adaptation of the *doctrinal paradox* familiar from the literature on JA (Kornhauser and Sager, 1993; List and Puppe, 2009). Suppose three agents share a common TBox that consists of two formulas:

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

That is, concept C_3 is defined as the intersection of C_1 and C_2, and any object belonging to C_4 does not also belong to C_3. Furthermore, suppose there is just a single object a in the domain under consideration, i.e., the ABox formulas that can be constructed are $C_1(a)$, $C_2(a)$, $C_3(a)$ and $C_4(a)$. Suppose the ABoxes of our three agents 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

That is, even though all individual ontologies are consistent, the ontology obtained by applying the absolute majority rule is not: by accepting $C_1(a)$ and $C_2(a)$, the group accepts that a belongs to the intersection of C_1 and C_2 and thus also to C_3, which contradicts the collective acceptance of $C_4(a)$.

The standard doctrinal paradox is slightly simpler than our example above and would more closely correspond to a situation where the common TBox only consists of the formula $C_3 \equiv C_1 \sqcap C_2$ and the individual ABoxes are as follows:

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

Observe that this is not a paradox in the same (strong) sense as the earlier example. As before, the group explicitly accepts that a belongs to both C_1 and C_2. Given $C_3 \equiv C_1 \sqcap C_2$, we can now *infer* that the group also accepts $C_3(a)$ to be true, even if this fact is not explicitly recorded in the collective ontology. That is, the only "paradox" we encounter here is that, even though the three individual ontologies are deductively closed (with respect to the set of four formulas under consideration here), this is not the case for the collective ontology we obtain when we apply the absolute majority rule.

3 Properties of Ontology Aggregators

In this section we shall define a number of properties that a given ontology aggregator may or may not satisfy and we analyse the consequences of imposing those properties as design requirements. Most of the properties considered 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 (Gaertner, 2006; List and Puppe, 2009). As in SCT, we refer to these properties as *axioms*.

Most of the axioms we shall introduce are natural requirements, but we stress that we do *not* impose them in general. Some may be more desirable than others for a given problem domain (but all should certainly be considered when designing an ontology aggregator).

3.1 Standard Axioms

The first set of properties we introduce are identical to some of the standard axioms that have been formulated in the literature on JA (List and Puppe, 2009).

The axiom of *unanimity* postulates that when all individual ontologies include φ, then so should the collective ontology. This clearly is a desirable property in any kind of domain.

Definition 4 (Unanimity). *An aggregator F is called unanimous if $O_1 \cap \cdots \cap O_n \subseteq F(\boldsymbol{O})$ for every profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$.*

An aggregator F is *anonymous* if it is symmetric with respect to 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.

Definition 5 (Anonymity). *An aggregator F is called anonymous if for any profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ and any permutation $\pi : \mathcal{N} \to \mathcal{N}$ we have that $F(O_1, \ldots, O_n) = F(O_{\pi(1)}, \ldots, O_{\pi(n)})$.*

F is *independent* if the inclusion of φ in the collective ontology only depends on the pattern of acceptance in the individual ontologies, i.e., if its inclusion is independent of which other formulas may or may not have been included. Independence is a more demanding axiom than the previous two; whether or not it should be imposed certainly is debatable.

Definition 6 (Independence). *An aggregator F is called independent if for any formula $\varphi \in \Phi$ and any two profiles $\boldsymbol{O}, \boldsymbol{O}' \in \mathrm{On}(\Phi)^{\mathcal{N}}$, we have that $\varphi \in O_i \Leftrightarrow \varphi \in O_i'$ for all agents $i \in \mathcal{N}$ implies $\varphi \in F(\boldsymbol{O}) \Leftrightarrow \varphi \in F(\boldsymbol{O}')$.*

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.

Definition 7 (Monotonicity). *An aggregator F is called monotonic if for any agent $i \in \mathcal{N}$, formula $\varphi \in \Phi$, and profiles $\boldsymbol{O}, \boldsymbol{O}' \in \mathrm{On}(\Phi)^{\mathcal{N}}$ 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 at least two possible interpretations of this notion, including these:

Definition 8 (Neutrality). *An aggregator F is called neutral if for any two formulas $\varphi, \psi \in \Phi$ and any profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ we have that $\varphi \in O_i \Leftrightarrow \psi \in O_i$ for all agents $i \in \mathcal{N}$ implies $\varphi \in F(\boldsymbol{O}) \Leftrightarrow \psi \in F(\boldsymbol{O})$.*

Definition 9 (Acceptance-Rejection Neutrality). *An aggregator F is called acceptance-rejection neutral if for any two formulas $\varphi, \psi \in \Phi$ and any profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ we have that $\varphi \in O_i \Leftrightarrow \psi \notin O_i$ for all agents $i \in \mathcal{N}$ 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, which we term *acceptance-rejection neutrality*,[4] is closer to the original neutrality axiom in voting theory proposed by May (1952). It says that if those patterns of acceptance are complementary, then exactly one of the two formulas should be accepted. The reason why we do not consider acceptance-rejection neutrality to be appropriate for ontology aggregation is that it makes the implicit assumption

---

[4]Dietrich and List (2007a) use the name "acceptance-rejection neutrality" for a slightly different axiom: for any $\varphi, \psi \in \Phi$ and $\boldsymbol{O}, \boldsymbol{O}' \in \mathrm{On}(\Phi)^{\mathcal{N}}$, 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.

that not explicitly including a formula into one's knowledge base amounts to actively rejecting the validity of that formula. That is, adopting acceptance-rejection neutrality as a reasonable principle of aggregation presupposes acceptance of the closed world assumption. This is an appropriate assumption in JA, but not here.

