Autonomous Agents and Multi-Agent Systems (2021) 35:18
https://doi.org/10.1007/s10458-021-09493-y

Arguing and negotiating using incomplete negotiators profiles

Yannis Dimopoulos[1] · Jean-Guy Mailly[2] · Pavlos Moraitis[2,3]

Accepted: 7 January 2021 / Published online: 19 April 2021

Abstract
Computational argumentation has taken a predominant place in the modeling of negotiation dialogues over the last years. A competent agent participating in a negotiation process is expected to decide its next move taking into account an, often incomplete, model of its opponent. This work provides a complete computational account of argumentation-based negotiation under incomplete opponent profiles. After the agent identifies its best option, in any state of a negotiation, it looks for suitable arguments that support this option in the theory of its opponent. As the knowledge on the opponent is uncertain, the challenge is to find arguments that, ideally, support the selected option despite the uncertainty. We present a negotiation framework based on these ideas, along with experimental evidence that highlights the advantages of our approach.

Keywords Argumentation · Automated negotiation · Multi-agent systems

1 Introduction

During the last years computational argumentation has taken a predominant place in the modeling of negotiation dialogues (for a survey see Dimopoulos and Moraitis [18], Rahwan et al. [42]). The goal of a negotiation dialogue is to allow interacting agents to resolve conflicts and reach a mutually accepted agreement, which in this work is a mutually accepted offer (*e.g.* the price of a product, the mode of payment). In an argumentation-based negotiation (ABN), agents choose offers that are likely to be accepted by the opponent and exchange arguments that support these offers, either based on their own theories (see *e.g.* Amgoud et al.

✉ Yannis Dimopoulos
yannis@cs.ucy.ac.cy

Jean-Guy Mailly
jean-guy.mailly@u-paris.fr

Pavlos Moraitis
pavlos.moraitis@u-paris.fr; pavlos@argument-theory.com

1 Department of Computer Science, University of Cyprus, Nicosia, Cyprus

2 LIPADE, University of Paris, Paris, France

3 Argument Theory, Paris, France

[4], Amgoud and Kaci [3], Kakas and Moraitis [31], Dung et al. [22], Parsons et al. [40], Hadidi et al. [27]), or based on the opponent's profile (see *e.g.* Hadidi et al. [28], Pilotti et al. [41], Bonzon et al. [15]).

The modeling of the opponent profile is an important issue in negotiation dialogues (and more generally other types of dialogue such as persuasion). As explained in Baarslag et al. [7], although there are important differences between opponent models, there are strong reasons justifying their use, such as the *minimization of negotiation cost*, *the adaptation to the opponent* and *the capacity to reach win-win agreements*, especially in cooperative environments. Learning the opponent profile means learning its acceptance and bidding strategies, the deadlines and its preference profile [7]. In most of the proposed works, the (online) opponent modeling is based on learning techniques (see *e.g.* Baarslag et al. [6] for a survey). Apart from the fact that learning the opponent profile with traditional learning techniques is not an easy task, as pointed out by Zafari and Mofakham [49], those techniques seem better suited to game-theoretic (or utility-based) negotiations, rather than argumentation-based negotiations. Other works (although they concern persuasion dialogues and legal disputes), have proposed a probabilistic approach for dealing with the uncertainty about the opponent profile. In these works (*e.g.* Hadjinikolis et al. [29], Riveret et al. [45], Hunter [30]), probabilities are used in different ways for finding the arguments that are most likely to be accepted by the opponent. Finally, some works (*e.g.* Rienstra et al. [44], Oren and Norman [39], Black and Atkinson [14]) investigate other approaches to modeling the opponent profile in argumentation-based dialogues.

In this work we completely adopt the aforementioned reasons justifying the importance to use an opponent model in automated negotiations. However, we propose an alternative approach to the traditional learning-based approaches (suitable essentially to game-theoretic negotiations) for opponent modelling, which is particularly suited to argumentation-based negotiations. Our work proposes an original approach to deal with uncertainty in knowledge representation that allows to represent and assume from the beginning of the dialogue, different possible profiles of the opponent. Then our approach allows the proponent agent to choose its best offer (with respect to its own arguments) and to look for supporting arguments that increase the chance of agreement about this offer, despite possible counter-arguments the opponent may have. This is a major difference from other approaches and more particularly those based on learning methods where a certain number of rounds is necessary in order to start learning the profile of the opponent. The advantage provided by our approach may be particularly useful in negotiations with time limits (or number of rounds) constraints. Moreover, our approach allows for a proponent agent to progressively refine (as the dialogue evolves) the initially considered profiles for identifying (learning) the profile that is closer to the real one, without the necessity to use a learning method. This is done based on the arguments and attacks that the opponent agent uses as justification to rejecting an offer. That allows for a proponent agent to progressively adapt its bidding strategy as the dialogue evolves. This aspect of our approach is also a novelty with respect to the other approaches in the literature. More importantly, as we have experimentally shown, these features of our approach have a positive impact on the number of reached agreements.

Thus this work advances the state of the art in argumentation-based negotiation by making original contributions to the *opponent modeling*, and the associated *acceptance strategy* (*i.e.* what offers are most likely to be accepted) as well as *bidding strategy* (*i.e.* the strategy that an agent applies for choosing the next offer). For opponent modeling, it builds on the work of Dimopoulos et al. [19] on *Control Argumentation Frameworks (CAFs)*, a formalism for modeling the uncertainty about the *opponent profile*. More specifically, it borrows the concepts of "on/off" arguments (*i.e.* arguments we don't know whether they are present or not in a theory), and the three different categories of attacks (*i.e.* attacks we know their existence and direction, attacks we know the existence but not the direction, attacks we don't know the existence but we know the direction). This allows generating different profiles modeled as completions of the known part of the opponent's theory, and seeking offers that satisfy all possible profiles (or as many as possible). Regarding the *bidding* and *acceptance strategies*, the originality of this work lies in the assumption that in argumentation-based negotiation, a central challenge for an agent is to lead, by means of appropriate arguments, its counter party to change its theory, and eventually accept the offer it proposes, hence influencing its *acceptance strategy*. Thus, in our approach, we propose a *bidding strategy* that relies on the previous assumption. More precisely, the idea is that a proponent agent uses first its own theory for choosing the best offer to propose, but next, it uses the incomplete theory of its opponent to find the arguments to support it. Then, it seeks and puts forward a set of arguments called *control configuration*, that could reinstate the supporting arguments, if these are rejected in the current state of the argumentative negotiation theories of all (or most) of the generated opponent profiles. Once the arguments of the control configuration are inserted in the opponent theory, they would, ideally, allow it to reach an agreement with the proponent, thus they alter its *acceptance decision*. The integration of control configurations in the opponents theories could be considered as a *persuasion dialogue* embedded within the negotiation dialogue (see *e.g.* van Laar and Krabbe [47]).

This paper is based on a previous publication [20]. The current version has been extended in several ways:

- we provide a deeper discussion of the contribution of this work, and its novelty with respect to existing works;
- we have added a proper background section, introducing formally propositional logic and QBFs, abstract argumentation, and CAFs;
- we have added a new section that presents interesting theoretical results about our negotiation framework, in particular concerning the optimality of the provided solutions and the completeness of the negotiation dialogue;
- we describe in depth a new set of experiments, that highlight the interest of our approach in handling uncertainty in argument-based negotiations.

The paper is organized as follows. Section 2 presents the background knowledge on propositional logic, Quantified Boolean Formulas (QBFs) and abstract argumentation. Then Sect. 3 presents in details the components of the original negotiation framework we propose in this work for argumentation-based negotiation. Section 4 presents some important theoretical results that characterize our framework while Sect. 5 presents an experimental evaluation of our framework by using different parameters that put in evidence the added value of our approach. Finally, Sect. 6 describes the related work, while Sect. 7 concludes the paper and presents some interesting tracks for future research.

2 Background

2.1 Propositional logic and quantified boolean formulas

We first introduce background notions and notations regarding propositional logic and Quantified Boolean Formulas (QBFs).

We consider a set of Boolean variables V, *i.e.* each of these variables can receive a value in $\mathbb{B} = \{0, 1\}$, where 0 stands for *false* and 1 for *true*. A well-formed propositional formula is:

- an atomic formula x, where $x \in V$;
- a negation $\neg\phi$, where ϕ is a well-formed formula;
- a conjunction $\phi_1 \wedge \phi_2$, where ϕ_1 and ϕ_2 are well-formed formulas;
- a disjunction $\phi_1 \vee \phi_2$, where ϕ_1 and ϕ_2 are well-formed formulas.

An interpretation $\omega : V \to \mathbb{B}$ is a valuation of the Boolean variables, that can be extended to formulas as follows:

- $\omega(\neg\phi) = 1 - \omega(\phi)$;
- $\omega(\phi_1 \wedge \phi_2) = \min(\omega(\phi_1), \omega(\phi_2))$;
- $\omega(\phi_1 \vee \phi_2) = \max(\omega(\phi_1), \omega(\phi_2))$.

We can define additional connectives, *e.g.* the implication $(\omega(\phi_1 \Rightarrow \phi_2) = \omega(\neg\phi_1 \vee \phi_2))$ and the equivalence $(\omega(\phi_1 \Leftrightarrow \phi_2) = \omega((\phi_1 \Rightarrow \phi_2) \wedge (\phi_2 \Rightarrow \phi_1)))$.

When $\omega(\phi) = 1$, we say that ϕ is *satisfied* by ω, or alternatively that ω is a *model* of ϕ, written $\omega \models \phi$.

Quantified Boolean Formulas (QBFs) are an extension of propositional formulas with the universal and existential quantifiers.

Formally, a well-formed QBF is:

- ϕ, where ϕ is a well-formed propositional formula;
- $\exists x, \Phi$, where $x \in V$ and Φ is a well-formed QBF;
- $\forall x, \Phi$, where $x \in V$ and Φ is a well-formed QBF.

If $X = \{x_1, \dots, x_n\}$ is a set of variables, we write $\exists X, \Phi$ as a shortcut for $\exists x_1, \dots, \exists x_n, \Phi$; similarly $\forall X, \Phi$ means $\forall x_1, \dots, \forall x_n, \Phi$.

Any QBF can be transformed into a prenex normal form QBF $\mathcal{Q}_1 X_1\ \mathcal{Q}_2 X_2 \dots \mathcal{Q}_n X_n \phi$ where

- ϕ is a propositional formula called the *matrix*,
- $\mathcal{Q}_1 X_1 \dots \mathcal{Q}_n X_n$ is called the *prefix*,
- $\forall i \in \{1, \dots, n\}, \mathcal{Q}_i \in \{\exists, \forall\}$,
- $\forall i \in \{1, \dots, n-1\}, \mathcal{Q}_i \neq \mathcal{Q}_{i+1}$,
- and $X_1, X_2, \dots, X_n$ are disjoint sets of propositional variables such that $X_1 \cup X_2 \cup \dots \cup X_n = V$.[1]

For instance, the formula $\exists x \forall y (x \vee \neg y) \wedge (\neg x \vee y)$ is satisfied if there is a value for x such that for all values of y the proposition $(x \vee \neg y) \wedge (\neg x \vee y)$ is true.

For more details about propositional logic and QBFs, we refer the reader to, *e.g.*, Biere et al. [13], Kleine Büning and Bubeck [32].

2.2 Abstract argumentation

We introduce the basic notions of abstract argumentation, defined by Dung [21].

Definition 1 (*Argumentation Framework*) An abstract argumentation framework (AF) is a pair $\mathcal{AF} = \langle A, R \rangle$, where A is a set of *arguments*, and $R \subseteq A \times A$ is an *attack relation*.

[1] If some variable $x \in V$ does not explicitly belong to any X_i, *i.e.* $X_1 \cup \dots \cup X_n \subset V$, then it implicitly means that x can be existentially quantified at the rightmost level.

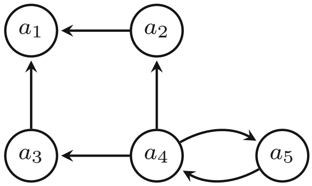


Fig. 1 Example of abstract AF

The intuitive meaning of a attacking b, denoted by $(a, b) \in R$, is that a is a counter-argument for b. We say that a set of arguments S attacks an argument b if $\exists a \in S$ such that $(a, b) \in R$.

Different acceptability semantics were also introduced in Dung [21]. While our approach is generic, we focus on the stable semantics in this paper for exemplifying our negotiation approach.

Definition 2 (*Stable semantics*) Given an $\mathcal{AF} = \langle A, R \rangle$, a set of arguments $S \subseteq A$ is a stable extension if

- S is conflict-free: $\forall a, b \in S, (a, b) \notin R$;
- S attacks every argument not contained in S: $\forall b \in A \setminus S, \exists a \in S$ such that $(a, b) \in R$.

The set of stable extensions of $\mathcal{AF}$ is denoted by $\mathrm{st}(\mathcal{AF})$.

The full catalogue of extension-based semantics is out of the scope of this paper. We refer the interested reader to Dung [21], Baroni et al. [8] for an overview.

Based on the acceptability semantics, we can define the status of any argument, namely *skeptically accepted, credulously accepted* and *rejected arguments*.

Definition 3 (*Argument Acceptance*) Given an $\mathcal{AF} = \langle A, R \rangle$ and an extension-based semantics σ, an argument $a \in A$ is:

- skeptically accepted (with respect to σ) if $\forall E \in \sigma(\mathcal{AF}), a \in E$;
- credulously accepted (with respect to σ) if $\exists E \in \sigma(\mathcal{AF})$ such that $a \in E$;
- rejected (with respect to σ) otherwise.

Example 1 Figure 1 depicts $\mathcal{AF} = \langle A, R \rangle$, with $A = \{a_1, a_2, a_3, a_4, a_5\}$ and $R = \{(a_2, a_1), (a_3, a_1), (a_4, a_2), (a_4, a_3), (a_4, a_5), (a_5, a_4)\}$. Its stable extensions are $\mathrm{st}(\mathcal{AF}) = \{\{a_1, a_4\}, \{a_2, a_3, a_5\}\}$. All the arguments are credulously accepted, and no arguments are skeptically accepted or rejected.

Many efficient computational methods for abstract argumentation are based on logical encodings of the problem (see *e.g.* Dvorák et al. [24], Lagniez et al. [33]). Such encodings have been originally defined by Besnard and Doutre [11]. In the following, we use the logical encoding of stable semantics. This encoding will be helpful for defining the computational method for reasoning with Control Argumentation Frameworks.

Definition 4 (Besnard and Doutre [11]) Let $\mathcal{AF} = \langle A, R \rangle$ be an AF. For each argument $x_i \in A$, we define a propositional variable acc_{x_i} to represent the acceptability of the argument x_i with respect to a particular extension (*i.e.* its membership to the extension).[2]

[2] Since we use the extension-based semantics defined by Dung, we consider binary acceptability statuses for arguments: an argument that is not accepted is rejected.

The formula $\phi_{\text{st}}(\mathcal{AF})$ is defined by

$$\bigwedge_{x_i \in A} (acc_{x_i} \Leftrightarrow (\bigwedge_{(x_j, x_i) \in R} \neg acc_{x_j}))$$

This encoding is such that models of $\phi_{\text{st}}(\mathcal{AF})$ exactly correspond to stable extensions of $\mathcal{AF}$. Moreover, to obtain a stable extension from a model $\omega \models \phi_{\text{st}}(\mathcal{AF})$, we just need to select the arguments $\{x_i \mid \omega(acc_{x_i}) = 1\}$. This means that an argument x_i is accepted (with respect to one of the extensions) if $acc_{x_i} = 1$ in the model corresponding to this extension. Let us notice that, for an AF $\mathcal{AF}$ with no stable extension, $\phi_{\text{st}}(\mathcal{AF})$ has no model.

Example 2 Continuing with $\mathcal{AF}$ given at Fig. 1,

$$\phi_{\text{st}}(\mathcal{AF}) = \begin{cases} (acc_{a_1} \Leftrightarrow (\neg acc_{a_2} \wedge \neg acc_{a_3})) \\ \wedge(acc_{a_2} \Leftrightarrow \neg acc_{a_4}) \\ \wedge(acc_{a_3} \Leftrightarrow \neg acc_{a_4}) \\ \wedge(acc_{a_4} \Leftrightarrow \neg acc_{a_5}) \\ \wedge(acc_{a_5} \Leftrightarrow \neg acc_{a_4}) \end{cases}$$

This formula has exactly two models: $\omega_1(a_1) = \omega_1(a_4) = 1$ and $\omega_1(a_2) = \omega_1(a_3) = \omega_1(a_5) = 0$ on the one hand; and $\omega_2(a_1) = \omega_2(a_4) = 0$ and $\omega_2(a_2) = \omega_2(a_3) = \omega_2(a_5) = 1$ on the other hand. Each of these models correspond to a stable extension of $\mathcal{AF}$.

This encoding can be generalized to represent the relation between any $\mathcal{AF}$ and its stable extensions.

Definition 5 Let A be a set of arguments. For each argument $x_i \in A$, we define a propositional variable acc_{x_i}. For each pair of arguments $(x_i, x_j) \in A \times A$, we define a propositional variable att_{x_i,x_j}. The generalized version of ϕ_{st} is defined by

$$\phi_{st}^{att}(A) = \bigwedge_{x_i \in A} [acc_{x_i} \Leftrightarrow (\bigwedge_{x_j \in A} (att_{x_j,x_i} \Rightarrow \neg acc_{x_j}))]$$

A model of this formula corresponds to the structure of an argumentation framework (given by the att_{x_i,x_j} variables) and one extension of this argumentation framework (given by the acc_{x_i} variables). This means that this formula encodes the stable semantics for any AF built on the set of arguments A, which explains why there is no reference to an attack relation in it.

In order to represent the stable semantics for a particular $\mathcal{AF} = \langle A, R \rangle$, we can use the formula:

$$\phi_{st}^{att}(A) \wedge (\bigwedge_{(x_i,x_j) \in R} att_{x_i,x_j}) \wedge (\bigwedge_{(x_i,x_j) \in (A \times A) \setminus R} \neg att_{x_i,x_j})$$

This formula is equivalent to $\phi_{\text{st}}(\mathcal{AF})$ from Definition 4: each model corresponds to an extension of $\mathcal{AF}$ when only its acc-variables are considered.

