

An informant-based approach to argument strength in Defeasible Logic Programming

Andrea Cohen*, Sebastian Gottifredi, Luciano H. Tamargo, Alejandro J. García and Guillermo R. Simari

Institute for Computer Science and Engineering (CONICET-UNS), Department of Computer Science and Engineering, Universidad Nacional del Sur, San Andrés 800 – Campus Palihue, Bahía Blanca, Buenos Aires, Argentina

E-mails: ac@cs.uns.edu.ar, sg@cs.uns.edu.ar, lt@cs.uns.edu.ar, ajg@cs.uns.edu.ar, grs@cs.uns.edu.ar

Abstract. This work formalizes an informant-based structured argumentation approach in a multi-agent setting, where the knowledge base of an agent may include information provided by other agents, and each piece of knowledge comes attached with its informant. In that way, arguments are associated with the set of informants corresponding to the information they are built upon. Our approach proposes an informant-based notion of argument strength, where the strength of an argument is determined by the credibility of its informant agents. Moreover, we consider that the strength of an argument is not absolute, but it is relative to the resolution of the conflicts the argument is involved in. In other words, the strength of an argument may vary from one context to another, as it will be determined by comparison to its attacking arguments (respectively, the arguments it attacks). Finally, we equip agents with the means to express reasons for or against the consideration of any piece of information provided by a given informant agent. Consequently, we allow agents to argue about the arguments' strength through the construction of arguments that challenge (respectively, defeat) or are in favour of their informant agents.

Keywords: Informants credibility, argument strength, Defeasible Logic Programming

1. Introduction

Argumentation has proved to be a powerful paradigm for conceptualizing commonsense reasoning and a promising research area within the field of Artificial Intelligence [7,11,14,40]. One of the main issues an argumentation system has to address is the selection of the acceptable arguments, whose conclusions can be considered to be justified. For this purpose, it is necessary to account for the conflicts between the arguments in the system, and how those conflicts are to be resolved. At this point, a crucial notion comes into play: the notion of argument strength. In general, given an attack from an argument \mathcal{A} to an argument \mathcal{B} , the attack will only succeed if \mathcal{A} is stronger than \mathcal{B} . However, there is no unique way of establishing the strength of the arguments, as evidenced by the wide variety of approaches existing in the literature of argumentation. For instance, in [3,5] the arguments' strength is established by a general preference relation. On the other hand, [13] defines the strength of an argument through a formula yielding a numerical value, based on the intuition that the strength of an argument depends on the strength of its attackers. Finally, for instance, approaches to value-based argumentation [10] consider

*Corresponding author. E-mail: ac@cs.uns.edu.ar.

that the strength of an argument depends on the social values it advances, and determining whether the attack of one argument on another succeeds depends on the comparative strength of the values advanced by the arguments concerned. In contrast to these, in this work we will adopt a symbolic informant-based approach to argument strength in structured argumentation, where the strength of an argument is established in terms of the strength or credibility of its informant agents. Moreover, the strength of arguments in our approach will not be absolute and, furthermore, could be challenged.

In this paper we assume a multi-agent setting where agents share domain knowledge with one another. Hence, each deliberative agent may obtain pieces of information from other informant agents, which can have different levels of credibility. In order to reach conclusions to establish their beliefs, agents will build arguments based on the information in their knowledge bases, which include the pieces of information received from a set of their informant agents. In this context, we introduce an informant-based argumentative approach where the credibility of the information sources (*i.e.*, the informant agents) will be used for determining the arguments' strength. On the one hand, as it is usual in argumentation systems, arguments could be challenged due to the information they use. On the other hand, since the pieces of information used for building the arguments are associated with their informant agents, our approach is such that arguments can also be challenged on their credibility, in other words, on their strength.

The knowledge representation and reasoning capabilities of our approach will take their basis from DeLP [25]. Specifically, our proposal greatly extends and further develops the approach introduced in [21] by providing a formal account of the arguments' strength based on the credibility of their informant sources. In particular, we extend DeLP's representational language by including backing and detracting rules, respectively enabling to express reasons for and against the consideration of the information provided by some informant agents. Then, these new types of rules will be used for building backing and detracting arguments, allowing to argue about the arguments' strength. As a result, in a situation where some informants are challenged whereas others are not, the existence of multiple informants for a given piece of information in an agent's knowledge base may strengthen its position (hence, of the arguments using them).

For example, let us consider an agent I_1 that has to decide whether to buy a used smartphone or not; this agent receives information from agents I_2 , I_3 , I_4 and I_5 . Let us also assume that agent I_1 regards I_3 as less credible than I_2 , and I_4 as more credible than I_5 . Now, suppose I_2 informs I_1 that "the phone is a good option", whereas I_1 receives information from I_3 expressing that "the phone is not a good option". With this information, agent I_1 will be able to build arguments for buying the phone and for not buying it, and these two arguments will attack each other. Then, taking the credibility of the informants into account, the former argument will prevail over the latter and agent I_1 will arrive to the conclusion that "the phone is a good option" based on information provided by the more credible agent. However, it could be the case that there exist reasons for or against the consideration of information received from some informant. For instance, suppose agent I_5 provides information saying that I_2 is an expert on electronic devices. In addition, suppose I_4 informs I_1 that I_2 should not be trusted when it comes to recommending an item it owns. This new information could be encoded through a backing and a detracting rule, respectively, leading to the construction of their homonym arguments. As a result, the backing and detracting arguments would be in conflict and, taking the credibility of their informants into account, the detracting argument would prevail over the backing argument (because I_4 is more credible than I_5). In particular, the approach proposed in this paper will be able to handle situations like these, where the strength of the argument expressing that the phone was a good option is being challenged by the detracting argument.

To summarize, the contribution of this paper is two-fold. On the one hand, we formalize an informant-based structured argumentation approach where the strength of an argument is determined by the strength (*i.e.*, the credibility) of its informant agents. Furthermore, as will become clear later, the strength of an argument in our approach is not absolute, but it is relative to the resolution of the conflicts the argument is involved in. In other words, the strength of an argument will vary from one context to another, as it will be determined by comparison to its attacking arguments (respectively, the arguments it attacks). As a result, it could be the case that some informant providing information for building an argument is relevant for establishing the argument's strength in a given context, but not in others. On the other hand, through the incorporation of backing and detracting rules (and their homonym arguments) we provide the means for reasoning about the arguments' strength. In particular, the former will allow us to express reasons for the consideration of any piece of information provided by a given informant agent; analogously, the latter enable to express reasons against the consideration of an informant and thus, will be challenging the strength of arguments making use of information provided by that informant.

The rest of this paper is organized as follows. In Section 2 we will introduce the elements that will be used to represent the agents' knowledge. Section 3 shows how an agent can build different kinds of arguments and identify different kinds of attacks between them. Then, in Section 4, we will introduce the ways in which the conflicts between arguments are resolved, leading to defeats. Moreover, taking those defeats into account, in Section 5 we formalize the agents' reasoning mechanism enabling to determine their warranted beliefs. Finally, Sections 6 and 7 discuss relevant related work, draw some conclusions and comment on future lines of work.

2. Knowledge representation

In this section we will introduce our proposal for agents' knowledge representation that will be defined as an *informant-based DeLP program*. We assume that agents may share tentative information in the form of defeasible rules, and that each agent can obtain information from other informant agents that have different degrees of credibility. Also, backing and detracting rules will be proposed in our formalization, allowing to express reasons for and against the consideration of the informant agents providing knowledge used for building the arguments.

The assumption of the existence of a total credibility order over informants is not quite realistic in many multi-agent application domains, and a similar observation applies to the existence of a global order shared by all agents. With this observation in mind, the approach to be introduced below will consider that every agent has its own partial order defined over the set of informant agents, representing the credibility it assigns to each informant. Each agent has a knowledge base where every piece of information is attached with an agent identifier representing that the corresponding agent is the source of that piece of information. In addition, agents could communicate with their peers for obtaining new information or for sharing their beliefs. Clearly, as agents may disagree with one another, the beliefs an agent has may be in conflict with another agent's knowledge. Then, when sharing conflicting information, the credibility order among the informant agents can be used to decide which information prevails. Next, we briefly introduce the notion of *credibility order*, which will be used throughout the rest of the paper.

We assume a finite set \mathbb{I} of identifiers for naming informant agents, shared by all agents. Agent identifiers will be denoted with an uppercase typewriter letter "I" that can have letters and natural numbers as subscripts (*i.e.*, $\mathbb{I} = \{I_a, I_b, \dots, I_z, I_1, I_2, \dots, I_n, \dots\}$), and each identifier is unequivocally associated with a single agent.

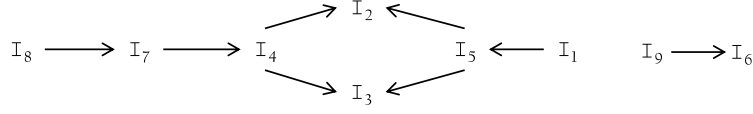


Fig. 1. Graphical representation of the credibility order $\prec_{co}^{I_6}$ from Example 1.

Each agent will have its own credibility order, represented by an irreflexive, asymmetric and transitive binary relation over \mathbb{I} , denoted $\prec_{co}^{I_x}$ (i.e., $\prec_{co}^{I_x}$ is a strict partial order over \mathbb{I}), where $I_x \in \mathbb{I}$ stands for the agent identifier this order belongs to; for instance, $\prec_{co}^{I_a}$ is the credibility order of the agent identified as I_a . In this paper we will assume that the credibility order relates agents that are sources of information about the same topic; multi-topic or multi-context credibility orders will be considered in future work. The notation $I_b \prec_{co}^{I_a} I_c$ means that agent I_a deems agent I_b as less credible than agent I_c (equivalently, under I_a 's perspective, I_c is more credible than I_b). Then, the notation $I_b \not\prec_{co}^{I_a} I_c$ is used to express that, for agent I_a , I_b is not less credible than I_c . Furthermore, if $I_b \not\prec_{co}^{I_a} I_c$ and $I_c \not\prec_{co}^{I_a} I_b$, then agent I_a regards I_b and I_c to be incomparable, noted as $I_b \sim_{co}^{I_a} I_c$. The following example illustrates a credibility order over agents, which will be used as part of the running example throughout the rest of the paper.

Example 1. Consider the set of informants $\{I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8, I_9\} \subseteq \mathbb{I}$. Suppose agent I_6 has the following credibility order $\prec_{co}^{I_6} = \{I_2 \prec_{co}^{I_6} I_5, I_3 \prec_{co}^{I_6} I_5, I_5 \prec_{co}^{I_6} I_1, I_2 \prec_{co}^{I_6} I_4, I_3 \prec_{co}^{I_6} I_4, I_4 \prec_{co}^{I_6} I_7, I_7 \prec_{co}^{I_6} I_8, I_6 \prec_{co}^{I_6} I_9\}$. Figure 1 shows a graphical representation of the credibility order $\prec_{co}^{I_6}$. Following this notation, an edge from a node N_1 to a node N_2 represents that $N_2 \prec_{co}^{I_6} N_1$. For instance, the edge from I_7 to I_4 in Fig. 1 denotes the credibility relation $I_4 \prec_{co}^{I_6} I_7$.