3.2   Groundedness and Exhaustiveness

We now introduce a number of properties that are specific to ontology aggregation and that do not have a counterpart in standard SCT or JA. The first such property is *groundedness*, which states that 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 $(O_1, \ldots, O_n)$.

Definition 10 (Groundedness). *An aggregator F is called grounded if $F(\boldsymbol{O}) \subseteq O_1 \cup \cdots \cup O_n$ for every profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$.*

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 (together with consistency) and does not require a separate axiom. Indeed, unanimity and groundedness are closely related properties: an aggregation procedure is grounded if unanimous rejection of a formula always results in collective rejection.

The second axiom we propose is *exhaustiveness:* it should not be possible to add any further formula from the support to the collective ontology without rendering the latter inconsistent. 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.

Definition 11 (Exhaustiveness). *An aggregator F is called exhaustive if for no profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ there exists a formula $\varphi \in O_1 \cup \cdots \cup O_n \setminus F(\boldsymbol{O})$ such that $F(\boldsymbol{O}) \cup \{\varphi\}$ is consistent.*

In other words, exhaustiveness requires collective judgment sets to be maximally consistent with respect to formulas in the support. Our axiom of exhaustiveness plays a similar role as (collective) *completeness* plays in standard JA. In JA, an aggregator is called complete if it ensures that, for any formula φ in the agenda, the group will accept either φ or its complement. Clearly, any such aggregator must also be exhaustive, as accepting any further formulas from the agenda (and thus also the support) would immediately lead to an inconsistency.

Some work in JA has replaced the completeness requirement by the weaker requirement of deductive closure (see, e.g., Gärdenfors, 2006). An aggregator is *deductively closed* if any collective judgment set it produces is deductively closed with respect to formulas *in the agenda*. The corresponding thing to do here would be to require deductive closure only with respect to formulas *in the support*. We call such aggregators *group-closed*.

Definition 12 (Group Closure)**.** *An aggregator F is called group-closed if for no profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ there exists a formula $\varphi \in O_1 \cup \cdots \cup O_n \setminus F(\boldsymbol{O})$ such that $F(\boldsymbol{O}) \models \varphi$.*

As is immediately apparent from the definitions, group closure is a weak form of exhaustiveness (just as, in standard JA, deductive closure is a weak form of completeness):

Fact 1. *Any consistent aggregator F that is exhaustive must also be group-closed.*

Observe that the restriction to *consistent* aggregators is required, because in case $F(\boldsymbol{O})$ is inconsistent the exhaustiveness condition will be satisfied vacuously, but the group closure condition does not have to be satisfied.

3.3 Axioms Expressed in Terms of Inferred Knowledge

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. (2006) in terms of *implicitly* and *explicitly* represented knowledge. We therefore formulate "semantic" (or "implicit") variants of the properties discussed earlier, in which we refer to the closures of ontologies rather than the ontologies themselves. Note that the existing literature on JA only deals with axioms that relate to formulas that occur explicitly in either the collectively accepted set of formulas or in the sets provided by the individual agents. Indeed, as most work in JA postulates that any such set should be deductively closed, there is no need to introduce this kind of distinction there.

Broadly speaking, we obtain the implicit version of an axiom if we replace occurrences of terms referring to sets of formulas by their closure. But in fact there will usually be several ways of doing this. As an initial case study, let us consider the axiom of unanimity, which requires $O_1 \cap \cdots \cap O_n \subseteq F(\boldsymbol{O})$. We define semantic unanimity as follows:

Definition 13 (Semantic unanimity)**.** *An aggregator 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 $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$.*

That is, any formula that can be inferred from each of the individual ontologies should also be derivable from the collective ontology. We also could have chosen a weaker definition: $\mathrm{Cl}(O_1 \cap \cdots \cap O_n) \subseteq \mathrm{Cl}(F(\boldsymbol{O}))$, i.e., any formula that can be inferred from that part of the individual

ontologies that all agents agree on should also be derivable from the collective ontology. In an even weaker formulation, we only apply the closure-operator on the righthand side: $O_1 \cap \cdots \cap O_n \subseteq \mathrm{Cl}(F(\boldsymbol{O}))$, i.e., any unanimously accepted formula should be at least derivable from the collective ontology. The following example demonstrates these two formulations are too weak to capture the intuitive concept of unanimity with respect to inferred knowledge:

$$
\begin{aligned}
O_1 &= \{C_1 \equiv C_2,\, C_1(a)\} \\
O_2 &= \{C_1 \equiv C_2,\, C_2(a)\}
\end{aligned}
$$

Any aggregator that is semantically unanimous in the sense of Definition 13 will include at least one of $C_1(a)$ and $C_2(a)$ in the collective ontology (as well as $C_1 \equiv C_2$, i.e., the second ABox formula will always be derivable), which intuitively is certainly desirable. On the other hand, neither unanimity in the sense of Definition 4 nor the two weaker forms of semantic unanimity sketched above will ensure the same.