Example 3 We continue the previous example, with $\mathcal{AF} = \langle A, R \rangle$ from Fig. 1. The conjunction of literals that represents the attacks of this AF is:

$$att_{\mathcal{AF}} = att_{a_2,a_1} \wedge att_{a_3,a_1} \wedge att_{a_4,a_2} \wedge att_{a_4,a_3} \wedge att_{a_4,a_5} \wedge att_{a_5,a_4} \\ \wedge \bigwedge_{(a_i,a_j) \in (A \times A) \setminus R} \neg att_{a_i,a_j}$$

When the conjunction $\phi_{st}^{att}(A) \wedge att_{\mathcal{AF}}$ is made, all the parts of the formula that involve an att-variable can be simplified:

– if att_{a_i,a_j} appears in $att_{\mathcal{AF}}$, then $att_{a_i,a_j} \Rightarrow \neg acc_{a_i}$ can be replaced by $\neg acc_{a_i}$;
– if $\neg att_{a_i,a_j}$ appears in $att_{\mathcal{AF}}$, then $att_{a_i,a_j} \Rightarrow \neg acc_{a_i}$ is a tautology, and can be completely removed from the formula.

These simplifications yield the formula given at Example 2.

Such a generalized encoding of stable semantics has been used *e.g.* for revising argumentation frameworks in Coste-Marquis et al. [17]. We will need it later for defining the computational approach for control argumentation frameworks. More precisely, the representation of attacks by Boolean variables will allow to represent uncertain attacks, *i.e.* attacks that may or may not actually exist. For instance, in Example 3, attacks that certainly exist or certainly do not exist are encoded as unit clauses in the formula (*e.g.* att_{a_2,a_1} expresses that a_2 attacks a_1, and $\neg att_{a_1,a_2}$ expresses that a_1 does not attack a_2). So, if there is no unit clause att_{x_i,x_j} nor $\neg att_{x_i,x_j}$, the formula admits models where this Boolean variable is true (representing AFs where x_i attacks x_j), as well as models where this Boolean variable is false (corresponding to AFs where x_i does not attack x_j).

A similar approach can be used to work with the complete semantics, again using the encoding from Besnard and Doutre [11] as a starting point. Other semantics have a higher complexity, and thus cannot be easily encoded in propositional formula (unless the polynomial hierarchy collapses). However, these semantics can be encoded thanks to QBF formulas (see *e.g.* Egly and Woltran [25] for the preferred semantics).

2.3 Control argumentation frameworks

This section introduces briefly the control argumentation frameworks (CAFs) proposed in Dimopoulos et al. [19], and discusses how they capture the knowledge of an agent on its opponents. On a high level, a CAF is an argumentation framework where arguments are divided in three parts, *fixed*, *uncertain* and *control*.

The $fixed$ part of the theory concerns the certain knowledge that an agent holds about its opponent. This includes arguments as well as attacks that undoubtedly belong to the argumentation theory of the opponent. For instance, a seller agent knows that the customer agent prefers European cars, that safety is an important issue for it and that it prefers electric or gasoline-powered cars than diesel cars.

The $uncertain$ part captures the uncertainty about the presence of arguments in a theory (expressed by the "on/off" arguments as shown below), as well as the presence and the direction of attacks between arguments in this theory. It reflects the uncertainty that arises due to lack of complete information on the current state of the world that determines the decisions of the opponent, but also its beliefs and preferences. For example, the seller agent may not know the income of the customer agent, whether a car is a social status symbol for it, the highest price that it is ready to pay, or whether it is willing to pay more if some extras are included, and payment by installments is accepted.

Let us discuss the different types of uncertainty embedded in CAFs. First, we consider uncertain arguments, *i.e.* arguments that may (or may not) actually be in the framework. There are different reasons for justifying the nature of these arguments:

– in a context of logic-based argumentation [12], the agent might be able to build an argument, without being sure whether the argument premises are true in the current state of the world;

– in a context of dialogue, it is reasonable to consider that the agent has some uncertain information about the arguments that other agents may use or not.

Then, the existence of an attack from an argument x_i to an argument x_j can also be uncertain. For instance, an agent can be uncertain about the existence of a preference of x_j over x_i [2] that would "cancel" the attack. Uncertain arguments and attacks appear in Incomplete AFs [10]. We also consider another kind of uncertainty regarding the attack relation: there may be situations where the agent is sure that there is a conflict between two arguments x_i and x_j, without being sure of the actual direction of the attack(s) (either (x_i, x_j), or (x_j, x_i), or both at the same time). Unknown preferences can also explain this kind of uncertainty in the argumentation framework.

Finally, the *control* part contains arguments that can be used against arguments of the fixed or uncertain parts that attack arguments that are in favour of some offer of the proponent. Therefore, the control part serves to ensure that arguments in the fixed part that support some offer of the seller that is not adequate with some certain (*i.e.* European car) or uncertain (*e.g.* max price, preferred mode of payment) preferences of the customer, can be accepted under some circumstances. For instance, a control argument could allow a seller agent to propose a car from abroad Europe (which is against the known preference of the customer agent and represented in the fixed part) by proposing some interesting options (*e.g.* five airbags knowing that safety is an important issue for the customer and also represented in the fixed part) and in a price that is probably higher than the highest price the customer is intended to pay (this is part of the uncertain knowledge) but which allows the seller to accept a payment by installments, if this is the preferred payment mode for the customer (this is also part of the uncertain knowledge).

Formally, a CAF is defined as follows:

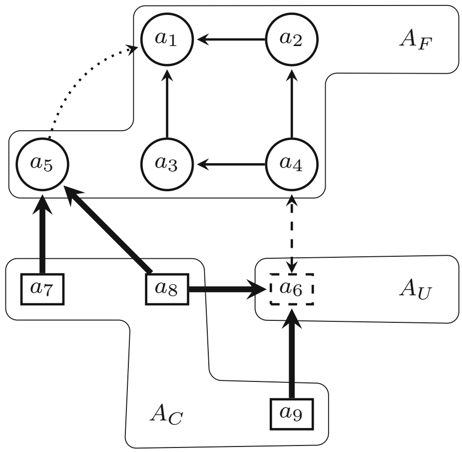
Definition 6 (Dimopoulos et al. [19]) Let $\mathcal{L}$ be a language from which we can build arguments, and let $\texttt{Args}(\mathcal{L})$ be the set which contains all those arguments. A Control Argumentation Framework (*CAF*) is a triple $\mathcal{CAF} = \langle F, C, U \rangle$ where F is the *fixed part*, U is the *uncertain part* and C is the *control part* of $\mathcal{CAF}$ with:

– $F = \langle A_F, \rightarrow \rangle$ where A_F is a set of arguments that we know they belong to the system and $\rightarrow \subseteq (A_F \cup A_U) \times (A_F \cup A_U)$ is an attack relation representing a set of attacks for which we are aware both of their existence and their direction.
– $U = \langle A_U, (\rightleftarrows \cup \dashrightarrow) \rangle$ where A_U is a set of arguments for which we are not sure that they belong to the system, $\rightleftarrows \subseteq (((A_U \cup A_F) \times (A_U \cup A_F)) \setminus \rightarrow)$ is an attack relation representing a set of attacks for which we are aware of their existence but not of their direction, and $\dashrightarrow \subseteq (((A_U \cup A_F) \times (A_U \cup A_F)) \setminus \rightarrow)$ is an attack relation representing a set of attacks for which we are not aware of their existence but we are aware of their direction, with $\rightleftarrows \cap \dashrightarrow = \emptyset$.
– $C = \langle A_C, \Rrightarrow \rangle$ where A_C is a set of arguments, called *control arguments*, that the agent can choose to use or not, and $\Rrightarrow \subseteq \{(a_i, a_j) \mid a_i \in A_C,\ a_j \in A_F \cup A_C \cup A_U\}$ is an attack relation.

A_F, A_U and A_C are disjoint subsets of $\texttt{Args}(\mathcal{L})$.

A CAF features a set of distinct attack relations that capture different sorts of information. Its simplest part is $\langle A_F, \rightarrow \cap (A_F \times A_F) \rangle$, which is a classical AF that contains the indisputable knowledge of the agent on its opponent. The idea of CAFs essentially extends this basic argumentation framework with additional attack relations defined on arguments from the sets A_U and A_C. For instance, there is an attack $(a_i, a_j) \in \rightleftarrows$, with $a_i, a_j \in A_F$ when it is

Fig. 2 Example of a CAF



certain that both arguments exist and are in conflict (*e.g.* because they make mutually exclusive claims), but the direction of the attack(s) is unknown (*e.g* because of lack of information on the intrinsic strength of arguments, or on the preference relation between arguments). An attack $(a_i, a_j) \in \dashrightarrow$, with $a_i \in A_U$ and $a_j \in A_F$, represents a situation where it is unknown whether a_i is present in the system (*e.g.* some of its premises could be false at the current time), but *if a_i is in the system*, then a_i definitely attacks a_j.

Example 4 Let $\mathcal{CAF} = \langle F, C, U \rangle$ be the CAF given at Fig. 2. We use circle nodes to represent the fixed arguments A_F, dashed nodes for the uncertain arguments A_U, and plain rectangle nodes for the control arguments A_C. Similarly, different kinds of arrows represent the different attack relations of the CAF. Plain arrows represent the fixed attacks (*e.g.* $(a_2, a_1) \in \rightarrow$). Dotted arrows represent the uncertain attacks, *i.e.* attacks for which we are not sure of the existence, but if it exists then we are sure of the direction. Here, we have $(a_5, a_1) \in \dashrightarrow$. The two-heads dashed arrows are used for non-directed attacks, that are situations where we are sure that arguments are conflicting, but the actual direction of the attack is uncertain ($(a_4, a_6) \in \rightleftarrows$). Finally, the bold attacks are the control attacks (*e.g.* $(a_7, a_5) \in \Rrightarrow$).

Central to controllability is the notion of *completion* of a CAF. Intuitively, a completion is a classical AF which is built from the CAF, by choosing one of the possible options for each uncertain argument or attack.

Definition 7 (Dimopoulos et al. [19]) Given a CAF $\langle F, C, U \rangle$, a completion of this CAF is an AF $\langle A, R \rangle$, s.t.

- $A = A_F \cup A_C \cup A_{comp}$ where $A_{comp} \subseteq A_U$;
- $R \subseteq \rightarrow \cup \rightleftarrows \cup \dashrightarrow \cup \Rrightarrow$;
- if $(a, b) \in \rightarrow$ and $a, b \in A$, then $(a, b) \in R$;
- if $(a, b) \in \rightleftarrows$ and $a, b \in A$, then $(a, b) \in R$ or $(b, a) \in R$;
- if $(a, b) \in \Rrightarrow$ and $a, b \in A$, then $(a, b) \in R$.

If $(a, b) \in \rightleftarrows$, then in the completion, there is either (a, b), or (b, a), or both at the same time. Note that the definition of a completion leaves the attacks from $\dashrightarrow$ unspecified, as these attacks may not appear in the theory.

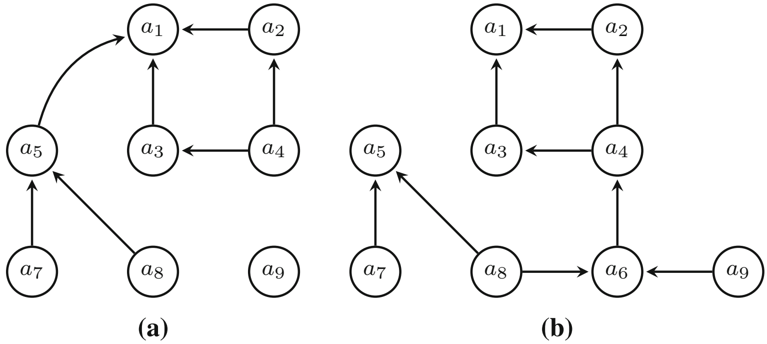


Fig. 3 Some completions of the previous CAF

Example 5 We continue Example 4, and exhibit some completions of the given CAF. In the completion given at Fig. 3a, the attack from a_5 to a_1 is present, while the argument a_6 is absent (thus, the attacks concerning a_6 are also missing). On the opposite, in the completion depicted at Fig. 3b, the attack (a_5, a_1) is missing, and a_6 belongs to the completion. Since a_6 is there, the control attack (a_9, a_6) is present in the completion, and some attack between a_6 and a_4 must exist (in this case, the attack (a_6, a_4)).

Controllability means that we can select a subset $A_{conf} \subseteq A_C$ (called a control configuration) and the corresponding attacks $\{(a_i, a_j) \in \Rightarrow \mid a_i \in A_C,\ a_j \in (A_F \cup A_C \cup A_U)\}$ such that whatever the completion of $\mathcal{CAF}$, a given target is always reached. We focus on two kinds of targets: credulous acceptance of a set of arguments (this is reminiscent of extension enforcement [9]), and skeptical acceptance of a set of arguments. However in the context of negotiation in the current paper we use only credulous acceptance.

Definition 8 (Dimopoulos et al. [19]) A control configuration of $\mathcal{CAF} = \langle F, C, U \rangle$ is a subset $A_{conf} \subseteq A_C$. Given a set of arguments $T \subseteq A_F$ and a semantics σ, we say that T is skeptically (respectively credulously) reached by the configuration A_{conf} under σ if T is included in every (respectively at least one) σ-extension of every completion of $\mathcal{CAF}' = \langle F, C', U \rangle$, with $C' = \langle A_{conf}, \{(a_i, a_j) \in \Rightarrow \mid a_i \in A_C,\ a_j \in (A_F \cup A_C \cup A_U)\}\rangle$. We say that $\mathcal{CAF}$ is skeptically (respectively credulously) *controllable* with respect to T and σ.

In a nutshell, CAFs are a powerful enabler of advanced negotiation techniques, that blend together a number of desirable features such as the qualitative representation of uncertainty, simultaneous reasoning with different profiles through completions, simultaneous consideration of both certain and uncertain knowledge of the opponent, the use of control arguments (corresponding to a persuasion phase embedded in negotiation, allowing for the reinstatement of rejected arguments), along with a computational model based on QBFs (see Sect. 3.4 for details about the QBF encoding).

3 The negotiation framework

This section presents a new argumentation-based negotiation framework that relies on CAFs (Dimopoulos et al. [19]) for representing the incomplete information that agents have about their opponents. Agents communicate through the exchange of messages (or dialogue moves, see *e.g.* Dimopoulos and Moraitis [18]). We assume that agents play the roles of the proponent and opponent in a turn-taking round-based protocol (*e.g.* similar to the alternating offers protocol of Hadidi et al. [27]), where a proponent initiates a round and passes the token to its opponent when it is unable to defend an offer rejected by the opponent. The opponent may accept an offer when one of the supporting arguments is an acceptable argument for it, or reject an offer if it cannot accept any of the different supporting arguments sent by the proponent. We build on the works of Amgoud et al. [4], Hadidi et al. [27], and in the following, $\mathcal{L}$ denotes a logical language, and $\equiv$ an equivalence relation associated with it. From $\mathcal{L}$, a set $\mathcal{O} = \{o_1, \ldots, o_n\}$ of n *offers* is identified, such that $\nexists o_i, o_j \in \mathcal{O}$ such that $o_i \equiv o_j$. This means that the offers are different. Offers correspond to the different alternatives (*e.g.* prices for a product) that can be exchanged during the negotiation dialogue. We assume that agents share the same set of offers $\mathcal{O}$ but those offers can be supported by different arguments (although not necessarily) in the theories of the negotiating agents. By argument, we mean a *reason* in believing (called epistemic arguments) or doing something (called practical arguments). The set $Args(\mathcal{L})$, made of all the arguments built from $\mathcal{L}$, is then divided into two subsets: a subset $Args_p(\mathcal{L})$ of practical arguments supporting offers, and a subset $Args_e(\mathcal{L})$ of epistemic arguments supporting beliefs. Thus, $Args(\mathcal{L}) = Args_p(\mathcal{L}) \cup Args_e(\mathcal{L})$, with $Args_p(\mathcal{L}) \cap Args_e(\mathcal{L}) = \emptyset$. Now let us formally introduce the agent's negotiation theory. We start with the agent's personal theory in Definition 9, and then the agent's opponent modelling in Definition 10.

Definition 9 (*Agent's personal theory*) The *personal theory* of an agent α is $\mathcal{T}^\alpha = \langle A^\alpha, \rightarrow_\alpha \rangle$ with

- $A^\alpha \subseteq Args(\mathcal{L})$ a set of arguments such that $A^\alpha = A^\alpha_p \cup A^\alpha_e$ where A^α_p are practical arguments, and A^α_e are epistemic arguments;
- $\rightarrow_\alpha = \rightarrow^\alpha_p \cup \rightarrow^\alpha_e \cup \rightarrow^\alpha_m$, where
  - $\rightarrow^\alpha_p \subseteq A^\alpha_p \times A^\alpha_p$ are attacks between practical arguments,
  - $\rightarrow^\alpha_e \subseteq A^\alpha_e \times A^\alpha_e$ are attacks between epistemic arguments,
  - $\rightarrow^\alpha_m \subseteq A^\alpha_e \times A^\alpha_p$ are attacks from epistemic to practical arguments *i.e.* $(a, \delta) \in \rightarrow^\alpha_m$, if $a \in A^\alpha_e$ and $\delta \in A^\alpha_p$ (see Amgoud et al. [5], Hadidi et al. [27]).

Following Amgoud et al. [4] and Hadidi et al. [27] these personal theories are also enhanced with three preferences relations noted as $\succeq_e$, $\succeq_p$ and $\succeq_m$ for building along with the respective attack relations above, three defeat relations noted as $\rhd_e$, $\rhd_p$ and $\rhd_m$ respectively. So agents have also a *preference-based* theory $T^\alpha = \langle A^\alpha, \rhd^\alpha \rangle$ with $\rhd^\alpha = \rhd_e \cup \rhd_p \cup \rhd_m$.