It should be noted that agents may change their assessment of one another; as a result, this change would impact their credibility orders, resulting in an update. The dynamic nature of credibility orders is outside the scope of this paper. However, for instance, the formalism proposed in [42], which provides a mechanism for handling the dynamics of a credibility order, can be a complement to our proposal.

As stated before, agents can receive information from different sources that are not equally credible. Thus, the information they provide, which may be contradictory, will be considered to be tentative. In this setting we need a mechanism for dealing with uncertain and conflicting information, able to ultimately determine the beliefs an agent should commit to. In particular, when deciding between contradictory conclusions, the reasoning mechanism will rely on the credibility order the agent has. In our approach, knowledge representation and reasoning is inspired on Defeasible Logic Programming (DeLP) [25]. Our interest in this particular computational tool is that its language provides the declarative capability of representing weak information in the form of *defeasible rules* and *presumptions*, and its defeasible argumentation inference mechanism allows warranting conclusions in the presence of contradictory information. Below we will introduce the elements that will be used for representing an agent's knowledge, some of which are taken from DeLP's representational language.

Let \mathcal{L} be a set of ground atoms. A *ground literal* L (or *literal*, for short) is an atom A or a negated atom $\sim A$, where $A \in \mathcal{L}$ and the symbol " \sim " represents strong negation.

Definition 1 (Defeasible rule). A *defeasible rule* is an ordered pair, denoted "*Head* \prec *Body*", where the first element (*Head*) is a ground literal and the second element (*Body*) is a finite set of ground literals. A defeasible rule with head L_0 and body $\{L_1, \dots, L_n\}$ ($n \geq 0$) will also be written as $L_0 \prec L_1, \dots, L_n$.

A defeasible rule “ $Head \prec Body$ ” expresses that “reasons to believe in the antecedent $Body$ give reasons to believe in the consequent $Head$ ”. For instance, the defeasible rule “ $closed_roads \prec snowing$ ” represents that “*reasons to believe that it is snowing provide reasons to believe that the roads are closed*”, whereas “ $\sim closed_roads \prec snowplows$ ” represents that “*reasons to believe that snowplows are working provide reasons to believe that the roads are not closed*”. In particular, a defeasible rule with an empty body is noted as “ $Head \prec$ ” and is called a *presumption* [37]. Hence, for instance, the presumption “ $snowplows \prec$ ” expresses that “*there are defeasible reasons to believe that snowplows are working*”. We associate defeasible rules (including presumptions) with their informant agents; for this, we introduce the notion of *defeasible domain object*.

Definition 2 (Defeasible domain object). Let \mathbb{I} be a finite set of agent identifiers. A *defeasible domain object* is a tuple $(\mathbb{I} : R)$, where $\mathbb{I} \in \mathbb{I}$ and R is a defeasible rule.

In addition to defeasible rules, we introduce *backing rules* and *detracting rules* to express reasons for and against the consideration of informants, respectively. Formally:

Definition 3 (Backing rule). Let \mathbb{I} be a finite set of agent identifiers. A *backing rule* is an ordered pair, denoted “ $(Head \leftarrow \oplus Body)$ ”, where $Head \in \mathbb{I}$ is the identifier of an informant agent and $Body$ is a finite and non-empty set of literals. A backing rule with head \mathbb{I} and body $\{L_1, \dots, L_n\}$ ($n \geq 1$) will also be written as $(\mathbb{I} \leftarrow \oplus L_1, \dots, L_n)$.

Definition 4 (Detracting rule). Let \mathbb{I} be a finite set of agent identifiers. A *detracting rule* is an ordered pair, denoted “ $(Head \leftarrow \ominus Body)$ ”, where $Head \in \mathbb{I}$ is the identifier of an informant agent and $Body$ is a finite and non-empty set of literals. A detracting rule with head \mathbb{I} and body $\{L_1, \dots, L_n\}$ ($n \geq 1$) will also be written as $(\mathbb{I} \leftarrow \ominus L_1, \dots, L_n)$.

Syntactically, the only difference between backing and detracting rules is the use of \oplus and \ominus ; however, these two types of rules are semantically opposite. A backing rule “ $(\mathbb{I} \leftarrow \oplus Body)$ ” expresses that “*the antecedent $Body$ gives reasons for considering information provided by the informant \mathbb{I}* ”. On the contrary, a detracting rule “ $(\mathbb{I} \leftarrow \ominus Body)$ ” expresses that “*the antecedent $Body$ gives reasons against considering information provided by the informant \mathbb{I}* ”. Also, note that backing and detracting rules do not have an associated agent (*i.e.*, their source is not identified), meaning that they correspond to the agent’s own knowledge. Nevertheless, since these rules have a set of literals in their body, it is also possible to argue about the reasons for or against the consideration of the informants in their heads. When convenient, we will refer to backing and detracting rules as *informant rules*.

The entire knowledge of an agent, which will be used to make inferences and construct arguments, is composed of a set of defeasible domain objects and a set of informant rules. This knowledge base will be called *informant-based DeLP program*.

Definition 5 (Informant-based DeLP program). An *informant-based DeLP program* (IBDP for short) is a pair (Δ, Σ) , where Δ is a finite set of defeasible domain objects and Σ is a finite set of informant rules.

The following example introduces an IBDP that will serve as a running example, to illustrate the different notions proposed in this paper.

Example 2. Let us consider the IBDP $P_{\mathcal{I}_6} = (\Delta_{\mathcal{I}_6}, \Sigma_{\mathcal{I}_6})$ of an agent \mathcal{I}_6 . Note that $P_{\mathcal{I}_6}$ encodes agent \mathcal{I}_6 's own information as well as information obtained from the following informant agents: $\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathcal{I}_4, \mathcal{I}_5, \mathcal{I}_7, \mathcal{I}_8$.

$$\Delta_{\mathcal{I}_6} = \left\{ \begin{array}{l} (\mathcal{I}_1 : h \multimap a), \quad (\mathcal{I}_6 : c \multimap), \quad (\mathcal{I}_4 : p \multimap), \\ (\mathcal{I}_2 : a \multimap b, g), \quad (\mathcal{I}_5 : d \multimap), \quad (\mathcal{I}_6 : w \multimap x), \\ (\mathcal{I}_3 : b \multimap), \quad (\mathcal{I}_2 : g \multimap), \quad (\mathcal{I}_3 : q \multimap i), \\ (\mathcal{I}_4 : \sim a \multimap c, d), \quad (\mathcal{I}_3 : i \multimap), \quad (\mathcal{I}_7 : j \multimap), \\ (\mathcal{I}_7 : z \multimap), \quad (\mathcal{I}_8 : z \multimap) \end{array} \right\}$$

$$\Sigma_{\mathcal{I}_6} = \left\{ \begin{array}{l} (\mathcal{I}_5 \leftarrow \oplus p), \quad (\mathcal{I}_3 \leftarrow \oplus x), \\ (\mathcal{I}_5 \leftarrow \ominus g, q), \quad (\mathcal{I}_4 \leftarrow \ominus x), \\ (\mathcal{I}_5 \leftarrow \ominus w), \quad (\mathcal{I}_5 \leftarrow \ominus j), \\ (\mathcal{I}_8 \leftarrow \ominus j, p), \quad (\mathcal{I}_6 \leftarrow \ominus j) \end{array} \right\}$$

The program $P_{\mathcal{I}_6}$ has a set of defeasible domain objects in $\Delta_{\mathcal{I}_6}$ that include defeasible rules and presumptions obtained from different informants. It is worth mentioning that two different defeasible domain objects can have the same informant: for instance, $(\mathcal{I}_7 : j \multimap)$ and $(\mathcal{I}_7 : z \multimap)$. Also, two different defeasible domain objects can have the same defeasible rule with different informant agents: for instance, $(\mathcal{I}_7 : z \multimap)$ and $(\mathcal{I}_8 : z \multimap)$. On the other hand, the set $\Sigma_{\mathcal{I}_6}$ includes backing and detracting rules. For instance, the detracting rule $(\mathcal{I}_5 \leftarrow \ominus g, q)$ expresses that “ g and q give reasons against the consideration of any piece of information provided by \mathcal{I}_5 ”, and the backing rule $(\mathcal{I}_5 \leftarrow \oplus p)$ expresses that “ p gives reasons for considering the information provided by \mathcal{I}_5 ”.

As evidenced in Example 2, an IBDP can have two or more defeasible domain objects with the same defeasible rule R but different informant agents. This does not mean that the agent's knowledge base is redundant; rather, it encodes the fact that the same piece of information was received from different sources. This feature can be considered as an advantage of our approach since, as mentioned before, the credibility order of an agent may change (even dynamically); hence, at any moment, the more credible informant of R could be considered. Furthermore, given the existence of detracting rules, a specific informant of R could be challenged, whereas others are not. Again, a situation like this is illustrated by Example 2, where there exists a detracting rule for the informant \mathcal{I}_8 of “ $z \multimap$ ”, but not for the informant \mathcal{I}_7 . As a result, the existence of multiple informants may strengthen the position of a defeasible rule by allowing for different defeasible domain objects within the agent's knowledge base (therefore, of the arguments using them). Finally, it is important to remark that the existence of backing and detracting rules in an IBDP is not mandatory.

In our approach, an agent will be specified in terms of four components: its own *agent identifier*, an *informant-based DeLP program* used to store its knowledge, a *credibility order* among informants, and an *informant-based comparison criterion*. The first three elements have been introduced above. On the other hand, the informant-based comparison criterion will make use of the credibility order to establish strict preferences among sets of informants. Briefly, sets of informant agents will be compared as part of the warranting process with the aim of deciding between conflicting arguments, to ultimately determine the accepted arguments and the justified conclusions of the agent. Consequently, since the notion of argument and the way in which they become in conflict (attack) have not been formalized yet, we will postpone the characterization of informant-based comparison criteria until Section 4; that is, we will

formally introduce that notion after providing the formal context in which such criteria will be applied. An agent is defined as follows:

Definition 6 (Agent). Let \mathbb{I} be a finite set of agent identifiers. An *agent* is a tuple $(\mathbb{I}, P_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ where $\mathbb{I} \in \mathbb{I}$, $P_{\mathbb{I}} = (\Delta_{\mathbb{I}}, \Sigma_{\mathbb{I}})$ is an informant-based DeLP program, $<_{co}^{\mathbb{I}} \subseteq \mathbb{I} \times \mathbb{I}$ is a credibility order over \mathbb{I} , and $<_{\mathbb{I}} \subseteq 2^{\mathbb{I}} \times 2^{\mathbb{I}}$ is an informant-based comparison criterion over sets of informants.