An interesting feature of our model is that it allows for stating precisely the relationship between implicitly and explicitly represented knowledge, by investigating the relationship between standard "syntactic" axioms and their semantic counterparts. For unanimity, for instance, we can show that the syntactic version does not entail the semantic version, nor *vice versa*. First, consider the following example (a minor variant from the previous one), which shows that there are syntactically unanimous aggregators that are not semantically unanimous. Suppose three agents share a common TBox including the formulas $C_1 \equiv C_2$ and $C_2 \equiv C_3$, and suppose the ABox of the first agent includes only $C_1(a)$, the second only $C_2(a)$, and the third only $C_3(a)$. Now the majority rule will produce an empty ABox. This violates semantic unanimity, as $C_1(a)$ can be inferred from all three individual ontologies, but not from the collective ontology. 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 inconsistent 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 speaking, semantic unanimity is the (much) stronger property. This intuition can be confirmed for "well-behaved" aggregators:

Proposition 2. *Any consistent and group-closed aggregator that is semantically unanimous must also be unanimous.*

Proof. Take any F that is consistent, group-closed, 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 consistency of F, the outcome $F(\boldsymbol{O})$ is consistent 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, $\varphi \in \mathrm{Cl}(F(\boldsymbol{O}))$. Hence, $\varphi \in O_1 \cup \cdots \cup O_n \setminus F(\boldsymbol{O})$ and $F(\boldsymbol{O}) \models \varphi$. But this violates group closure, and we are done. □

By Fact 1, this entails also that any consistent and exhaustive aggregator that is semantically unanimous will be unanimous.

For independence and neutrality, we propose the following semantic variants:

Definition 14 (Semantic independence). *An aggregator F is called semantically independent if for any formula $\varphi \in \Phi$ and any two profiles $\boldsymbol{O}, \boldsymbol{O}' \in \mathrm{On}(\Phi)^{\mathcal{N}}$, we have that $\varphi \in O_i \Leftrightarrow \varphi \in O_i'$ for all agents $i \in \mathcal{N}$ implies $\varphi \in \mathrm{Cl}(F(\boldsymbol{O})) \Leftrightarrow \varphi \in \mathrm{Cl}(F(\boldsymbol{O}'))$.*

Definition 15 (Semantic neutrality). *An aggregator F is called semantically neutral if for any two formulas $\varphi, \psi \in \Phi$ and any profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ we have that $\varphi \in O_i \Leftrightarrow \psi \in O_i$ for all agents $i \in \mathcal{N}$ implies $\varphi \in \mathrm{Cl}(F(\boldsymbol{O})) \Leftrightarrow \psi \in \mathrm{Cl}(F(\boldsymbol{O}))$.*

Both of these axioms are strictly weaker than the originals. This is attractive: independence is usually considered too demanding a property anyway, and regarding neutrality, it may be considered questionable whether two syntactically distinct formulas always need to be treated symmetrically. For instance, if one formula entails fewer agenda formulas than another (i.e., when it is less likely to cause an inconsistency), we might be inclined to accept the former, but not the latter, even if they have the same support. Our axiom of semantic neutrality does not give up on this condition, but it does weaken the consequences.

Next, we turn to the axiom of exhaustiveness. In its plain syntactic variant, exhaustiveness postulates that we should amend $F(\boldsymbol{O})$ with formulas from the support $O_1 \cup \cdots \cup O_n$ as long as this does not render $F(\boldsymbol{O})$ inconsistent. What would be an appropriate definition of semantic exhaustiveness, i.e., to which parts of Definition 11 should we apply the closure operator? First, note that $F(\boldsymbol{O}) \cup \{\varphi\}$ is consistent if and only if its closure is consistent, so this set is not a good candidate for application of the closure operator. In the light of our discussions of possible ways of defining semantic unanimity, there are two natural ways of defining semantic exhaustiveness. One of them would require that there never is a formula $\varphi \in O_1 \cup \cdots \cup O_n \setminus \mathrm{Cl}(F(\boldsymbol{O}))$ such that $F(\boldsymbol{O}) \cup \{\varphi\}$ is consistent. This property is strictly weaker than exhaustiveness (which already is a very weak requirement), and we shall not discuss it any further here. Instead, we opt for this definition:

Definition 16 (Semantic exhaustiveness). *An aggregator F is called semantically exhaustive if for no profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ there exists a formula $\varphi \in \mathrm{Cl}(O_1) \cup \cdots \cup \mathrm{Cl}(O_n) \setminus \mathrm{Cl}(F(\boldsymbol{O}))$ such that $F(\boldsymbol{O}) \cup \{\varphi\}$ is consistent.*

That is, an aggregator is semantically exhaustive if adding any formula derivable from one of the individual ontologies but not the collective ontology to the collective ontology would render it inconsistent.

13

Finally, let us consider the axiom of groundedness. We shall argue that there is no interesting notion of semantic groundedness. Recall that groundedness requires $F(\boldsymbol{O}) \subseteq O_1 \cup \cdots \cup O_n$. We could apply the closure operator to each of the O_i, to their union, and to $F(\boldsymbol{O})$. First, applying it to the full righthand side of the condition is not interesting: $O_1 \cup \cdots \cup O_n$ will be an inconsistent set of formulas for any nontrivial example, i.e., $\mathrm{Cl}(O_1 \cup \cdots \cup O_n)$ would be the set of all formulas. Second, the formulation $F(\boldsymbol{O}) \subseteq \mathrm{Cl}(O_1) \cup \cdots \cup \mathrm{Cl}(O_n)$ is even weaker than the plain groundedness axiom, which already is a very weak requirement. Third, $\mathrm{Cl}(F(\boldsymbol{O})) \subseteq O_1 \cup \cdots \cup O_n$ is clearly not a good choice: it would only be satisfied in the most extreme cases, e.g., when $F(\boldsymbol{O})$ only includes formulas from, say, O_1 and O_1 is deductively closed. This leaves $\mathrm{Cl}(F(\boldsymbol{O})) \subseteq \mathrm{Cl}(O_1) \cup \cdots \cup \mathrm{Cl}(O_n)$, which 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).