Let us give an example of such an agents' personal theories.

Example 6 Agent α's personal theory $\mathcal{T}^\alpha$ is pictured in Fig. 4, while $\mathcal{T}^\beta$, the personal theory of agent β, is given in Fig. 4b. For agent α, we suppose that $A^\alpha_p = \{X\}$ and $A^\alpha_e = \{B, E, K\}$, thus the attack relations are $\rightarrow^\alpha_p = \emptyset$, $\rightarrow^\alpha_e = \{(E, K), (K, E)\}$ and $\rightarrow^\alpha_m = \{(B, X), (E, X)\}$.

Regarding β, we have $A^\beta_p = \{Y\}$ and $A^\beta_e = \{B, E, D, F\}$. The attack relations are as follows: $\rightarrow^\beta_p = \emptyset$, $\rightarrow^\beta_e = \{(D, B), (F, E)\}$ and $\rightarrow^\beta_m = \{(B, Y), (E, Y)\}$.

Definition 10 (*Agent's opponent modelling*) Let α and β be two negotiating agents, with (respectively) $\mathcal{T}^\alpha$ and $\mathcal{T}^\beta$ their personal theories (as introduced in Definition 9). We define the following sets of arguments for agent α:

- $A^{\alpha,\beta}_{F_e} \subseteq A^\beta_e$ (respectively $A^{\alpha,\beta}_{F_p} \subseteq A^\beta_p$) are epistemic (respectively practical) arguments such that agent α is certain about their existence in β's personal theory;

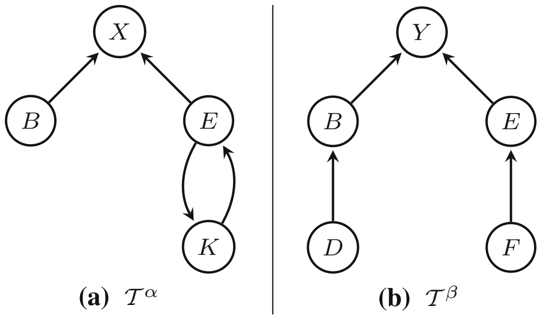


Fig. 4 The personal theories of agents α and β

- $A^{\alpha,\beta}_{U_e} \subseteq A^{\beta}_{e}$ (respectively $A^{\alpha,\beta}_{U_p} \subseteq A^{\beta}_{p}$) are epistemic (respectively practical) arguments such that agent α is not certain about their existence in β's personal theory;

Agent α's *opponent modelling* is a control argumentation framework $\mathcal{CAF}^{\alpha,\beta} = \langle F^{\alpha,\beta}, U^{\alpha,\beta}, C^{\alpha,\beta}\rangle$ where

- $F^{\alpha,\beta}$=$\langle A^{\alpha,\beta}_{F}, \rightarrow_{\alpha,\beta}\rangle$ with $A^{\alpha,\beta}_{F}$=$A^{\alpha,\beta}_{F_e} \cup A^{\alpha,\beta}_{F_p}$, $\rightarrow_{\alpha,\beta}$=$\rightarrow^{\alpha,\beta}_{e} \cup \rightarrow^{\alpha,\beta}_{p}$, and $\langle A^{\alpha,\beta}_{F_e}, \rightarrow^{\alpha,\beta}_{e}\rangle$ defining the epistemic arguments subpart such that $\rightarrow^{\alpha,\beta}_{e} \subseteq (A^{\alpha,\beta}_{F_e} \cup A^{\alpha,\beta}_{U_e}) \times (A^{\alpha,\beta}_{F_e} \cup A^{\alpha,\beta}_{U_e})$. The above also hold for the practical arguments subpart.
- $U^{\alpha,\beta}$=$\langle A^{\alpha,\beta}_{U}, \rightleftarrows_{\alpha,\beta} \cup \dashrightarrow_{\alpha,\beta})\rangle$ with $A^{\alpha,\beta}_{U}$=$A^{\alpha,\beta}_{U_e} \cup A^{\alpha,\beta}_{U_p}$, $\rightleftarrows_{\alpha,\beta}$=$\rightleftarrows_{e} \cup \rightleftarrows_{p}$, $\dashrightarrow_{\alpha,\beta}$=$\dashrightarrow_{e} \cup \dashrightarrow_{p}$, and $\langle A^{\alpha,\beta}_{U_e}, \rightleftarrows_{e} \cup \dashrightarrow_{e})\rangle$, $\rightleftarrows_{e} \subseteq ((A^{\alpha,\beta}_{U_e} \cup A^{\alpha,\beta}_{F_e}) \times (A^{\alpha,\beta}_{U_e} \cup A^{\alpha,\beta}_{F_e}))\setminus \rightarrow^{\alpha,\beta}_{e})$, $\dashrightarrow_{e} \subseteq ((A^{\alpha,\beta}_{U_e} \cup A^{\alpha,\beta}_{F_e}) \times (A^{\alpha,\beta}_{U_e} \cup A^{\alpha,\beta}_{F_e}))\setminus \rightarrow^{\alpha,\beta}_{e}$, defining the epistemic arguments subpart. The same hold for the practical arguments subpart. $\rightleftarrows_{e} \cap \dashrightarrow_{e} = \emptyset$.
- $C^{\alpha,\beta}$=$\langle A^{\alpha}_{c}, \Rrightarrow_{\alpha,\beta}\rangle$ where $A^{\alpha}_{c} \subseteq A^{\alpha}_{e}$ is the set of control arguments and $\Rrightarrow_{\alpha,\beta} \subseteq \{(a_i, a_j) \mid a_i \in A^{\alpha}_{c}$ and $a_j \in A^{\alpha}_{c} \cup A^{\alpha,\beta}_{F_e} \cup A^{\alpha,\beta}_{U_e}\} \setminus (\rightarrow^{\alpha,\beta}_{e} \cup \rightleftarrows_{e} \cup \dashrightarrow_{e}))$.

Of course, $\mathcal{CAF}^{\beta,\alpha}$ can be defined analogously for the opponent modelling of agent β. Notice that, except the control arguments that come from the agent's personal theory, all the other arguments (and attacks) come from the opponent's personal theory. Indeed we make the assumption that there is no "mistake" in the opponent's modelling, meaning that no argument or attack appears in $\mathcal{CAF}^{\alpha,\beta}$ if it has no counterpart in $\mathcal{T}^{\beta}$. But there may be some ignorance (*i.e.* some arguments or attacks from $\mathcal{T}^{\beta}$ that do not appear in $\mathcal{CAF}^{\alpha,\beta}$, neither in the fixed part nor in the uncertain part).

Example 7 We continue the previous example. Based on $\mathcal{T}^{\beta}$, we define $\mathcal{CAF}^{\alpha,\beta}$, *i.e.* the (partially uncertain) knowledge of agent α about agent β. We see on Fig. 5a that $A^{\alpha,\beta}_{F_e} = \{E\}$, $A^{\alpha,\beta}_{F_p} = \{Y\}$, and $A^{\alpha,\beta}_{U_e} = \{B\}$. The attack (B, Y) is fixed (so α is sure that this attack appears in $\mathcal{T}^{\beta}$), and on the contrary (E, Y) is uncertain. There is no control argument in $\mathcal{CAF}^{\alpha,\beta}$.

Now, $\mathcal{CAF}^{\beta,\alpha}$ (Fig. 5b) represents β's uncertain knowledge about α (as well as control arguments that can be used to persuade α about the acceptance of some argument). The fixed and uncertain parts are based on $\mathcal{T}^{\alpha}$: $A^{\beta,\alpha}_{F_e} = \{B, K\}$, $A^{\beta,\alpha}_{F_p} = \{X\}$, and $A^{\beta,\alpha}_{U_e} = \{E\}$. The attack relations in $\mathcal{CAF}^{\beta,\alpha}$ are $\rightarrow_{\beta,\alpha} = \{(E, X)\}$, $\dashrightarrow_{\beta,\alpha} = \{(B, X)\}$ and $\rightleftarrows_{\beta,\alpha} = \{(E, K), (K, E)\}$. Finally, the control arguments and attacks are coming from β's personal theory: $A^{\beta}_{c} = \{D, F\}$ and $\Rrightarrow_{\beta,\alpha} = \{(D, B), (F, E)\}$.

Now we can use the previous definitions to introduce a negotiating agent theory:

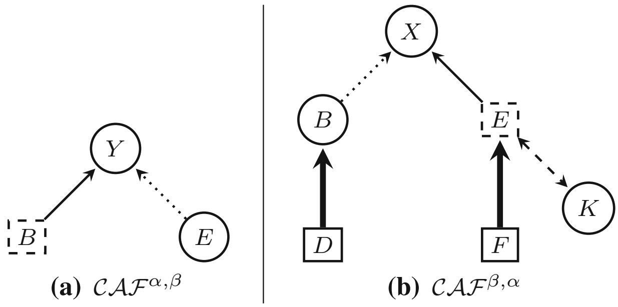


Fig. 5 The opponent modelling of agents α and β

Definition 11 (*Negotiating agent theory*) Let $\mathcal{O}$ be a set of n offers. A *negotiating theory* of an agent α is a tuple $\mathcal{T}^{\alpha}{}_{NT} = \langle \mathcal{O}, T^{\alpha}, \mathcal{T}^{\alpha}, \mathcal{CAF}^{\alpha,\beta}, \mathcal{F}^{\alpha} \rangle$ with $T^{\alpha} = \langle A^{\alpha}, \rhd^{\alpha} \rangle$, $\mathcal{T}^{\alpha} = \langle A^{\alpha}, \rightarrow_{\alpha} \rangle$ and $\mathcal{CAF}^{\alpha,\beta} = \langle F^{\alpha,\beta}, U^{\alpha,\beta}, C^{\alpha,\beta} \rangle$ as introduced in Definition 9 and Definition 10 respectively, and where $\mathcal{F}^{\alpha}$ is a function that returns the practical arguments supporting offers in $\mathcal{O}$. Formally:

– $\mathcal{F}^{\alpha}: \mathcal{O} \rightarrow 2^{A_p^{\alpha}}$ such that $\forall i, j$ with $i \neq j$, $\mathcal{F}^{\alpha}(o_i) \cap \mathcal{F}^{\alpha}(o_j) = \emptyset$.

Let us define ${A_p}^{\alpha}_{\mathcal{O}} = \bigcup_{i=1}^{n} \mathcal{F}^{\alpha}(o_i)$ the set of practical arguments in agent α's personal theory that support some offer in $\mathcal{O}$.

As in Amgoud et al. [5], Hadidi et al. [27] we assume that the same practical argument cannot be equally good for two different offers. In fact the idea is that a supporting argument must emphasizes a unique characteristic of an offer which distinguishes it from other offers and provides an added value with respect to the other offers and so it cannot be used for several offers. This also allows computing a ranking on the offers based on a ranking on the supporting arguments (see Sect. 3.2). However, this restriction can be dropped.

By adopting the preference-based argumentation (PAF) framework proposed in Amgoud et al. [5] for representing the negotiating agents theories (that has also already been used in Hadidi et al. [27]), we make also some assumptions that are very well suited to the negotiation dialogues context. More particularly based on the theoretical results proposed in Amgoud et al. [5] we assume that in the $AF_p = \langle A_p, \rhd_p \rangle$ subpart of the argumentation system of the agents: a) there are no odd cycles (i.e. that avoids indecision among several possible offers during the negotiation) b) the system is coherent i.e. each preferred extension is a stable one c) the system has at least one non empty preferred/stable extension (i.e. each agent has at least one offer to propose otherwise it is useless to enter into a negotiation dialogue with another agent) d) there are no self-attacking arguments (in our negotiation system this also concerns the $AF_e = \langle A_e, \rhd_e \rangle$ subpart of the argumentation system). We do believe that in a negotiation dialogue it is incoherent for an agent to use self-attacking arguments when negotiating with another agent.

In the following sub-sections we present the different steps that the agents apply in a negotiation dialogue when they behave as proponents or opponents and the procedures that implement the components of agents architectures as well as the new negotiation protocol we propose in this paper. More precisely the first step for an agent acting as proponent is the selection of the best offer by using its own theory (see Sect. 3.2). The next step is to choose the argument that supports its offer (probably the same as in its theory but not necessarily) in the (incomplete) knowledge of its opponent he disposes through the use of $CAFs$ (see

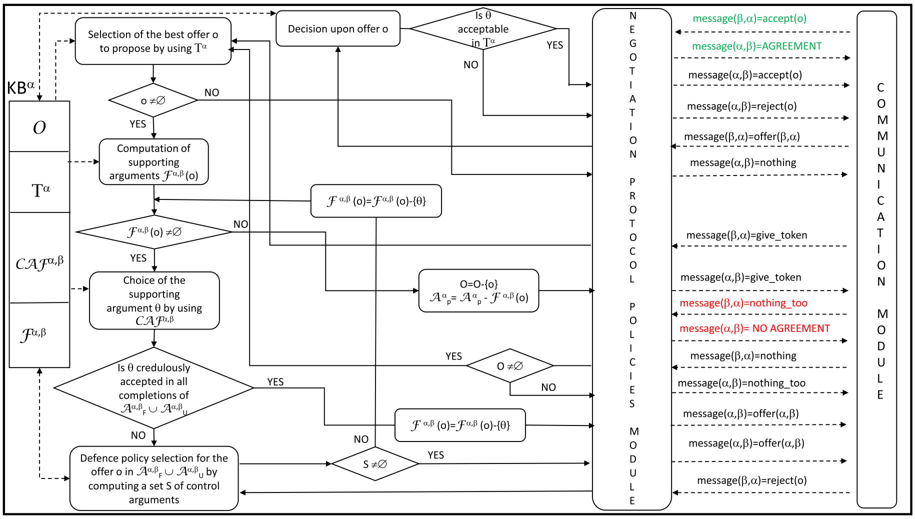


Fig. 6 Architecture of a negotiating agent α

Sect. 3.3). Then the next step is to verify whether this is an acceptable argument in the opponent's (incomplete) theory and to find a defending set of control arguments in case it is not, for allowing its reinstatement (see Sect. 3.4). These steps constitute the bidding strategy of an agent.

Then we present the behavior of an agent acting as opponent by presenting the so called acceptance strategy (see Sect. 3.5) and the policies of the negotiation protocol that implement the interaction rules (see Sect. 3.6) of agents in our negotiation framework. Finally we present an extended example for illustrating the whole negotiation process (see Sect. 3.7).

3.1 Negotiating agent architecture

Figure 6 presents the different components of a negotiating agent architecture that implement the different steps presented below. This architecture provides a global view of an agent internal operation acting either as proponent or opponent according to the different phases of a negotiation dialogue. Each of the described procedures as well as the negotiation protocol policies are presented in details in the following sub-sections. The communication module is responsible (as in any classical agent architecture) of sending the messages (as they are decided by the negotiation protocol policies module) to the opponent agent and of receiving and forwarding the corresponding messages sent by the opponent agent to the negotiation protocol policies module for further treatment. Green and red messages present the pairs of messages exchanged between negotiating agents in the cases of termination of the dialogue with and without agreement respectively.

3.2 Best offers selection

Algorithm 1 is the procedure invoked by the proponent agent α in order to compute, first, its best offer, based on its own theory, and it is implemented through function *comp_next_offer*. This function looks for the best offer supported by an acceptable practical argument, by using a ranking on the supporting arguments based on a partial preorder, that allows to

Algorithm 1: choose-best-offer(O, T^{α}, $\mathcal{CAF}^{\alpha,\beta}$, $\mathcal{F}^{\alpha,\beta}(o)$)

```
o ← comp_next_offer(O, T^α)
if o ≠ ∅ then
    F^{α,β}(o) ← compute_sup_arg(o, A^{α,β}_{F_p} ∪ A^{α,β}_{U_p})
    call choose-support-arg(o, F^{α,β}(o), CAF^{α,β})          // Algorithm 2
else
    message(α, β)=nothing
    send(message(α, β))
```

choose each time during a negotiation the best current argument and consequently the best offer to propose (supported by this argument). This preorder can be given (which is the case in the current implementation of our system) or computed by using different methods (*e.g.* a multi-criteria decision making approach [16], a ranking-based semantics approach [1], etc.). More precisely in our framework this ranking is based on a given partial preorder noted $\succeq_p$. This preference relation is used along with the attack relation $\rightarrow_p$ (see definition 9) in a defeat relation noted as $\rhd_p$ that computes the acceptability among practical arguments considering that a practical argument θ defeats another practical argument θ' noted $\theta \rhd_p \theta'$ if and only if θ and θ' are in conflict (they attack each other as they support only one offer at a time) *i.e.* $\{(\theta, \theta'), (\theta', \theta)\} \subseteq \rightarrow_p$, $\theta \succeq_p \theta'$ and $\theta' \not\succ_p \theta$ (where $\theta' \not\succ_p \theta$ means $\theta \succeq_p \theta'$ and $\theta' \not\succeq_p \theta$). The function *comp_next_offer* also uses a defeat relation noted as $\rhd_e$ (constructed in a similar way than $\rhd_p$) that allows to compute the acceptability among epistemic arguments and a defeat relation noted as $\rhd_m$ (called mixed and constructed as the two previous ones) that computes the acceptability of epistemic arguments with respect to practical arguments considering that an epistemic argument can defeat a practical argument but not vice-versa. For more details on this reasoning mechanism the reader can see [5,27].

As each practical argument supports only one offer at a time we consider that there exists a partial preorder noted as $\succeq_{off}$ which expresses a preference relation between offers and induces a ranking between them in O based on $\succeq_p$ and consequently on $\rhd_p$. We can therefore have the following definition:

Definition 12 Let o^i_j and o^k_l, $k \neq i$ two different offers where j and l are their respective ranks in O. If $\rhd_m$=∅, then $o^i_j \succeq_{off} o^k_l$ with $j < l$ if and only if $\exists \delta \in \mathcal{F}(o^i_j)$ s.t. $\forall \delta' \in \mathcal{F}(o^k_l)$, then $\delta \rhd_p \delta'$.

