Epistemic Argumentation Framework: Theory and Computation

Chiaki Sakama   SAKAMA@WAKAYAMA-U.AC.JP
Wakayama University
930 Sakaedani, Wakayama 640-8510 Japan

Tran Cao Son   TSON@CS.NMSU.EDU
New Mexico State University
Las Cruces, NM 88003, USA

Abstract

The paper introduces the notion of an *epistemic argumentation framework* (EAF) as a means to integrate the beliefs of a reasoner with argumentation. Intuitively, an EAF encodes the beliefs of an agent who reasons about arguments. Formally, an EAF is a pair of an argumentation framework and an *epistemic constraint*. The semantics of the EAF is defined by the notion of an ω-*epistemic labelling set*, where ω is complete, stable, grounded, or preferred, which is a set of ω-labellings that collectively satisfies the epistemic constraint of the EAF. The paper shows how EAF can represent different views of reasoners on the same argumentation framework. It also includes representing preferences in EAF and multi-agent argumentation. Finally, the paper discusses complexity issues and computation using epistemic logic programming.

1. Introduction

Rational agents often claim that they make their decision based on their knowledge and beliefs when facing alternative and conflicting choices. Consider two examples:

- On January 15, 2019, British Prime Minister's Theresa May suffered a humiliating defeat in the vote on the Brexit deal; 432 Members of Parliament (MPs) voted against the deal while 202 were for it.[1] The MPs who voted against the deal believe that the deal is bad for Britain. Those who voted for the deal believe that the deal is the best that Britain can get.

- In the US presidential election, a voter selects one candidate from a set of candidates (often only two candidates). Everyone claims that he/she has made the "right choice."

In each scenario above, an agent (an MP or a voter) listens to various arguments, which either support or reject a potential decision, and then opts for one among the possibilities, which he/she believes is the right choice. In each situation, the arguments supporting/against a choice, their counter-arguments, etc. can be easily encoded in an *abstract argumentation framework* (AF) introduced by Dung (1995). For instance, $AF = (\{(a)ccept, (r)eject\}, \{(a,r),(r,a)\})$, having two arguments mutually attacking each other, represents (in its most condensed form) the AF that the MPs have for making their choice about the Brexit's deal. Given arguments made by each agent in each scenario, an argumentation semantics of the corresponding AF provides the result of rational reasoning. The stable semantics of the above AF supports two alternative choices, $\{a\}$ or $\{r\}$, while the grounded semantics of the AF supports "no decision", i.e., the empty extension. As such,

---

1. "*Brexit vote*", Jan. 15th, 2019. `washingtonpost.com`

it would likely result in the unanimous choice by all agents who participate in argumentation and claim that they are rational.

The above discussion raises the question "how to express an agent's opinion for supporting an argument among conflicting arguments in the outcome of an AF?" Arguably, there are two possibilities: the agent modifies the AF so that the new AF supports his/her choice or the agent is simply biased towards his/her conclusion.

In the first case, nothing other than the agent's beliefs could influence his/her choice of arguments and/or attacks that lead to the new AF, which ultimately leads to his/her conclusion. In this approach, a modified AF represents objective evidences and subjective beliefs indistinguishably. If one merges objective evidences (normally invariant) and subjective beliefs (possibly variant) in a single AF, however, it must be revised whenever an agent changes its own belief. Moreover, it would become hard to distinguish subjective beliefs from objective evidences in a personally customized AF. In this respect, it is desirable to have a mechanism that can distinguishably represent subjective beliefs (or biases) of agents as well as objective evidences as an AF.

In the second case, biases, reflecting beliefs of agents, could be viewed as agents' preferences. Furthermore, there is a huge amount of literature in AF on dealing with preferences in argumentation. It is therefore instructive to consider whether previously developed approaches to dealing with preferences would be sufficient to capture biases. In most approaches in abstract AF, the key idea is to extend an AF with a syntactic component that records the preferences such as a preference relation among arguments or an attack relation between arguments and attacks, and then define a new semantics for this extended AF (detailed discussion is in Section 5). Indirectly, these approaches also merge the objective evidences (the original AF) with subjective biases and define new notions of extensions for AF with preferences. In addition, approaches to dealing with preferences have thus far only considered biases/preferences between arguments (e.g., prefer an argument over another one) or preferences between arguments and attacks. However, preference may change in accordance with the change of beliefs. In June 2016, 52% of UK voters support Brexit and 48% oppose it, while the rate changes after 3 years of debate and negotiation. Such a dynamic change of preference is hard to specify using static preference relations among individual arguments. Then we represent preferences as a formula over epistemic literals, and realize a change of preference as revision of beliefs. When an argumentation framework has multiple extensions (or labellings) and agents support different conclusions, *judgment aggregation* focuses on resolution of disagreement between agents (e.g., Caminada & Pigozzi, 2011; Rahwan & Tohmé, 2010). However, most studies do not provide any mechanism explaining why an agent chooses particular extensions. We represent private belief of an agent towards an AF, which explains why an agent supports particular conclusions. We then argue agreement and majority voting in multi-agent argumentation.

In this paper, we propose an approach to incorporate agents' beliefs into an argumentation framework. Specifically, we propose an extension of AF, called *epistemic argumentation framework* (EAF). EAF introduces the third component to an AF, *an epistemic constraint*, that represents the belief of an agent given an AF. We study formal properties of EAF and show that it can be used in representing preferences and decision making in multi-agent environments. We also investigate computational complexity and provide a procedure for computing EAF.

The rest of the paper is organized as follows. Section 2 reviews basic notions of argumentation frameworks used in this paper. Section 3 introduces epistemic argumentation frameworks and addresses its applications. Section 4 presents computation of EAF. Section 5 discusses related issues and Section 6 concludes the paper. This paper is an extended version of the paper by Sakama and

Son (2019) which does not include proofs of propositions. Revision is made throughout the paper and new considerations are often added. Section 4 is also new in this paper.

2. Argumentation Framework

This paper uses *(abstract) argumentation frameworks* introduced by Dung (1995).

An *argumentation framework* (AF) is a pair (Ar, att) where Ar is a (finite) set of *arguments* and $att \subseteq Ar \times Ar$. We write $a \to b$ (say, *a attacks b*) iff $(a, b) \in att$. We say that a *indirectly attacks* b if there is a finite sequence $x_0, ..., x_{2n+1}$ $(n \geq 1)$ such that $a = x_0$ and $b = x_{2n+1}$ and for each $0 \leq i \leq 2n$, $(x_i, x_{i+1}) \in att$.

For the semantics of AFs, we use the labelling-based semantics introduced by Caminada and Gabbay (2009). A *labelling* of (Ar, att) is a (total) function $\mathcal{L} : Ar \to \{ \text{in, out, und} \}$. When $\mathcal{L}(a) = \text{in}$ (resp. $\mathcal{L}(a) = \text{out}$ or $\mathcal{L}(a) = \text{und}$) for an argument $a \in Ar$, it is written as $\text{in}(a)$ (resp. $\text{out}(a)$ or $\text{und}(a)$). In this case, the argument a is said to be *accepted* (resp. *rejected* or *undecided*) in $\mathcal{L}$. Given $AF = (Ar, att)$ and a labelling $\mathcal{L}$, define $\text{in}(\mathcal{L}) = \{ x \mid \mathcal{L}(x) = \text{in for } x \in Ar \}$, $\text{out}(\mathcal{L}) = \{ x \mid \mathcal{L}(x) = \text{out for } x \in Ar \}$, and $\text{und}(\mathcal{L}) = \{ x \mid \mathcal{L}(x) = \text{und for } x \in Ar \}$. A labelling $\mathcal{L}$ of (Ar, att) is also represented as a set $S(\mathcal{L}) = \{ \lambda(x) \mid \mathcal{L}(x) = \lambda \text{ for } x \in Ar \}$. We say that $\lambda(x)$ represents the *justification state* of $x \in Ar$.

A labelling $\mathcal{L}$ of $AF = (Ar, att)$ is a *complete labelling* if for each argument $a \in Ar$, it holds that:

- $\mathcal{L}(a) = \text{in}$ iff $\mathcal{L}(b) = \text{out}$ for every $b \in Ar$ such that $(b, a) \in att$.

- $\mathcal{L}(a) = \text{out}$ iff $\mathcal{L}(b) = \text{in}$ for at least one $b \in Ar$ such that $(b, a) \in att$.

- $\mathcal{L}(a) = \text{und}$, otherwise.

Let $\mathcal{L}$ be a complete labelling of AF. Then,

- $\mathcal{L}$ is a *stable labelling* iff $\text{und}(\mathcal{L}) = \emptyset$.

- $\mathcal{L}$ is a *grounded labelling* iff $\text{in}(\mathcal{L}) \subseteq \text{in}(\mathcal{L}')$ for any complete labelling $\mathcal{L}'$ of AF.

- $\mathcal{L}$ is a *preferred labelling* iff there is no complete labelling $\mathcal{L}'$ of AF such that $\text{in}(\mathcal{L}) \subset \text{in}(\mathcal{L}')$.

We often abbreviate complete, stable, grounded, and preferred labelling as *co, st, gr*, and *pr*, respectively.

3. Epistemic Argumentation Framework

In this section, we introduce the notion of epistemic argumentation frameworks and address its applications.

3.1 Epistemic Labelling Set

Given $AF = (Ar, att)$, define $\mathcal{A}_{AF} = \{ \text{in}(a), \text{out}(a), \text{und}(a) \mid a \in Ar \}$. An *epistemic atom* over AF is of the form $\mathbf{K} \, \varphi$ or $\mathbf{M} \, \varphi$ where φ is a propositional formula over $\mathcal{A}_{AF}$. An *epistemic literal* is an epistemic atom or its negation. An *epistemic formula* (over $\mathcal{A}_{AF}$) is a propositional

formula constructed over epistemic literals together with $\top$ (true) and $\bot$ (false). Intuitively, $\mathbf{K}\,\varphi$ (resp. $\mathbf{M}\,\varphi$) states that the agent believes that φ is *true* (resp. *possibly true*).[2] We will use epistemic formulas to represent the epistemic side of an agent given an AF.

Let φ be a propositional formula over $\mathcal{A}_{AF}$ and $\mathcal{L}$ be a labelling over AF. Then $S(\mathcal{L})$ is considered an interpretation of φ. We say that φ is true in $\mathcal{L}$, denoted by $\mathcal{L} \models \varphi$, if φ is interpreted to be true under $S(\mathcal{L})$.

Definition 1 (satisfaction) A set SL of labellings *satisfies* an epistemic formula φ, denoted by $SL \models \varphi$, if one of the following conditions holds:

(i) $\varphi = \top$,

(ii) $\varphi = \mathbf{K}\,\psi$ and $\mathcal{L} \models \psi$ for every $\mathcal{L} \in SL$,

(iii) $\varphi = \mathbf{M}\,\psi$ and $\mathcal{L} \models \psi$ for some $\mathcal{L} \in SL$,

(iv) $\varphi = \neg\psi$ and $SL \not\models \psi$,

(v) $\varphi = \varphi_1 \wedge \varphi_2$ and $(SL \models \varphi_1$ and $SL \models \varphi_2)$,

(vi) $\varphi = \varphi_1 \vee \varphi_2$ and $(SL \models \varphi_1$ or $SL \models \varphi_2)$.