3.4 Impossibility Results

We have seen a range of, mostly, desirable properties for ontology aggregators, but we have not yet discussed whether it is actually possible to design aggregators that will satisfy those properties. In SCT, an important line of work has been about establishing *impossibility results*, which show either that for a certain combination of properties there exists no aggregator that satisfies them all or that the only aggregators that do are obviously unattractive (e.g., they might be *dictatorial* in the sense of only taking into account the view of a single fixed individual). We are now going to see a number of results of this kind.

First, we have argued that exhaustiveness is a very natural requirement to ask for in an aggregator and that group closure is a particularly weak form of this requirement. We are now in a position to make our objection to the axiom of acceptance-rejection neutrality more precise, by showing that it is impossible to design an aggregator that is both group-closed (or exhaustive) and acceptance-rejection neutral.

Proposition 3. *Any aggregator that satisfies acceptance-rejection neutrality violates both exhaustiveness and group closure.*

Proof. Suppose the agenda includes two logically equivalent but syntactically distinct formulas φ and ψ. Furthermore, suppose $\varphi \in O_i \Leftrightarrow \psi \notin O_i$ for all agents $i \in \mathcal{N}$, and φ and ψ are each accepted by at least one agent. Then any aggregator satisfying acceptance-rejection neutrality cannot accept both φ and ψ. On the other hand, any exhaustive or group-closed

aggregator must accept both of them.   □

Next, we turn to impossibilities triggered by the axiom of semantic unanimity. As we have seen, semantic unanimity is, broadly speaking, *stronger* than plain unanimity (see Proposition 2). We shall now illustrate that it is actually *much* stronger. To position the next result, recall that the seminal impossibility result in the field of JA, due to List and Pettit (2002), states that for agendas with a certain minimal structural complexity (such as including at least two distinct propositions and their conjunction), there exists no aggregator that is anonymous, neutral, independent, consistent and complete. The crucial requirements here are independence and consistency, which are both very demanding. In our next result, we remove the requirement of consistency and instead work with semantic unanimity.[5] As we shall see, this does not render the task of defining an aggregator impossible, but it restricts us to two simplistic and fairly unattractive options. One of them is the union aggregator. The other is what we call the *indiscriminate aggregator:* for any profile, it will accept every single formula in the agenda.

Proposition 4. *The only aggregators that are anonymous, neutral, independent and semantically unanimous are the union aggregator and the indiscriminate aggregator.*

Proof. Let F be an ontology aggregator that is anonymous, neutral, independent and semantically unanimous. By anonymity and independence, collective acceptance of a formula φ only depends on the cardinality of the coalition of agents accepting φ. By neutrality, these acceptable cardinalities do not depend on φ, but are the same for all formulas. That is, F must be definable in terms of a function $f : \Phi \to \mathbb{N} \cup \{0\}$, mapping each formula in the agenda to the number of agents accepting it in a given profile, and a set of numbers $Accept \subseteq \mathbb{N} \cup \{0\}$, as follows: $\varphi \in F(\boldsymbol{O})$ if and only if $f(|N_\varphi^{\boldsymbol{O}}|) \in Accept$. Note that F is the union aggregator exactly in case $Accept = \{1, \ldots, n\}$ and that F is the indiscriminate aggregator exactly in case $Accept = \{0, \ldots, n\}$

For the sake of contradiction, assume there exists a $k \in \{1, \ldots, n\}$ with $k \notin Accept$. We shall construct a particular agenda Φ and a particular profile $\boldsymbol{O}$ where F is bound to violate semantic unanimity: Introduce n atomic concepts C_1, ..., C_n, and one object name a. Suppose all agents accept the TBox formulas $C_1 \equiv C_2$, $C_2 \equiv C_3$, ..., $C_{n-1} \equiv C_n$. Furthermore, suppose each ABox formula of the form $C_i(a)$ (for $i = 1, \ldots, n$) is accepted by exactly k agents; specifically, agents $i, i+1, \ldots, (i+k-1 \bmod n)$ all accept $C_i(a)$. Observe that this means that—whatever the value of k may be exactly—every agent accepts at least one of $C_1(a)$, ..., $C_n(a)$. Hence, by semantic unanimity, every agent should accept all of

---

these ABox formulas. But that would mean that $k \in Accept$, i.e., we have obtained a contradiction. $\square$

If we add the axiom of groundedness to Proposition 4, then we obtain a full characterisation of the union aggregator:

Corollary 5. *The only aggregator that is grounded, anonymous, neutral, independent and semantically unanimous is the union aggregator.*

Proof. Immediate from Proposition 4 and the fact that the indiscriminate aggregator is not grounded. $\square$

As is clear from the proof of Proposition 4, our results here show that there *exist* agendas (namely those agendas including all of the agenda formulas used in the proof) for which any aggregator satisfying the properties stated must agree with the union aggregator or the indiscriminate aggregator. The proof also shows that the requirements on an agenda to allow us to derive these negative results are very weak: all that is required are the $\equiv$-connective and a sufficiently large number of atomic concepts. This is similar to the impossibility theorem of List and Pettit (2002), who prove their impossibility for any agenda including at least two propositions and their conjunction. The result of List and Pettit has later been strengthened in the form of *agenda characterisation theorems* that characterise the exact class of agendas on which a particular impossibility occurs (List and Puppe, 2009). Whether our results can be extended in an analogous way is an interesting question that remains open at this stage.