However other approaches can be also applied here for choosing each time the best argument and therefore the best offer. Thus different methods can be used for implementing the *comp_next_offer* function and the way the choice of the next offer is made does not affect the overall functioning of the negotiation system.

Then, based on its $\mathcal{CAF}^{\alpha,\beta}$, the proponent agent α computes the practical arguments that support this offer in its opponent theory through function *compute_sup_arg*$(o, A^{\alpha,\beta}_{F_p} \cup A^{\alpha,\beta}_{U_p})$ and calls a procedure, implemented by Algorithm 2, that selects the supporting argument to be sent. If the proponent agent has no (other) offer to propose, the opponent of the agent is informed by a suitable message (*i.e.* nothing).

3.3 Supporting argument selection

Algorithm 2 selects through function $choose - arg$, the argument that the proponent agent α sends to its opponent agent to support its offer from the set of the supporting arguments computed by the function *compute_sup_arg*$(o, A_{F_p}^{\alpha,\beta} \cup A_{U_p}^{\alpha,\beta})$. This choice can be random (as in the current implementation of our system) or by using different criteria [16] depending on the type of the application domain or the opponent's profile. Again, as noted previously, this choice does not affect the overall functioning of the negotiation system. Moreover, another procedure finds the arguments that defend this supporting argument whenever this argument is currently rejected by the opponent. This task is carried out by the procedure implemented by Algorithm 3. If there is no other available argument that supports the current offer, the agent abandons this offer and passes the negotiation token to the opponent agent.

Algorithm 2: choose-support-arg$(o, \mathcal{F}^{\alpha,\beta}(o), \mathcal{CAF}^{\alpha,\beta})$

```
if F^{α,β}(o) ≠ ∅ then
    θ ← choose-arg(F^{α,β}(o))
    call defend-offer(o, θ, F^{α,β}(o), CAF^{α,β})        // Algorithm 3
else
    O=O − {o}
    A_p^α = A_p^α − F^α(o)
    message(α, β)=give_token
    send(message(α, β));
```

3.4 The bidding strategy

The *bidding strategy* of the proponent agent is implemented by Algorithm 3. The main task here is to defend the proposed offer by an argument that (as said before) supports the offer in the opponent's theory. Consider for instance a car seller agent who proposes an expensive luxury SUV of a prestigious brand to a customer who, as the agent understands, seems to afford it. The reason (argument) that the seller agent has chosen this particular car is probably the high sales commission that it brings. However, this is not an argument it can use to convince its customer. The pool of appropriate arguments could include the smooth ride, fast acceleration, high top speed, off-road capabilities, safety features, or even the high social status associated with the brand. In fact, the discovery of those arguments takes place inside Algorithms 1 and 2. The role of the bidding strategy algorithm is to determine whether such a supporting argument is already acceptable in the opponent's theory, or to search for a *control configuration* that can defend the selected supporting argument under all possible opponent profiles.

More precisely, acceptance in the context of incomplete theories is based on the notion of *completion* which represents a possible profile (see Definition 7). The computation in line 1 of the algorithm relies on reasoning with Quantified Boolean Formulas (QBFs), as described in Dimopoulos et al. [19], that is carried out by the `quantom` solver [43]. The credulous controllability with respect to the theory $A_F^{\alpha,\beta} \cup A_U^{\alpha,\beta}$ (*i.e.* arguments in A_c^{α} are not considered in this case) is computed by using the following $Formula$ 1:

$$\forall\{on_{x_i} \mid x_i \in A_U^{\alpha,\beta}\}\forall\{att_{x_i,x_j} \mid (x_i, x_j) \in \dashrightarrow_{\alpha,\beta} \cup \rightleftarrows_{\alpha,\beta}\}$$
$$\exists\{acc_{x_i} \mid x_i \in \mathbf{A}\}[\phi_{st}^{cr}(\mathcal{CAF}, \theta)$$
$$\vee(\bigvee_{(x_i,x_j)\in \dashrightarrow_{\alpha,\beta}}(\neg att_{a_i,a_j} \wedge \neg att_{a_j,a_i}))]$$

Algorithm 3: defend-offer$(o, \theta, \mathcal{F}^{\alpha,\beta}(o), \mathcal{CAF}^{\alpha,\beta})$

1 **if** *θ is credulously accepted in all completions of the theory* $A_F^{\alpha,\beta} \cup A_U^{\alpha,\beta}$
2 **then**
3 offer$(\alpha, \beta) = \langle o, \theta, \langle \emptyset, \emptyset \rangle \rangle$
4 $\mathcal{F}^{\alpha,\beta}(o) = \mathcal{F}^{\alpha,\beta}(o) - \{\theta\}$
5 message(α, β)=offer(α, β)
6 send(message(α, β))
7 **else**
8 $S \leftarrow comp\_contr\_conf(\mathcal{CAF}^{\alpha,\beta}, \theta)$
9 **if** $S \neq \emptyset$ **then**
10 $\mathcal{R} = \{(a_i, a_j) | a_i \in S,\ a_j \in A_F^{\alpha,\beta} \cup A_U^{\alpha,\beta}\}$
11 offer$(\alpha, \beta) = \langle o, \theta, \langle S, \mathcal{R} \rangle \rangle$
12 message(α, β)=offer(α, β)
13 send(message(α, β))
14 **else**
15 $\mathcal{F}^{\alpha,\beta}(o) = \mathcal{F}^{\alpha,\beta}(o) - \{\theta\}$
16 call choose-support-arg$(o, \mathcal{F}^{\alpha,\beta}(o), \mathcal{CAF}^{\alpha,\beta})$ `// Algorithm 2`

where $\mathbf{A} = A_F^{\alpha,\beta} \cup A_{comp}$ with $A_{comp} \subseteq A_U^{\alpha,\beta}$.

The on_{x_i} variable means that the argument x_i currently belongs to the system; it is used for making the differentiation between the completions where x_i is included and those where it is not. Similarly, att_{x_i,x_j} is true when there is an attack from x_i to x_j. This variable has to be true if (x_i, x_j) is a fixed attack of $\mathcal{CAF}$. Otherwise, the truth value of this variable allows to distinguish between the completions where (x_i, x_j) is included and those where it is not. Finally acc_{x_i} is a propositional variable representing the acceptance status of the argument x_i. The propositional part $\phi_{st}^{cr}(\mathcal{CAF}, \theta)$ of the formula is satisfiable when θ belongs to at least one extension of a completion of $\mathcal{CAF}$ (more details about this part are given later). Straightforwardly, the prefix of the formula corresponds to an enumeration of every completion (by the $\forall$ quantifiers); for every such completion, we have to search for at least one extension (represented by the existentially quantified part) such that θ belongs to it.

Now, in case this computation succeeds, θ is acceptable in all possible opponent profiles (completions), and agent α sends to agent β the offer o, along with θ.

In case θ is not acceptable with respect to the above theory, agent α reacts as depicted in lines 7-13 of Algorithm 3. First, it uses its CAF to seek a *control configuration* S, that defends θ. This is again a problem on QBFs that is solved by a call to `quantom` solver (line 7 of the algorithm). However, this time, arguments in A_c^α are considered and credulous controllability is computed by using the following $Formula$ 2:

$$\exists\{on_{x_i} \mid x_i \in A_c^\alpha\}\forall\{on_{x_i} \mid x_i \in A_U^{\alpha,\beta}\}\forall\{att_{x_i,x_j} \mid (x_i, x_j) \in \dashrightarrow_{\alpha,\beta} \cup \rightleftarrows_{\alpha,\beta}\}\exists\{acc_{x_i} \mid x_i \in \mathbf{A}\}[\phi_{st}^{cr}(\mathcal{CAF}, \theta) \vee (\bigvee_{(x_i,x_j)\in \dashrightarrow_{\alpha,\beta}} (\neg att_{a_i,a_j} \wedge \neg att_{a_j,a_i}))]$$

where $\mathbf{A} = A_F^{\alpha,\beta} \cup A_c^\alpha \cup A_{comp}$ with $A_{comp} \subseteq A_U^{\alpha,\beta}$.

Note that this formula is very similar to the previous one. This time, the existential quantifier over the on_{x_i} variables, for $x_i \in A_c^\alpha$, corresponds to the search for one control configuration. So the whole formula corresponds to the definition of credulous controllability: the formula is true if there is a control configuration such that, for every completion, θ belongs to at least one extension.

In both above cases we use the formula $\phi_{st}^{cr}(\mathcal{CAF}, \theta) = \phi_{st}(\mathcal{CAF}) \wedge acc_\theta$, which is based on

$$\phi_{st}(\mathcal{CAF}) = \bigwedge_{x_i \in A_F^{\alpha,\beta}}[acc_{x_i} \Leftrightarrow \bigwedge_{x_j \in \mathbf{A}}(att_{x_j,x_i} \Rightarrow \neg acc_{x_j})] \wedge \bigwedge_{x_i \in A_C^{\alpha,\beta} \cup A_U^{\alpha,\beta}}[acc_{x_i} \Leftrightarrow (on_{x_i} \wedge \bigwedge_{x_j \in \mathbf{A}}(att_{x_j,x_i} \Rightarrow \neg acc_{x_j}))] \wedge \bigwedge_{(x_i,x_j) \in \rightarrow_{\alpha,\beta} \cup \Rrightarrow_{\alpha,\beta}} att_{x_i,x_j} \bigwedge_{(x_i,x_j) \in \rightleftarrows_{\alpha,\beta}} att_{x_i,x_j} \vee att_{x_j,x_i} \bigwedge_{(x_i,x_j) \notin \mathbf{R}} \neg att_{x_i,x_j}$$

where $\mathbf{R} = \rightarrow_{\alpha,\beta} \cup \Rrightarrow_{\alpha,\beta} \cup \dashrightarrow_{\alpha,\beta} \cup \rightleftarrows_{\alpha,\beta}$.

Moreover, in the first case, where the control arguments are not used (in *Formula* 1), $\bigwedge_{x_i \in A_C^{\alpha,\beta} \cup A_U^{\alpha,\beta}}$ becomes $\bigwedge_{x_i \in A_U^{\alpha,\beta}}$.

This formula is a generalization of the encoding of stable semantics defined in Besnard and Doutre [11], in the same line than the encoding given in Definition 5. When every *att*-variable and every *on*-variable is assigned a truth value, this assignment corresponds to a completion. Then, the consistent truth assignments of the *acc*-variables correspond to the set of stable extensions of the completion. This means that if $\phi_{st}(\mathcal{CAF}) \wedge acc_\theta$ is satisfiable, then θ belongs to at least one stable extension of the completion which is represented by the *att* and *on*-variables.

Now if in this second case the call succeeds, agent α sends offer o to agent β, along with the supporting argument θ, the set of arguments S, and the associated attacks R. Otherwise, the agent abandons this argument and picks another from $\mathcal{F}^{\alpha,\beta}(o)$ in order to continue defending o. This is done by function *choose-support-arg*. Recall that our approach looks for control configurations that work for all possible profiles of agent β (*i.e.* all possible completions of the CAF). However, if there is no such solution, the QBF optimization techniques of `quantom` [43] can find configurations that work for as many configuration as possible. This means that even if agent α cannot be sure that his argument will be accepted by β, he can still maximize its chances of success. Further study of this issue is out of the scope of this paper, and is kept for future work.

In the following, we define an operator $\oplus$ that is used in Algorithms 4 and 5.

Definition 13 Let A_1, A_2, A_3 be sets. We define $(A_1, A_2) \oplus A_3$ as the pair (A_1', A_2') such that $A_1' = A_1 \setminus (A_1 \cap A_3)$ and $A_2' = A_2 \cup (A_1 \cap A_3)$.

At the beginning of the negotiation each agent has in its theory (*i.e.* A^α and A^β respectively) only a part of the possible epistemic arguments (with respect to a specific application). That means that some arguments are in A^α and not in A^β (and vice-versa). However, when an agent will use arguments (and the associated attacks) that do not belong to the opponent's theory, the opponent agent will add them (as well as the associated attacks) in its own theory, and it will be able to use them from that point onward in the negotiation. This situation may take place in the Algorithms 4 and 5.

3.5 The acceptance strategy

This section discusses Algorithm 4, that implements the *acceptance strategy* of an agent. Upon receiving an offer and its supporting arguments (and the associated attacks) sent by a proponent agent, the algorithm updates the theory as well as the $\mathcal{CAF}$ of the receiving agent by integrating the supporting arguments, the defending arguments (*i.e.* the control configuration), and the associated attacks into both theories (*i.e.* the receiving agent own theory and its $\mathcal{CAF}$). Then, the receiver agent either accepts the offer (*i.e.* if the supporting arguments are acceptable) and informs the proponent accordingly, or sends to the proponent the reasons for rejecting its offer.

Algorithm 4: decide-upon-offer($\mathcal{T}^{\alpha}$, T^{α}, $\mathcal{CAF}^{\alpha,\beta}$, offer($\beta,\alpha$))

$\langle o, \theta, \langle S, \mathcal{R}\rangle\rangle$=offer$(\beta, \alpha)$
if $S \neq \emptyset$ **then**
    $\mathcal{T}^{\alpha}$=$(A^{\alpha} \cup S, \rightarrow_{\alpha} \cup \mathcal{R})$
    $(A_U^{\alpha,\beta}, A_F^{\alpha,\beta}) = (A_U^{\alpha,\beta}, A_F^{\alpha,\beta}) \oplus S$
    $(\dashrightarrow_{\alpha,\beta}, \rightarrow_{\alpha,\beta}) = (\dashrightarrow_{\alpha,\beta}, \rightarrow_{\alpha,\beta}) \oplus \mathcal{R}$
    $(\rightleftarrows_{\alpha,\beta}, \rightarrow_{\alpha,\beta}) = (\rightleftarrows_{\alpha,\beta}, \rightarrow_{\alpha,\beta}) \oplus \mathcal{R}$
if *θ is a credulous conclusion of theory T^{α}* **then**
    message(α, β)=Accept(o)
    send(message(α, β))
else
    Compute Q $\subseteq \mathcal{E}$ where $\mathcal{E}$ is an extension of T^{α} and Q is the set of arguments from which θ is reachable in the attack graph
    Reasons=$\{(p, \theta) | (p, \theta) \in \rightarrow_{\alpha}$ and $p \in Q\ \}$
    message(α, β)=Reject$(o, \theta, \langle Q, \mathit{Reasons}\rangle)$
    send(message(α, β));

3.6 The negotiation protocol

The Algorithm 5 described below implements the core procedure that drives the overall negotiation between the two negotiating agents through the necessary updates of their negotiation theories and calls to appropriate functions. This algorithm differentiates the behavior of the agents according to the role (*i.e.* proponent or opponent) they are playing during a negotiation round. The first part of algorithm (lines 1–2) implements the behavior of an agent when it is the proposer of the first offer, whereas the second part (lines 3–24) is concerned with its reaction when it receives an answer from another agent (*i.e.* the opponent). While the first part is straightforward as it concerns the selection of the best offer to propose, the second part is more involved and breaks down to several subcases. Those cases concern different situations that may arise during a negotiation, such as the rejection of an offer by the opponent, the acceptance of an offer (that terminates the negotiation with an agreement), the situation where the opponent informs that it has no other offer to propose, the situation where the opponent responds that it has no offer to propose too in a received similar message by the (proponent) agent (this ends the negotiation without agreement), the situation where an agent informs that it gives the token, and the situation where an offer is received and the receiver agent has to decide upon its acceptance or rejection. The example below explains how the protocol works.

3.7 A negotiation example

In the following we run an example of negotiation for illustrating our framework. We consider again the agents personal theories described in Example 6, as well as the associated opponent modelling (Example 7). Figure 7 recalls the agents α and β theories before the negotiation and their associated CAF respectively. Thus in the current example we have $A_p^{\alpha} = \{X\}$ and $A_e^{\alpha} = \{B, E, K\}$ for agent α and $A_p^{\beta} = \{Y\}$ and $A_e^{\beta} = \{B, E, D, F\}$ for agent β. The arguments $\{D, F\}$ are ignored by agent α. We have also the common set of offers $\mathcal{O}^{\alpha} = \mathcal{O}^{\beta} = \{o\}$. $\mathcal{F}^{\alpha}(o) = \{X\}$ and $\mathcal{F}^{\beta}(o) = \{Y\}$ represent the practical arguments supporting offer o in the agents α and β theories respectively. For their CAF we have $\mathcal{F}(o)^{\alpha,\beta} = \{Y\}$ and $\mathcal{F}(o)^{\beta,\alpha} = \{X\}$ respectively. Regarding the uncertainty, for $\mathcal{CAF}^{\alpha,\beta}$ we have $A^{\alpha,\beta}_{U_e}$=$\{B\}$, $\dashrightarrow_{\alpha,\beta}$=$\{(E, Y)\}$ and for $\mathcal{CAF}^{\beta,\alpha}$ we have $A^{\beta,\alpha}_{U_e}$=$\{E\}$, $\dashrightarrow_{\beta,\alpha}$=$\{(B, X)\}$, $\rightleftarrows_{\beta,\alpha}$=$\{(K, E), (E, K)\}$, $\Rrightarrow_{\beta,\alpha}$=$\{(F, E), (D, B)\}$ and control arguments $A^{\beta}_{c} = \{D, F\}$.