We write $\varphi \supset \psi$ as the abbreviation of $\neg\varphi \vee \psi$ as usual. An epistemic formula φ is *consistent* if there exists a (non-empty) set SL of labellings such that $SL \models \varphi$; otherwise, φ is *inconsistent*. Some basic properties are addressed.

Proposition 1 *Let SL be a set of labellings. For any propositional formula φ and ψ over $\mathcal{A}_{AF}$,*

(i) $SL \models \neg\mathbf{M}\,\varphi$ *iff* $SL \models \mathbf{K}\,\neg\varphi$,

(ii) $SL \models \neg\mathbf{K}\,\varphi$ *iff* $SL \models \mathbf{M}\,\neg\varphi$,

(iii) $SL \models \mathbf{M}\,(\varphi \vee \psi)$ *iff* $SL \models \mathbf{M}\,\varphi \vee \mathbf{M}\,\psi$,

(iv) $SL \models \mathbf{K}\,(\varphi \wedge \psi)$ *iff* $SL \models \mathbf{K}\,\varphi \wedge \mathbf{K}\,\psi$.

Proof: (i) $SL \models \neg\mathbf{M}\,\varphi$ iff $SL \not\models \mathbf{M}\,\varphi$ iff φ is false in every $\mathcal{L} \in SL$ iff $\neg\varphi$ is true in every $\mathcal{L} \in SL$ iff $SL \models \mathbf{K}\,\neg\varphi$. (ii) $SL \models \neg\mathbf{K}\,\varphi$ iff $SL \not\models \mathbf{K}\,\varphi$ iff φ is false in some $\mathcal{L} \in SL$ iff $\neg\varphi$ is true in some $\mathcal{L} \in SL$ iff $SL \models \mathbf{M}\,\neg\varphi$. (iii) $SL \models \mathbf{M}\,(\varphi \vee \psi)$ iff $\mathcal{L} \models \varphi \vee \psi$ for some $\mathcal{L} \in SL$ iff $\mathcal{L} \models \varphi$ or $\mathcal{L} \models \psi$ for some $\mathcal{L} \in SL$ iff $SL \models \mathbf{M}\,\varphi \vee \mathbf{M}\,\psi$. (iv) $SL \models \mathbf{K}\,(\varphi \wedge \psi)$ iff $\mathcal{L} \models \varphi \wedge \psi$ for every $\mathcal{L} \in SL$ iff $\mathcal{L} \models \varphi$ and $\mathcal{L} \models \psi$ for every $\mathcal{L} \in SL$ iff $SL \models \mathbf{K}\,\varphi \wedge \mathbf{K}\,\psi$. $\qquad\square$

Definition 2 (epistemic argumentation framework) An *epistemic argumentation framework* (EAF) is a triple (Ar, att, φ) where $AF = (Ar, att)$ is an argumentation framework and φ is an epistemic formula (called an *epistemic constraint*). We also refer to an EAF by (AF, φ) whenever it is clear from the context what AF refers to.

---

2. By the meaning, it might be better to write $\mathbf{B}\,\varphi$ rather than $\mathbf{K}\,\varphi$, but we use $\mathbf{K}$ because we implement it using epistemic logic programs in which $\mathbf{K}$ and $\mathbf{M}$ are used (see Section 4).

Intuitively, an EAF (Ar, att, φ) represents the view of an agent who, given $AF = (Ar, att)$, believes that φ is true. An EAF consists of two different types of information: an objective evidence AF and a subjective belief φ of an agent. An epistemic constraint φ represents an individual view on AF. Different agents may have different views (or bias) towards the same AF. The situation is represented by different EAFs (AF, φ_1), ..., (AF, φ_n) for n agents. Moreover, an agent may change its own belief and the situation is realized by revising φ. We will see such a situation later in Example 2.

Example 1 Consider an AF with the set of arguments $\{a, b, c, d\}$ where

a: The UK should remain in the EU.

b: The UK should exit from the EU.

c: EU membership gives Britain access to the European single market.

d: The effect of single market has been exaggerated and the supposed benefits of EU membership are largely imaginary.

Suppose that a and b mutually attack, c and d mutually attack, and c attacks b.

Then some EAFs are defined as follows:

- $EAF_1 = (AF, \mathbf{M} \text{ in}(b))$ represents the view of an agent who believes that b is possibly accepted.

- $EAF_2 = (AF, \mathbf{K} \text{ in}(a) \vee \mathbf{K} \text{ in}(b))$ represents the view of an agent who believes that either a or b should be accepted.

- $EAF_3 = (AF, \mathbf{M} \text{ in}(c) \supset \mathbf{K} \text{ in}(a))$ represents the view of an agent who believes that if $\text{in}(c)$ is possibly true then $\text{in}(a)$ should be accepted.

Next we define the semantics of an EAF.

Definition 3 (epistemic labelling set) Let $EAF = (AF, \varphi)$ and $\omega \in \{co, st, gr, pr\}$. A set SL of labellings is an ω-*epistemic labelling set* of (AF, φ) if (i) each $\mathcal{L} \in SL$ is an ω-labelling of AF, and (ii) SL is a $\subseteq$-maximal set of ω-labellings of AF that satisfies φ. An EAF possibly has multiple ω-epistemic labelling sets.

Intuitively, an ω-epistemic labelling set is a collection of ω-labellings that reflects the belief of an agent. In particular, $EAF = (AF, \top)$ has the unique ω-epistemic labelling set that coincides with the set of ω-labellings of AF. In what follows, we assume $\omega \in \{co, st, gr, pr\}$ unless stated otherwise. By definition, EAF always has an ω-epistemic labelling set (possibly as an empty set).

Proposition 2 $EAF = (AF, \bot)$ *has the ω-epistemic labelling set $\emptyset$.*

Proof: When φ is inconsistent, there is no set of labellings satisfying φ. Hence, the result holds. $\square$

Our primary interest is an EAF that has non-empty ω-epistemic labelling sets.

Example 2 Consider the EAFs of Example 1. First, AF in the EAFs has three stable labellings:

$$L_1 = \{\, \texttt{in}(a),\, \texttt{out}(b),\, \texttt{in}(c),\, \texttt{out}(d) \,\},$$
$$L_2 = \{\, \texttt{in}(a),\, \texttt{out}(b),\, \texttt{out}(c),\, \texttt{in}(d) \,\},$$
$$L_3 = \{\, \texttt{out}(a),\, \texttt{in}(b),\, \texttt{out}(c),\, \texttt{in}(d) \,\},$$

and the grounded labelling:

$$L_4 = \{\, \texttt{und}(a),\, \texttt{und}(b),\, \texttt{und}(c),\, \texttt{und}(d) \,\}.$$

This implies that EAF_1 has a single stable epistemic labelling set $\{L_1, L_2, L_3\}$; EAF_2 has two stable epistemic labelling sets $\{L_1, L_2\}$ and $\{L_3\}$; and EAF_3 has two stable epistemic labelling sets $\{L_1, L_2\}$ and $\{L_2, L_3\}$. By contrast, EAF_1 and EAF_2 have the grounded epistemic labelling set $\emptyset$, while EAF_3 has a single grounded epistemic labelling set $\{L_4\}$.

Suppose an agent, who previously has EAF_3, introduces a new belief $\mathbf{K}\ \texttt{in}(d)$. The situation is represented by

$$EAF_4 = (AF, (\mathbf{M}\ \texttt{in}(c) \supset \mathbf{K}\ \texttt{in}(a)) \wedge \mathbf{K}\ \texttt{in}(d)\,).$$

Then EAF_4 has a unique stable epistemic labelling set $\{L_2, L_3\}$, while it has the grounded epistemic labelling set $\emptyset$.

As shown in the above example, EAF can represent belief change of an agent by revising an epistemic constraint without modifying AF. The revised EAF would produce new epistemic labelling sets that reflect new belief states of an agent. In Example 2, EAF_4 introduces an additional constraint $\mathbf{K}\ \texttt{in}(d)$ to EAF_3, which results in eliminating $\{L_1, L_2\}$ from the stable epistemic labelling sets of EAF_3. Moreover, if another belief $\mathbf{K}\ \texttt{out}(a)$ is introduced to EAF_4, then the revised EAF has the single stable epistemic labelling set $\{L_3\}$. In this way, EAF could represent dynamic change of opinions resulting from belief revision of an agent. In argumentation frameworks, if an AF has a single stable labelling, it is also the grounded labelling, while in EAF the single stable epistemic labelling set does not coincide with the grounded epistemic labelling set in general.

For two epistemic formulas φ_1 and φ_2, we say that φ_1 is *stronger* than φ_2 if for every set SL of labellings, $SL \models \varphi_1$ implies $SL \models \varphi_2$. We write $\varphi_1 \models \varphi_2$ [3] to denote that φ_1 is stronger than φ_2. Introducing a stronger constraint to EAF eliminates elements of SL in general.

Proposition 3 *Let $EAF_1 = (AF, \varphi_1)$ and $EAF_2 = (AF, \varphi_2)$ be two EAFs such that φ_1 is stronger than φ_2. Then, for each ω-epistemic labelling set SL_1 of EAF_1 there exists some ω-epistemic labelling set SL_2 of EAF_2 such that $SL_1 \subseteq SL_2$.*

Proof: Suppose an ω-epistemic labelling SL_1 of EAF_1. Each $\mathcal{L} \in SL_1$ is an ω-labelling of AF satisfying φ_1. Since φ_1 is stronger than φ_2, $SL_1 \models \varphi_1$ implies $SL_1 \models \varphi_2$. Then there is an ω-epistemic labelling SL_2 of EAF_2 such that $\mathcal{L} \in SL_2$. Since $\mathcal{L} \in SL_1$ implies $\mathcal{L} \in SL_2$, $SL_1 \subseteq SL_2$ holds. □

---

3. Note that $\varphi_1 \models \varphi_2$ can be proved using the set of inference rules consisting of the modus ponens rule and the following rules: $\mathbf{K}\,\varphi \vdash \mathbf{K}\,\psi$, $\mathbf{M}\,\varphi \vdash \mathbf{M}\,\psi$, $\mathbf{K}\,\varphi \vdash \mathbf{M}\,\psi$, and $\mathbf{K}\,\varphi \vdash \psi$ where φ and ψ are propositional formulas such that $\varphi \models \psi$.

In argumentation frameworks, stable, grounded, or preferred labellings are complete labellings. Also stable labellings are preferred labellings. Similar relations hold in epistemic argumentation frameworks.