Finally, note that, if we restrict attention to group-closed aggregators, then the semantic versions of neutrality and independence coincide with their usual (syntactic) counterparts. Thus, we can rephrase in our setting the impossibility results in JA that characterise oligarchic rules (Dietrich and List, 2008). In particular, the only consistent aggregator that satisfies anonymity, semantic neutrality, semantic independence and group closure is the *unanimous aggregator*, which accepts a formula if and only if every individual does, i.e., it is the oligarchic rule that takes the full set $\mathcal{N}$ as the oligarchy. Hence, weakening independence and neutrality to their semantic variants does in fact not allow us to circumvent the problems highlighted by existing impossibility results.

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 consistent 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

16

for real-world ontology aggregation. Rather, our intention is to provide a catalogue of basic aggregators that can serve as building blocks for constructing more sophisticated procedures in the future. Fully understanding the properties of these basic aggregators is a necessary step towards designing more advanced procedures.

4.1 The Absolute Majority Rule

We have already introduced the (absolute) majority rule in Section 2.3. We have seen that the majority rule can produce inconsistent collective ontologies. Following Endriss et al. (2010), we call an agenda Φ *safe* for a given aggregator F if $F(\boldsymbol{O})$ is consistent for any profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$. We will now identify necessary and sufficient conditions for the safety of Φ under the majority rule.

Adapting the terminology from JA (List and Puppe, 2009), we recall that an agenda Φ satisfies the *median property* if and only if every inconsistent set $X \subseteq \Phi$ contains itself an inconsistent set Y with cardinality at most 2. Now a simple reformulation of a known result due to Nehring and Puppe shows that an agenda Φ is safe for the majority rule if and only if it satisfies the median property (Nehring and Puppe, 2007; List and Puppe, 2009; Endriss et al., 2010). This result can be refined if we put restrictions on the range of profiles on Φ that we consider. The use of description logics to specify 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 Φ satisfies the $\mathcal{T}$-*median property* if and only if for every set of ABox formulas $X \subseteq \Phi^A$ such that $\mathcal{T} \cup X$ is inconsistent there exists a set $Y \subseteq X$ with cardinality at most 2 such such $\mathcal{T} \cup Y$ is also inconsistent. We obtain the following characterisation:

Proposition 6. *The absolute majority rule will return a consistent ontology for any profile with a common TBox $\mathcal{T}$ if and only if the agenda Φ satisfies the $\mathcal{T}$-median property.*

Proof. One direction follows immediately from the (first) variant of the doctrinal paradox discussed in Section 2.3. The agenda used in that example violates the $\mathcal{T}$-median property[6] and the absolute majority does indeed not always return a consistent ontology.

For the other direction, for the sake of contradiction, assume the $\mathcal{T}$-median property holds but $F_m(\boldsymbol{O})$ is inconsistent. By definition of the absolute majority rule, the TBox of $F_m(\boldsymbol{O})$ is exactly the common TBox $\mathcal{T}$. Thus, by the $\mathcal{T}$-median property, there must be a set Y of ABox formulas in $F_m(\boldsymbol{O})$ with $|Y| \leqslant 2$ such that $\mathcal{T} \cup Y$ is inconsistent. Now consider the possible cardinalities of Y:

---

[6]No subset of two formulas of the set of ABox formulas $\{C_1(a), C_2(a), C_4(a)\}$ together with the common TBox $\mathcal{T} = \{C_3 \equiv C_1 \sqcap C_2,\ C_4 \sqsubseteq \neg C_3\}$ is inconsistent, while $\mathcal{T} \cup \{C_1(a), C_2(a), C_4(a)\}$ is inconsistent.

- First, Y cannot be empty, as that would mean that $\mathcal{T}$ is inconsistent, contradicting our assumption that individual ontologies are consistent.

- 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 consistent.

- Third, suppose that $|Y| = 2$ with $Y = \{\varphi, \psi\}$. These formulas could only have been accepted by F_m if $|N^{\boldsymbol{O}}_\varphi| > \frac{n}{2}$ and $|N^{\boldsymbol{O}}_\psi| > \frac{n}{2}$. 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 consistent.

That is, we obtain a contradiction in every possible case. $\qquad\square$

We stress that the technical insight connecting the median property with the inconsistency of the majority rule is entirely due to Nehring and Puppe (2007). Besides giving a particularly simple proof here, our only novel addition to this insight is that it is possible to apply the concept of the median property to only a part of the agenda (here the ABox) and obtain the same kind of result. In fact, the result we obtain is, technically speaking, stronger (as it includes the special case where the part of the agenda to which we apply the median property is the full agenda).