Example 3. Consider the IBDP $P_{\mathbb{I}_6}$ from Example 2 and the credibility order $<_{co}^{\mathbb{I}_6}$ presented in Example 1. Then, agent \mathbb{I}_6 can be specified as $(\mathbb{I}_6, P_{\mathbb{I}_6}, <_{co}^{\mathbb{I}_6}, <_{\mathbb{I}_6})$. The informant-based comparison criterion $<_{\mathbb{I}_6}$ will be introduced in Section 4.

Given the characterization of an agent, it can be the case that two different agents $(\mathbb{I}_1, P_{\mathbb{I}_1}, <_{co}^{\mathbb{I}_1}, <_{\mathbb{I}_1})$ and $(\mathbb{I}_2, P_{\mathbb{I}_2}, <_{co}^{\mathbb{I}_2}, <_{\mathbb{I}_2})$ have the same IBDP and the same informant-based comparison criterion (thus, $P_{\mathbb{I}_1} = P_{\mathbb{I}_2}$ and $<_{\mathbb{I}_1} = <_{\mathbb{I}_2}$) but a different credibility order. In such a case, even though the two agents share the same knowledge, since the comparison criterion is defined in terms of the credibility order, the conflicts arising from the consideration of inconsistent information may be resolved differently for the two agents. Hence, in such a case, the agents' inferences (warranted conclusions or beliefs) may differ. On the other hand, even though two agents $(\mathbb{I}_3, P_{\mathbb{I}_3}, <_{co}^{\mathbb{I}_3}, <_{\mathbb{I}_3})$ and $(\mathbb{I}_4, P_{\mathbb{I}_4}, <_{co}^{\mathbb{I}_4}, <_{\mathbb{I}_4})$ cannot have the same agent identifier, they can share every other component (*i.e.*, it can be the case that $P_{\mathbb{I}_3} = P_{\mathbb{I}_4}$, $<_{co}^{\mathbb{I}_3} = <_{co}^{\mathbb{I}_4}$ and $<_{\mathbb{I}_3} = <_{\mathbb{I}_4}$); clearly, in such a case, the two agents will draw the same conclusions.

In the following section we will introduce the three kinds of arguments that can be built using the knowledge represented in an IBDP. After that, the different kinds of conflicts that may arise between those arguments (referred to as attacks) will be identified. Then, in Section 4, we will introduce the notion of valid informant-based comparison criterion, used for determining the successful attacks.

3. Arguments and attacks

A central piece of this formalism that will allow agents to handle contradictory domain information is the notion of *argument*. Intuitively, an argument is a structure whose conclusion is obtained from a set of premises, informed by some agents, through the use of a reasoning mechanism. In particular, the claims of the arguments will be the tentative beliefs of the agents. When analyzing an argument for a particular belief, an agent can find other arguments, referred to as counter-arguments, that are in conflict with it. Specifically, a conflict may arise because the counter-argument contradicts some information (*i.e.*, a premise, an intermediate conclusion or the final claim) in the argument or because it challenges one of its informants. In this situation, it is necessary to have a mechanism for comparing the conflicting arguments to decide which one prevails. This analysis leads to a dialectical process seeking to validate the arguments in conflict. The arguments that survive all possible attacks from their counter-arguments will be said to warrant their conclusions or claims, and these will be the agent's beliefs.

Next, we will show how an agent can build different types of arguments using the defeasible domain objects and the informant rules stored in its informant-based DeLP program. As a preliminary notion, before formally defining the arguments, we introduce the concept of *defeasible derivation*.

Definition 7 (Defeasible derivation). Let $P = (\Delta, \Sigma)$ be an IBDP and L a literal. A defeasible derivation of L from $S \subseteq \Delta$, denoted $S \vdash_P L$, consists of a finite sequence $L_1, L_2, \dots, L_n = L$ of literals, where each literal L_i ($1 \leq i \leq n$) is in the sequence because there exists a defeasible domain object $(\mathbb{I} : R)$ in

S such that: $R = L_i \multimap$; or $R = L_i \multimap B_1, \dots, B_k$, and every B_t ($1 \leq t \leq k$) is an element L_j of the sequence appearing before L_i ($j < i$).

A derivation for a literal L is called “defeasible” because, as we will show next, there may exist information in contradiction with L or any other literal appearing in the sequence, and under certain conditions this could prevent the acceptance of L as a warranted belief. It should be noted that rules from different informants can be combined to derive a literal. Also, note that the set S contains all the defeasible domain objects available for obtaining the derivation. Finally, it is important to observe that from the same IBDP it is possible to obtain several, distinct, defeasible derivations for a given literal. Furthermore, as the following example shows, from a given IBDP it is possible to obtain defeasible derivations for complementary literals (w.r.t. the strong negation “ \sim ”).

Example 4. Let us consider the IBDP P_{I_6} from Example 2. From that program it is possible to obtain the following defeasible derivation for the literal “ a ”: the sequence ‘ b, g, a ’ that makes use of the defeasible domain objects $(I_2 : a \multimap b, g)$, $(I_3 : b \multimap)$ and $(I_2 : g \multimap)$. Also, P_{I_6} allows to derive the literal “ $\sim a$ ” using the defeasible domain objects $(I_4 : \sim a \multimap c, d)$, $(I_6 : c \multimap)$ and $(I_5 : d \multimap)$ to obtain the sequence ‘ $c, d, \sim a$ ’. Thus, from P_{I_6} it is possible to obtain defeasible derivations for complementary literals.

As shown in Example 2, an IBDP may contain defeasible domain objects such that their heads correspond to complementary literals. This is reasonable because different informants may provide pieces of knowledge providing reasons for or against a given conclusion. Moreover, it could be the case that the same informant, under different conditions, gives reasons for or against a literal. Consequently, as illustrated in Example 4, defeasible derivations for complementary literals can be obtained from the same IBDP. In order to be able to identify coherent sets of elements within an IBDP, the notion of a *contradictory set* of defeasible domain objects of an IBDP is defined next.

Definition 8 (Contradictory set). Let $P = (\Delta, \Sigma)$ be an IBDP and $S \subseteq \Delta$. We say that the set S is *contradictory* if and only if there exist two complementary literals L and $\sim L$ such that $S \vdash_P L$ and $S \vdash_P \sim L$.

To illustrate this notion, let P_{I_6} be the IBDP from Example 2. Then, the set $\mathcal{A}_x = \{(I_2 : a \multimap b, g), (I_3 : b \multimap), (I_2 : g \multimap), (I_4 : \sim a \multimap c, d), (I_6 : c \multimap), (I_5 : d \multimap)\}$ is contradictory whereas the set $\mathcal{A}_y = \{(I_2 : a \multimap b, g), (I_4 : \sim a \multimap c, d), (I_6 : c \multimap), (I_5 : d \multimap)\}$ is not. As a result, for instance, every superset of \mathcal{A}_x will also be contradictory.

Observe that backing and detracting rules are not used for obtaining defeasible derivations. As will be shown next, they will only be used to build arguments that, respectively, support or challenge the consideration of informant agents. Moreover, as will be formalized in Section 4, these new types of argument will allow to argue about the arguments’ strength. The usual definition of argument is then extended to consider backing and detracting rules, when required. As a result, we will distinguish among three different types of arguments: the first type regards arguments that conclude literals, whereas the other two deal with arguments for or against informant agents, respectively.

Definition 9 (Claim argument). Let $P = (\Delta, \Sigma)$ be an IBDP and L a literal. $\langle \mathcal{A}, L \rangle$ is a *claim argument* for the literal L , built from P , if the following conditions hold:

- (1) $\mathcal{A} \subseteq \Delta$,
- (2) $\mathcal{A} \vdash_P L$,

- (3) \mathcal{A} is non-contradictory, and
- (4) \mathcal{A} is minimal: there is no $\mathcal{B} \subset \mathcal{A}$ satisfying conditions 2 and 3.

Definition 10 (Backing argument). Let \mathbb{I} be a finite set of agent identifiers, $P = (\Delta, \Sigma)$ an IBDP and $\mathbb{I} \in \mathbb{I}$. $\langle \mathcal{A}, \mathbb{I} \rangle$ is a *backing argument* for the informant \mathbb{I} , built from P , if $\mathcal{A} = \{(\mathbb{I} \leftarrow \oplus L_1, \dots, L_n)\} \cup \mathcal{A}'$, with $\mathcal{A}' \subseteq \Delta$, and the following conditions hold:

- (1) $\mathcal{A} \subseteq (\Delta \cup \Sigma)$,
- (2) $\mathcal{A}' \vdash_P L_i$ ($1 \leq i \leq n$),
- (3) $\nexists (\mathbb{I} : R) \in \mathcal{A}'$,
- (4) \mathcal{A}' is non-contradictory, and
- (5) \mathcal{A}' is minimal: there is no $\mathcal{B} \subset \mathcal{A}'$ satisfying conditions 2, 3 and 4.

Definition 11 (Detracting argument). Let \mathbb{I} be a finite set of agent identifiers, $P = (\Delta, \Sigma)$ an IBDP and $\mathbb{I} \in \mathbb{I}$. $\langle \mathcal{A}, \mathbb{I} \rangle$ is a *detracting argument* for the informant \mathbb{I} , built from P , if $\mathcal{A} = \{(\mathbb{I} \leftarrow \ominus L_1, \dots, L_n)\} \cup \mathcal{A}'$, with $\mathcal{A}' \subseteq \Delta$, and the following conditions hold:

- (1) $\mathcal{A} \subseteq (\Delta \cup \Sigma)$,
- (2) $\mathcal{A}' \vdash_P L_i$ ($1 \leq i \leq n$),
- (3) $\nexists (\mathbb{I} : R) \in \mathcal{A}'$,
- (4) \mathcal{A}' is non-contradictory, and
- (5) \mathcal{A}' is minimal: there is no $\mathcal{B} \subset \mathcal{A}'$ satisfying conditions 2, 3 and 4.

Briefly, as specified by the last clause in Definitions 9–11, all arguments share the characteristic of having a minimal and non-contradictory set of rules that allows to defeasibly derive a conclusion or the conditions for or against the consideration of an informant agent. In particular, this minimality requirement aligns with the definition of argument structures in [25], and has the aim of avoiding irrelevant information in the arguments (consequently, minimizing the points of attack). Also, the first clause in Definitions 10 and 11 is meant to ensure that the backing and detracting rules (as well as the corresponding defeasible domain objects) used for building the arguments actually belong to the IBDP from which the arguments are built. Another feature shared by backing and detracting arguments is that, when obtaining the required derivations, they cannot make use of defeasible domain objects provided by the informant they support or challenge, respectively. This constraint is meant to maintain a kind of coherence within the arguments that is not captured by the requirement of them being non-contradictory, as discussed below.