Algorithm 5: Procedure negotiate($\langle \mathcal{O}, \mathcal{T}^{\alpha}, T^{\alpha}, \mathcal{CAF}^{\alpha,\beta}, \mathcal{F}^{\alpha} \rangle$)

```
1  if agent α proposes first then
2      call choose-best-offer(O, T^α, CAF^{α,β}, F^{α,β}(o))          // Algorithm 1
3  while true do
4      get message(β, α)
5      switch message(β, α) do
6          case Reject(o, θ, ⟨Q, Reasons⟩)
7              (A_U^{α,β}, A_F^{α,β}) = (A_U^{α,β}, A_F^{α,β}) ⊕ Q
8              (⇢_{α,β}, →_{α,β}) = (⇢_{α,β}, →_{α,β}) ⊕ Reasons
9              (⇄_{α,β}, →_{α,β}) = (⇄_{α,β}, →_{α,β}) ⊕ Reasons
10             call defend-offer(o, θ, F^{α,β}(o), CAF^{α,β})        // Algorithm 3
11         case Accept(o)
12             End of negotiation with agreement on offer o
13         case nothing
14             if O ≠ ∅ then
15                 call choose-best-offer(O, T^α, CAF^{α,β}, F^{α,β}(o))  // Algorithm 1
16             else
17                 answer(α, β)=nothing_too
18                 send(answer(α, β))
19         case nothing_too
20             End of negotiation without agreement
21         case give_token
22             call choose-best-offer(O, T^α, CAF^{α,β}, F^{α,β}(o))    // Algorithm 1
23         case offer(β, α)=⟨o, θ, ⟨S, R⟩⟩
24             call decide-upon-offer(T^α, T^α, CAF^{α,β}, offer(β, α))  // Algorithm 4
25
```

The negotiation starts with agent α as proponent (see Fig. 7) by invoking Algorithm 5. Following line 2 there is a call of Algorithm 1. This algorithm computes the next (best) offer (line 1) to propose that is supported by an acceptable argument. In our example there is offer o but the supporting argument X is rejected as it is attacked by arguments B and E that belong into the two stable extensions namely $\{B, K\}$ and $\{B, E\}$. Agent α has no offer to

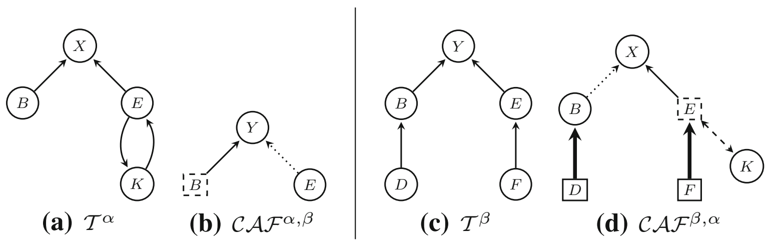


Fig. 7 The theories of agents α and β before the negotiation

propose to agent β and following line 6 it prepares a $message(\alpha, \beta) = nothing$ and sends it to agent β.

Agent β acts now as proponent (see Fig. 7). By using Algorithm 5 (line 13) it checks whether $\mathcal{O}^{\beta} \neq \emptyset$ (line 14) which is the case and calls Algorithm 1. This algorithm computes (as previously) the next (best) offer (line 1) that is supported by an acceptable argument. In the current situation we have the offer o which is now supported by the acceptable argument Y as it belongs to the (only) stable extension $\{Y, D, F\}$. Then (line 3), it computes the supporting practical arguments in the uncertain theory of agent α namely $\mathcal{F}(o)^{\beta,\alpha} = \{X\}$ by using its CAF. Then (line 4), there is a call to Algorithm 2. This algorithm selects a supporting argument (line 2). In our case there is only one, the argument X. Then there is a call (line 3) of Algorithm 3. This algorithm allows to check firstly (line 1) whether X is credulously accepted in the uncertain theory of agent α without the use of a control configuration (see $Formula$ 1).

Argument X is attacked by the uncertain argument E (*i.e.* see attack (E, X)). That means that there is a completion (or profile) where this argument is present in the theory. Moreover the type of uncertain attack between arguments K and E informs us that an attack is indeed present but the direction is unknown. That means that there are two completions (profiles) (among the three possible ones) where we have $\{(K, E), (E, K)\}$ and $\{(E, K)\}$ as possible attacks. In one of these completions argument E defends itself against the attack from K and in the other it attacks K. Therefore, in both cases E will be an acceptable argument and X will be rejected (as there is no defence against this attack).

Argument X is also attacked by argument B through the uncertain attack (B, X). That means that there is a completion (profile) where this attack is present in the theory and in that case X will also be rejected as B is an acceptable argument and there is no defence for X against the attack (B, X). Therefore, X cannot be accepted without the use of a control configuration.

By looking at the real theory of agent α, we may observe that the profile with the attacks $\{(K, E), (E, K)\}$ is the right one but agent β ignores this information. Then the algorithm tries to check whether it can find (see $Formula$ 2) a control configuration S (line 7). As we may observe such a set exists (see line 9) that can defend X no matter the real profile (*i.e.* for all the completions) of agent α. More precisely we have $S = \{D, F\}$ and $\mathcal{R}$=$\{(F, E), (D, B)\}$ and an offer(β, α)=$\langle o, X, \langle \{D, F\}, \{(F, E), (D, B)\}\rangle\rangle$ is built. Then, following line 10, a message(β, α)=offer(β, α) is prepared and sent to agent α. Agent α acts as receiver now. By using Algorithm 5 (see line 23) it calls Algorithm 4 (see line 24). By using Algorithm 4 agent α updates its theory and CAF (see lines 3–6), by using $S = \{D, F\}$ and $\mathcal{R}$=$\{(F, E), (D, B)\}$ (see Fig. 8). Then it checks whether it can accept X (see line 7). As shown in Fig. 8, the integration of agent's β control arguments $\{D, F\}$ (and the associated attacks) in agent's α theory, allows this agent to accept argument X as $\{X, D, F, K\}$ is a stable extension and therefore to accept offer o. Thus, following lines 8–9 it prepares a message(α, β)=accept(o) and sends it to agent β. Agent β acts as receiver by using Algorithm 5 (see line 11) and the negotiation ends successfully (line 12) with an agreement on offer o.

4 Theoretical results

In this section we present some interesting theoretical properties of our negotiation framework. We start with the definition of several notions.

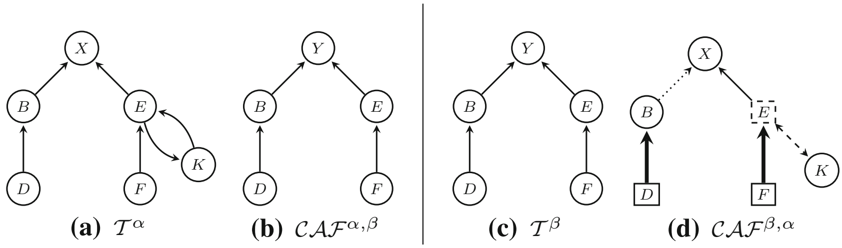


Fig. 8 The theories of agents α and β after the negotiation

Intuitively, an offer is acceptable if it is supported by a practical argument that is acceptable at some point during the negotiation dialogue. Formally:

Definition 14 Let a negotiation dialogue $\mathcal{N}$ between two agents α and β and an offer $o \in O$ the common set of offers. Then o is called an acceptable offer for agent α (respectively β) if $\exists \delta \in \mathcal{F}^\alpha(o)$ (respectively $\mathcal{F}^\beta(o)$) s.t. δ is a skeptical or credulous conclusion of T^α (respectively T^β) at some stage of $\mathcal{N}$.

An important concept in a context of negotiation dialogue is the agreement, *i.e.* the fact that the negotiating agents have found a solution that is acceptable to both of them.

Definition 15 Let a negotiation dialogue $\mathcal{N}$ between two agents α and β. We consider that agents α and β have reached an agreement on offer o if there is a message(α, β)=offer(α, β) sent from α to β with offer$(\alpha, \beta) = \langle o, \theta, \langle S, R\rangle\rangle$ where o is an acceptable offer for agent α and a message(β, α)=Accept(o) sent from β to α meaning that o is an acceptable offer for agent β too.

Finally, we introduce the notion of optimal solution, that corresponds to a situation where a specific offer is acceptable to both agents, and no better offer for one of them could have been accepted by the other, given their theories and preferences.

Definition 16 Let O^α (respectively O^β with $O^\alpha = O^\beta$) be a partially ordered set of n offers shared between two agents α and β and $o^i_j \in O^\alpha$ (respectively $o^i_{j'} \in O^\beta$) a specific offer where j (resp j') represents its current rank in O^α (respectively O^β). Offer o^i_j (respectively $o^i_{j'}$) is an optimal solution for agent α (respectively β) in a negotiation dialogue $\mathcal{N}$, if α (respectively β) has reached an agreement with β (respectively α) on o^i and there is no acceptable offer $o^k_l \in O^\alpha$ (respectively $o^k_{l'} \in O^\beta$), $k \neq i$ s.t. $l < j$ (respectively $l' < j'$) in the current state of its theory $\mathcal{T}^\alpha_{NT}$ (respectively $\mathcal{T}^\beta_{NT}$).

In the following we prove that our approach can guarantee the property of optimality, meaning that if the agents reach an agreement, then they have agreed on the optimal solution.

Proposition 1 *Let a negotiation dialogue $\mathcal{N}$ between two agents α and β. If α and β have reached an agreement on offer o then o is the optimal solution for both agents in the current state of their negotiating theories $\mathcal{T}^\alpha_{NT}$ and $\mathcal{T}^\beta_{NT}$ respectively.*

Proof Let's consider a dialogue $\mathcal{N}$ between two agents α and β where agent α is the proponent and β the opponent and an agreement reached on an offer o^i. Following Definition 15 and Algorithm 3 that means that agent α did send to agent β a message(α, β)=offer(α, β) with offer(α, β) $= \langle o^i, \theta, \langle \emptyset, \emptyset \rangle \rangle$ (line 5) or offer(α, β) $= \langle o^i, \theta, \langle S, R \rangle \rangle$ (line 10) where $\langle S, \mathcal{R} \rangle$ is a control configuration that defends argument θ in agent's β theory T^β (if necessary). Following Algorithm 1 and more particularly function *comp_next_offer* (line 1) and Definition 12, we know that o^i is an acceptable offer for which it exists a supporting practical argument say δ that defeats through $\rhd_p^\alpha$ all the other practical arguments supporting the other available offers in O^α and that are not defeated by epistemic arguments through $\rhd_m^\alpha$ in the current state of the theory T^α. We also know that the defeat relation $\rhd_p^\alpha$ is updated each time an offer o is rejected by agent β in a round of negotiation, as following Algorithm 2 (line 5) α removes this offer from the set O^α and along with all the supporting arguments $\mathcal{F}^\alpha(o)$ from its set of practical arguments A_p^α before giving the token to agent β (line 6). That means that the acceptable practical arguments that were supporting the offers proposed by α in previous rounds (and rejected by agent β till the current round) and that were defeating in a previous state of theory T^α the supporting practical argument δ of currently chosen offer o_j^i, don't belong anymore in A_p^α. As δ is an acceptable argument in T^α we also know that it does not exist an epistemic argument $\gamma \in A_e^\alpha$ s.t. $(\gamma, \delta) \in \rhd_m^\alpha$. Therefore if offer o_j^i is the acceptable offer that is chosen by agent α at the current round of the dialogue and j its rank in O^α, we know by Definition 12, that there is no acceptable offer $o_l^k \in O^\alpha$ s.t. $l < j$ because otherwise this should signify that o_l^k is supported by an acceptable practical argument say δ' s.t. $(\delta', \delta) \in \rhd_p^\alpha$ as δ is not defeated by an epistemic argument. However, we know that this is not possible because if it was the case, o_l^k would have been chosen by α (through function *comp_next_offer* of Algorithm 1) as the offer to be proposed in the current round. Thus offer o_j^i is an optimal solution for agent α. Let us now examine what is happening with agent β. Argument θ is a practical argument that supports o_j^i in $\mathcal{CAF}^{\alpha,\beta}$ and is credulously accepted in all completions of the theory $A_F^{\alpha,\beta} \cup A_U^{\alpha,\beta}$ as follows from Algorithm 3, either defended by a control configuration (line 1) or not (lines 7-10). As o^i is an agreement between agents α and β we know according Definition 15, that agent β did send to agent α a message(β, α)=Accept(o^i) as it follows from Algorithm 4 (line 8) and that means that argument θ is a credulous conclusion of its theory T^β (lines 7-9). Therefore θ is an acceptable argument for agent β. As agent β disposes a reasoning mechanism similar to the one of the proponent agent α, that means that argument θ defeats through $\rhd_p^\beta$ all the other practical arguments supporting other offers and that are not defeated by epistemic arguments through $\rhd_m^\beta$ in the current state of the theory T^β. Argument θ is also not defeated by an epistemic argument. So $o_{j'}^i$ is an acceptable offer for agent β and if j' is its rank in O^β at this round of the negotiation dialogue, this means that there is no acceptable offer $o_{l'}^k \in O^\beta$ s.t. $l' < j'$. Otherwise, as said previously, this should mean that there is an acceptable practical argument say θ' supporting $o_{l'}^k$ s.t. $(\theta', \theta) \in \rhd_p^\beta$, which we know that cannot hold. Offer o_l^i is therefore an optimal solution for agent β. Thus o^i is the optimal solution for both agents and this concludes the proof. □

In the following we prove that our negotiation method is complete, *i.e.* the agents will certainly reach an agreement if this is possible.

Proposition 2 *A negotiation dialogue $\mathcal{N}$ between two agents α and β is complete.*

Proof A complete negotiation dialogue between two agents α and β means that if there is a possible agreement on an offer $o \in O$, the negotiation framework guarantees that the agents will find this offer. According to Definition 15, that means that it will obligatory exist a round during the negotiation where agents α and β will exchange the messages offer(α, β) $= \langle o, \theta, \langle S, R\rangle\rangle$ (following Algorithm 1, line 5 or line 10) and message(β, α)=Accept(o) (following Algorithm 4, line 8) assuming that α is the proponent and β the opponent (or the other way around). Let us therefore consider a negotiation dialogue $\mathcal{N}$ between agents α and β, an offer o that can be an agreement between the two agents and assume that this agreement was not reached. That means that either agent α did not send a message offer(α, β) $= \langle o^i, \theta, \langle S, R\rangle\rangle$ or opponent agent β didn't reply with a message(β, α) =Accept(o). We will examine the two situations. Let us start by considering the first situation namely the assumption that agent α did not sent such a message to agent β. Different possible reasons could validate this assumption. Following Algorithm 1 a first possible reason is that there is no practical argument say δ supporting o which is skeptically or credulously accepted in theory T^α (*i.e.* o it is not an acceptable offer for α). Following Algorithm 3 a second possible reason is that there is no practical argument say θ (or δ itself) supporting o in $\mathcal{CAF}^{\alpha,\beta}$ that is credulously accepted (with or without the defence of a control configuration) in $A_F^{\alpha,\beta} \cup A_U^{\alpha,\beta}$. However as we know that o is a possible agreement we know from Proposition 1 that o is an optimal solution for both agents and therefore we know that there exists an acceptable practical argument (possibly not the same) in $T^\alpha \cup T^\beta$ that supports o. Therefore these two reasons cannot hold. A third possible reason is that agent α could not find the supporting argument θ (among possibly several ones in $\mathcal{F}^{\alpha,\beta}(o)$) that supports o in $\mathcal{CAF}^{\alpha,\beta}$ and is acceptable in $A_F^{\alpha,\beta} \cup A_U^{\alpha,\beta}$. However following Algorithm 2 we know that agent α is testing (lines 1-3) the acceptability of all practical arguments that are in $\mathcal{F}^{\alpha,\beta}(o)$ before abandoning this offer and giving the token to agent β (line 6). Thus if such an argument exists it cannot be ignored. Finally a fourth possible reason is that the offer o was ignored. However following Algorithm 1 this cannot happen as agent α is testing (lines 2-4) the acceptability of all the possible offers in O before sending the message(α, β)=nothing (line 6). Therefore the first situation we have considered cannot occur and we are sure that agent α did send a message offer(α, β) $= \langle o, \theta, \langle S, R\rangle\rangle$ to agent β. Let us now consider the second situation namely the assumption that opponent agent β didn't send the message(β, α) =Accept(o). In this situation, following Algorithm 4 (lines 1-9), that means that none of the practical arguments supporting offer o (including θ) could be accepted by agent β. However this cannot happen because o is an optimal solution for β (*i.e.* as we know it is a possible agreement) and consequently an acceptable offer for it. Therefore this situation cannot occur either and therefore we are sure that agent β did send a message(β, α)=Accept(o). The conclusions of the analysis of the two above situations lead to a contradiction with our initial assumption that the agreement on offer o was not reached as either agent α didn't send a message offer(α, β) $= \langle o, \theta, \langle S, R\rangle\rangle$ or agent β didn't reply with a message(β, α)=Accept(o) and this concludes the proof. □

An important property of a negotiation dialogue is termination. Our approach can guarantee termination under the assumption that there exists at least one acceptable offer in its personal theory when it plays the role of proponent.

Proposition 3 *A negotiation dialogue $\mathcal{N}$ between two agents α and β always terminates either with an agreement on some offer o or without agreement provided that there exists at least one acceptable offer to propose for each agent during the dialogue.*

Proof We assume that in real world negotiations the agents are endorsed by the designers with negotiation theories where there exists at least one acceptable offer for each agent according to the current state of the negotiation theories. That means that the theories T^α and T^β of two negotiating agents α and β respectively, have at least one σ-extension $\mathcal{E} \neq \emptyset$ (where σ is the acceptability semantics used) during the dialogue, when they are playing the role of the proponent. Let us therefore consider a negotiation dialogue $\mathcal{N}$ between two agents α and β. From Proposition 2 we know that a negotiation dialogue $\mathcal{N}$ is complete. So if there exists an offer o that can be an agreement between the two agents, this offer will be found and the dialogue will be ending following Algorithm 5 with an agreement on offer o (lines 11-12). Let's now consider that such an offer does not exist. Following Algorithms 1, 2 and 3 we know that a proponent agent α is sending to the opponent agent β (each time it has the token) any acceptable offer $o \in O^{\alpha}$ that is supported by an acceptable practical argument in $A_F^{\alpha,\beta} \cup A_U^{\alpha,\beta}$ with the goal that one of these offers is also an acceptable offer for agent β. However the set O^{α} (respectively O^{β}) is finite and the set of practical arguments $\mathcal{F}^{\alpha}(o)$ supporting the offer o as well. So when there is not anymore an available offer to propose, following Algorithm 1, a proponent agent α is sending a message(α,β)=nothing (line 6). When agent β is found as well at the same position when acting as proponent (*i.e.* the set of available offers O^{β} is empty), it responds following Algorithm 5 with a message(β, α)=nothing_too (lines 14-18). Thus following Algorithm 5 the dialogue terminates without agreement (lines 19-20). This concludes the proof that a dialogue $\mathcal{N}$ always terminates either with an agreement on some offer o or without agreement. □

Finally, we observe that, if an agreement is reached, then the solution of the negotiation dialogue is supported by an acceptable argument in both agents theories.