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) and 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 17 (Quota rules)**.** *Let $q \in [0,1]$. The quota rule with quota q is the ontology aggregator F_q with $F_q(\boldsymbol{O}) = \{\varphi \in \Phi \mid |N^{\boldsymbol{O}}_\varphi| \geqslant q \cdot n\}$ for all $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$.*

We could also generalise further and allow different quotas for different formulas; Dietrich and List (2007b) make a distinction between general and *uniform* quota rules. Observe that we obtain the absolute majority rule for $q = \frac{1}{2} + \epsilon$ for any positive $\epsilon < \frac{1}{n}$. Also observe that

for $0 < q \leqslant \frac{1}{n}$ the aggregator F_q is equal to the union aggregator F_u. The indiscriminate aggregator is the trivial quota rule F_q with $q = 0$.

As noted by Dietrich and List (2007b), any quota rule of the kind defined above is (syntactically) anonymous, neutral, independent, and monotonic. Hence, by Proposition 4, the only semantically unanimous quota rules are those with a quota $q \leqslant \frac{1}{n}$.

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 inconsistent. 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 Φ such that $\varphi \gg_{\boldsymbol{O}} \psi$ implies $|N_\varphi^{\boldsymbol{O}}| \geqslant |N_\psi^{\boldsymbol{O}}|$ for all $\varphi, \psi \in \Phi$.[7]

Definition 18 (Support-based procedure). *Given a priority rule $\gg$, the support-based procedure with $\gg$ is the ontology aggregator $\mathrm{SBP}_\gg$ mapping any profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ to $\mathrm{SBP}_\gg(\boldsymbol{O}) := \Delta$ for the unique set $\Delta \subseteq \Phi$ for which $\varphi \in \Delta$ if and only if*

(i) $N_\varphi^{\boldsymbol{O}} \neq \emptyset$ *and*

(ii) $\{\psi \in \Delta \mid \psi \gg_{\boldsymbol{O}} \varphi\} \cup \{\varphi\}$ *is consistent.*

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). As we have discussed before, both neutrality and independence have been argued to be overly restrictive before and are central to a number of known impossibility results, so dropping them from our list of desiderata may be considered acceptable.

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 Φ. For JA, such rules where formulas are considered in some arbitrary but fixed order have been discussed by List (2004). This is sometime referred to as the *sequential rule* in JA. Each choice of $\gg$ corresponds to a different *greedy procedure* that attempts to

---

[7]Independently from our initial work on this procedure (Porello and Endriss, 2011), Lang et al. (2011) have proposed the same kind of rule and furthermore pointed out that it is a natural counterpart to Tideman's "ranked pairs" rule in voting theory (Tideman, 1987).

accept as many formulas as possible without violating consistency, in order of priority as specified by $\gg_\varphi$. For instance, take a priority rule $\gg$ for which $\varphi \gg_{\boldsymbol{O}} \psi$ holds whenever $N_\varphi^{\boldsymbol{O}} \supseteq N_\psi^{\boldsymbol{O}}$ is the case—but not necessarily whenever $|N_\varphi^{\boldsymbol{O}}| \geqslant |N_\psi^{\boldsymbol{O}}|$. It will be appropriate for aggregating 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 $N_\varphi^{\boldsymbol{O}}$. 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 An Asymmetric Distance-Based Procedure

In voting theory, many voting rules can be defined using a notion of distance (i.e., a real-valued function satisfying symmetry, non-negativity, identity of indiscernibles, and the triangular inequality) between a profile and a particular outcome. The well-known *Kemeny rule* is a natural example (Gaertner, 2006). Similar ideas have also been used in belief merging (Konieczny and Pino Pérez, 2002) and then exported to JA (Pigozzi, 2006).

We will now define an aggregation procedure that chooses from a class of acceptable ontologies (namely the consistent 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 $d : 2^\Phi \times 2^\Phi \to \mathbb{N} \cup \{0\}$ with $d(X, Y) := |\{\varphi \mid \varphi \in X \text{ and } \varphi \notin Y\}|$ a distance.

Definition 19 (Distance-based procedure). *The distance-based procedure is the (irresolute) ontology aggregator* DBP *mapping any profile* $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ *to the following set of consistent ontologies:*

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

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 consistent by construction.

Note that if we choose a tie-breaking rule that selects a maximal set (with respect to set-inclusion), then the DBP will always return a maximally consistent 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 (List and Pettit, 2002). 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 20 (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:*

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

That is, we first use some aggregator of choice on the ABoxes alone. A natural choice for the procedure used in the first step would be the absolute majority rule. The result is a collective ABox, which we impose on every individual. Finally, we use the distance-based rule on these new individual ontologies (with the collective ABox and the individual TBoxes).

Observe that it is possible that the collective TBox obtained in this manner is empty. An interesting variant of this approach would be to allow agents to revise their TBoxes themselves after the collective ABox has been fixed.

An assertion-based procedure stresses the information coming from the ABox. Similarly, we may want to give priority to TBox information and first aggregate TBoxes, then fix a TBox, and finally aggregate ABoxes.

5 Related Work

The problem of modelling ontology change touches on a broad range of issues, and research in this area has dealt with a vast number of interrelated phenomena, such as updating after new information has arrived, revision, or debugging for inconsistencies (Bruijn et al., 2004; Flouris et al., 2008). Contributions to ontology merging range from sophisticated engineering solutions (see, e.g., McGuinness et al., 2000; Noy and Musen, 2000, 2003), to work in belief revision based on mathematical logic (see, e.g., Ribeiro and Wassermann, 2009).