Proposition 4 *Let (AF, φ) be an EAF. If a non-empty set SL of labellings is a stable, grounded, or preferred epistemic labelling set of (AF, φ), then $\mathcal{L} \in SL$ is an element of a complete epistemic labelling set of (AF, φ). Also if SL is a stable labelling set of (AF, φ), then $\mathcal{L} \in SL$ is an element of a preferred epistemic labelling set of (AF, φ).*

Proof: Suppose a non-empty set SL that is a stable epistemic labelling set of (AF, φ). Then each $\mathcal{L} \in SL$ is a stable labelling of AF, and SL is a maximal set of stable labellings of AF that satisfies φ. Since a stable labelling of AF is a complete labelling of AF, SL is a set of complete labellings of AF that satisfies φ. Hence, the result holds. Relations between grounded/preferred epistemic labelling sets and complete epistemic labelling sets, and the relation between stable epistemic labelling sets and preferred epistemic labelling sets are shown in similar ways. $\square$

We next consider a sufficient condition for the uniqueness of ω-epistemic labelling sets.

Lemma 5 *Let φ be a conjunction of epistemic literals over $\mathcal{A}_{AF}$. If two sets of labellings SL_1 and SL_2 satisfy φ (i.e., $SL_1 \models \varphi$ and $SL_2 \models \varphi$), then $SL_1 \cup SL_2 \models \varphi$.*

Proof: (i) For any epistemic literal $\mathbf{K}\psi$ in φ, $SL_1 \models \varphi$ and $SL_2 \models \varphi$ imply that $\mathcal{L} \models \psi$ for every $\mathcal{L} \in SL_1 \cup SL_2$. Then $SL_1 \cup SL_2 \models \mathbf{K}\psi$. (ii) For any epistemic literal $\mathbf{M}\psi$ in φ, $SL_1 \models \varphi$ and $SL_2 \models \varphi$ imply that $\mathcal{L} \models \psi$ for some $\mathcal{L} \in SL_1 \cup SL_2$. Then $SL_1 \cup SL_2 \models \mathbf{M}\psi$. (iii) For any epistemic literal $\neg\mathbf{K}\psi$ in φ, $SL_1 \models \varphi$ and $SL_2 \models \varphi$ imply that $\mathcal{L} \not\models \psi$ for some $\mathcal{L} \in SL_1 \cup SL_2$. Then $SL_1 \cup SL_2 \models \neg\mathbf{K}\psi$. (iv) For any epistemic literal $\neg\mathbf{M}\psi$ in φ, $SL_1 \models \varphi$ and $SL_2 \models \varphi$ imply that $\mathcal{L} \not\models \psi$ for every $\mathcal{L} \in SL_1 \cup SL_2$. Then $SL_1 \cup SL_2 \models \neg\mathbf{M}\psi$. Since φ is a conjunction of epistemic literals of the forms (i)–(iv), the result holds. $\square$

Using the lemma, we can prove the next result.

Proposition 6 *Let (AF, φ) be an EAF such that φ is a conjunction of epistemic literals. Then (AF, φ) has a unique ω-epistemic labelling set. .*

Proof: Suppose that (AF, φ) has two ω-epistemic labelling sets SL_1 and SL_2 ($SL_1 \neq SL_2$). Then $SL_1 \models \varphi$ and $SL_2 \models \varphi$ imply $SL_1 \cup SL_2 \models \varphi$ (by Lemma 5). This contradicts the assumption that SL_1 (or SL_2) is a $\subseteq$-maximal set of ω-labellings of AF satisfying φ. $\square$

Assume that φ is a DNF in which each disjunct is a conjunction of epistemic literals. Due to Proposition 1, we can assume that each disjunct in φ is of the form $\mathbf{K}\,\psi_0 \wedge \mathbf{M}\,\psi_1 \wedge \cdots \wedge \mathbf{M}\,\psi_n$[4] where ψ_i ($0 \leq i \leq n$) is a propositional formula over $\mathcal{A}_{AF}$, which will be denoted by $EC(\psi_0; \psi_1, \ldots, \psi_n)$. By definition, the next result holds.

Lemma 7 *Let SL be a set of labellings such that $SL \models EC(\psi_0; \psi_1, \ldots, \psi_n)$. Then, for each $i = 1, \ldots, n$, there exists some $\mathcal{L} \in SL$ such that $\mathcal{L} \models \psi_0 \wedge \psi_i$.*

---

4. $\neg\mathbf{M}\,\psi$ (resp. $\neg\mathbf{K}\,\psi$) is converted to $\mathbf{K}\,\neg\psi$ (resp. $\mathbf{M}\,\neg\psi$), and $\mathbf{K}\,\psi_1 \wedge \mathbf{K}\,\psi_2$ is converted to $\mathbf{K}\,(\psi_1 \wedge \psi_2)$.

This leads to the following proposition.

Proposition 8 *Let* $\varphi = \bigvee_{j=1}^{k} EC(\psi_j; \psi_1^j, \ldots, \psi_{n_j}^j)$ $(k \geq 1)$ *be an epistemic formula. Then,* $EAF{=}(AF, \varphi)$ *has a non-empty ω-epistemic labelling set if there exists an integer j $(1 \leq j \leq k)$ such that for each $1 \leq i \leq n_j$, AF has an ω-labelling $\mathcal{L}$ and $\mathcal{L} \models \psi_j \wedge \psi_i^j$.*

Proof: Assume that there exists j $(1 \leq j \leq k)$ such that for each i $(1 \leq i \leq n_j)$, AF has an ω-labelling $\mathcal{L}_i$ and $\mathcal{L}_i \models \psi_j \wedge \psi_i^j$. By Lemma 7 it is clear that $SL = \{\mathcal{L}_i \mid i = 1, \ldots, n_j\} \models EC(\psi_j; \psi_1^j, \ldots, \psi_{n_j}^j)$, i.e., $SL \models \varphi$. Hence, the result follows. $\qquad\square$

Each AF semantics imposes some specific condition on every argument, e.g., the stable semantics allows no argument to be undecided, while the grounded semantics keeps controversial arguments undecided. EAF is useful for selecting intended labellings from the set of all possible labellings.

Example 3 Consider the AF in Example 1 in which an agent does not want to decide yet whether he/she will support Brexit or not. The situation is specified as the epistemic constraint $\varphi = \mathbf{K}\,\mathrm{und}(b)$. Then (AF, φ) has the single preferred epistemic labelling set $\{\{\mathrm{und}(a), \mathrm{und}(b), \mathrm{out}(c), \mathrm{in}(d)\}\}$.

3.2 Representing Preference

Preference among arguments can be specified in EAF as follows. Let $\succeq$ be a pre-order (i.e., reflexive and transitive) relation over $Ar \times Ar$ such that $(x, y) \in \succeq$ implies that x indirectly attacks y or vice versa. $x \succeq y$ means that an argument x is at least as preferred as y. We write $x \succ y$ if $x \succeq y$ and $y \not\succeq x$.

Definition 4 (preference over arguments) Given $AF = (Ar, att)$ and a preorder relation $\succeq \subseteq Ar \times Ar$, define $EAF = (AF, \varphi_A)$ where

$$\varphi_A = \bigwedge_{x \succ y} \mathbf{K}(\mathrm{in}(y) \supset \mathrm{in}(x)).$$

Intuitively speaking, φ_A represents that an argument x should be accepted whenever another argument y of lower preference is accepted. Note that the preference is specified as $x \succ y$ but not as $x \succeq y$ in φ_A. When both $x \succeq y$ and $y \succeq x$ exist, there is no reason to prefer one of them. In this case, the conjunct involved x and y in φ_A is $\top$. By definition, the next proposition holds.

Proposition 9 *Let $EAF = (AF, \varphi_A)$ be an EAF defined as above. Then, for any ω-epistemic labelling set SL of EAF, there is no $\mathcal{L} \in SL$ such that $\mathcal{L} \not\models \mathrm{in}(x)$ and $\mathcal{L} \models \mathrm{in}(y)$ for any $x \succ y$.*

Example 4 Consider $AF = (\{a, r\}, \{(a, r), (r, a)\})$ with $r \succ a$. Then $EAF = (AF, \varphi_A)$ with $\varphi_A = \mathbf{K}(\mathrm{in}(a) \supset \mathrm{in}(r))$ has the unique stable epistemic labelling set $\{\{\mathrm{in}(r), \mathrm{out}(a)\}\}$, and the unique complete epistemic labelling set $\{\{\mathrm{in}(r), \mathrm{out}(a)\}, \{\mathrm{und}(r), \mathrm{und}(a)\}\}$.

In Example 4, the complete epistemic labelling set contains $\{\mathrm{und}(r), \mathrm{und}(a)\}$. This can be eliminated by introducing the constraint $\varphi_A = \mathbf{K}(\mathrm{in}(a) \vee \mathrm{und}(a) \supset \mathrm{in}(r))$. In EAF preferences could be revised by belief change of an agent. For instance, if one wants to neutralize the preference

$r \succ a$, it could be done by changing φ_A to $\varphi'_A = \mathbf{K}(\mathtt{in}(a) \supset \mathtt{in}(r)) \vee \mathbf{K}(\mathtt{in}(r) \supset \mathtt{in}(a))$. In this case, (AF, φ'_A) has two stable epistemic labelling sets $\{\{\mathtt{in}(r), \mathtt{out}(a)\}\}$ and $\{\{\mathtt{in}(a), \mathtt{out}(r)\}\}$.

Preference over arguments is generalized to preference over justification states of arguments as follows. A pre-order relation $\sqsupseteq$ over justification states of arguments is a collection of elements of the form $\lambda(x) \sqsupseteq \mu(y)$ where $\lambda, \mu \in \{\mathtt{in}, \mathtt{out}, \mathtt{und}\}$, meaning that $\lambda(x)$ is at least as preferred as $\mu(y)$ for arguments x and y. We write $\lambda(x) \sqsupset \mu(y)$ if $\lambda(x) \sqsupseteq \mu(y)$ and $\mu(y) \not\sqsupseteq \lambda(x)$.

Definition 5 (preference over justification states) Given $AF = (Ar, att)$ and a preorder relation $\sqsupseteq \subseteq \mathcal{A}_{AF} \times \mathcal{A}_{AF}$, define $EAF = (AF, \varphi_J)$ where

$$\varphi_J = \bigwedge_{\lambda(x) \sqsupset \mu(y)} \mathbf{K}\left(\mu(y) \supset \lambda(x)\right).$$

φ_J states that if the justification state $\lambda(x)$ is preferred to $\mu(y)$ for $x, y \in Ar$, then $\mathcal{L} \models \mu(x)$ implies $\mathcal{L} \models \lambda(x)$ for any $\mathcal{L} \in SL$ where SL is any ω-epistemic labelling set of EAF.

By definition, Def. 4 is considered a special case of Def. 5 with $\mu = \lambda = \mathtt{in}$.