Proposition 4 *Let a negotiation dialogue $\mathcal{N}$ between two agents α and β, an agreement on offer o and $\theta \in \mathcal{F}(o)$ a supporting practical argument for o. Then θ is a skeptical or credulous conclusion of the theory T^{α} or T^{β}.*

Proof This proof follows directly from Algorithms 1, 4 and Definition 14. □

To conclude this section, let us briefly discuss the computational issues related to our negotiation protocol. As can be seen in the algorithms, the number of steps (negotiation rounds, message exchanges,. . .) is bounded by the number of arguments that appear in the agents theories. Thus, the hard part of the protocol are the verification whether an argument is accepted in an AF (because the agent's offer must be supported by an acceptable argument), and the search for a control configuration. The complexity of the reasoning depends on the chosen extension semantics, and is already well-known (see *e.g.* Dvorák and Dunne [23] for the complexity of reasoning with AFs, and Dimopoulos et al. [19], Niskanen et al. [37] for the complexity of CAFs). Although the complexity can be high (it ranges from polynomial time to completeness for the third level of the polynomial hierarchy), efficient computational techniques have been proposed (see for instance the results of the ICCMA competition [26,46] for the classical reasoning tasks with AFs, and the experiments by Niskanen et al. [37] for reasoning with CAFs).

5 Experimental evaluation

The proposed framework has been implemented in the JADE platform, and evaluated on negotiations with random argumentation theories. We first describe the generation of those theories, and then report on the experimental evaluation of negotiating with these theories.

5.1 Random theory generation

The experimental evaluation of the proposed framework is based on a system, implemented in Java, that generates pairs of random negotiation theories and associated CAFs, with different user specified characteristics.

Each negotiation experiment involves a pair of random theories $\mathcal{T}^\alpha = \langle A^\alpha, \rightarrow_\alpha \rangle$ and $\mathcal{T}^\beta = \langle A^\beta, \rightarrow_\beta \rangle$ that share a (non-empty) common part, *i.e.* there exists $N_{\alpha,\beta} = \langle A^{N_{\alpha,\beta}}, \rightarrow_{N_{\alpha,\beta}} \rangle$, such that $A^{N_{\alpha,\beta}} = A^\alpha \cap A^\beta$ and $(a, b) \in \rightarrow_{N_{\alpha,\beta}}$ if and only if $(a, b) \in \rightarrow_\alpha \cap \rightarrow_\beta$. Moreover, control arguments are only attacked by other control arguments, *i.e.* $((A^\alpha \setminus A^\alpha_c) \times A^\alpha_c) \cap \rightarrow_\alpha = \emptyset$.

The structure of the generated theories depends on a number of user supplied parameter values that are explained briefly below.

The user needs to define the number of epistemic, practical and control arguments of theories $\mathcal{T}^\alpha$ and $\mathcal{T}^\beta$, as well as their `density`, defined as the ratio of attacks present in the theory to the number of all possible attacks between the arguments of the theory. Moreover, the instance generation system receives as input the number of epistemic, practical and control arguments of the shared part $N_{\alpha,\beta}$.

From theory $\mathcal{T}^\beta$, $\mathcal{CAF}^{\alpha,\beta} = \langle \langle A^{\alpha,\beta}_F, \rightarrow_{\alpha,\beta} \rangle, \langle A^{\alpha,\beta}_U, \rightleftarrows_{\alpha,\beta} \cup \dashrightarrow_{\alpha,\beta} \rangle, \langle A^\alpha_c, \Rrightarrow \rangle \rangle$ is built (similarly for $\mathcal{T}^\alpha$ and $\mathcal{CAF}^{\beta,\alpha}$), which is the theory that agent α holds about agent β. $\mathcal{CAF}^{\alpha,\beta}$ satisfies the following conditions:

1. $A^{\alpha,\beta}_F \cup A^{\alpha,\beta}_U = A^\beta \cup A^\alpha_p$,
2. $A^\alpha_p \subseteq A^{\alpha,\beta}_F$.

The attack relation $\rightarrow_{\alpha,\beta} \cup \rightleftarrows_{\alpha,\beta} \cup \dashrightarrow_{\alpha,\beta}$ of $\mathcal{CAF}^{\alpha,\beta}$, is generated so that it satisfies the following conditions:

1. $\rightarrow_{\alpha,\beta} \subseteq \rightarrow_\beta$,
2. $\rightarrow_\beta \cap (A^{\alpha,\beta}_F \times A^{\alpha,\beta}_F) \subseteq \rightarrow_{\alpha,\beta}$,
3. $(\rightleftarrows_{\alpha,\beta} \cup \dashrightarrow_{\alpha,\beta}) \subseteq (\rightarrow_\beta \setminus \rightarrow_{\alpha,\beta})$,
4. $\rightleftarrows_{\alpha,\beta} \cap \dashrightarrow_{\alpha,\beta} = \emptyset$.

The main consequence of the above requirements is that the attack relation of $\mathcal{CAF}^{\alpha,\beta}$ is a subset of the attack relation of $\mathcal{T}^\beta$. The rationale for this restriction, in this initial experimental evaluation, is to focus on negotiation experiments where agents possess an "accurate" model of their opponent. One way to formalize the model accuracy is via the above relation between individual theories and CAFs. Moreover, it is interesting to study how the framework behaves when this restriction is removed. Indeed, the next section provides initial evidence that the method of this paper can cope with the relaxation of this restriction.

As with the individual agent theories $\mathcal{T}^\alpha$ and $\mathcal{T}^\beta$, the random instance generation software accepts as input a number of parameter values that determine various features of the CAFs of the agents. Most of them concern the uncertainty of an agent profile on its opponent, as captured by the corresponding CAF. The first is parameter `rateUncertArgs` that defines the ratio of uncertain arguments to all (fixed and uncertain) arguments of the theory. That is,

$$\texttt{rateUncertArgs} = \frac{|A^{\alpha,\beta}_U|}{|A^{\alpha,\beta}_F \cup A^{\alpha,\beta}_U|}$$

for agent α, and similarly for agent β.

Other parameters of the system include `rateUncertAtt`, that defines the ratio of uncertain attacks over all attacks, as well as `rateUndirAtt` that defines the ratio of uncertain

Table 1 Combinations of parameter values

Parameter	comb1	comb2	comb3	comb4
rateUncertArgs	0	0.10	0.25	0.50
rateUndirAtt	0	0.05	0.125	0.25
rateUncertAtt	0	0.05	0.125	0.25

attacks to all attacks. That is,

$$\texttt{rateUncertAtt} = \frac{|\dashrightarrow_{\alpha,\beta}|}{|\rightarrow_{\alpha,\beta} \cup \rightleftarrows_{\alpha,\beta} \cup \dashrightarrow_{\alpha,\beta}|},$$

and

$$\texttt{rateUndirAtt} = \frac{|\rightleftarrows_{\alpha,\beta}|}{|\rightarrow_{\alpha,\beta} \cup \rightleftarrows_{\alpha,\beta} \cup \dashrightarrow_{\alpha,\beta}|}.$$

Moreover, parameter `densContrAtt` defines the ratio of attacks from the control arguments of the agent to the arguments of its opponent that are included in its CAF to all possible such attacks from control arguments. For instance, `densContrAtt`=0.1 for $\mathcal{CAF}^{\alpha,\beta}$, means that 10% of all possible attacks from arguments of A_c^α to arguments in $A_F^{\alpha,\beta} \cup A_U^{\alpha,\beta}$ are included in the particular $CAF^{\alpha,\beta}$.

Finally, the instance generation system receives as input the number of offers, *i.e.* $|O^\alpha|$ and $|O^\beta|$, as well as the number of practical arguments that support each offer.

5.2 Experimental results

This section reports on selected results of the experimental evaluation of the framework. As the negotiation theory generation system accepts several parameter values, it is outside the scope of this work to provide exhaustive experimental results for all possible value combinations. Instead, we present results for selected runs that reveal important factors that influence the working of the negotiation algorithm, and highlight its merits and limitations. In all experiments we fix $|A^\alpha| = |A^\beta| = 40$, $|A_p^\alpha| = |A_p^\beta| = 6$, and $A_c^\alpha \cap A^{N_{\alpha,\beta}} = A_c^\beta \cap A^{N_{\alpha,\beta}} = \emptyset$.

The experimental evaluation is centered around 12 sets of agent theories, and associated CAFs, that differ in the uncertainty of these CAFs and the size of the shared part of agent theories. More specifically, four combinations of parameter values concerning the CAFs are considered, same for both agents. These combinations are listed in Table 1.

The combination `comb1` corresponds to the case where both agents have complete knowledge of their opponent. Then, uncertainty increases, with `comb4` the case where the agents have the highest uncertainty about their opponents among all the experiments.

Each of the above set of values for the three CAF parameters is combined with one of the three possible values {0.25, 0.50, 0.75} for the ratio $|A^{N_{\alpha,\beta}}|/|A^\alpha|$ that capture different degrees of similarity between agent theories.

The experimental results of this section focus on investigating the influence of control arguments on the *agreement rate* of the negotiations, *i.e.* the ratio of the number of negotiations terminated with agreement over their total number. Each experiment is composed of 600 negotiations consisting of 50 randomly generated experiments for each of the 12 parameter values combinations described above. Therefore, each experiment is an amalgamation of

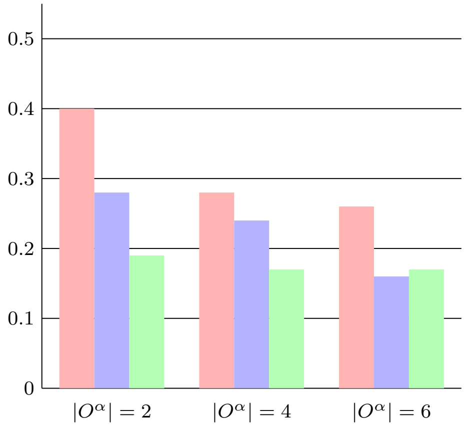


Fig. 9 Agreement rate (rounded to 10^{-2}) for negotiations with theories without control arguments, for $|O^\alpha| \in \{2, 4, 6\}$, and `density` is 0.10 (red), 0.15 (blue) or 0.20 (green)

negotiation theories of various types as far as the values of the 12 value parameters are concerned.

Negotiation without control arguments The first set of experiments serves as a "baseline" and concerns negotiations where agents possess no control arguments. The results are reported in Fig. 9 for various values for the number of offers (with $|O^\alpha| = |O^\beta|$) and the density of the theories.

In all these experiments the number of practical arguments remains constant (the actual value is $|A_p^\alpha| = |A_p^\beta| = 6$), therefore increasing the number of offers (while keeping their support uniform) decreases the number of supporting arguments per offer. More specifically, in the case of two offers, each offer is supported by two arguments, in the case of four offers, two of them are supported by two arguments and the other two by one argument, whereas in the case of six offers each of them is supported by one practical argument. The decrease in the number of supporting arguments for each offer that comes with the increase in the number of offers, leads to a decrease in the agreement rate. Moreover, the number of agreements seems to decrease as the density of the theory increases, a trend that is also present in negotiations with theories with control arguments.

Impact of the density of control attacks Figure 10 presents the first results concerning negotiations between agents with theories that contain control arguments, with various values for the ratio of attacks from these arguments, captured by parameter `densContrAtt`, as well the number of offers $|O^\alpha| = |O^\beta|$. In all experiments the density of the theories is 0.15. It seems that the presence of control arguments (and attacks) has an important influence on the number of successful negotiations: while the agreement rate was at most 0.40 without control arguments (Fig. 9), it is now between 0.55 and 0.82.

It also seems that the agreement rate diminishes if the number of supporting argument for offers is decreased, and increases with the attacks from the control arguments.

Impact of the theories density For all the experiments that follow, we fix the numbers of offers $|O^\alpha| = |O^\beta| = 4$. The results of Fig. 11 highlight the way the number of control arguments (`numContrArg`), and the ratio of control attacks (`densContrAtt`), affect the agreement rate. The left (respectively right) side of the Figure refer to the agreement rates achieved when the `density` of the individual theories of the agents participating in the negotiations is fixed to 0.15 (respectively 0.2). Clearly, the density of the theories plays a role in the agreement rate.

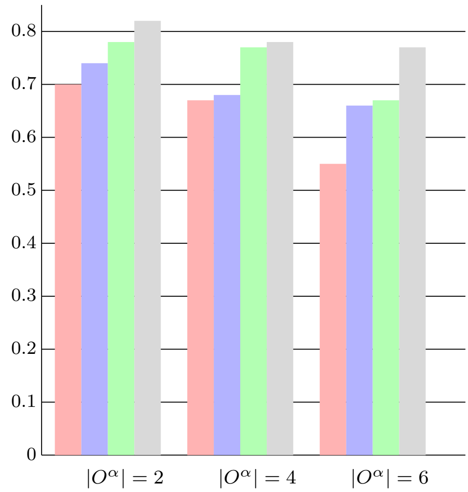


Fig. 10 Agreement rate (rounded to 10^{-2}) for $|O^{\alpha}| \in \{2, 4, 6\}$, and `densContrAtt` is 0.03 (red), 0.05 (blue), 0.10 (green) or 0.20 (gray), for theories with `density` = 0.15

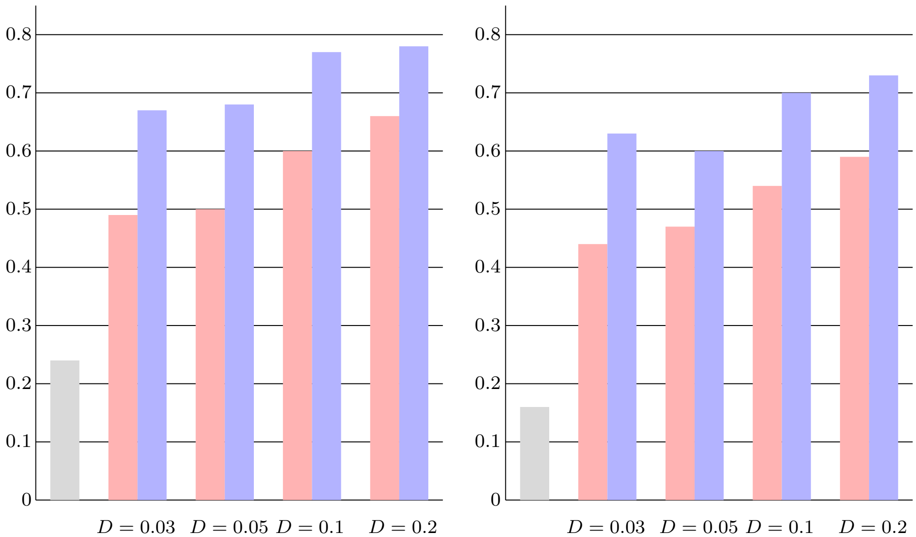


Fig. 11 Agreement rate (rounded to 10^{-2}) for negotiations with theories with `density` 0.15 (left) or 0.20 (right), `numContrArg` is 0 (gray), 3 (red) or 6 (blue), and `densContrAtt` is $D \in \{0.03, 0.05, 0.1, 0.2\}$

Observe that the data plotted in gray correspond to the case where the agents have no control arguments and corresponds to the results described in Fig. 9. Again, the important conclusion that can be readily drawn from Fig. 11 is that the presence of control arguments increases significantly the number of negotiations that terminate with agreement. Indeed, as also noted above, for theories with `density`= 0.15, the agreement rate almost doubles from 0.23 to 0.49 for cases where there are relatively few control arguments (`numContrArg`= 3) and attacks (`densContrAtt`= 0.03) from those arguments, and triples to 0.78 in the experiments with the highest number of control arguments and attacks. Similar are the results when the density of individual theories of the participating agents is set to 0.2 (`density`=0.2 on the right side of the Figure).

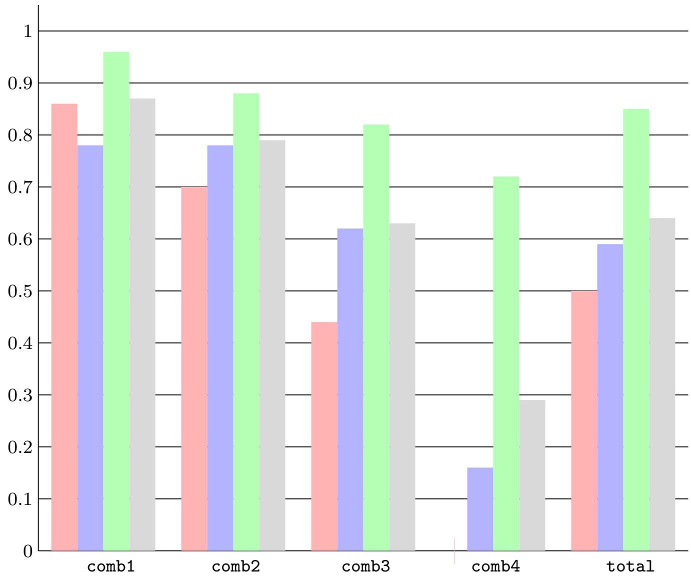


Fig. 12 Agreement rate (rounded to 10^{-2}) for the different combinations of parameters, for `density` = 0.15, 6 offers, and `densContrAtt` = 0.05, the size of the shared part of theories is 0.25 (red), 0.50 (blue), or 0.75 (green), and the aggregation of the three different sizes (gray)

Impact of the uncertainty and shared part of theories Figure 12 shows the agreement rate for specific value combinations for the parameters that refer to the CAF uncertainty (`comb1` to `comb4`) and the relative size of the shared part of the individual theories. The data plotted in gray refers to the aggregate agreement rates over the different shared theory sizes for a given combination of parameters (*i.e.* a degree of uncertainty in the CAFs). The last set of data (`total`) corresponds to the aggregate agreement rates of the different combinations.