The very phrase *ontology merging* might refer to several types of issues in ontology change. For a precise analysis of the types of operations that are viewed as ontology merging, we refer to the surveys by Pinto et al. (1999) and Flouris et al. (2008). For example, ontology merging might refer to heterogeneity resolution as well as integration of compatible sources. Moreover, ontology merging is usually viewed as a part of a more articulated process that includes ontology alignment and mapping (Chalupsky, 2000). These operations are usually performed by checking correspondences or overlaps between ontologies, by means of manual or semi-automated tools: the tool suggests what part of the ontology has to be manipulated, but the final choice is performed by the ontology engineer.

The study of ontology aggregators that we have initiated here is more closely related to the application of belief merging and revision to ontologies defined in the language of description logics. In particular, we view the study of operations on ontologies as an abstract problem of defining suitable aggregation functions, rather than dealing with concrete instances of particular diverging, conflicting, or complementary ontologies. Applications of AGM belief revision to ontology merging and debugging have been discussed, for instance, by Ribeiro and Wassermann (2009). The use of belief revision for description logics has also been discussed by Meyer et al. (2005). An important difference with respect to our approach is that we do not assume any information concerning the reliability of the sources, as we believe that the cost of retrieving such information might be too high in many scenarios.

The relationship between positive and negative information and the open world assumption is closely related to the interpretation of negation in description logics. In particular, Flouris et al. (2006) present a deep analysis of several types of negation and the related notions of inconsistency. Here we decided to work with a standard description-logical language and to use rationality constraints and properties of aggregators to model the relationship between positive and negative information, without enriching the language.

Moreover, the distinction we have formalised between semantic and syntactic axioms is inspired by the analysis of implicit *vs.* explicit knowledge developed by Flouris et al. (2006). We believe that our approach provides a clear formal understanding of this distinction.

The application of social choice-theoretic insights to belief merging has been investigated for example by Everaere et al. (2007). The notion of strategy-proof merging operator points

22

at an axiomatic analysis of merging operators that is close to the methodology we have explored here. However, the results obtained by Everaere et al. (2007) do not directly deal with ontologies; that is, the operators do not work on agendas defined in description logics. Moreover, they do not deal with the distinction between implicit and explicit knowledge. Still, we do believe that our axiomatic approach is compatible with the approach taken in belief merging and that a further investigation of the connections could prove fruitful.

Our approach is of course closely connected to the JA literature, as reviewed, for instance, by List and Puppe (2009). Thus, it is important to point out some significant differences. Our discussion of the open world assumption is related to the rationality assumptions on the sets of judgments in JA. The weakening of completeness of the collective judgment set to deductive closure has been discussed by Gärdenfors (2006) and Dietrich and List (2008), amongst others. As the results of these authors show, even when we weaken completeness to deductive closure, the impossibility results are still compelling: instead of characterising dictatorships, aggregators that return consistent and deductively closed judgments sets are oligarchic. Weaker rationality conditions on both individual and collective sets have been discussed by Dietrich and List (2007a). However, their framework does include a powerful axiom that requires symmetric treatment of accepted and rejected propositions across profiles, which, as we have argued in the context of the discussion of our own variant of acceptance-rejection neutrality, presupposes a closed world assumption, as it forces us to give equal weight to negative and positive information. Other possible attitudes towards individual information have been discussed by Dokow and Holzman (2010), who introduce the possibility of abstention concerning an issue. The model generalises the binary case by introducing a third possible value besides accepting and rejecting. Even if the abstention attitude might model a lack of information, we believe that this approach still endorses a notion of negative information that is not compatible with the open world assumption.

Our definitions of distance-based procedures are related to the work on distance-based JA (Pigozzi, 2006; Miller and Osherson, 2009) and belief merging (Konieczny and Pino Pérez, 2002; Konieczny et al., 2004). The most significant difference is that we need to deal with asymmetric distances in order to cope with the open world assumption.

Another important aspect of the JA approach that is related to our work is the distinction between premises and conclusions of a decision problem (List and Pettit, 2002; Dietrich and Mongin, 2010). In particular, the doctrinal paradox can be viewed as a situation where a premise-based procedure and a conclusion-based procedure lead to different (mutually incompatible) outcomes. However, when we deal with agendas of formulas in standard propositional logic, it is not clear how to actually define premises and conclusions independently of the decision problem at issue. By working with description logics and ontologies, on the other hand, we already have a clear partition of the agenda into premises (the TBox) and

conclusions (the ABox). This feature allows us to present a clear distinction between *doctrinal paradoxes* (where the legal doctrine, i.e., the logical connections, are shared amongst the agents) and instances of the *discursive dilemma* (where the logical connections are not unanimously accepted) in terms of the distinction between shared and non-shared TBoxes.

6 Conclusion

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, expressed in terms of making appropriate use of the formulas in the support of an aggregation problem. We have then presented relevant results concerning those axioms and we have introduced several simple ontology aggregation procedures, 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.

We stress that we view our methodology as providing a complementary tool to existing work on ontology merging, intended to provide the ontology engineer with new desiderata (i.e., the axioms) that can guide the choice between alternative solutions, at least in the case of coarse merging. To be sure, the axioms we have presented define properties that might only be desirable for particular instances of ontology merging. Moreover, the aggregators we presented only provide a first base line for procedures that satisfy such properties. Their great simplicity both makes them ideal candidates for studying fundamental properties of ontology merging *and* certainly means that in their own right they will not be sufficiently sophisticated to yield good results in practice. We leave investigations regarding the interfacing of more practice-oriented ontology merging tools with our abstract axiomatic analysis for future work.