Proposition 10 *Let $EAF = (AF, \varphi_J)$ be an EAF defined as above. Then, for any ω-epistemic labelling set SL of EAF, there is no $\mathcal{L} \in SL$ such that $\mathcal{L} \not\models \lambda(x)$ and $\mathcal{L} \models \mu(y)$ for any $\lambda(x) \sqsupset \mu(y)$. In particular, $\mathcal{L} \not\models \mu(y)$ for any $\mathcal{L} \in SL$ if $x = y$.*

Proof: If there is $\mathcal{L} \in SL$ such that $\mathcal{L} \not\models \lambda(x)$ and $\mathcal{L} \models \mu(y)$, then SL does not satisfy the constraint φ_J. When $x = y$, if $\mathcal{L} \models \mu(y)$ then $\mathcal{L} \models \lambda(y)$ for $\mu \neq \lambda$. This does not happen because $\mathcal{L}$ is a function. Hence, the result holds. $\square$

Example 5 Suppose that in Example 4, an MP prefers keeping the decision undecided if possible. This is represented by $\sqsupset = \{(\mathtt{und}(x), \mathtt{in}(x)), (\mathtt{und}(x), \mathtt{out}(x)) \mid x \in \{a, r\}\}$ which is translated to $\varphi_J = \bigwedge_{x \in \{a,r\}} \mathbf{K}(\mathtt{in}(x) \supset \mathtt{und}(x)) \wedge \mathbf{K}(\mathtt{out}(x) \supset \mathtt{und}(x))$. Then $EAF = (AF, \varphi_J)$ has the unique complete epistemic labelling set $\{\{\mathtt{und}(r), \mathtt{und}(a)\}\}$. Furthermore, $\emptyset$ is the stable epistemic labelling set, since there is no choice to make a and r undecided.

In this way, EAF enables us to specify preference over not only arguments but also justification states of arguments.

Definition 5 is extended to specify conditional preferences. For instance, we could specify a preference $\psi \supset \varphi_J$ where ψ is an epistemic formula. It represents that if ψ is true then the preference φ_J is applied. In Example 5, if $\varphi_C = \mathbf{M}(\mathtt{und}(a) \vee \mathtt{und}(r)) \supset \varphi_J$ is used instead of φ_J, then $EAF = (AF, \varphi_C)$ has the unique complete epistemic labelling set $\{\{\mathtt{und}(r), \mathtt{und}(a)\}\}$, while it still has the unique stable epistemic labelling set $\{\{\mathtt{in}(a), \mathtt{out}(r)\}, \{\mathtt{in}(r), \mathtt{out}(a)\}\}$.

Furthermore, it could also be useful to introduce preferences among epistemic formulas. For instance, we could write $\mathbf{K}\,\lambda(x) > \mathbf{K}\,\mu(x)$ for some argument x to indicate that we prefer SL_1 over SL_2 whenever $SL_1 \models \mathbf{K}\,\lambda(x)$ and $SL_2 \models \mathbf{K}\,\mu(x)$ for two arbitrary ω-epistemic labelling sets SL_1 and SL_2. Such a preference relation is specified as the epistemic constraint $\mathbf{K}\,\mu(x) \supset \mathbf{K}\,\lambda(x)$ whose intended meaning is clear. Generally, given two epistemic formulas ψ_1 and ψ_2, the preference relation $\psi_1 > \psi_2$ (ψ_1 is preferred to ψ_2) is specified as the epistemic constraint $\varphi_F = \psi_2 \supset \psi_1$. Then, $EAF = (AF, \varphi_F)$ would select ω-epistemic labelling sets SL such that $SL \models \psi_2$ implies $SL \models \psi_1$.

3.3 Multiple Agents

Suppose that two agents share $AF = (\{a, r\}, \{(a, r), (r, a)\})$. If they have the same belief represented by the epistemic constraint $\varphi = \mathbf{K}\, \mathtt{in}(a)$, the EAF (AF, φ) has the single epistemic complete labelling set $\{\{\mathtt{in}(a), \mathtt{out}(r)\}\}$ and the agents agree on accepting a. On the other hand, if two agents have conflicting beliefs $\varphi_1 = \mathbf{K}\, \mathtt{in}(a)$ and $\varphi_2 = \mathbf{K}\, \mathtt{in}(r)$ respectively, then they do not agree on accepting a or r. In this section, we assume multiple agents who share the same AF while having different beliefs in general. The situation is represented by the collection of EAFs (AF, φ_i) $(1 \leq i \leq n)$. First, we define two different types of agreements.

Definition 6 (agreement) Let $AF = (Ar, att)$ and $EAF_1 = (AF, \varphi_1), \ldots, EAF_n = (AF, \varphi_n)$ $(n \geq 1)$. Then $EAF_1, \ldots, EAF_n$ *credulously agree on* $\lambda(a)$ for $a \in Ar$ where $\lambda \in \{\mathtt{in}, \mathtt{out}, \mathtt{und}\}$ *under ω-epistemic labelling* if each EAF_i $(i = 1, \ldots, n)$ has an ω-epistemic labelling set SL_i such that $SL_i \models \mathbf{M}\, \lambda(a)$. In contrast, $EAF_1, \ldots, EAF_n$ *skeptically agree on* $\lambda(a)$ *under ω-epistemic labelling* if for any ω-epistemic labelling set SL_i of EAF_i $(i = 1, \ldots, n)$ $SL_i \models \mathbf{K}\, \lambda(a)$.

The above definition characterizes two different situations (credulous or skeptical) in which agents reach an agreement on $\lambda(a)$. For simplicity reasons, Def.6 assumes that different agents employ the same ω-epistemic labelling, but the definition is easily extended to a case in which agents employ different ω-labellings.

Proposition 11 Let $AF = (Ar, att)$ and $EAF_1 = (AF, \varphi_1), \ldots, EAF_n = (AF, \varphi_n)$ $(n \geq 1)$. Then, $EAF_1, \ldots, EAF_n$ skeptically agree on $\lambda(a)$ for $a \in Ar$ under ω-epistemic labelling iff EAF_i and $EAF_i' = (AF, \varphi_i \wedge \mathbf{K}\, \lambda(a))$ $(i = 1, \ldots, n)$ have the same ω-epistemic labelling sets.

Proof: The if-part is clear. We show the only-if part. If $EAF_1, \ldots, EAF_n$ skeptically agree on $\lambda(a)$, then $SL_i \models \mathbf{K}\, \lambda(a)$ for any ω-epistemic labelling set SL_i of EAF_i $(i = 1, \ldots, n)$. Since every ω-epistemic labelling set SL_i of EAF_i satisfies $\mathbf{K}\, \lambda(a)$, SL_i is also an ω-epistemic labelling set of $EAF_i' = (AF, \varphi_i \wedge \mathbf{K}\, \lambda(a))$. $\qquad\square$

Proposition 12 Let $AF = (Ar, att)$ and $EAF_1 = (AF, \varphi_1), \ldots, EAF_n = (AF, \varphi_n)$ $(n \geq 1)$. If $EAF_1, \ldots, EAF_n$ credulously agree on $\lambda(a)$ for $a \in Ar$ under ω-epistemic labelling, then $(AF, \varphi_1 \vee \cdots \vee \varphi_n)$ has an ω-epistemic labelling set SL such that $SL \models \mathbf{M}\, \lambda(a)$. Conversely, if $(AF, \varphi_1 \wedge \cdots \wedge \varphi_n)$ has an ω-epistemic labelling set SL such that $SL \models \mathbf{M}\, \lambda(a)$, then $EAF_1, \ldots, EAF_n$ credulously agree on $\lambda(a)$ under ω-epistemic labelling.

Proof: Suppose that $EAF_1, \ldots, EAF_n$ credulously agree on $\lambda(a)$ for $a \in Ar$ under ω-epistemic labelling. Then, EAF_i $(1 \leq i \leq n)$ has an ω-epistemic labelling set SL_i such that $SL_i \models \mathbf{M}\, \lambda(a)$, and there is $\mathcal{L} \in SL_i$ such that $\mathcal{L} \models \lambda(a)$. By Proposition 3, for each ω-epistemic labelling set SL_i of (AF, φ_i), there is some ω-epistemic labelling set SL of $(AF, \varphi_1 \vee \cdots \vee \varphi_n)$ such that $SL_i \subseteq SL$. Then, $\exists \mathcal{L} \in SL$ such that $\mathcal{L} \models \lambda(a)$. Hence, $SL \models \mathbf{M}\, \lambda(a)$. Conversely, if $(AF, \varphi_1 \wedge \cdots \wedge \varphi_n)$ has an ω-epistemic labelling set SL such that $SL \models \mathbf{M}\, \lambda(a)$, then there is some ω-epistemic labelling set SL_j of (AF, φ_j) such that $SL \subseteq SL_j$ for each EAF_j $(1 \leq j \leq n)$ by Proposition 3. By $SL_j \models \mathbf{M}\, \lambda(a)$, $EAF_1, \ldots, EAF_n$ credulously agree on $\lambda(a)$ under ω-epistemic labelling. $\qquad\square$

Algorithm 1: Existence(EAF, ω)

Input: ω, $EAF = (AF, \varphi)$.

Output: true if EAF has a (non-empty) ω-epistemic labelling set; **false** otherwise.

1 Convert to DNF: $\varphi = \vee_{j=1}^{k} EC(\psi_j; \psi_1^j, \ldots, \psi_{n_j}^j)$

2   where $EC(\psi; \psi_1, \ldots, \psi_k) = \mathbf{K}\psi \wedge \bigwedge_{i=1}^{k} \mathbf{M}\psi_i$

3 **for** $j = 1$ *to* k **do**

4      $num\_labelling := 0$

5   **for** $i = 1$ *to* n_j **do**

6   **if** $\mathtt{D}(\omega, AF, \psi_j \wedge \psi_i^j) = $ **true then**

7              $num\_labelling := num\_labelling + 1$

8   **if** $num\_labelling = n_j$ **then return true**

9 **return false**

We next show that EAF can be used for formalizing majority voting. In the presence of $EAF_i = (AF, \varphi_i)$ $(1 \leq i \leq n)$, define:

$$M_\psi^\omega = \{\, i \mid EAF_i \text{ has an } \omega\text{-epistemic labelling set } SL \text{ s.t. } SL \models \mathbf{M}\,\psi \,\},$$
$$N_\psi^\omega = \{\, i \mid \text{for each } \omega\text{-epistemic labelling set } SL \text{ of } EAF_i,\ SL \models \mathbf{K}\,\psi \,\}.$$

Definition 7 (majority voting) Let $AF = (Ar, att)$ and $EAF_i = (AF, \varphi_i)$ for $(1 \leq i \leq n)$. For $a \in Ar$, $\lambda(a)$ is *credulously* (resp. *skeptically*) adopted by *majority voting under ω-epistemic labelling* iff the cardinality of the set $M_{\lambda(a)}^\omega$ (resp. $N_{\lambda(a)}^\omega$) is greater than the cardinality of the set $M_{\mu(a)}^\omega$ (resp. $N_{\mu(a)}^\omega$) where $\lambda, \mu \in \{\mathtt{in}, \mathtt{out}, \mathtt{und}\}$ and $\lambda \neq \mu$.

When $|M_{\lambda(a)}^\omega| = n$ (resp. $|N_{\lambda(a)}^\omega| = n$) in Def. 7, $EAF_1, \ldots, EAF_n$ credulously (resp. skeptically) agree on $\lambda(a)$.