On the one hand, the third clause of Definition 10 has the aim of preventing the construction of backing arguments for a given informant, which are based on information provided by that same informant.¹ On the other hand, for detracting arguments, this constraint might not seem as intuitive as it is for backing arguments. This is because, in some cases, one might take into account pieces of information provided by agent \mathbb{I} in order to distrust that agent.² For instance, suppose the existence of a medical application where knowledge is expressed through an IBDP. Now, consider the existence of a detracting rule $(\mathbb{I}_X \leftarrow \ominus \sim \text{physician}_{\mathbb{I}_X})$, expressing reasons against the consideration of informant \mathbb{I}_X if he is not a physician. Then, if agent \mathbb{I}_X says he is not a physician (for instance, by providing a defeasible domain object $(\mathbb{I}_X : \sim \text{physician}_{\mathbb{I}_X} \leftarrow)$), one could think of using this piece of knowledge for building,

¹Such arguments can be seen as an instance of the fallacy known as *begging the question*, which refers to its own assertion to prove the assertion.

²We thank one of the reviewers for pointing out this example.

together with the previous detracting rule, a detracting argument for agent \mathbb{I}_X .³ Nonetheless, it should be noted that such a detracting argument for \mathbb{I}_X would not only be against other arguments built using the information provided by \mathbb{I}_X but would also be challenging itself, since its reasoning is based on information stated by \mathbb{I}_X . Hence, the third clause in Definition 11 is meant to avoid situations like these, where detracting arguments would end up being self-attacking (see Definition 14 below) and therefore, inconsistent.

It is worth mentioning that backing and detracting arguments (also, their homonym rules) resemble the backing and undercutting arguments (respectively, rules) of [19]. However, their intended meanings differ: whereas backing and undercutting arguments in [19] respectively provide reasons for and against the use of specific defeasible rules (with no notion of information source), the aim of backing and detracting arguments here is to allow arguing about the arguments' strength by providing reasons for and against the consideration of any piece of information provided by some informant agents. Also, differently from [19], in our approach there is no explicit notion of support between arguments, as it is usually done in the literature of bipolar argumentation (see [20] for an overview).

Finally note that, since informant rules are not used for obtaining defeasible derivations, claim arguments will not include informant rules. In contrast, a backing (respectively, detracting) argument will only include one backing (respectively, detracting) rule. Nevertheless, despite their common features, the three argument types are mutually exclusive. When convenient, we will abstract from an argument's type, referring to it just as an argument. Then, given an argument built from an IDBP, we can define the notion of *sub-argument* as follows.

Definition 12 (Sub-argument). Let $\mathbf{P} = (\Delta, \Sigma)$ be an IDBP and $\langle \mathcal{A}_1, L_1 \rangle, \langle \mathcal{A}_2, L_2 \rangle$ two arguments built from \mathbf{P} . We say that $\langle \mathcal{A}_2, L_2 \rangle$ is a *sub-argument* of $\langle \mathcal{A}_1, L_1 \rangle$ iff $\mathcal{A}_2 \subseteq \mathcal{A}_1$.

It should be noted that every proper sub-argument of a backing or a detracting argument is a claim argument. In general, if $\langle \mathcal{A}_2, Y \rangle$ is a proper sub-argument of $\langle \mathcal{A}_1, X \rangle$ (i.e., $\mathcal{A}_2 \subset \mathcal{A}_1$) then $\langle \mathcal{A}_2, Y \rangle$ is a claim argument.

Example 5. Given the agent's specification introduced in Example 3, agent \mathbb{I}_6 will be able to build the following arguments from $\mathbf{P}_{\mathbb{I}_6}$, among others:

- $\langle \mathcal{A}_1, h \rangle$, with $\mathcal{A}_1 = \{(\mathbb{I}_1 : h \multimap a), (\mathbb{I}_2 : a \multimap b, g), (\mathbb{I}_3 : b \multimap \cdot), (\mathbb{I}_2 : g \multimap \cdot)\}$
- $\langle \mathcal{A}_2, a \rangle$, with $\mathcal{A}_2 = \{(\mathbb{I}_2 : a \multimap b, g), (\mathbb{I}_3 : b \multimap \cdot), (\mathbb{I}_2 : g \multimap \cdot)\}$
- $\langle \mathcal{A}_3, \sim a \rangle$, with $\mathcal{A}_3 = \{(\mathbb{I}_4 : \sim a \multimap c, d), (\mathbb{I}_6 : c \multimap \cdot), (\mathbb{I}_5 : d \multimap \cdot)\}$
- $\langle \mathcal{A}_4, \mathbb{I}_5 \rangle$, with $\mathcal{A}_4 = \{(\mathbb{I}_5 \leftarrow \ominus g, q), (\mathbb{I}_2 : g \multimap \cdot), (\mathbb{I}_3 : q \multimap i), (\mathbb{I}_3 : i \multimap \cdot)\}$
- $\langle \mathcal{A}_5, \mathbb{I}_5 \rangle$, with $\mathcal{A}_5 = \{(\mathbb{I}_5 \leftarrow \oplus p), (\mathbb{I}_4 : p \multimap \cdot)\}$
- $\langle \mathcal{A}_6, \mathbb{I}_5 \rangle$, with $\mathcal{A}_6 = \{(\mathbb{I}_5 \leftarrow \ominus j), (\mathbb{I}_7 : j \multimap \cdot)\}$
- $\langle \mathcal{A}_7, \mathbb{I}_6 \rangle$, with $\mathcal{A}_7 = \{(\mathbb{I}_6 \leftarrow \ominus j), (\mathbb{I}_7 : j \multimap \cdot)\}$

Note that $\langle \mathcal{A}_1, h \rangle$, $\langle \mathcal{A}_2, a \rangle$ and $\langle \mathcal{A}_3, \sim a \rangle$ are claim arguments. Also, $\langle \mathcal{A}_5, \mathbb{I}_5 \rangle$ is a backing argument, whereas $\langle \mathcal{A}_4, \mathbb{I}_5 \rangle$, $\langle \mathcal{A}_6, \mathbb{I}_5 \rangle$ and $\langle \mathcal{A}_7, \mathbb{I}_6 \rangle$ are detracting arguments. In addition, observe that $\langle \mathcal{A}_2, a \rangle$ is a proper sub-argument of $\langle \mathcal{A}_1, h \rangle$.

³Recall that in this work we are assuming that the knowledge encoded through an IDBP belongs to the same topic. In this example all the information being modeled, in particular, the one claiming that \mathbb{I}_X is not a physician, pertains to the medical domain.

The above example illustrates that from a given IBDP it is possible to build arguments that are in conflict with one another, such as the claim arguments $\langle \mathcal{A}_2, a \rangle$ and $\langle \mathcal{A}_3, \sim a \rangle$. These conflicts become evident since the arguments' conclusions correspond to complementary literals. On the other hand, the existence of backing and detracting arguments may lead to new kinds of conflict, that will allow to argue about the arguments' strength in the resolution of other conflicts. The following definitions formalize the different kinds of conflict that may occur between arguments built from an IBDP, from hereon referred to as *attacks*.

The first kind of attack, called *conclusion attack* (or *c-attack*) captures the usual conflict in an argumentation system, where the claim of one argument contradicts a conclusion (premise, intermediate or final claim) of another.

Definition 13 (Conclusion attack). Let $P = (\Delta, \Sigma)$ be an IBDP, $\langle \mathcal{A}_1, L_1 \rangle$ a claim argument built from P , and $\langle \mathcal{A}_2, L_2 \rangle$ any argument built from P . We say that $\langle \mathcal{A}_1, L_1 \rangle$ *c-attacks* $\langle \mathcal{A}_2, L_2 \rangle$ at the literal L iff there exists a claim sub-argument $\langle \mathcal{A}, L \rangle$ of $\langle \mathcal{A}_2, L_2 \rangle$ such that L_1 and L are complementary literals with respect to “ \sim ”. We refer to $\langle \mathcal{A}, L \rangle$ as the *disagreement sub-argument* in the attack.

The second kind of attack we consider, called *strength attack* (or *s-attack*) aims at capturing the intuition that a detracting argument for a given informant is challenging any argument that makes use of a defeasible domain object provided by that informant agent. Then, since the strength of an argument is determined in terms of the strength of its informant agents, according to the adopted informant-based comparison criterion, the detracting argument is somehow attacking the other argument's strength.

Definition 14 (Strength attack). Let \mathbb{I} be a finite set of agent identifiers, $\mathbb{I} \in \mathbb{I}$, $P = (\Delta, \Sigma)$ an IBDP, $\langle \mathcal{A}_1, \mathbb{I} \rangle$ a detracting argument built from P , and $\langle \mathcal{A}_2, L_2 \rangle$ any argument built from P . We say that $\langle \mathcal{A}_1, \mathbb{I} \rangle$ *s-attacks* $\langle \mathcal{A}_2, L_2 \rangle$ at the informant \mathbb{I} iff there exists a defeasible domain object $(\mathbb{I} : R) \in \mathcal{A}_2$.

The next kind of attack we consider, referred to as *strength-defense attack* (or *sd-attack*), corresponds to the situation where a backing argument for a given informant exists, and the backing argument attacks a detracting argument for that informant. In such a situation, we consider that the backing argument is defending the strength of the argument being attacked by the detracting argument.

Definition 15 (Strength-defense attack). Let \mathbb{I} be a finite set of agent identifiers, $\mathbb{I} \in \mathbb{I}$, $P = (\Delta, \Sigma)$ an IBDP, $\langle \mathcal{A}_1, \mathbb{I} \rangle$ a backing argument built from P , and $\langle \mathcal{A}_2, \mathbb{I} \rangle$ a detracting argument built from P . We say that $\langle \mathcal{A}_1, \mathbb{I} \rangle$ *sd-attacks* $\langle \mathcal{A}_2, \mathbb{I} \rangle$.

Given the existence of a backing argument for an informant \mathbb{I} , who might originate a strength-defense attack, any detracting argument for \mathbb{I} can be considered as providing a counter-defense for the strength of the argument for whom \mathbb{I} was being challenged. This kind of attack, called *strength-counter-defense attack* (or *scd-attack*), is defined as follows.

Definition 16 (Strength-counter-defense attack). Let \mathbb{I} be a finite set of agent identifiers, $\mathbb{I} \in \mathbb{I}$, $P = (\Delta, \Sigma)$ an IBDP, $\langle \mathcal{A}_1, \mathbb{I} \rangle$ a detracting argument built from P , and $\langle \mathcal{A}_2, \mathbb{I} \rangle$ a backing argument built from P . We say that $\langle \mathcal{A}_1, \mathbb{I} \rangle$ *scd-attacks* $\langle \mathcal{A}_2, \mathbb{I} \rangle$.