The highest agreement rate, that reaches 96%, is achieved for negotiations with complete knowledge of the opponent theory (`comb1`) and individual theories with 75% similarity.

Moreover, the agreement rate, as presented by the data plotted in gray, decreases monotonically with the increase of the uncertainty on the opponent theory, starting with 88% for the combination of values `comb1`, and ending with 29% for `comb4`.

Figure 12 also highlights the effect of the size of the shared theories on the outcome of the negotiations, and reveals the positive effects of the high similarity between agent theories on the outcome of their negotiation. Indeed, in the set of data, labeled `total`, the agreement rate increases from 0.50 to 0.85 when the shared part goes from 25% to 75% of the theory size.

In order to understand better the effects of uncertainty on the outcome of negotiations, a series of experiments was run, with various values for the parameters that relate to the uncertainty of the theories. The results are reported in Fig. 13, that depicts agreement rates for sets of 100 negotiations.

We provide information about the degree of uncertainty of the random theories, as well as where this uncertainty is present the theory. More specifically, `unc` is the uncertainty rate, whereas `Arg` (plotted in red), `Att` (in blue), and `Both` (in green) indicate the part of the theory where this uncertainty appears. Data with `Arg` correspond to negotiations with theories where the uncertainty concerns the arguments (there is no uncertainty related to the attack relation), whereas data with `Att` refers to theories where the uncertainty concerns entirely the attack relation. Finally, data labelled with `Both`, concern theories where the uncertainty is equally divided between arguments and attacks. For instance, `unc=0.20,Both` refers to experiments with theories with 10% uncertainty related to the arguments, and 10% uncertainty related to the attack relation, equally distributed between relations $\rightleftarrows$ and $\dashrightarrow$. That is, 5% of the attacks belong to the set $\rightleftarrows$ and another 5% to $\dashrightarrow$. The size of the shared part of the theories is 25% (Fig. 13a), 50% (Fig. 13b), or 75% (Fig. 13c), and the aggregate is described at Fig. 13d.

The experimental results show that the uncertainty decreases the chances of an agreement in a negotiation. However, this effect is less severe for agents with theories that share a large common part, and theories where the uncertainty is in the arguments rather than the attack relation.

Impact of the unknown Recall that the negotiation experiments described so far are generated so that $A_F^{\alpha,\beta} \cup A_U^{\alpha,\beta} = A^\beta \cup A_p^\alpha$ *i.e.* agent α CAF about β contains all the arguments of its opponents. In the last set of experiments, whose results are described at Fig. 14, this assumption is removed by allowing agent β to possess arguments that are not part of the CAF of agent α. The number of these arguments is determined by the value of parameter `unknown` defined as

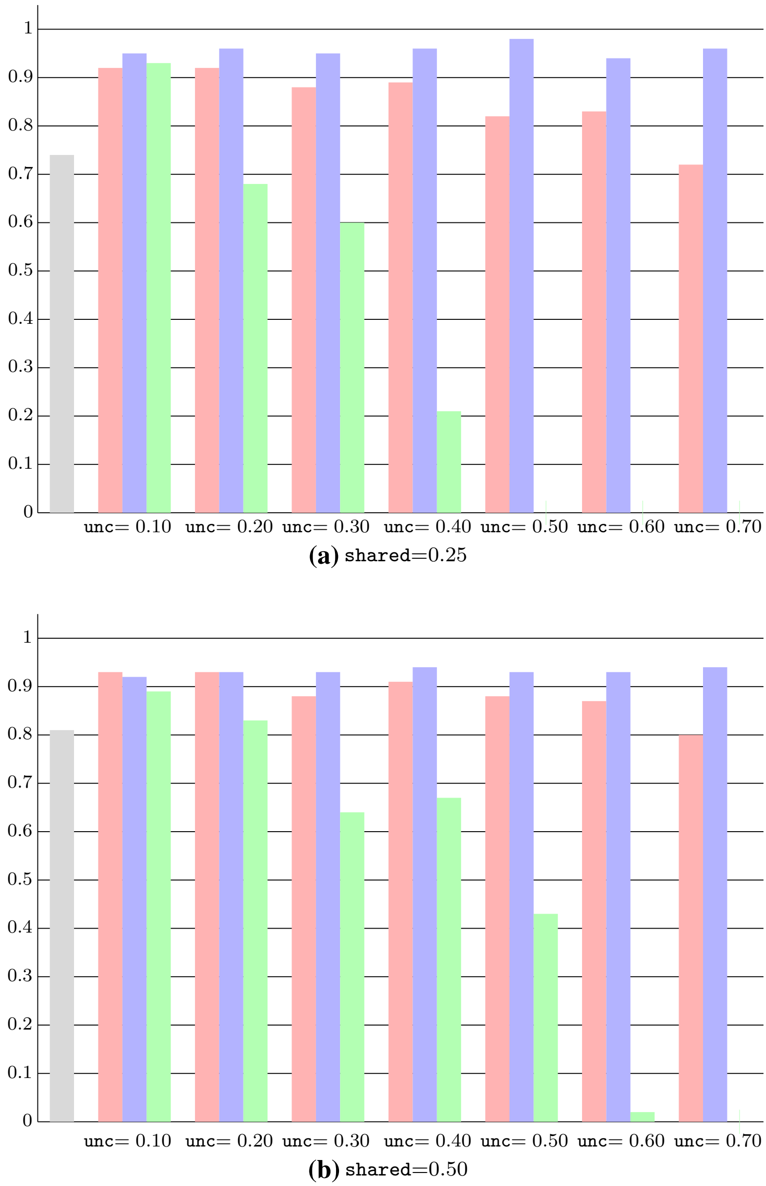
$$\texttt{unknown} = \frac{|(A^\beta - (A_F^{\alpha,\beta} \cup A_U^{\alpha,\beta})|)}{|(A_F^{\alpha,\beta} \cup A_U^{\alpha,\beta})|}$$

The experiments of Fig. 14 highlight the (little) effect of unknown arguments on the number of agreements achieved. Clearly, increasing the number of unknown arguments leads to a general reduction of the agreement rate. However, this decrease is less significant for theories with more control attacks, and globally even theories with a high number of such unknown arguments (`unknown`= 0.50, Fig. 14e) exhibit a satisfying agreement rate, between 0.42 and 0.78 (depending on the values of `numContrArg` and `densContrAtt`.

General conclusion of the experiments The experimental evaluation leads to a number of general conclusions. The first is that, not surprisingly, the effectiveness of the approach with respect to the rate of agreements depends on a number of parameters including the density of the individual theories, the number of attacks from control arguments, etc. Moreover, it seems that, in all cases, for “reasonably good” opponent profiles, the method leads to a significant increase in the number of negotiations that terminate with agreement. In a nutshell, the results clearly demonstrate that effective negotiation requires providing the “right offer” to the opponent, which in turn implies good knowledge of the opponent.

6 Related work

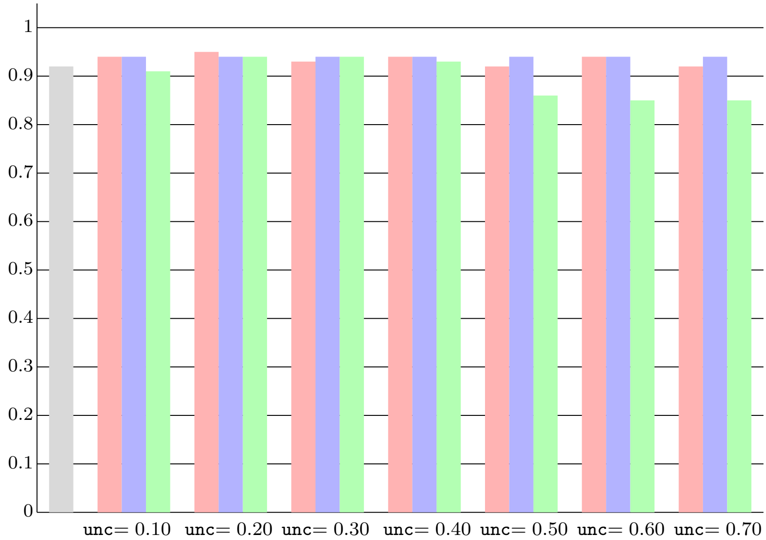
In this paper we presented an original argumentation-based negotiation framework that exploits a recent work proposed in Dimopoulos et al. [19] on control argumentation frameworks for modeling the uncertainty about the opponent profile and also the acceptance and bidding strategies of the negotiating agents. Compared to previous works proposed in the literature on argumentation-based negotiation, this new framework introduces and combines together a number of original ideas, with most notable a qualitative representation of uncertainty that enables simultaneous consideration of several different profiles, the bidding strategy that allows an agent to use arguments that do not belong to its theory, along with the notion of control arguments that facilitates persuasion and utilizes arguments that



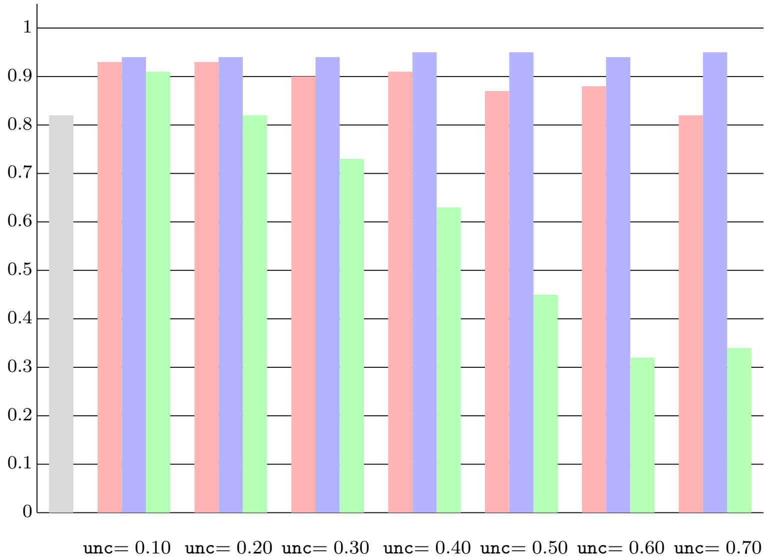
(a) shared=0.25

(b) shared=0.50

Fig. 13 Agreement rate (rounded to 10^{-2}) of negotiations for theories with density $=$ 0.15, densContrAtt = 0.05, and 4 offers, unc$\in$ {0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70} with parameters Arg in red, Att in blue and Both in green, and the aggregate in gray



(c) shared=0.75

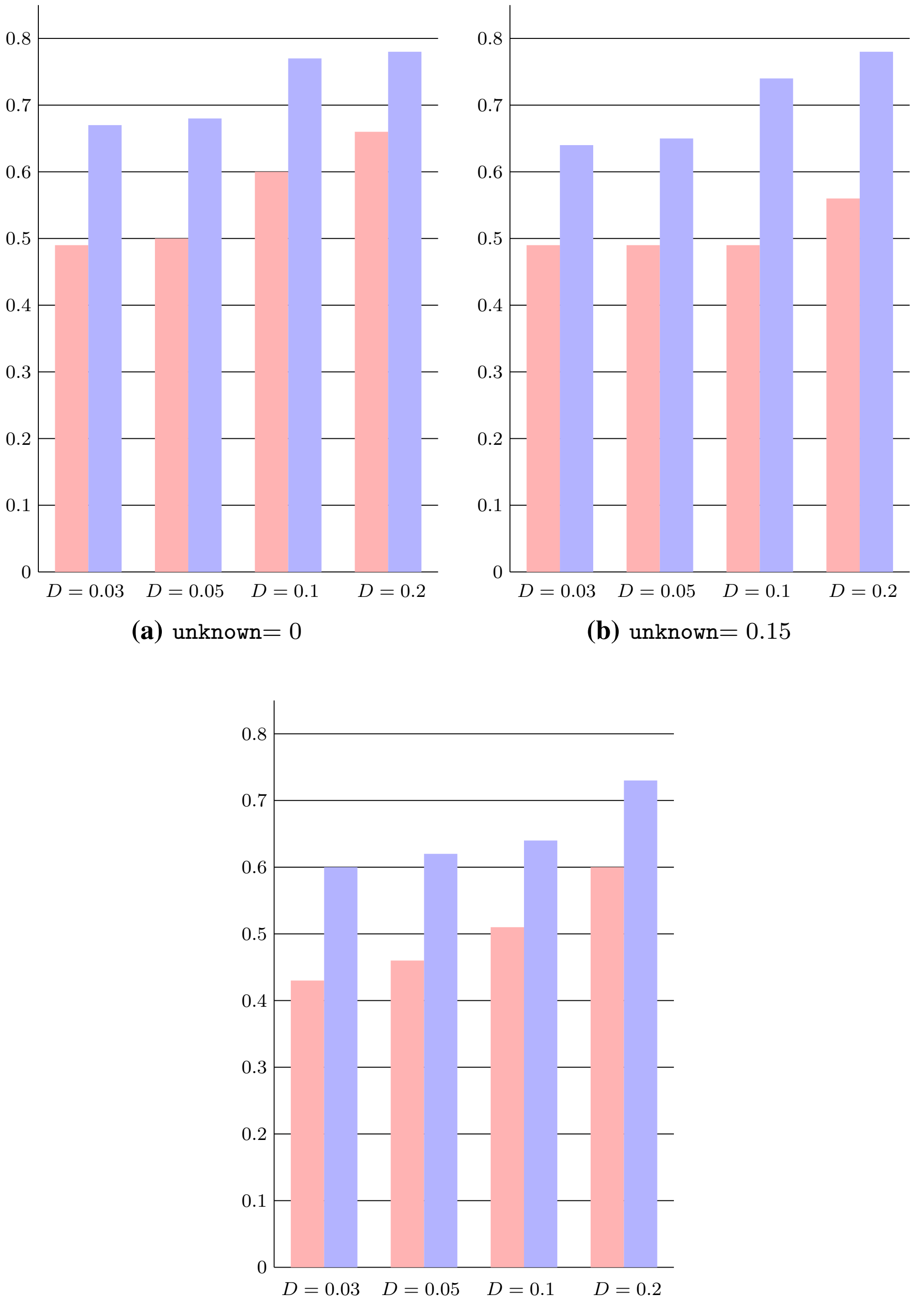


(d) total

Fig. 13 continued

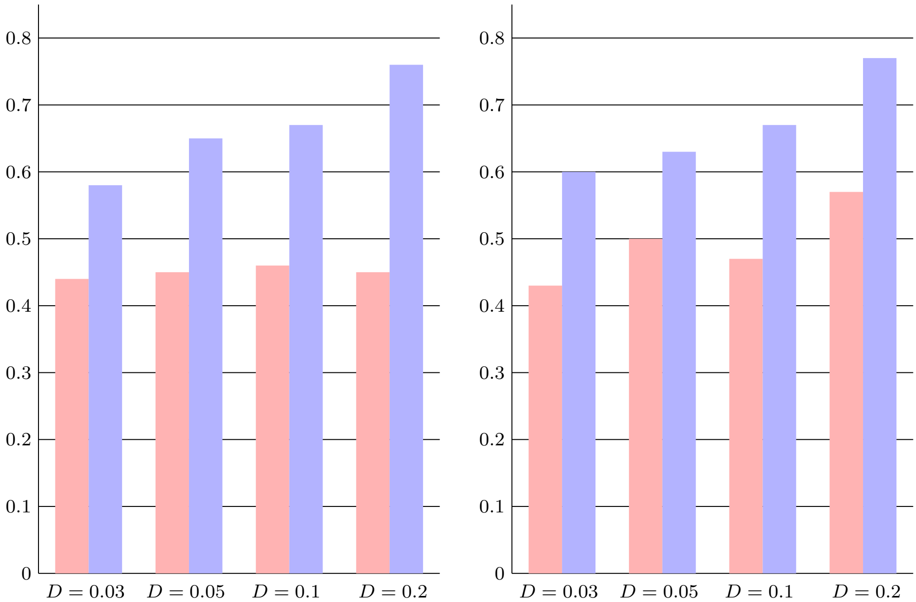
defend against all the possible attacks at once, hence minimizing the number of exchanged messages.

More specifically, the main difference with works such as Amgoud et al. [4], Amgoud and Kaci [3], Kakas and Moraitis [31], Dung et al. [22], Parsons et al. [40], Hadidi et al. [27], Marey et al. [35], Mancini [34] is that in these works the negotiation theories of the agents do not contain any kind of information on the opponents profiles. Therefore the bidding strategies



(a) unknown= 0 **(b)** unknown= 0.15 **(c)** unknown= 0.25

Fig. 14 Agreement rate (rounded to 10^{-2}) for negotiations with theories with density 0.15, parameter unknown is 0 (Fig. 14a), 0.15 (Fig. 14b), 0.25 (Fig. 14c), 0.35 (Fig. 14d) or 0.50 (Fig. 14e), numContrArg is 3 (red) or 6 (blue), and densContrAtt is $D \in \{0.03, 0.05, 0.1, 0.2\}$



(d) `unknown`$= 0.35$ **(e)** `unknown`$= 0.50$

Fig. 14 continued

(by using different policies) are based on the proponents agents own theories incrementally enhanced with the arguments sent by the opponents during the negotiation.

However, there are some works that integrate some information (in one way or another) on the opponent agent in the proponent's theory such as Monteserin and Amandi [36]. More specifically, in this work there are two types of information concerning the opponent namely the trust that denotes the level of trust (which takes three values *i.e.* low, medium and high) between sender and receiver, and the authority (which also takes three values *i.e.* subordinated, peer and superior) that indicates the relation of authority between sender and receiver. Both parameters are taken into consideration (amongst others) by the argument selection policy that looks for the most appropriate argument to utter during the negotiation. This policy is based on a reinforcement learning approach. Our work has many differences with this work. One important difference concerns for example the bidding strategy. In their case, the bidding strategy is based on the theory of the proponent but it tries to improve the argument selection effectiveness by updating the selection policy based on machine learning in order to adapt it to the different negotiation contexts as the agent gains experience. In our case the bidding strategy based on the CAFs first uses the proponent theory for finding the best offer to propose, and then it uses the uncertain opponent profile (represented as a CAF) either for finding whether there exists an acceptable argument supporting this offer, or for finding a control configuration that could reinstate a supporting argument which is rejected. Moreover, in our case the information on the opponent is evolving as the negotiation is progressing and concerns the whole profile of the opponent, and not only particular features such as trust and authority (in our case no assumption about the relationship between the negotiating agents is made).