Example 6 Consider $AF = (\{a, r\}, \{(a, r), (r, a)\})$ and three EAFs: $EAF_1 = (AF, \mathbf{K}\,\mathtt{in}(a))$, $EAF_2 = (AF, \neg\mathbf{M}\,\mathtt{und}(a))$, and $EAF_3 = (AF, \mathbf{K}\,\mathtt{und}(a))$. Then $\mathtt{in}(a)$ is credulously adopted by majority voting under the complete epistemic labelling, while it is not skeptically adopted.

It is known that judgment aggregation may lead to an inconsistent group outcome where the reasons do not support the conclusion by Caminada and Pigozzi (2011). Majority voting in this section concerns the justification state of a single argument, and the issue of handling inconsistent collective judgment of arguments is outside the scope of this paper.

3.4 Complexity

In this section, we consider a decision problem of whether an EAF has a non-empty ω-epistemic labelling set. We assume that the readers are familiar with the well-known notations in computational complexity (e.g., P-c, NP-c, coNP-c, etc.). Let $\omega \in \{gr, st, co, pr\}$ and $EAF = (AF, \varphi)$. Due to Proposition 8, we can check for the existence of a non-empty ω-epistemic labelling set using Algorithm 1, assuming the existence of a procedure $\mathtt{D}(\omega, AF, \psi)$ that determines the existence of an ω-labelling $\mathcal{L}$ of AF such that $\mathcal{L} \models \psi$.

In essence, Algorithm 1 shows that checking whether EAF has a non-empty ω-epistemic labelling set can be reduced to checking whether a labelling $\mathcal{L}$ of AF satisfies a formula over $\mathcal{A}_{AF}$. In line 1 we assume that φ has at most k disjuncts, and each contains at most p conjuncts, where p and k are polynomial in the size of the AF and refer to φ as a (k,p)-DNF.[5] Under this assumption, Algorithm 1 will call $\mathrm{D}(\omega, AF, \psi)$ at most $k \times p$ times. Consider the following decision problem:

$\mathrm{Exists}_{\omega}^{(k,p)}$: Given an $AF = (Ar, att)$ and a (k,p)-DNF epistemic formula φ over $\mathcal{A}_{AF}$, does (AF, φ) have a non-empty ω-epistemic labelling set?

The above discussion gives us the next result.

Proposition 13 $\mathrm{Exists}_{\omega}^{(k,p)}$ *is P-c for $\omega = gr$ and NP-c for $\omega \in \{co, st, pr\}$.*

Proof: Suppose $EAF = (AF, \varphi)$ with $\varphi = \bigvee_{j=1}^{k} EC(\psi_j; \psi_1^j, \ldots, \psi_{n_j}^j)$.

(i) Consider $\omega = gr$. Since the grounded labelling $\mathcal{L}_{gr}$ of AF is unique and can be computed in polynomial time, there is only a single potential grounded epistemic labelling set $SL_{gr} = \{\mathcal{L}_{gr}\}$ of EAF. Since $SL_{gr} \models \varphi$ iff there exists some j such that $\mathcal{L}_{gr} \models \psi_j \wedge \bigwedge_{i=1}^{n_j} \psi_i^j$, which can be done in polynomial time, the result follows immediately.

(ii) Consider $\omega \in \{co, st, pr\}$. Hardness of the problem follows from the fact that if $\varphi = \mathbf{K}\,\mathrm{in}(a)$ for some a then $\mathrm{Exists}_{\omega}^{(k,p)}$ is the same as checking whether or not AF has an ω-labelling containing $\mathrm{in}(a)$ which is NP-c (or the problem Cred_{σ} in Dvorák, 2011; Dvorák & Dunne, 2018).

Assume that we have an oracle $\mathcal{O}$ that determines whether or not AF has an ω-labelling satisfying a propositional formula over $\mathcal{A}_{AF}$. Then, the following polynomial algorithm can be used to decide $\mathrm{Exists}_{\omega}^{(k,p)}$.

- for each $j = 1, \ldots, k$ and $i = 1, \ldots, n_j$, use $\mathcal{O}$ to determine whether or not an ω-labelling $\mathcal{L}_{j,i}$ satisfying $\psi_j \wedge \psi_i^j$ exists;

- if there exists some j such that $\mathcal{L}_{j,i}$ for $i = 1, \ldots, n_j$ then the answer to $\mathrm{Exists}_{\omega}^{(k,p)}$ is yes; otherwise, the answer is no.

This shows that the complexity of $\mathrm{Exists}_{\omega}^{(k,p)}$ is $P^{\mathcal{C}}$ where $\mathcal{C}$ is the complexity class of Cred_{ω}, i.e., determining whether or not there exists an ω-labelling $\mathcal{L}_{j,i}$ satisfying a formula over $\mathcal{A}_{AF}$. Because $\mathcal{C}$ is NP-c for $\omega \in \{st, co, pr\}$, we can conclude that $\mathrm{Exists}_{\omega}^{(k,p)}$ is NP-c. $\qquad\square$

The above results show that the decision problem $\mathrm{Exists}_{\omega}^{(k,p)}$ has the same complexity as Cred_{σ} in abstract argumentation.

4. Computation

In this section, we describe a system for computing epistemic labelling sets of an EAF.

---

5. The DNF of a formula φ might have exponential number of disjuncts in general, however, it would be a rare case that belief of an agent is expressed by an exponential formula.

4.1 Epistemic Logic Programs

A *logic program* (LP) is a finite set of *rules* of the form:

$$L_1 \; or \; \cdots \; or \; L_l \leftarrow \quad L_{l+1}, \ldots, L_m, \mathbf{not} \; L_{m+1}, \ldots, \mathbf{not} \; L_n \tag{1}$$

$(n \geq m \geq l \geq 0)$ where each L_i is a ground literal (an atom a or its negation $\neg a$). **not** is *negation as failure* (NAF) and **not** L is called an *NAF-literal*. The left-hand side of $\leftarrow$ is the *head*, and the right-hand side is the *body*. For each rule r of the above form, $head(r)$, $body^+(r)$, and $body^-(r)$ denote the sets of literals $\{L_1, \ldots, L_l\}$, $\{L_{l+1}, \ldots, L_m\}$, and $\{L_{m+1}, \ldots, L_n\}$, respectively. A rule r is a *constraint* if $head(r) = \emptyset$; and r is a *(disjunctive) fact* if $body^+(r) = body^-(r) = \emptyset$. We often write a rule with variables as a shorthand of its ground instances. A logic program is simply called a *program*. The semantics of a program is defined by the *answer set semantics*, introduced by Gelfond and Lifschitz (1991).

Let $lit(P)$ be the set of all ground literals of a program P. A set S of literals is *consistent* if $a \in S$ implies $\neg a \notin S$ for any atom a. A consistent set of literals $S \subset lit(P)$ *satisfies* a rule r of the form (1) if $body^+(r) \subseteq S$ and $body^-(r) \cap S = \emptyset$ imply $head(r) \cap S \neq \emptyset$. In particular, S satisfies a constraint r such that $head(r) = \emptyset$ if $body^+(r) \setminus S \neq \emptyset$ or $body^-(r) \cap S \neq \emptyset$. S is a *model* of the program P if it satisfies every rule in P. For a program P with $body^-(r) = \emptyset$ for every rule r in P, a model S of P is *minimal* if there is no model T of P such that $T \subset S$. Given a program P, a consistent set of literals S is an *answer set* of P if it coincides with a minimal model of P^S, the *reduct* of P with respect to S, defined as follows:

$$P^S = \big\{ head(r) \leftarrow body^+(r) \, \big| \; r \in P \text{ and } body^-(r) \cap S = \emptyset \; \big\}.$$

A program may have no, one, or multiple answer sets in general. A program is *consistent* if it has at least one answer set; otherwise, the program is *inconsistent*. The set of answer sets of a program P is denoted by $\text{AS}(P)$. Representing knowledge by logic programs under the answer set semantics is called *answer set programming* (ASP) (Brewka, Eiter, & Truszczynski, 2011). We use the terms logic programming and answer set programming interchangeably.

An *epistemic logic program* (ELP), introduced by Gelfond (1991), is a collection of rules of the form (1) with the following extended syntax: each L_j $(l+1 \leq j \leq n)$ is either a ground literal, or a *subjective literal* of the form $\mathbf{K} \, \ell$ or $\mathbf{M} \, \ell$ where ℓ is a ground literal (also referred to as an *objective literal*). The semantics of epistemic logic programs is defined by *world views*. The semantics used in this paper is due to Kahl, Leclerc, and Son (2016) and Shen and Eiter (2016).

Let W be a non-empty set of consistent sets of ground literals, and ℓ be a literal.

- $\mathbf{K} \, \ell$ is satisfied by W if $\forall A \in W : \ell \in A$.

- $\mathbf{not} \, \mathbf{K} \, \ell$ is satisfied by W if $\exists A \in W : \ell \notin A$.

- $\mathbf{M} \, \ell$ is satisfied by W if $\exists A \in W : \ell \in A$.

- $\mathbf{not} \, \mathbf{M} \, \ell$ is satisfied by W if $\forall A \in W : \ell \notin A$.

We write $W \models \mathbf{K} \, \ell$ ($W \models \mathbf{M} \, \ell$) to mean that $\mathbf{K} \, \ell$ ($\mathbf{M} \, \ell$) is satisfied by W.

Let Π be an epistemic logic program and W a non-empty set of consistent sets of ground literals. We denote by Π^W the *modal reduct of Π with respect to W* defined as the ASP program[6] obtained

---

6. with nested expressions of the form $\;\mathbf{not} \; \mathbf{not} \, \ell$ as defined by Lifschitz, Tang, and Turner (1999).

from Π by replacing/removing subjective literals and/or deleting associated rules in Π as follows: for any subject literal φ in a rule,

- If φ is $\mathbf{K}\,\ell$ and $W \models \varphi$ then replace $\mathbf{K}\,\ell$ with ℓ; otherwise, delete the rule from Π;

- If φ is $\mathbf{not}\,\mathbf{K}\,\ell$ and $W \models \varphi$ then remove $\mathbf{not}\,\mathbf{K}\,\ell$ from the rule; otherwise, replace $\mathbf{not}\,\mathbf{K}\,\ell$ with $\mathbf{not}\,\ell$;

- If φ is $\mathbf{M}\,\ell$ and $W \models \varphi$ then remove $\mathbf{M}\,\ell$ from the rule; otherwise, replace $\mathbf{M}\,\ell$ with $\mathbf{not}\,\mathbf{not}\,\ell$;

- If φ is $\mathbf{not}\,\mathbf{M}\,\ell$ and $W \models \varphi$ then replace $\mathbf{not}\,\mathbf{M}\,\ell$ with $\mathbf{not}\,\ell$; otherwise, delete the rule from Π.

Let Π be an epistemic logic program and W a non-empty set of consistent sets of literals. Then W is a *world view* of Π if $W = \mathrm{AS}(\Pi^W)$. An ELP might have multiple world views, e.g., the program

$$\{\, p \ or \ q, \quad r \leftarrow \mathbf{not}\,\mathbf{M}\,p, \quad \neg p \leftarrow \mathbf{M}\,r, \mathbf{not}\,q \,\}$$

has two world views $\{\{p\},\{q\}\}$ and $\{\{q,r\}\}$.