Consequently, whenever an sd-attack exists, an scd-attack will also exist and vice-versa. This is because, as explained above, the latter is meant to provide a counter-defense to the defense provided by a backing argument. More generally, whenever a backing argument and a detracting argument for the same informant \mathbb{I} exist, a 2-cycle of attacks will exist between such arguments (respectively, an sd-attack

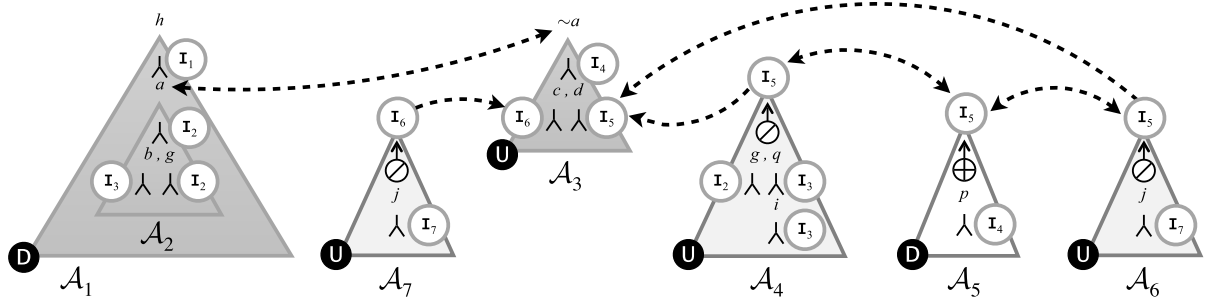


Fig. 2. Arguments and attacks from Example 6.

from the backing argument to the detaching argument, and an scd-attack from the detaching argument to the backing argument).

Finally, if an argument $\langle \mathcal{A}_1, L_1 \rangle$ attacks (either c-attacks, s-attacks, sd-attacks or scd-attacks) another argument $\langle \mathcal{A}_2, L_2 \rangle$, we will say that $\langle \mathcal{A}_1, L_1 \rangle$ is a *counter-argument* of $\langle \mathcal{A}_2, L_2 \rangle$.

Example 6. Given the arguments listed in Example 5, some attacks are identified next: argument $\langle \mathcal{A}_3, \sim a \rangle$ c-attacks arguments $\langle \mathcal{A}_1, h \rangle$ and $\langle \mathcal{A}_2, a \rangle$ at the literal “ a ”; the disagreement sub-argument in both cases is $\langle \mathcal{A}_2, a \rangle$. Similarly, $\langle \mathcal{A}_2, a \rangle$ c-attacks $\langle \mathcal{A}_3, \sim a \rangle$ at the literal “ $\sim a$ ”, where the disagreement sub-argument is $\langle \mathcal{A}_3, \sim a \rangle$. In addition, the detaching arguments $\langle \mathcal{A}_4, I_5 \rangle$ and $\langle \mathcal{A}_6, I_5 \rangle$ s-attack $\langle \mathcal{A}_3, \sim a \rangle$ because $(I_5 : d -) \in \mathcal{A}_3$. Then, the detaching argument $\langle \mathcal{A}_7, I_6 \rangle$ s-attacks $\langle \mathcal{A}_3, \sim a \rangle$ because $(I_6 : c -) \in \mathcal{A}_3$. On the other hand, for instance, the backing argument $\langle \mathcal{A}_5, I_5 \rangle$ sd-attacks the detaching arguments $\langle \mathcal{A}_4, I_5 \rangle$ and $\langle \mathcal{A}_6, I_5 \rangle$, and both $\langle \mathcal{A}_4, I_5 \rangle$ and $\langle \mathcal{A}_6, I_5 \rangle$ scd-attack argument $\langle \mathcal{A}_5, I_5 \rangle$.

These arguments and attacks are illustrated in Fig. 2. In particular, arguments are depicted with triangles, and attacks are represented with dashed arrows between the arguments. Furthermore, the circles beside the symbol of the defeasible rule “ $-$ ” within the arguments indicate the informant associated with each defeasible domain object in those arguments.

As the preceding example shows, conclusion attacks are symmetric in a sense. That is, if $\langle \mathcal{A}_1, L_1 \rangle$ c-attacks $\langle \mathcal{A}_2, L_2 \rangle$, then the disagreement sub-argument $\langle \mathcal{A}, L \rangle$ of $\langle \mathcal{A}_2, L_2 \rangle$ is such that it c-attacks $\langle \mathcal{A}_1, L_1 \rangle$. In particular, if $\langle \mathcal{A}_1, L_1 \rangle$ c-attacks $\langle \mathcal{A}_2, L_2 \rangle$ on its final conclusion L_2 , then $\langle \mathcal{A}_2, L_2 \rangle$ c-attacks $\langle \mathcal{A}_1, L_1 \rangle$; these situations were illustrated by arguments $\langle \mathcal{A}_3, \sim a \rangle$, $\langle \mathcal{A}_1, h \rangle$ and $\langle \mathcal{A}_2, a \rangle$ in Example 6. In contrast, strength attacks need not be symmetric. This is due to the fact that the attacking argument is a detaching argument, and no particular attacked sub-argument is identified. However, this does not prevent detaching arguments from attacking each other. For instance, it could be the case that $\langle \mathcal{A}_1, I_1 \rangle$ s-attacks $\langle \mathcal{A}_2, I_2 \rangle$ at the informant I_1 , and $\langle \mathcal{A}_2, I_2 \rangle$ s-attacks $\langle \mathcal{A}_1, I_1 \rangle$ at the informant I_2 ; what would occur in such a case is that there exist two defeasible domain objects $(I_1 : R)$ and $(I_2 : R')$ such that $(I_1 : R) \in \mathcal{A}_2$ and $(I_2 : R') \in \mathcal{A}_1$. On the other hand, as mentioned before, whenever a backing argument \mathcal{B} sd-attacks a detaching argument \mathcal{D} , \mathcal{D} scd-attacks \mathcal{B} ; moreover, any other detaching argument for the same informant will also scd-attack \mathcal{B} (consequently, will be sd-attacked by \mathcal{B}). That is, the existence of a backing argument and a detaching argument for the same informant leads to a two-way conflict between those arguments (the former being an sd-attack and the latter an scd-attack); again, these situations were illustrated by arguments $\langle \mathcal{A}_5, I_5 \rangle$, $\langle \mathcal{A}_4, I_5 \rangle$ and $\langle \mathcal{A}_6, I_5 \rangle$ in Example 6.

Given that an agent may build multiple arguments, which in turn may have several counter-arguments, in order to determine the agent’s beliefs we need to determine the undefeated arguments. To estab-

lish whether an argument $\langle \mathcal{A}, L \rangle$ is undefeated, it is necessary to explicitly account for all its counter-arguments. Let $\langle \mathcal{A}_1, L_1 \rangle, \langle \mathcal{A}_2, L_2 \rangle, \dots, \langle \mathcal{A}_k, L_k \rangle$ be the counter-arguments of $\langle \mathcal{A}, L \rangle$. If any counter-argument $\langle \mathcal{A}_i, L_i \rangle$ is (according to the informant-based comparison criterion) better than or unrelated to $\langle \mathcal{A}, L \rangle$, then $\langle \mathcal{A}_i, L_i \rangle$ is a candidate for defeating $\langle \mathcal{A}, L \rangle$. However, if argument $\langle \mathcal{A}, L \rangle$ is better than $\langle \mathcal{A}_i, L_i \rangle$, then this counter-argument will not be taken in consideration as a defeater for $\langle \mathcal{A}, L \rangle$. Therefore, in order to determine the defeaters of an argument, we will make use of the informant-based comparison criterion. Then, once all the successful attackers (*i.e.*, the defeaters) of an argument are identified, we will be able to determine its acceptance status and establish whether both the argument and its conclusion are warranted or not. These issues will be addressed in the following section.

4. Defeats

Agents build arguments from their knowledge bases with the aim of establishing their beliefs. However, as discussed in Section 3, the existence of conflicting information within an agent's knowledge base leads to the existence of attacks between those arguments. Furthermore, since attacks could succeed or fail, it is necessary to have a comparison criterion to determine whether the attacking argument in a conflict prevails, in which case it becomes a defeater. In general, an attack will be considered to be effective when the set of informants from the attacked argument is not better than that of the attacking argument, with respect to an informant-based comparison criterion.

When comparing sets of informants corresponding to arguments built from an agent's IBDP, the comparison criterion " $<$ " will be the one provided in the agent's specification. In other words, the informant-based comparison criterion is modular. However, in order to be considered as valid, the criterion has to satisfy some constraints. The notion of a *valid informant-based comparison criterion* is formalized below. Following the usual convention, $IS_2 <_{\mathbb{I}} IS_1$ means that the set of informants IS_1 is strictly better than (or preferred to) the set IS_2 , and $IS_2 \not<_{\mathbb{I}} IS_1$ means that the set IS_1 is not better than (or not preferred to) the set IS_2 .

Definition 17 (Valid informant-based comparison criterion). Let \mathbb{I} be a finite set of agent identifiers and $(\mathbb{I}, P_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$. We say that $<_{\mathbb{I}}$ is a *valid informant-based comparison criterion* iff it holds that: (1) for every $IS_1, IS_2 \subseteq \mathbb{I}$, if $IS_2 <_{\mathbb{I}} IS_1$, then there exist $I_1 \in IS_1$ and $I_2 \in IS_2$ such that $I_2 <_{co}^{\mathbb{I}} I_1$ (based on $<_{co}^{\mathbb{I}}$); (2) for every $IS \subseteq \mathbb{I}$, $IS \not<_{\mathbb{I}} IS$ (irreflexive); and (3) for every $IS_1, IS_2 \subseteq \mathbb{I}$, if $IS_2 <_{\mathbb{I}} IS_1$, then $IS_1 \not<_{\mathbb{I}} IS_2$ (asymmetric).

Following the preceding definition, a valid informant-based comparison criterion should be based on the agent's credibility order, and it should be irreflexive and asymmetric. We do not intend these conditions to be the only ones that can be satisfied by an informant-based comparison criterion. In contrast, we aim at establishing the minimum requirements surrounding such criteria. Specifically, the first condition is meant to link an informant-based comparison criterion with the credibility order it is based on, so that it actually makes use of the information provided by that order and does not contradict it when considering the corner case of singleton sets. That is, a valid informant-based comparison criterion is such that, when comparing unitary sets of informants $\{I_1\}$ and $\{I_2\}$, if $\{I_2\} <_{\mathbb{I}} \{I_1\}$ then $I_2 <_{co}^{\mathbb{I}} I_1$; in other words, if a unitary set of informants is better than another unitary set of informants, then it holds that the informant in the latter set is less credible than the one in the former set (according to the credibility order). On the other hand, by requiring the valid criteria to be irreflexive and asymmetric

(second and third conditions), we are establishing the characteristic of them being *strict*.⁴ Finally, as stated above, a valid informant-based comparison criterion might impose additional constraints as long as it complies with the conditions established in Definition 17.