Concerning future work, we believe that our approach grounded in social choice theory may also hold useful insights 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 amongst several possible ways of building concept definitions. We also believe that our work can provide an interesting starting point for future research within 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 choice.

References

F. Baader and W. Nutt. Basic description logics. In *The Description Logic Handbook*. Cambridge University Press, 2003.

J. d. Bruijn, F. Martín-Recuerda, D. Manov, and M. Ehrig. State-of-the-art survey on ontology merging and aligning v1. Deliverable D4.2.1, SEKT: EU-IST Integrated Project IST-2003-506826, 2004.

H. Chalupsky. OntoMorph: A translation system for symbolic knowledge. In *Proc. 7th International Conference on Principles of Knowledge Representation and Reasoning (KR-2000)*. Morgan Kaufmann, 2000.

F. Dietrich and C. List. Judgment aggregation with consistency alone. Working paper, London School of Economics, 2007a.

F. Dietrich and C. List. Judgment aggregation by quota rules: Majority voting generalized. *Journal of Theoretical Politics*, 19(4):391–424, 2007b.

F. Dietrich and C. List. Judgment aggregation without full rationality. *Social Choice and Welfare*, 31(1):15–39, 2008.

F. Dietrich and P. Mongin. The premiss-based approach to judgment aggregation. *Journal of Economic Theory*, 145(2):562–582, 2010.

E. Dokow and R. Holzman. Aggregation of binary evaluations with abstentions. *Journal of Economic Theory*, 145(2):544–561, 2010.

U. Endriss, U. Grandi, and D. Porello. Complexity of judgment aggregation: Safety of the agenda. In *Proc. AAMAS-2010*, 2010.

P. Everaere, S. Konieczny, and P. Marquis. The strategy-proofness landscape of merging. *Journal of Artificial Intelligence Research*, 28:49–105, 2007.

G. Flouris, Z. Huang, J. Z. Pan, D. Plexousakis, and H. Wache. Inconsistencies, negations and changes in ontologies. In *Proc. 21st National Conference on Artificial Intelligence (AAAI-2006)*, 2006.

G. Flouris, D. Manakanatas, H. Kondylakis, D. Plexousakis, and G. Antoniou. Ontology change: Classification and survey. *Knowledge Engineering Review*, 23(2):117–152, 2008.

W. Gaertner. *A Primer in Social Choice Theory*. Oxford University Press, 2006.

P. Gärdenfors. A representation theorem for voting with logical consequences. *Economics and Philosophy*, 22(2):181–190, 2006.

S. Konieczny and R. Pino Pérez. Merging information under constraints: A logical framework. *Journal of Logic and Computation*, 12(5):773–808, 2002.

S. Konieczny, J. Lang, and P. Marquis. DA^2 merging operators. *Artificial Intelligence*, 157 (1–2):49–79, 2004.

L. A. Kornhauser and L. G. Sager. The one and the many: Adjudication in collegial courts. *California Law Review*, 81(1):1–59, 1993.

J. Lang, G. Pigozzi, M. Slavkovik, and L. van der Torre. Judgment aggregation rules based on minimization. In *Proc. 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2011)*, 2011.

C. List. A model of path-dependence in decisions over multiple propositions. *Amerisan Political Science Review*, 98(3):495–513, 2004.

C. List and P. Pettit. Aggregating sets of judgments: An impossibility result. *Economics and Philosophy*, 18(1):89–110, 2002.

C. List and C. Puppe. Judgment aggregation: A survey. In *Handbook of Rational and Social Choice*. Oxford University Press, 2009.

K. O. May. A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica*, 20(4):680–684, 1952.

D. McGuinness, R. Fikes, J. Rice, and S. Wilder. An environment for merging and testing large ontologies. In *Proc. 7th International Conference on Principles of Knowledge Representation and Reasoning (KR-2000)*. Morgan Kaufmann, 2000.

T. A. Meyer, K. Lee, and R. Booth. Knowledge integration for description logics. In *Proc. 20th National Conference on Artificial Intelligence (AAAI-2005)*, 2005.

M. K. Miller and D. Osherson. Methods for distance-based judgment aggregation. *Social Choice and Welfare*, 32(4):575–601, 2009.

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.

N. F. Noy and M. A. Musen. Algorithm and tool for automated ontology merging and alignment. In *Proc. 17th National Conference on Artificial Intelligence (AAAI-2000)*, 2000.

N. F. Noy and M. A. Musen. The PROMPT suite: Interactive tools for ontology merging and mapping. *International Journal of Human-Computer Studies*, 59(6):983–1024, 2003.

G. Pigozzi. Belief merging and the discursive dilemma: An argument-based account of paradoxes of judgment aggregation. *Synthese*, 152(2):285–298, 2006.

H. S. Pinto, A. Gómez-Pérez, and J. P. Martins. Some issues on ontology integration. In *Proc. IJCAI-1999 Workshop on Ontologies and Problem-Solving Methods (KRR5)*. CEUR Workshop Proceedings, Vol. 18, 1999.

D. Porello and U. Endriss. Ontology merging as social choice. In *Proc. 12th International Workshop on Computational Logic in Multiagent Systems (CLIMA-2011)*, volume 6814 of *LNAI*, pages 157–170. Springer-Verlag, 2011.

M. M. Ribeiro and R. Wassermann. Base revision for ontology debugging. *Journal of Logic and Computation*, 19(5):721–743, 2009.

N. Tideman. Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, 4(3):185–206, 1987.