4.2 Computing Epistemic Labelling Sets

An $EAF = (AF, \varphi)$ is called a *normal* EAF if φ is a CNF form $\varphi = \psi_1 \wedge \cdots \wedge \psi_k$ in which $\psi_i\,(1 \leq i \leq k)$ is a disjunction of *simple epistemic literals* of the form $\mathbf{E}\lambda(x)$ or $\neg\mathbf{E}\lambda(x)$ where $\mathbf{E}$ is either $\mathbf{K}$ or $\mathbf{M}$ and $\lambda \in \{\texttt{in}, \texttt{out}, \texttt{und}\}$.[7] In the following, we assume that an argumentation framework is represented by atoms of the forms $arg(x)$ (x is an argument) or $att(x,y)$ (x attacks y).

We first introduce a set of rules for argumentative reasoning in logic programming. These rules are similar to the rules introduced by Sakama and Rienstra (2017), with some modification so that they are generic in the sense that they are independent from the input AF. The program makes use of the extended syntax of answer set programming such as aggregate and choice atoms which are implemented in available answer set solvers such as $\texttt{clingo}$ or $\texttt{dlv}$[8].

Definition 8 (AF rules) The set Γ_{AF} consists of the following *AF rules*:

$$n\_atts(X, N) \leftarrow arg(X), N = \#count\{Y : att(Y, X)\}. \tag{2}$$
$$\texttt{in}(X) \leftarrow arg(X), n\_atts(X, 0). \tag{3}$$
$$\texttt{in}(X) \leftarrow n\_atts(X, N), N == \#count\{Y : \texttt{out}(Y), att(Y, X)\}. \tag{4}$$
$$\texttt{out}(X) \leftarrow arg(X), \texttt{in}(Y), att(Y, X). \tag{5}$$
$$\leftarrow arg(X), \texttt{in}(X), att(Y, X), \mathbf{not}\,\texttt{out}(Y). \tag{6}$$
$$\leftarrow arg(X), \texttt{out}(X), n\_atts(X, N), N == \#count\{Y : \mathbf{not}\,\texttt{in}(Y), att(Y, X)\}. \tag{7}$$

In the above rules, the extended syntax of ASP is used as follows. For an argument X,

- the atom $\#count\{Y : att(Y, X)\}$ encodes the number of attackers of X;

---

7. By Prop. 1, $\mathbf{M}\,\neg\lambda(x)$ (resp. $\mathbf{K}\,\neg\lambda(x)$) is rewritten by $\neg\mathbf{K}\,\lambda(x)$ (resp. $\neg\mathbf{M}\,\lambda(x)$).
8. https://potassco.org or https://www.mat.unical.it/DLV2/

- $\#count\{Y : \mathtt{out}(Y), att(Y, X)\}$ (resp. $\#count\{Y : \mathbf{not}\ in(Y), att(Y, X)\}$) encodes the number of attackers of X which are labelled $\mathtt{out}$ (resp. not labelled $\mathtt{in}$);

- $==$ tests for equality between two values.

Rule (2) computes the number of attacks against an argument X. Rule (3) implies that if there is no attacker of X then X is labelled $\mathtt{in}$. Rule (4) states that if all attackers of X is labelled $\mathtt{out}$ then X is labelled $\mathtt{in}$. Rule (5) states that an argument X is labelled $\mathtt{out}$ if one of its attackers, Y, is labelled $\mathtt{in}$. The constraint (6) states that every $\mathtt{in}$-labelled argument X has no attacker Y which is not labelled $\mathtt{out}$. The constraint (7) states that every $\mathtt{out}$-labelled argument X has at least one attacker which is labelled $\mathtt{in}$.

Definition 9 (EC rules) Suppose an epistemic constraint $\varphi = \psi_1 \wedge \cdots \wedge \psi_k$ where $\psi_i = \neg E_1 \vee \cdots \vee \neg E_m \vee E_{m+1} \vee \cdots \vee E_n$ $(1 \leq i \leq k)$ and $(\neg)E_j$ $(1 \leq j \leq n)$ is a simple epistemic literal. Then the set Γ_φ of *EC rules* are defined as

$$\Gamma_\varphi = \{ \leftarrow E_1, \ldots, E_m, \mathbf{not}\ E_{m+1}, \ldots, \mathbf{not}\ E_n \mid \psi_i \text{ is in } \varphi \}.$$

Definition 10 (EAF program) A normal EAF (Ar, att, φ) is transformed to an *epistemic AF program* (*EAF program*, for short)

$$\Pi^\omega_{EAF} = L(AF) \cup \Gamma_{AF} \cup \Gamma_\varphi \cup \Gamma^\omega$$

where $L(AF) = \{ arg(x) \mid x \in Ar \} \cup \{ att(x, y) \mid (x, y) \in att \}$, $\omega \in \{co, st, gr, pr\}$, and Γ^ω is defined as follows:

1. Γ^{co} consists of the following rule:

$$1\{\mathtt{in}(X); \mathtt{out}(X); \mathtt{und}(X)\}1 \leftarrow arg(X)$$

   where the choice atom $1\{\mathtt{in}(X); \mathtt{out}(X); \mathtt{und}(X)\}1$ indicates that exactly one of the atoms $\mathtt{in}(X)$, $\mathtt{out}(X)$, and $\mathtt{und}(X)$ must be true.

2. Γ^{st} consists of the following rule:

$$1\{\mathtt{in}(X); \mathtt{out}(X)\}1 \leftarrow arg(X).$$

3. Γ^{gr} consists of the following rule:

$$\mathtt{und}(X) \leftarrow arg(X), \mathbf{not}\ \mathtt{in}(X), \mathbf{not}\ \mathtt{out}(X).$$

4. Γ^{pr} consists of the following rules:

$$1\{\mathtt{in}(X); \mathtt{out}(X)\} \leftarrow arg(X)$$
$$\mathtt{p\text{-}in}(X) \leftarrow arg(X), \mathtt{in}(X), \mathbf{not}\ \mathtt{out}(X)$$
$$\mathtt{p\text{-}out}(X) \leftarrow arg(X), \mathbf{not}\ \mathtt{in}(X), \mathtt{out}(X)$$
$$\mathtt{p\text{-}und}(X) \leftarrow arg(X), \mathtt{in}(X), \mathtt{out}(X)$$

   where $\mathtt{p\text{-}in}$, $\mathtt{p\text{-}out}$, and $\mathtt{p\text{-}und}$ are new unary predicates. $1\{\mathtt{in}(X); \mathtt{out}(X)\}$ is a choice atom requiring that at least one of the atoms in $\{\mathtt{in}(X), \mathtt{out}(X)\}$ is true.

With this setting, we have the next result.

Proposition 14 *Let $EAF = (Ar, att, \varphi)$ be a normal EAF, and Π^ω_{EAF} its transformed ELP. Then there is 1-1 correspondence between the (non-empty) ω-epistemic labelling sets of EAF and the world-views of Π^ω_{EAF} for $\omega \in \{co, st, pr, gr\}$.*

Proof: There is 1-1 correspondence between answer sets of $L(AF) \cup \Gamma_{AF} \cup \Gamma^\omega$ and ω-labellings of AF as shown by Sakama and Rienstra (2017). Since the world views of Π^ω_{EAF} are non-empty collections of answer sets satisfying the constraints in Γ_φ, they correspond to the (non-empty) ω-epistemic labelling sets of $EAF = (Ar, att, \varphi)$. $\qquad\qquad\qquad\qquad\qquad\square$

The system presented in this section could be realized by an epistemic logic program solver such as the one presented by Son, Le, Kahl, and Leclerc (2017).

5. Related Work

EAF could be viewed as an approach to limiting the set of extensions (or labellings) of an argumentation framework for semantical consideration and this is similar, at least in the spirit, to argumentation with preferences and probabilistic argumentation. By introducing epistemic constraints, it is similar to works focusing on a reasoner's belief. In this section, we compare EAF and related studies.

5.1 Constrained Argumentation Framework

Constrained argumentation frameworks (CAF) proposed in (Coste-Marquis, Devred, & Marquis, 2006) are syntactically similar to EAF. Both are of the form $\langle A, R, C \rangle$ where (A, R) is an AF and C is a propositional formula (over A) in a CAF whilst it is an epistemic formula (over $\mathcal{A}_{AF}$) in an EAF. The key distinction between CAF and EAF lies in the use of the constraint. In CAF, C is imposed on the admissibility of sets of arguments, effectively changing the set of extensions of the AF and leading to a new set of extensions of the original AF. In contrast, φ does not change the labellings of the original AF in an EAF (AF, φ). For instance, a preferred $\mathcal{C}$-extension of a CAF $\langle A, R, C \rangle$ is a *subset* of a preferred extension of $\langle A, R \rangle$ in general. In EAF, on the other hand, every element in an ω-epistemic labelling set is an ω-labelling of the original AF. A CAF represents constraints as propositional formulas that specify conditions on arguments at the objective level. In contrast, an EAF represents constraints as epistemic formulas that specify conditions on arguments at the subjective level. Generally, an epistemic constraint of the type $\mathbf{K}\,F$ (an agent believes F) is also represented by CAF as a constraint F. On the other hand, an epistemic constraint of the type $\mathbf{M}\,F$ (an agent possibly believes F) is not represented by CAF. A more detailed comparison is given below.

Example 7 Consider the AF from Example 1.

- *Semantical differences*: Consider $CAF = (AF, b)$ and $EAF = (AF, \mathbf{K}\,\mathtt{in}(b))$ and the stable semantics of AF. In principle, both CAF and EAF indicate a bias of b.

  The set $\{a, c\}$ is not a stable $\mathcal{C}$-extension in CAF. In contrast, its equivalent labelling, $\{\mathtt{in}(a), \mathtt{in}(c), \mathtt{out}(b), \mathtt{out}(d)\}$, is still a stable labelling of AF in EAF.

In this case, both CAF and EAF accept the same set of conclusions. CAF has only one stable $\mathcal{C}$-extension $\{b, d\}$ and EAF has only one stable epistemic labelling set $\{\{\texttt{in}(b), \texttt{in}(d), \texttt{out}(a), \texttt{out}(c)\}\}$.

- **$\mathbf{M}\varphi$ *has no equivalent representation in CAF*:** for the sake of discussion, let us consider $EAF' = (AF, \mathbf{M}\,\texttt{in}(b))$. Intuitively, EAF' states that b is possibly accepted. As such, EAF' has a stable epistemic labeling set which consists of all possible stable labellings of AF, since AF has one stable labelling containing b. To obtain this result, i.e., to keep all stable extensions of AF as $\mathcal{C}$-stable extensions in the CAF representation, one would likely have to use $(AF, \top)$ as the corresponding CAF. However, this translation would not be adequate for situation where the original AF does not have a stable extension containing b. In that case, $EAF' = (AF, \mathbf{M}\,\texttt{in}(b))$ would have $\{\,\}$ as its stable epistemic labelling set whereas $(AF, \top)$ will still have the same set of stable extensions.