Next, we will introduce a valid informant-based comparison criterion that will be used in our examples for determining the successful attacks between the arguments built from an IBDP. This criterion, called *single informant credibility criterion*, is an adapted version of the *single rule criterion* from [43], modified to account for sets of informant agents. Intuitively, it prefers a set of informants IS_1 over a set of informants IS_2 if there is at least one informant in IS_1 that is more credible than an informant in IS_2 , and no informant in IS_2 is more credible than an informant in IS_1 . Formally:

Definition 18 (Single informant credibility criterion). Let \mathbb{I} be a finite set of agent identifiers, $<_{co}$ a credibility order over \mathbb{I} and $IS_1, IS_2 \subseteq \mathbb{I}$. We say that IS_1 is preferred to IS_2 , denoted $IS_2 <_S IS_1$, iff it holds that: (1) there exist $I_i \in IS_1$ and $I_j \in IS_2$ such that $I_j <_{co} I_i$; and (2) there is no $I_k \in IS_1$ and no $I_t \in IS_2$ such that $I_k <_{co} I_t$.

Proposition 1. Let \mathbb{I} be a finite set of agent identifiers and $(\mathbb{I}, P_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$, where $<_{\mathbb{I}} = <_S$ is the single informant credibility criterion. It holds that $<_S$ is a valid informant-based comparison criterion.

Proof. We have to show that for every $IS_1, IS_2 \subseteq \mathbb{I}$ it holds that:

- (1) If $IS_2 <_S IS_1$, then there exist $I_1 \in IS_1$ and $I_2 \in IS_2$ such that $I_2 <_{co}^{\mathbb{I}} I_1$ (based on $<_{co}^{\mathbb{I}}$). This follows directly from the first clause in Definition 18.
- (2) $IS_1 \not<_S IS_1$ (irreflexive). Suppose by contradiction that $IS_1 <_S IS_1$. Then, it would be the case that there exist $I_k, I_t \in IS_1$ such that $I_k <_{co}^{\mathbb{I}} I_t$, contradicting the second clause of Definition 18.
- (3) If $IS_2 <_S IS_1$, then $IS_1 \not<_S IS_2$ (asymmetric). Suppose by contradiction that $IS_1 <_S IS_2$. Then, by the second clause of Definition 18 it would be the case that there is no $I_k \in IS_2$ and no $I_t \in IS_1$ such that $I_k <_{co}^{\mathbb{I}} I_t$. However, this would violate the first clause of Definition 18, contradicting the hypothesis that $IS_2 <_S IS_1$. \square

It is worth remarking that, as specified by Definition 17, a valid informant-based comparison criterion is not required to be transitive. As discussed before, this does not imply that any valid criteria cannot be transitive; thus, transitive as well as non-transitive criteria can be considered as long as they meet the requirements imposed by Definition 17. Then, by not imposing such a constraint on valid criteria, we allow for a wider family of criteria to be considered. Among others, this allows us to consider the single informant credibility criterion which, as introduced in Definition 18, is not transitive. To illustrate the fact that this criterion does not satisfy transitivity, let us consider the following example.⁵ Consider the set of agents $\{I_1, I_2, I_3, I_4\}$ and a credibility order $<_{co}$ such that $I_1 <_{co} I_2$ and $I_3 <_{co} I_4$. By applying the single informant credibility criterion $<_S$, it holds that $\{I_1\} <_S \{I_2, I_3\}$ and $\{I_2, I_3\} <_S \{I_4\}$. However, since I_1 and I_4 are incomparable w.r.t. $<_{co}$, it does not hold that $\{I_1\} <_S \{I_4\}$ and thus, the criterion is not transitive. Notwithstanding this, the non-transitivity of the single informant credibility criterion does not undermine its usefulness. In fact, this criterion is analogous to the *rules priorities criterion* from the standard literature of structured argumentation (see e.g., [25, Def. 3.7]). On the other hand, as

⁴As will be shown later in this section, this is a key aspect in the resolution of attacks into defeats.

⁵We thank one of the reviewers for suggesting this example.

an example of a transitive and valid informant-based comparison criterion, we can consider the *max-max* lifting criterion introduced in [9] when taking a credibility order as the basis of the \succ relation (see Section 6 for a detailed discussion).

We will now turn to establish the conditions under which the attacks introduced in Section 3 succeed and become *defeats* by making use of a valid informant-based comparison criterion. As mentioned before, in general, an attack will succeed if the set of informants from the attacking argument is not worse than that of the attacked argument, with respect to an informant-based comparison criterion. Hence, for each kind of attack, we will establish the sets of informants to be accounted for by the comparison criterion. For this, we need to formally characterize the set of informants associated with an argument.

Definition 19 (Argument informants set). Let \mathbb{I} be a finite set of agent identifiers and $\langle \mathcal{A}, L \rangle$ an argument built from an IDBP. The *informants set* of $\langle \mathcal{A}, L \rangle$ is $\text{Inf}(\langle \mathcal{A}, L \rangle) = \{\mathbb{I} \in \mathbb{I} \mid \exists(\mathbb{I} : R) \in \mathcal{A}\}$.

Having established a mechanism for identifying the set of informants associated with an argument, we now turn to formalize the different kinds of defeat that may occur between a pair of arguments built from an IDBP. In the following, whenever we want to refer to a generic argument (without caring for its conclusion or the informant it supports or challenges), for convenience we will sometimes write \mathcal{A} instead of $\langle \mathcal{A}, L \rangle$ for claim arguments (respectively, \mathcal{A} instead of $\langle \mathcal{A}, \mathbb{I} \rangle$ for backing and detracting arguments); this is because an argument can be unequivocally identified by its associated set of rules. Then, for instance, $\text{Inf}(\langle \mathcal{A}, L \rangle)$ and $\text{Inf}(\langle \mathcal{A}, \mathbb{I} \rangle)$ will sometimes be written as $\text{Inf}(\mathcal{A})$.

Definition 20 (Conclusion defeat). Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, P_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$, and $\mathcal{A}_1, \mathcal{A}_2$ two arguments built from $P_{\mathbb{I}}$ such that \mathcal{A}_1 c-attacks \mathcal{A}_2 at the literal L . We say that \mathcal{A}_1 *c-defeats* \mathcal{A}_2 (equivalently, \mathcal{A}_1 is a c-defeater of \mathcal{A}_2) iff the disagreement sub-argument $\langle \mathcal{A}, L \rangle$ of \mathcal{A}_2 is such that $\text{Inf}(\mathcal{A}_1) \not\prec_{\mathbb{I}} \text{Inf}(\mathcal{A})$. In particular, if $\text{Inf}(\mathcal{A}) <_{\mathbb{I}} \text{Inf}(\mathcal{A}_1)$, we say that \mathcal{A}_1 is a *proper c-defeater* of \mathcal{A}_2 ; otherwise, we say that \mathcal{A}_1 is a *blocking c-defeater* of \mathcal{A}_2 .

Definition 21 (Strength defeat). Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, P_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$, and $\mathcal{A}_1, \mathcal{A}_2$ two arguments built from $P_{\mathbb{I}}$ such that \mathcal{A}_1 s-attacks \mathcal{A}_2 . We say that \mathcal{A}_1 *s-defeats* \mathcal{A}_2 (equivalently, \mathcal{A}_1 is an s-defeater of \mathcal{A}_2) and, in particular, \mathcal{A}_1 is a *proper s-defeater* of \mathcal{A}_2 .

Definition 22 (Strength-defense defeat). Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, P_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$, and $\mathcal{A}_1, \mathcal{A}_2$ two arguments built from $P_{\mathbb{I}}$ such that \mathcal{A}_1 sd-attacks \mathcal{A}_2 . We say that \mathcal{A}_1 *sd-defeats* \mathcal{A}_2 (equivalently, \mathcal{A}_1 is an sd-defeater of \mathcal{A}_2) iff $\text{Inf}(\mathcal{A}_1) \not\prec_{\mathbb{I}} \text{Inf}(\mathcal{A}_2)$. In particular, if $\text{Inf}(\mathcal{A}_2) <_{\mathbb{I}} \text{Inf}(\mathcal{A}_1)$, we say that \mathcal{A}_1 is a *proper sd-defeater* of \mathcal{A}_2 ; otherwise, we say that \mathcal{A}_1 is a *blocking sd-defeater* of \mathcal{A}_2 .

Definition 23 (Strength-counter-defense defeat). Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, P_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$, and $\mathcal{A}_1, \mathcal{A}_2$ two arguments built from $P_{\mathbb{I}}$ such that \mathcal{A}_1 scd-attacks \mathcal{A}_2 . We say that \mathcal{A}_1 *scd-defeats* \mathcal{A}_2 (equivalently, \mathcal{A}_1 is an scd-defeater of \mathcal{A}_2) iff $\text{Inf}(\mathcal{A}_1) \not\prec_{\mathbb{I}} \text{Inf}(\mathcal{A}_2)$. In particular, if $\text{Inf}(\mathcal{A}_2) <_{\mathbb{I}} \text{Inf}(\mathcal{A}_1)$, we say that \mathcal{A}_1 is a *proper scd-defeater* of \mathcal{A}_2 ; otherwise, we say that \mathcal{A}_1 is a *blocking scd-defeater* of \mathcal{A}_2 .

There exist some differences in the way in which the different types of attack are resolved. Specifically, the differences rely on the sets of informants that are compared in each case. For the resolution of c-attacks, it suffices to compare the set of informants from the attacking argument with the set of informants from the disagreement sub-argument. Note that the resolution of c-attacks in this way is

analogous to the standard resolution of rebutting attacks in the literature of structured argumentation (see *e.g.*, [25,36]). When resolving s-attacks, it is important to recall the nature of the attacking argument, which provides reasons against the consideration of a given informant used by the attacked argument. To prevent the success of an s-attack, the attacked argument should somehow give reasons for the consideration of that informant, and they should be provided by informants that are not worse than the ones associated with the attacking argument. As a result, since the attacked argument does not give reasons for the consideration of the challenged informant, the s-attack will always succeed (*i.e.*, no comparison between sets of informants is made). Again, the resolution of s-defeats relates to the way in which undercutting attacks are resolved in the literature.

Then, sd-attacks and scd-attacks are handled analogously. In these two cases, there exists a conflict between a backing argument and a detracting argument, where they respectively provide reasons for and against the consideration of a given informant. Therefore, in the resolution of these types of attack, the entire sets of informants associated with the attacking and the attacked argument are compared. Given the resemblance between backing and detracting arguments as proposed here and the backing and undercutting arguments of [19], it can be noted that sd-defeats and scd-defeats somehow relate to the implicit defeats of [19], and are resolved analogously; notwithstanding this, there is a clear difference between them since our approach compares the sets of informant agents of the backing and detracting arguments whereas [19] makes use of a preference relation which might not take into account the information sources of arguments.