Our work is also different from the work proposed in Pilotti et al. [41] where the negotiating agents have also incomplete information on the opponent. More specifically, in this work an agent has a belief sets about which resources are available for the opponent and which goals it believes that his opponent has. This knowledge evolves based on the information contained in the exchanged offers during the negotiation through classical belief revision. This is a main difference with our work where the beliefs (epistemic arguments) an agent has on his opponent may concern any kind of knowledge the opponent has on the world (*e.g.* it prefers European cars than Japanese cars, safety is an important issue, etc. when considering a car seller's beliefs on some customer) and not only its resources and goals. Moreover in our work an agent disposes also an incomplete set of practical arguments of an opponent's theory which gives some information on the goals (or options) of the opponent agent but also on the arguments that support those goals. This allows the proponent to find (and choose) the argument that supports its best offer in the opponent's theory and if it is rejected to send the appropriate arguments (through a control configuration) for enhancing the opponent's theory and defending this argument. This issue as we already explained is on the basis of our bidding strategy which also differs from the one in Pilotti et al. [41] where the next offer to propose is based on a function that takes into consideration the history of the negotiation and an utility function that evaluates the proposal to be sent. The two functions can be combined in different ways (*e.g.*, using a weighted sum) for representing different agent behaviors.

Finally, other works that consider information on the opponents profiles are those proposed in Bonzon et al. [15] and Hadidi et al. [28]. We consider that our work generalizes these previous works. More particularly in Bonzon et al. [15] the bidding strategy is similar to the one of the current paper. However, the opponent modeling approach based on CAF used in the current paper is more efficient than in the above paper as far as the way this (partial) knowledge is used for building efficient negotiation strategies. This is due to the fact that the CAF based representation takes into consideration the uncertainty on the opponents profiles in a way (*i.e.* by using different kind of attacks and arguments (on/off)) that allows to generate simultaneously several possible profiles through the completions and also gives the possibility to search for an offer that could satisfy all those profiles through the QBF based reasoning mechanism.

The bidding strategy proposed in Hadidi et al. [28] is different as it focuses on profile, behavior and time constraints based tactics that can be combined together to implement complex strategies similar to those studied in game-theoretic negotiation. The two works (*i.e.* the current and the above) are using the same formalism proposed in Hadidi et al. [27]) for representing the agents behaviors (both proponents and opponents). More particularly in Hadidi et al. [28] the profile associates each opponent with a defeat relation, that provides information about the negotiation behavior of this agent. This information may be incomplete and given before the start of the negotiation, or acquired during the negotiation as it is also done in our paper through the exchange of arguments and counter-arguments. However, this opponent profile modeling, corresponds only in the certain part representation of a CAF and thus it misses all the information that can be assumed on the opponent, through the uncertain part representation of a CAF and the possible different profiles that can be generated via the completions. We do believe that the incorporation of the CAF based representation and the associated reasoning mechanism in Hadidi et al. [28] could improve the performances of the proposed tactics and concessions strategies.

7 Conclusion

In this paper, we have presented an original framework for argumentation-based negotiation using an incomplete representation of the opponent profiles. The originality of our work lies on several aspects such as the representation method we have adopted for the opponents modeling based on the control argumentation frameworks (CAFs) that allows us to generate several possible profiles through the notion of completion, and the associated QBF based reasoning mechanism that allows us to search for solutions satisfying all those profiles, the bidding and the acceptance strategies based on the advantages that the representation method and the reasoning mechanism provide us.

We have presented some interesting theoretical results that show that our approach can guarantee some very good properties such as optimality of the reached solutions and completeness and termination of the negotiation dialogue. We have also run a very important number of experiments (more than 25,000 negotiations) which have proved the added value of our approach. More particularly, our experimental results have shown that the outcome of an argumentation-based negotiation dialogue depends on different parameters of the argumentation theories of the agents, but in all cases the use of control arguments seems to have a positive impact on the number of agreements.

Our future work concerns different issues. First, while the general negotiation approach proposed in this paper is generic with respect to the underlying extension-based semantics, our computational method relies on the stable semantics. We plan to study the computation of control configurations for other semantics. SAT-based counter-example guided abstraction refinement (CEGAR) seems to be a promising line of research to do that, since it has already proven to be efficient for extension enforcement [48] or for reasoning with Incomplete AFs [38], two problems that are closely related to CAFs. CEGAR, as well as QBF encodings, have been successfully used for reasoning with CAFs [37], thus it is a promising approach to adapt our negotiation protocol to other semantics.

Another issue is the automated generation of negotiation theories from "controlled" natural language. This will allow users to build in a easy and rapid way negotiation theories. A second issue is the experimentation of different approaches for implementing the function "supporting argument selection" and investigating whether there is a real influence on the efficiency of the overall process and the negotiation outcome (*i.e.* number of agreements), while a third issue will be the use of reinforcement learning techniques for improving the process of the "real" opponent profile recognition through the exchange of arguments and control configurations. We also plan to apply our negotiation platform to real world applications in different domains (*e.g.* trading, risk management, etc.). Finally we do believe that our approach for opponent modeling based on $CAFs$ can also be used in other types of dialogues such as persuasion and deliberation and this also constitutes part of our future work.

Acknowledgements The authors would like to thank their students Toufik Ider and Mickael Lafages for their excellent work in the implementation of the proposed framework. The authors would like also to thank the reviewers for their very constructive comments that allowed to improve significantly the quality of the paper.

References

1. Amgoud, L., & Ben-Naim, J. (2013). Ranking-based semantics for argumentation frameworks. In *Scalable uncertainty management - 7th international conference, SUM 2013, Washington, DC, USA, September 16–18, 2013. Proceedings* (pp. 134–147)

2. Amgoud, L., & Cayrol, C. (2002). A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence, 34*(1–3), 197–215.
3. Amgoud, L., & Kaci, S. (2006). On the study of negotiation strategies. In *Agent Communication II, International Workshops on Agent Communication, AC 2005 and AC 2006, Utrecht, Netherlands, July 25, 2005 and Hakodate, Japan, May 9, 2006*, Selected and Revised Papers (pp. 150–163).
4. Amgoud, L., Dimopoulos, Y., & Moraitis, P. (2007). A unified and general framework for argumentation-based negotiation. In *Proceedings of the 6th international joint conference on autonomous agents and multiagent systems, (AAMAS07)*, p 158, https://doi.org/10.1145/1329125.1329317.
5. Amgoud, L., Dimopoulos, Y., & Moraitis, P. (2008). Making decisions through preference-based argumentation. In *Principles of knowledge representation and reasoning: proceedings of the eleventh international conference, KR 2008, Sydney, Australia, September 16–19, 2008* (pp. 113–123).
6. Baarslag, T., Hendrikx, M. J. C., Hindriks, K. V., & Jonker, C. M. (2016a). Learning about the opponent in automated bilateral negotiation: A comprehensive survey of opponent modeling techniques. *Autonomous Agents and Multi-Agent Systems, 30*(5), 849–898. https://doi.org/10.1007/s10458-015-9309-1.
7. Baarslag, T., Hendrikx, M. J. C., Hindriks, K. V., & Jonker, C. M. (2016b). A survey of opponent modeling techniques in automated negotiation. In Jonker, C. M., Marsella, S., Thangarajah, J. & Tuyls, K. (eds) *Proceedings of the 2016 international conference on autonomous agents & multiagent systems, Singapore, May 9–13, 2016* (pp. 575–576). ACM.
8. Baroni, P., Caminada, M., & Giacomin, M. (2018). Abstract argumentation frameworks and their semantics. In D. Gabbay, M. Giacomin, & L. van der Torre (Eds.), *Baroni P* (pp. 159–236). Handbook of Formal Argumentation: College Publications.
9. Baumann, R., & Brewka, G. (2010). Expanding argumentation frameworks: Enforcing and monotonicity results. In *Computational models of argument: proceedings of COMMA 2010, Desenzano del Garda, Italy, September 8–10, 2010* (pp. 75–86).
10. Baumeister, D., Neugebauer, D., Rothe, J., & Schadrack, H. (2018). Verification in incomplete argumentation frameworks. *Artificial Intelligence, 264,* 1–26.
11. Besnard, P., & Doutre, S. (2004). Checking the acceptability of a set of arguments. In *10th international workshop on non-monotonic reasoning (NMR 2004), Whistler, Canada, June 6–8, 2004, Proceedings* (pp. 59–64).
12. Besnard, P., & Hunter, A. (2008). *Elements of argumentation*. Cambridge: MIT Press.
13. Biere, A., Heule, M., van Maaren, H., & Walsh, T. (eds) (2009). *Handbook of Satisfiability, Frontiers in Artificial Intelligence and Applications*, vol 185, IOS Press.
14. Black, E., & Atkinson, K. (2011). Choosing persuasive arguments for action. In *10th international conference on autonomous agents and multiagent systems (AAMAS 2011), Taipei, Taiwan, May 2–6, 2011* (Vol. 1–3, pp. 905–912).
15. Bonzon, E., Dimopoulos, Y., & Moraitis, P. (2012). Knowing each other in argumentation-based negotiation. In *International conference on autonomous agents and multiagent systems, AAMAS 2012, Valencia, Spain, June 4–8, 2012* (3 Volumes) (pp. 1413–1414).
16. Bouyssou, D., Marchant, T., Pirlot, M., Tsoukiàs, A., & Vincke, P. (2006). Evaluation and decision models with multiple criteria: Stepping stones for the analyst, 1st edn. Springer, Boston, URL http://www.springer.com/sgw/cda/frontpage/0118554-40521-22-116132747-000.html.
17. Coste-Marquis, S., Konieczny, S., Mailly, J. G., & Marquis, P. (2014). A translation-based approach for revision of argumentation frameworks. In *Logics in artificial intelligence—14th European conference, JELIA 2014, Funchal, Madeira, Portugal, September 24–26, 2014. Proceedings* (pp. 397–411).
18. Dimopoulos, Y., & Moraitis, P. (2014). Advances in argumentation-based negotiation. *Negotiation and Argumentation in Multi-Agent Systems: Fundamentals* (pp. 82–125). Systems and Applications: Theories.
19. Dimopoulos, Y,, Mailly, J. G., & Moraitis, P. (2018). Control argumentation frameworks. In *Proceedings of the thirty-second AAAI conference on artificial intelligence AAAI 2018, New Orleans, USA, February 2–7, 2018* (pp. 4678–4685).
20. Dimopoulos, Y., Mailly, J. G., & Moraitis, P. (2019). Argumentation-based negotiation with incomplete opponent profiles. In *Proceedings of the 18th international conference on autonomous agents and multiagent systems, AAMAS '19, Montreal, QC, Canada, May 13–17, 2019* (pp. 1252–1260).
21. 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.
22. Dung, P. M., Thang, P. M., & Toni, F. (2008). Towards argumentation-based contract negotiation. In *Computational models of argument: proceedings of COMMA 2008, Toulouse, France, May 28–30, 2008* (pp. 134–146).
23. Dvorák, W., & Dunne, P. E. (2018). Computational problems in formal argumentation and their complexity. In D. Gabbay, M. Giacomin, & L. van der Torre (Eds.), *Baroni P* (pp. 631–688). Handbook of Formal Argumentation: College Publications.

24. Dvorák, W., Järvisalo, M., Wallner, J. P., & Woltran, S. (2014). Complexity-sensitive decision procedures for abstract argumentation. *Artificial Intelligence, 206,* 53–78.
25. Egly, U., & Woltran, S. (2006). Reasoning in argumentation frameworks using quantified boolean formulas. In *Computational models of argument: Proceedings of COMMA 2006, September 11–12, 2006, Liverpool, UK* (pp. 133–144).
26. Gaggl, S.A., Linsbichler, T., Maratea, M., & Woltran, S. (2020). Design and results of the second international competition on computational models of argumentation. *Artificial Intelligence* 279.
27. Hadidi, N., Dimopoulos, Y., & Moraitis, P. (2010). Argumentative alternating offers. In *9th international conference on autonomous agents and multiagent systems (AAMAS 2010), Toronto, Canada, May 10–14, 2010* (Vol. 1–3, pp. 441–448).
28. Hadidi, N., Dimopoulos, Y., & Moraitis, P. (2012). Tactics and concessions for argumentation-based negotiation. In *Computational models of argument—proceedings of COMMA 2012, Vienna, Austria, September 10–12, 2012* (pp. 285–296).
29. Hadjinikolis, C., Siantos, Y., Modgil, S., Black, E., & McBurney, P. (2013). Opponent modelling in persuasion dialogues. In *IJCAI 2013, proceedings of the 23rd international joint conference on artificial intelligence, Beijing, China, August 3–9, 2013* (pp. 164–170).
30. Hunter, A. (2015). Modelling the persuadee in asymmetric argumentation dialogues for persuasion. In *Proceedings of the twenty-fourth international joint conference on artificial intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25–31, 2015* (pp. 3055–3061).
31. Kakas, A. C., & Moraitis, P. (2006). Adaptive agent negotiation via argumentation. In *5th international joint conference on autonomous agents and multiagent systems (AAMAS 2006), Hakodate, Japan, May 8–12, 2006* (pp. 384–391).
32. Kleine Büning, H., & Bubeck, U. (2009). Theory of quantified boolean formulas. In *Handbook of satisfiability* (pp. 735–760). IOS Press.
33. Lagniez, J.M., Lonca, E., & Mailly, J.G. (2015). CoQuiAAS: A constraint-based quick abstract argumentation solver. In *27th IEEE international conference on tools with artificial intelligence, ICTAI 2015, Vietri sul Mare, Italy, November 9–11, 2015* (pp. 928–935).
34. Mancini, T. (2016). Now or never: Negotiating efficiently with unknown or untrusted counterparts. *Fundamenta Informaticae, 149*(1–2), 61–100. https://doi.org/10.3233/FI-2016-1443.
35. Marey, O., Bentahar, J., Asl, E. K., Mbarki, M., & Dssouli, R. (2014). Agents' uncertainty in argumentation-based negotiation: Classification and implementation. In *Proceedings of the 5th international conference on ambient systems, networks and technologies (ANT 2014)* (pp. 61–68).
36. Monteserin, A., & Amandi, A. (2013). A reinforcement learning approach to improve the argument selection effectiveness in argumentation-based negotiation. *Expert Systems with Applications, 40*(6), 2182–2188. https://doi.org/10.1016/j.eswa.2012.10.045.
37. Niskanen, A., Neugebauer, D., & Järvisalo, M. (2020a). Controllability of control argumentation frameworks. In *Proceedings of the twenty-ninth international joint conference on artificial intelligence, IJCAI, 2020* (pp. 1855–1861).
38. Niskanen, A., Neugebauer, D., Järvisalo, M., & Rothe, J. (2020b). Deciding acceptance in incomplete argumentation frameworks. In *Proceedings of AAAI 2020*.
39. Oren, N., & Norman, T. J. (2009). Arguing using opponent models. In *Argumentation in multi-agent systems, 6th international workshop, ArgMAS 2009, Budapest, Hungary, May 12, 2009*. Revised Selected and Invited Papers (pp. 160–174).
40. Parsons, S., Sierra, C., & Jennings, N. R. (1998). Agents that reason and negotiate by arguing. *Journal of Logic and Computation, 8*(3), 261–292. https://doi.org/10.1093/logcom/8.3.261.
41. Pilotti, P., Casali, A., & Chesñevar, C. I. (2015). A belief revision approach for argumentation-based negotiation agents. *Applied Mathematics and Computer Science, 25*(3), 455–470.
42. Rahwan, I., Ramchurn, S. D., Jennings, N. R., McBurney, P., Parsons, S., & Sonenberg, L. (2003). Argumentation-based negotiation. *Knowledge Engineering Review, 18*(4), 343–375. https://doi.org/10.1017/S0269888904000098.
43. Reimer, S., Sauer, M., Marin, P., & Becker, B. (2014). QBF with soft variables. In *14th international workshop on automated verification of critical systems (AVOCS)*.
44. Rienstra, T., Thimm, M., & Oren, N. (2013). Opponent models with uncertainty for strategic argumentation. In *IJCAI 2013, proceedings of the 23rd international joint conference on artificial intelligence, Beijing, China, August 3–9, 2013* (pp. 332–338).
45. Riveret, R., Rotolo, A., Sartor, G., Prakken, H., & Roth, B. (2007). Success chances in argument games: A probabilistic approach to legal disputes. In Lodder, A. R., Mommers, L. (eds) *JURIX 2007: The twentieth annual conference on legal knowledge and information systems, Leiden, The Netherlands, 2007*, (Vol. 165, pp. 99–108). IOS Press.

46. Thimm, M., & Villata, S. (2017). The first international competition on computational models of argumentation: Results and analysis. *Artificial Intelligence, 252,* 267–294.
47. van Laar, J. A., & Krabbe, E. C. W. (2018). The role of argument in negotiation. *Argumentation, 32*(4), 549–567. https://doi.org/10.1007/s10503-018-9458-x.
48. Wallner, J. P., Niskanen, A., & Järvisalo, M. (2017). Complexity results and algorithms for extension enforcement in abstract argumentation. *Journal of Artificial Intelligence Research, 60,* 1–40.
49. Zafari, F., & Mofakham, F. N. (2017). POPPONENT: Highly accurate, individually and socially efficient opponent preference model in bilateral multi issue negotiations (extended abstract). In *Proceedings of the twenty-sixth international joint conference on artificial intelligence, IJCAI 2017, Melbourne, Australia, August 19–25, 2017* (pp. 5100–5104).