In general, we can prove the following: let Ar be a set of arguments and $a \in Ar$.

- for every formula C over Ar, there exists an epistemic formula $\mathbf{K}C_e$ such that for every attack relation $att \subseteq Ar \times Ar$, $\mathcal{S}$ is the set of stable $\mathcal{C}$-extensions of (Ar, att, C) iff $\{\, S \mid S \in \mathcal{S} \,\}$ is the stable epistemic labelling set of $(Ar, att, \mathbf{K}C_e)$;

- there exists no formula C_f over Ar such that for every attack relation $att \subseteq Ar \times Ar$, $\mathcal{S}$ is the set of stable $\mathcal{C}$-extensions of (Ar, att, C_f) iff $\{\, S \mid S \in \mathcal{S} \,\}$ is the stable epistemic labelling set of $(Ar, att, \mathbf{M}\,\texttt{in}(a))$.

The first item shows that EAF could be used to represent CAF despite the semantical differences at the basic level, i.e., in the definition of admissible set. On the other hand, the second item shows that something expressible in EAF cannot be done using CAF. We note that similar results hold for the other types of semantics studied in this paper as well. In addition, EAF allows reasoning about undecided arguments (using $\texttt{und}(x)$ formulas), while in CAF an argument may take only two states, accepted or rejected, corresponding respectively to $\texttt{in}$ or $\texttt{out}$ labels of EAF.

It is worth mentioning that CAF can represent *bipolar argumentation frameworks* (BAF), which was introduced by Cayrol and Lagasquie-Schiex (2009), by representing support relations as the constraint $C_s = \bigwedge_{(a,b) \in Supp} (a \Rightarrow b)$ where $Supp$ is the set of support relations and $\Rightarrow$ represents material implication. The above consideration implies that BAF is also encoded in EAF as $(AF, \mathbf{K}\,C_s)$.

There are several approaches for selecting extensions of AF. Dauphin, Cramer, and van der Torre (2018) introduce a method of choosing an extension from multiple extensions. Given a partial extension that specifies arguments to be included/excluded in the final outcome, it constructs a decision graph for classifying other arguments to reach a total extension. It provides a step-wise decision making framework for choosing extensions, which is different from EAF that specifies constraints over arguments for choosing intended labellings. Dimopoulos, Mailly, and Moraitis (2018) introduce *control argumentation frameworks* (CTAF). A CTAF is defined as a triple (F, C, U) where F is a fixed AF, U is an AF that contains uncertain arguments or attacks, and C is an AF that a user can choose to use or not. Given a target set T of arguments, they consider whether an AF is built from a CTAF such that arguments in T are credulously/skeptically accepted (or *controllable*) by fixing uncertain arguments/attacks in U and choosing appropriate ones from C. They encode conditions for controllability using quantified Boolean formulas and investigate the decision problem

whether a CTAF is controllable or not. In EAF the controllability of arguments is represented using epistemic constraints, while EAF does not distinguish F, C and U as CTAF does.

5.2 Abstract Dialectical Framework

Another extension of Dung's AF is *abstract dialectical framework* (ADF) (Brewka & Woltran, 2010) where each argument has an associated acceptance condition expressed by a propositional formula over the existing arguments. For instance, the ADF "$a\,[\mathbf{t}]$, $b\,[\neg a \vee c]$, $c\,[\mathbf{t}]$" represents that the argument a and c are always accepted, and b is accepted if $\neg a \vee c$ is true. The ADF has the stable extension $\{a, b, c\}$ which is not stable in AF. Thus, an ADF acquires new meaning in the presence of support relations. Generally, an ADF will not be directly translated into EAF, since it changes extensions of an AF in general. In EAF individual arguments do not have acceptance conditions, while epistemic constraints specify beliefs concerning which extensions are to be selected in the final outcome. The ADF also handles preference in its extended syntax called a *prioritized* ADF (Brewka, Strass, Ellmauthaler, Wallner, & Woltran, 2013). A prioritized ADF expresses *dynamic preferences* in which preferences depend on arguments to be accepted. In this framework, an extension E is constructed in a way compatible with the preference information contained in E. In EAF, on the other hand, belief change of an agent may lead to constructing new epistemic labelling sets, while those sets just change the selection of labellings of the original AF. ADF is similar to CAF in the sense that constraints are represented by propositional formulas over arguments. As a result, it has limitation similar to CAF, that is, an epistemic constraint of the type $\mathbf{M}F$ in EAF is not represented in ADF.

5.3 Probabilistic Argumentation

Probabilistic argumentation as proposed in (Hunter, 2013; Hunter & Thimm, 2017) focuses on the uncertainty of arguments rather than reasoners' beliefs. This approach represents the beliefs of agents by a probability assignment to arguments (e.g., by Hunter, 2013) or an epistemic labelling (e.g., by Hunter & Thimm, 2017). It provides methods for computing epistemic extensions of an AF which contain arguments with probability greater than a certain threshold or assigning labels to arguments in accordance to the probability of the labelling, i.e., it merges an objective evidence and subjective beliefs in a single framework, which is in contrast to our approach. Moreover, it differs from EAFs significantly as beliefs are represented quantitatively using probabilities. On the other hand, there would be a connection between probabilistic argumentation and EAF. For instance, we consider that for each $EAF = (AF, \varphi)$ and $\omega \in \{co,\ st,\ gr,\ pr\}$, there would exist a probabilistic distribution P with respect to AF with the property that x is believed wrt P ($P(x) > 0.5$)) then $\mathrm{in}(x)$ is skeptically entailed by every ω-labelling set of EAF. We believe that the inverse could be true as well. We leave the precise formulation and proof of this interesting problem for future work. Recent work in this direction has introduced *epistemic attack semantics* that considers extended probability distribution, which assigns degrees of belief to arguments and attacks (Thimm, Polberg, & Hunter, 2018) which is then further investigated in dynamic setting (Hunter, Polberg, & Potyka, 2018). Whether formulas in EAF could sufficiently model this type of extension is an open question that we intend to pursuit as well.

5.4 Argumentation with Preference

Argumentation with preferences or priorities has been studied extensively in recent years. Preference over arguments is introduced as a preorder relation over arguments in (Amgoud & Cayrol, 1998; Amgoud & Vesic, 2010, 2011; Kaci & van der Torre, 2008), while a new attack relation that ranges from arguments to attacks is used in (Modgil, 2009). Our representation of preferences is close to the approach employing a preorder but there are differences from them. For instance, consider $AF = (\{a, b\}, \{(a, b)\})$ with the preference $b \succeq a$, where a attacks b while b has higher priority than a (such (a, b) is called a *critical attack*). In this case, Kaci and van der Torre (2008) provide its semantics by extensions of $AF_1 = (\{a, b\}, \{\})$, and Amgoud and Vesic (2010) convert AF to $AF_2 = (\{a, b\}, \{(b, a)\})$. Amgoud and Vesic (2011) consider $\{b\}$ as the acceptable extension under the postulate that preferences are privileged in critical attacks. In these studies, the structure of the original argumentation graph or semantics is changed, and as a result, extensions of the preference-based AF are not extensions selected from those of AF. Wakaki (2015) introduces *preference-based AF* (PAF) which, as we do, selects extensions based on preference relation over arguments. Our representation of preference in EAF is different from PAF in the sense that EAF can represent preference over not only arguments but justification states. Moreover, EAF can represent conditional preferences using epistemic formulas.

Value-based argumentation framework (VAF) (Bench-Capon, 2002) represents preference in AF by assigning values to arguments. In VAF acceptable arguments may change depending on the order of values. Arguments acceptable irrespective of any value order are called *objectively acceptable* and those acceptable for some order are called *subjectively acceptable*. In EAF justification states of arguments change depending on epistemic constraints, so the effect of epistemic constraints in EAF is similar to the effect of value in VAF. On the other hand, VAF may produce extensions that are not those of the original AF, while EAF produces labellings that are also labellings of the original AF.

Airiau, Bonzon, Endriss, Maudet, and Rossit (2017) consider the problem such that given a profile of argumentation frameworks $(AF_1, \ldots, AF_n)$, one for each agent, can this profile be explained in terms of a single master argumentation framework, an association of arguments with values, and a profile of preference orders over values $(\succeq_1, \ldots, \succeq_n)$, one for each agent? Their approach represents individual views of a common AF by preference orders over values, which is in contrast with our approach in which individual views are encoded by epistemic formulas over arguments. Visser, Hindriks, and Jonker (2012) introduce an *epistemic argumentation framework* for reasoning about preferences with uncertain information. They provide languages and inference schemes for instantiated AFs, which is in contrast with our framework for abstract argumentation. Some studies handle dynamics of preferences. Dimopoulos, Moraitis, and Amgoud (2008) consider negotiation theories based on preference-based argumentation. They formulate dynamics of preferences in the context of aggregate argumentation systems: taking into account the interests of the negotiating agents by the intersection of individual preference relations. In contrast to this, we formulate belief change of a single agent in an EAF. In (Kakas, Moraitis, & Spanoudakis, 2019), the authors specify changes of preferences in the context of "Scenario-based Preferences". Prakken and Sartor (1997) specify defeasible priorities in their argument-based logic programming. These frameworks handle structured argumentation, and they are significantly different from our framework in both representing and reasoning about preferences.

5.5 Multi-agent Argumentation

Judgment aggregation in multi-agent argumentation has been studied by several researchers. Caminada and Pigozzi (2011) study aggregation of individual labellings of a given argumentation framework. They introduce three operators for *skeptical aggregation*, *credulous aggregation* and *super-credulous aggregation*. Those aggregation operators produce outcomes that are compatible with the individual judgments, and are guaranteed to keep rationality in terms of admissibility and related argumentation-theoretic concepts. Awad, Booth, Tohmé, and Rahwan (2017) and Rahwan and Tohmé (2010) study judgment aggregation in abstract argumentation. In their setting agents have a shared argumentation framework while having different labellings of the AF. They give social-choice theoretic analyses of argument evaluation semantics to reach a collective decision. Ganzer-Ripoll, López-Sánchez, and Rodríguez-Aguilar (2017) study collective decision making in bipolar argumentation frameworks. They introduce the notion of *coherent labelling*, which is a relaxed version of the notion of complete labelling. Then they provide an aggregate function that guarantees the coherent collective rationality of the outcome as well as valuable social-choice theoretic properties. Chen and Endriss (2018) study whether properties of argumentation semantics, such as conflict-freeness, admissibility, or stability are preserved or not under aggregation. They show, for instance, that the majority rule always preserves conflict-freeness, while no quota rule can guarantee the preservation of either admissibility or stability unless imposing restrictions on argumentation framework.