Finally note that, in all cases where sets of informants are compared to determine the success of an attack (thus turning it into a defeat), the traditional form of resolution of [3] is adopted. Namely, in our approach, an attack succeeds if the set of informants from the attacked argument is not better (w.r.t. the informant-based comparison criterion) than the set of informants from the attacking argument. In contrast, works like [6,31] consider that preferences play an additional role, serving to repair the attack relation in order to account for conflicts derived from the preferences, even in cases where those conflicts might not have been originally expressed within the attack relation. Among others, they both consider the existence of a defeat from an argument \mathcal{A}_1 to an argument \mathcal{A}_2 in a scenario where \mathcal{A}_2 attacks \mathcal{A}_1 (and not vice-versa) but \mathcal{A}_1 is strictly preferred to \mathcal{A}_2 .

Example 7. Let us consider $(\mathbb{I}_6, \mathbb{P}_{\mathbb{I}_6}, <_{co}^{\mathbb{I}_6}, <_{\mathbb{I}_6})$, the specification of agent \mathbb{I}_6 given in Example 3, and the arguments listed in Example 5. Also, suppose that $<_{\mathbb{I}_6}$ is the single informant credibility criterion from Definition 18. Then, we identify the following defeats (among others): $\langle \mathcal{A}_3, \sim a \rangle$ c-defeats $\langle \mathcal{A}_1, h \rangle$ and $\langle \mathcal{A}_2, a \rangle$. In both cases, the disagreement sub-argument $\langle \mathcal{A}_2, a \rangle$ is such that $\text{Inf}(\langle \mathcal{A}_2, a \rangle) = \{\mathbb{I}_2, \mathbb{I}_3\}$. Then, since $\text{Inf}(\langle \mathcal{A}_3, \sim a \rangle) = \{\mathbb{I}_4, \mathbb{I}_5, \mathbb{I}_6\}$, $\text{Inf}(\langle \mathcal{A}_3, \sim a \rangle) \not<_{\mathbb{I}_6} \text{Inf}(\langle \mathcal{A}_2, a \rangle)$. In particular, $\text{Inf}(\langle \mathcal{A}_2, a \rangle) <_{\mathbb{I}_6} \text{Inf}(\langle \mathcal{A}_3, \sim a \rangle)$, so $\langle \mathcal{A}_3, \sim a \rangle$ is a proper c-defeater of $\langle \mathcal{A}_1, h \rangle$ and $\langle \mathcal{A}_2, a \rangle$. Continuing with the resolution of attacks from Example 6, the detracting arguments $\langle \mathcal{A}_4, \mathbb{I}_5 \rangle$ and $\langle \mathcal{A}_6, \mathbb{I}_5 \rangle$ proper s-defeat $\langle \mathcal{A}_3, \sim a \rangle$. Similarly, the detracting argument $\langle \mathcal{A}_7, \mathbb{I}_6 \rangle$ is a proper s-defeater of $\langle \mathcal{A}_3, \sim a \rangle$. On the other hand, the backing argument $\langle \mathcal{A}_5, \mathbb{I}_5 \rangle$ proper sd-defeats $\langle \mathcal{A}_4, \mathbb{I}_5 \rangle$ since $\text{Inf}(\langle \mathcal{A}_5, \mathbb{I}_5 \rangle) = \{\mathbb{I}_4\}$, $\text{Inf}(\langle \mathcal{A}_4, \mathbb{I}_5 \rangle) = \{\mathbb{I}_2, \mathbb{I}_3\}$ and $\text{Inf}(\langle \mathcal{A}_4, \mathbb{I}_5 \rangle) <_{\mathbb{I}_6} \text{Inf}(\langle \mathcal{A}_5, \mathbb{I}_5 \rangle)$. Finally, the detracting argument $\langle \mathcal{A}_6, \mathbb{I}_5 \rangle$ is a proper scd-defeater of $\langle \mathcal{A}_5, \mathbb{I}_5 \rangle$; this is because $\text{Inf}(\langle \mathcal{A}_6, \mathbb{I}_5 \rangle) = \{\mathbb{I}_7\}$, $\text{Inf}(\langle \mathcal{A}_5, \mathbb{I}_5 \rangle) = \{\mathbb{I}_4\}$ and $\text{Inf}(\langle \mathcal{A}_5, \mathbb{I}_5 \rangle) <_{\mathbb{I}_6} \text{Inf}(\langle \mathcal{A}_6, \mathbb{I}_5 \rangle)$.

The defeats are illustrated in Fig. 3. The notation for the arguments is the same as the one used in Fig. 2; on the other hand, defeats are represented with solid arrows between the arguments.

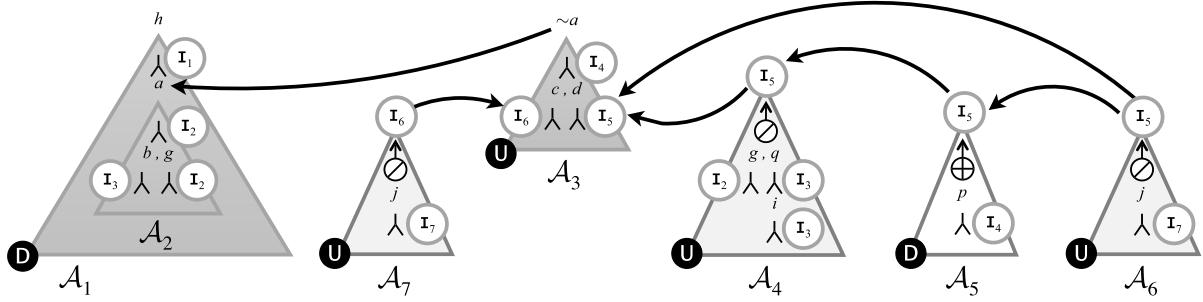


Fig. 3. Arguments and defeats from Example 7.

5. Warrant

To determine whether an agent \mathbb{I} can accept a literal L as a belief it is necessary to find out if, from its IBDP $P_{\mathbb{I}}$, it is possible to build an argument $\langle \mathcal{A}, L \rangle$ that ends up undefeated after all things considered. Naturally, this will require the consideration of all possible defeaters for $\langle \mathcal{A}, L \rangle$. Then, given a defeater $\langle \mathcal{B}, H \rangle$, defeaters for it will also have to be considered, as well as the defeaters for those defeaters, and so on. This dialectical analysis leads to characterizing the notion of an *argumentation line*, which constitutes a sequence of arguments where each argument is a defeater of its predecessor.

Definition 24 (Argumentation line). Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, P_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$, and \mathcal{A}_0 an argument built from $P_{\mathbb{I}}$. An *argumentation line* for \mathcal{A}_0 is a sequence of arguments built from $P_{\mathbb{I}}$, denoted $\Lambda = [\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n]$, where each element of the sequence \mathcal{A}_i is a defeater of its predecessor \mathcal{A}_{i-1} ($0 < i \leq n$).

Given an argumentation line Λ for an argument \mathcal{A}_0 , every argument in an odd position (e.g., $\mathcal{A}_2, \mathcal{A}_4$, if they exist) is said to be a *supporting argument* for \mathcal{A}_0 because they reinstate \mathcal{A}_0 ; the set of such arguments will be denoted as Λ_S . Analogously, arguments in even positions of Λ are called *interfering arguments*, and the corresponding set will be denoted as Λ_I .

Example 8. Given the specification of agent \mathbb{I}_6 introduced in Example 3, the arguments built from its IBDP $P_{\mathbb{I}_6}$ (Example 5) and the defeats between them listed in Example 7, we obtain (among others) the following argumentation lines for $\langle \mathcal{A}_1, h \rangle$: $\Lambda_1 = [\langle \mathcal{A}_1, h \rangle, \langle \mathcal{A}_3, \sim a \rangle, \langle \mathcal{A}_4, \mathbb{I}_5 \rangle, \langle \mathcal{A}_5, \mathbb{I}_5 \rangle, \langle \mathcal{A}_6, \mathbb{I}_5 \rangle]$, whose set of supporting arguments is $\Lambda_{S_1} = \{\langle \mathcal{A}_1, h \rangle, \langle \mathcal{A}_4, \mathbb{I}_5 \rangle, \langle \mathcal{A}_6, \mathbb{I}_5 \rangle\}$ and its set of interfering arguments is $\Lambda_{I_1} = \{\langle \mathcal{A}_3, \sim a \rangle, \langle \mathcal{A}_5, \mathbb{I}_5 \rangle\}$; $\Lambda_2 = [\langle \mathcal{A}_1, h \rangle, \langle \mathcal{A}_3, \sim a \rangle, \langle \mathcal{A}_7, \mathbb{I}_6 \rangle]$, with $\Lambda_{S_2} = \{\langle \mathcal{A}_1, h \rangle, \langle \mathcal{A}_7, \mathbb{I}_6 \rangle\}$ and $\Lambda_{I_2} = \{\langle \mathcal{A}_3, \sim a \rangle\}$; and $\Lambda_3 = [\langle \mathcal{A}_1, h \rangle, \langle \mathcal{A}_3, \sim a \rangle, \langle \mathcal{A}_6, \mathbb{I}_5 \rangle]$, whose set of supporting arguments is $\Lambda_{S_3} = \{\langle \mathcal{A}_1, h \rangle, \langle \mathcal{A}_6, \mathbb{I}_5 \rangle\}$ and its set of interfering arguments is $\Lambda_{I_3} = \{\langle \mathcal{A}_3, \sim a \rangle\}$.

There may exist argumentation lines that lead to fallacious chains of reasoning. In order to avoid those, we impose some restrictions on argumentation lines, to distinguish the *acceptable argumentation lines*. The first situation we want to avoid is to have infinite chains of reasoning; this is captured by the first and third clauses in Definition 26, by requiring an acceptable argumentation line to be a finite sequence, and to avoid the introduction of repeated arguments and disagreement sub-arguments. Another constraint imposed on acceptable argumentation lines is a kind of consistency within the sets of supporting and interfering arguments, in order to prevent an argument from being defended by another argument that is in conflict with it; this intuition is modeled in the second clause of Definition 26, requiring the sets of supporting and interfering arguments of an argumentation line to be non-contradictory.