In contrast to previous studies, we consider a multi-agent argumentation framework as the collection of EAFs. In our setting agents have different beliefs towards the single AF and those beliefs are explicitly represented by epistemic formulas in individual EAFs. We then considered whether those agents can reach an agreement for acceptance/rejection of a particular argument (not an extension) under credulous or skeptical reasoning. Further investigation is needed to verify whether such an agreement satisfies valuable social-choice properties or not. Schwarzentruber, Vesic, and Rienstra (2012) introduce a logical framework for reasoning about arguments owned by agents and their knowledge about other agents' arguments. They introduce epistemic logics to represent belief state of agents in dialogues and define Kripke semantics. For instance, they represent that "an agent 1 believes that there exists an argument about global warming (*gw*) owned by an agent 2" by the formula: $B_1(\langle U \rangle(gw \land ownedby(2)))$. Our approach is different from theirs in two ways: first EAF is an extension of AF and we do not use modal logic based on Kripke structures. Second, our primary interest in this paper is to represent an agent's own beliefs, and we do not consider reasoning about beliefs of other agents. An EAF realizes meta-level reasoning about arguments in abstract argumentation frameworks. In this sense, it could be viewed as a kind of *meta-level arguments* discussed in (Boella, Gabbay, van der Torre, & Villata, 2009), while the paper by Boella et al. (2009) does not argue for a method of introducing epistemic formulas as we do in this paper.

6. Conclusion

An epistemic argumentation framework introduces belief of agents to argumentation frameworks. A unique feature of EAF is that it can represent arguments and attacks as objective evidence in AF, while at the same time, it can encode subjective beliefs of individual agents by epistemic constraints over the outcome. By separating objective knowledge and subjective beliefs, individual agents could produce different conclusions based on their biases toward a common AF. Such a situation happens, for instance, in a court case where jurors share the same open AF while could reach different con-

clusions based on their biases. Moreover, the separation has an advantage that an individual agent can easily revise his/her belief without changing the structure of an AF. We introduce EAF for complete, stable, grounded, or preferred semantics, but the framework is applied to other semantics as well.

This paper focuses on representing an agent's own belief in EAF, while EAF could be extended to reasoning about beliefs of other agents and representing an agent's own belief based on beliefs of other agents. This type of belief contains a constraint such that "$\mathbf{K}_1\mathtt{in}(a) \supset \mathbf{M}_2\mathtt{in}(a)$" (if an agent 1 supports the acceptance of an argument a then an agent 2 would not argue against it). Then we could combine some axiomatic system for reasoning about beliefs in EAF. In this paper we consider EAF where AF is fixed and an epistemic constraint φ is variable. On the other hand, EAF is also usable by changing AF with a fixed φ. The *extension enforcement* (Baumann & Brewka, 2010) is the problem of determining whether an AF can be modified so that a given set of arguments becomes a subset of an extension of the AF, for instance, by adding new arguments that interact with existing ones. The problem could be characterized as finding an EAF $(AF', \bigwedge_{a \in S} \mathbf{M}\ \mathtt{in}(a))$ having a non-empty ω-epistemic labelling set where AF' is an expansion of AF and S is a desired set of arguments. As such, the enforcement problem would be characterized by changing AF with a given epistemic constraint. Formal connection to the issue is left for future work.

Acknowledgment

References

Airiau, S., Bonzon, E., Endriss, U., Maudet, N., & Rossit, J. (2017). Rationalisation of profiles of abstract argumentation frameworks: Characterisation and complexity. *J. Artif. Intell. Res.*, *60*, 149–177.

Amgoud, L., & Cayrol, C. (1998). On the acceptability of arguments in preference-based argumentation. In *UAI '98: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, University of Wisconsin Business School, Madison, Wisconsin, USA, July 24-26, 1998*, pp. 1–7.

Amgoud, L., & Vesic, S. (2010). On the role of preference in argumentation frameworks. In *Proc. 22nd IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 219–222.

Amgoud, L., & Vesic, S. (2011). A new approach for preference-based argumentation frameworks. *Ann. Math. Artif. Intell.*, *63*(2), 149–183.

Awad, E., Booth, R., Tohmé, F., & Rahwan, I. (2017). Judgement aggregation in multi-agent argumentation. *J. Log. Comput.*, *27*(1), 227–259.

Baumann, R., & Brewka, G. (2010). Expanding argumentation frameworks: enforcing and monotonicity results. In *Proc. 3rd Int. Conf. Computational Models of Argument, Frontiers in AI and Applications*, Vol. 216, pp. 75–86. IOS Press.

Bench-Capon, T. J. M. (2002). Value-based argumentation frameworks. In *Proc. 9th Int. Workshop on Non-Monotonic Reasoning (NMR)*, pp. 443–454.

Boella, G., Gabbay, D. M., van der Torre, L., & Villata, S. (2009). Meta-argumentation modelling I: methodology and techniques. *Studia Logica*, *93*(1), 297–355.

Brewka, G., Eiter, T., & Truszczynski, M. (2011). Answer set programming at a glance. *Commun. ACM*, *54*(12), 92–103.

Brewka, G., Strass, H., Ellmauthaler, S., Wallner, J. P., & Woltran, S. (2013). Abstract dialectical frameworks revisited. In *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, pp. 803–809.

Brewka, G., & Woltran, S. (2010). Abstract dialectical frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010, Toronto, Ontario, Canada, May 9-13*.

Caminada, M., & Pigozzi, G. (2011). On judgment aggregation in abstract argumentation. *Auton. Agents Multi Agent Syst.*, *22*(1), 64–102.

Caminada, M. W. A., & Gabbay, D. M. (2009). A logical account of formal argumentation. *Studia Logica*, *93*(2-3), 109–145.

Cayrol, C., & Lagasquie-Schiex, M. (2009). Bipolar abstract argumentation systems. In *Argumentation in Artificial Intelligence*, pp. 65–84.

Chen, W., & Endriss, U. (2018). Aggregating alternative extensions of abstract argumentation frameworks: Preservation results for quota rules. In Modgil, S., Budzynska, K., & Lawrence, J. (Eds.), *Computational Models of Argument - Proceedings of COMMA 2018, Warsaw, Poland, 12-14 September 2018*, Vol. 305 of *Frontiers in Artificial Intelligence and Applications*, pp. 425–436. IOS Press.

Coste-Marquis, S., Devred, C., & Marquis, P. (2006). Constrained argumentation frameworks. In *Proc. of KR 2006*, pp. 112–122.

Dauphin, J., Cramer, M., & van der Torre, L. W. N. (2018). A dynamic approach for combining abstract argumentation semantics. In *Dynamics, Uncertainty and Reasoning, The Second Chinese Conference on Logic and Argumentation, CLAR 2018, Hangzhou, China, 16-17 June 2018*, pp. 21–43.

Dimopoulos, Y., Mailly, J., & Moraitis, P. (2018). Control argumentation frameworks. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 4678–4685.

Dimopoulos, Y., Moraitis, P., & Amgoud, L. (2008). Theoretical and computational properties of preference-based argumentation. In *ECAI 2008 - 18th European Conference on Artificial Intelligence, Patras, Greece, July 21-25, 2008, Proceedings*, pp. 463–467.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, *77*, 321 – 357.

Dvorák, W. (2011). On the complexity of computing the justification status of an argument. In *Theorie and Applications of Formal Argumentation - First International Workshop, TAFA 2011. Barcelona, Spain, July 16-17, 2011, Revised Selected Papers*, pp. 32–49.

Dvorák, W., & Dunne, P. E. (2018). Computational Problems in Formal Argumentation and their Complexity. In *Handbook of Formal Argumentation*, pp. 631–688.

Ganzer-Ripoll, J., López-Sánchez, M., & Rodríguez-Aguilar, J. A. (2017). A multi-agent argumentation framework to support collective reasoning. In Aydogan, R., Baarslag, T., Gerding, E., Jonker, C. M., Julián, V., & Sánchez-Anguix, V. (Eds.), *Conflict Resolution in Decision Making - Second International Workshop, COREDEMA 2016, The Hague, The Netherlands, August 29-30, 2016, Revised Selected Papers*, Vol. 10238 of *Lecture Notes in Computer Science*, pp. 100–117.

Gelfond, M. (1991). Strong introspection. In *Proceedings AAAI-91*, pp. 386–391.

Gelfond, M., & Lifschitz, V. (1991). Classical negation in logic programs and disjunctive databases. *New Generation Computing*, 365–387.

Hunter, A. (2013). A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, *54*, 47–81.

Hunter, A., Polberg, S., & Potyka, N. (2018). Updating belief in arguments in epistemic graphs. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018.*, pp. 138–147.

Hunter, A., & Thimm, M. (2017). Probabilistic reasoning with abstract argumentation frameworks. *Journal of Artificial Intelligence Research*, *59*, 565–611.

Kaci, S., & van der Torre, L. (2008). Preference-based argumentation: arguments supporting multiple values. *International Journal of Approximate Reasoning*, *48*, 730–751.

Kahl, P. T., Leclerc, A. P., & Son, T. C. (2016). A parallel memory-efficient epistemic logic program solver: Harder, better, faster. In *ASPOCP*. `https://arxiv.org/pdf/1608.06910.pdf`.

Kakas, A. C., Moraitis, P., & Spanoudakis, N. I. (2019). *GORGIAS*: Applying argumentation. *Argument Comput.*, *10*(1), 55–81.

Lifschitz, V., Tang, L. R., & Turner, H. (1999). Nested expressions in logic programs. *Annals of Mathematics and Artificial Intelligence*, *25*, 369–389.

Modgil, S. (2009). Reasoning about preferences in argumentation frameworks. *Artif. Intell.*, *173*(9-10), 901–934.

Prakken, H., & Sartor, G. (1997). Argument-based extended logic programming with defeasible priorities. *J. Appl. Non Class. Logics*, *7*(1), 25–75.

Rahwan, I., & Tohmé, F. (2010). Collective argument evaluation as judgement aggregation. In *9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010), Toronto, Canada, May 10-14, 2010, Volume 1-3*, pp. 417–424.

Sakama, C., & Rienstra, T. (2017). Representing argumentation frameworks in answer set programming. *Fundam. Inform.*, *155*(3), 261–292.

Sakama, C., & Son, T. C. (2019). Epistemic argumentation framework. In *PRICAI 2019: Trends in Artificial Intelligence - 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, August 26-30, 2019, Proceedings, Part I*, pp. 718–732.

Schwarzentruber, F., Vesic, S., & Rienstra, T. (2012). Building an epistemic logic for argumentation. In *Proceedings of the 13th European Conference on Logics in Artificial Intelligence*, pp. 359–371. Springer, LNAI 7519.

Shen, Y.-D., & Eiter, T. (2016). Evaluating epistemic negation in answer set programming. *Artificial Intelligence*, *237*, 115–135.

Son, T. C., Le, T., Kahl, P. T., & Leclerc, A. P. (2017). On computing world views of epistemic logic programs. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pp. 1269–1275.

Thimm, M., Polberg, S., & Hunter, A. (2018). Epistemic attack semantics. In *Computational Models of Argument - Proceedings of COMMA 2018, Warsaw, Poland, 12-14 September 2018*, pp. 37–48.

Visser, W., Hindriks, K. V., & Jonker, C. M. (2012). Argumentation-based qualitative preference modeling with incomplete and uncertain information. *Group Decision and Negotiation*, *21*(1), 99–127.

Wakaki, T. (2015). Preference-based argumentation built from prioritized logic programming. *Journal of Logic and Computation*, *25*(2), 251–301.