Let us now analyze the inclusion of blocking defeaters in an argumentation line. The existence of a blocking defeat from \mathcal{A}_i to \mathcal{A}_{i-1} in an argumentation line implies that the sets of informants of the two arguments were incomparable under the adopted comparison criterion.⁶ In addition, if there exists an argument \mathcal{A}_{i+1} that is a blocking defeater of \mathcal{A}_i , it would imply that the comparison criterion could not resolve that conflict either. As a result, if \mathcal{A}_{i+1} were to be included in the corresponding argumentation line, it would somehow imply that \mathcal{A}_{i-1} prevails over \mathcal{A}_i , just because it has argument \mathcal{A}_{i+1} supporting it (which, in turn, was not better than \mathcal{A}_i). To avoid these issues, consecutive blocking defeaters are not allowed in an acceptable argumentation line, and this is captured in the fourth clause of Definition 26.

Next, let us move to considering the inclusion of s-defeats in an argumentation line. As expressed before, s-defeats (corresponding to s-attacks, which always succeed), are aimed at challenging the strength of the defeated argument by targeting one of its informant agents. This is because, as discussed earlier, we consider that the strength of an argument is linked to the credibility of its informant agents. However, it should be noted that the existence of an s-defeat towards an argument is not, on its own, sufficient to establish that the defeat actually took place because of challenging the defeated argument's strength. This is due to the fact that not every informant providing information used for building an argument is relevant for determining its strength. For instance, as expressed in Definition 20, the strength of a claim argument \mathcal{A} that c-attacks an argument \mathcal{B} , with \mathcal{C} being the disagreement sub-argument, will be determined by a subset of its informants $\text{IS}_{\mathcal{A}} \subseteq \text{Inf}(\mathcal{A})$ such that $\text{IS}_{\mathcal{A}} \not\prec \text{Inf}(\mathcal{C})$, where $\text{Inf}(\mathcal{C})$ is the entire set of informants of argument \mathcal{C} ; specifically, as shown by Definition 17, some informants might not be accounted for when comparing the two arguments. Consequently, the strength of an argument is determined in the context of the resolution of an attack and, as a result, we consider that the strength of an argument (possibly established by alternative subsets of its informants) is not absolute; rather, it is relative to the resolution of each attack it is involved in, as different sets of agents may be considered in each case to determine the argument's strength.

To capture these intuitions, we next introduce the notion of *strength-determining set of informants* of an argument in an argumentation line, which corresponds to a set of informants of the argument that provides the necessary strength for its inclusion as a defeater of its predecessor in the line. In particular, as will be shown later in Definition 26, this notion will serve to identify s-defeaters that indeed defeat an argument by challenging its strength, because they provide reasons against one of its strength-determining informants. For instance, consider the argumentation line $\Lambda'_1 = [\langle \mathcal{A}_1, h \rangle, \langle \mathcal{A}_3, \sim a \rangle]$ which is a sub-sequence of the line Λ_1 (also, of the line Λ_3) from Example 8, where $\langle \mathcal{A}_3, \sim a \rangle$ proper c-defeats argument $\langle \mathcal{A}_1, h \rangle$; in this case, the strength-determining sets of informants of $\langle \mathcal{A}_3, \sim a \rangle$ would be $\{\mathbb{I}_4\}$ and $\{\mathbb{I}_5\}$, since the informants in these sets are individually better than every informant used by the disagreement sub-argument $\langle \mathcal{A}_2, a \rangle$. Furthermore, since the strength-determining set of informants of an argument may not be unique, the following definition allows to obtain every set of informants meeting that requirement:

Definition 25 (Strength-determining sets of informants). Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, \text{P}_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$, \mathcal{A} an argument built from $\text{P}_{\mathbb{I}}$ and Λ an argumentation line ending in \mathcal{A} . We define the *strength-determining sets of informants of \mathcal{A} in Λ* as the elements in the set returned by the function StrDetInf defined in Fig. 4.

⁶Note that the two sets of informants cannot be equally preferred since valid informant-based comparison criteria are required to be asymmetric, in line with the credibility order they are based.

$$\text{StrDetInf}(\prec_I, \mathcal{A}, \Lambda) = \left\{ \begin{array}{ll} \{\text{Inf}(\mathcal{A})\} & \Lambda = [\mathcal{A}]; \text{ or} \\ & \Lambda = [\mathcal{A}_1, \dots, \mathcal{A}_n, \mathcal{A}] \ (n \geq 1) \\ & \text{and } \mathcal{A} \text{ proper s-defeats } \mathcal{A}_n. \\ \{\text{IS}_{\mathcal{A}} \subseteq \text{Inf}(\mathcal{A}) \mid \text{Inf}(\mathcal{B}) \prec_I \text{IS}_{\mathcal{A}}, \\ \quad \nexists \text{IS}'_{\mathcal{A}} \subsetneq \text{IS}_{\mathcal{A}} \text{ s.t. } \text{Inf}(\mathcal{B}) \prec_I \text{IS}'_{\mathcal{A}}\} & \Lambda = [\mathcal{A}_1, \dots, \mathcal{A}_n, \mathcal{A}] \ (n \geq 1), \\ & \mathcal{A} \text{ proper c-defeats } \mathcal{A}_n \\ & \text{and } \mathcal{B} \text{ is the disagreement} \\ & \text{sub-argument of } \mathcal{A}_n. \\ \{\text{IS}_{\mathcal{A}} \subseteq \text{Inf}(\mathcal{A}) \mid \text{IS}_{\mathcal{A}} \not\prec_I \text{Inf}(\mathcal{B}), \\ \quad \nexists \text{IS}'_{\mathcal{A}} \subsetneq \text{IS}_{\mathcal{A}} \text{ s.t. } \text{IS}'_{\mathcal{A}} \not\prec_I \text{Inf}(\mathcal{B})\} & \Lambda = [\mathcal{A}_1, \dots, \mathcal{A}_n, \mathcal{A}] \ (n \geq 1), \\ & \mathcal{A} \text{ blocking c-defeats } \mathcal{A}_n \\ & \text{and } \mathcal{B} \text{ is the disagreement} \\ & \text{sub-argument of } \mathcal{A}_n. \\ \{\text{IS}_{\mathcal{A}} \subseteq \text{Inf}(\mathcal{A}) \mid \text{Inf}(\mathcal{A}_n) \prec_I \text{IS}_{\mathcal{A}}, \\ \quad \nexists \text{IS}'_{\mathcal{A}} \subsetneq \text{IS}_{\mathcal{A}} \text{ s.t. } \text{Inf}(\mathcal{A}_n) \prec_I \text{IS}'_{\mathcal{A}}\} & \Lambda = [\mathcal{A}_1, \dots, \mathcal{A}_n, \mathcal{A}] \ (n \geq 1) \\ & \text{and } \mathcal{A} \text{ proper sd-defeats or} \\ & \text{proper scd-defeats } \mathcal{A}_n. \\ \{\text{IS}_{\mathcal{A}} \subseteq \text{Inf}(\mathcal{A}) \mid \text{IS}_{\mathcal{A}} \not\prec_I \text{Inf}(\mathcal{A}_n), \\ \quad \nexists \text{IS}'_{\mathcal{A}} \subsetneq \text{IS}_{\mathcal{A}} \text{ s.t. } \text{IS}'_{\mathcal{A}} \not\prec_I \text{Inf}(\mathcal{A}_n)\} & \Lambda = [\mathcal{A}_1, \dots, \mathcal{A}_n, \mathcal{A}] \ (n \geq 1) \\ & \text{and } \mathcal{A} \text{ blocking sd-defeats} \\ & \text{or blocking scd-defeats } \mathcal{A}_n. \end{array} \right.$$

Fig. 4. Function StrDetInf characterizing the strength-determining sets of informants.

Note that, for the first argument in an argumentation line, the StrDetInf function returns a unitary set containing the entire set of informant agents from the argument; this is so because the argument was not introduced in the line for being a defeater of its predecessor. Hence, since its strength was not accounted for in the resolution of a previous attack, we deem all its informants as able to determine its strength. The same situation occurs when the argument under consideration is a detracting argument that originates an s-defeat; in such a case, we consider its entire set of informants to be the only strength-determining set because no comparison is made for the resolution of the s-attack into an s-defeat. In every other case, each set of informants that is able to make the argument defeat its predecessor (through a proper or a blocking defeat, respectively) will be a strength-determining set of informants.

Taking the notion of strength-determining set of informants into account, we will only consider the inclusion of an s-defeater in an acceptable argumentation line in cases where the detracting argument targets an informant belonging to (at least) one set of strength-determining informants of its predecessor argument in the line; this intuition is captured in the last clause of Definition 26. Finally, regarding sd-defeaters and scd-defeaters, if they appear in an argumentation line it is as a consequence of the existence of a previous s-defeater. Therefore, no additional considerations have to be taken into account for their

inclusion in an acceptable argumentation line. As a result, the notion of *acceptable argumentation line* is formalized below:

Definition 26 (Acceptable argumentation line). Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, \mathbf{P}_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$, \mathcal{A}_0 an argument built from $\mathbf{P}_{\mathbb{I}}$ and $\Lambda = [\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n]$ an argumentation line for \mathcal{A}_0 . We say that Λ is an *acceptable argumentation line* iff it holds that:

- (1) Λ is a finite sequence.
- (2) The set Λ_S of supporting arguments and the set Λ_I of interfering arguments are non-contradictory.
- (3) No argument \mathcal{A}_i in Λ is an argument appearing earlier in Λ or a disagreement sub-argument of an argument \mathcal{A}_j appearing earlier in Λ ($j < i$).
- (4) For every argument \mathcal{A}_i of Λ such that \mathcal{A}_i is a blocking defeater of \mathcal{A}_{i-1} , if there exists \mathcal{A}_{i+1} in Λ , then \mathcal{A}_{i+1} is a proper defeater of \mathcal{A}_i .
- (5) For every detracting argument \mathcal{A}_i of Λ such that \mathcal{A}_i is a proper s-defeater of \mathcal{A}_{i-1} challenging the informant \mathbb{I}_i (hence, $\mathbb{I}_i \in \text{Inf}(\mathcal{A}_{i-1})$), there exists $\text{IS}_{\mathcal{A}_{i-1}} \in \text{StrDetInf}(<_{\mathbb{I}}, \mathcal{A}_{i-1}, \Lambda')$ such that $\mathbb{I}_i \in \text{IS}_{\mathcal{A}_{i-1}}$, where Λ' is the sub-sequence of Λ ending in \mathcal{A}_{i-1} .

Given the argumentation lines Λ_1 , Λ_2 and Λ_3 from Example 8, Λ_1 and Λ_3 are acceptable argumentation lines whereas Λ_2 is not because it does not satisfy condition 5 of Definition 26 (\mathbb{I}_6 does not belong to any of the strength-determining sets of informants of \mathcal{A}_3 in $\Lambda'_1 = [\langle \mathcal{A}_1, h \rangle, \langle \mathcal{A}_3, \sim a \rangle]$). The set of all acceptable argumentation lines for a given argument \mathcal{A} are then gathered to conform a tree structure containing all possible lines of reasoning starting with \mathcal{A} . Next, we formally define this structure by introducing the notion of *dialectical tree*.

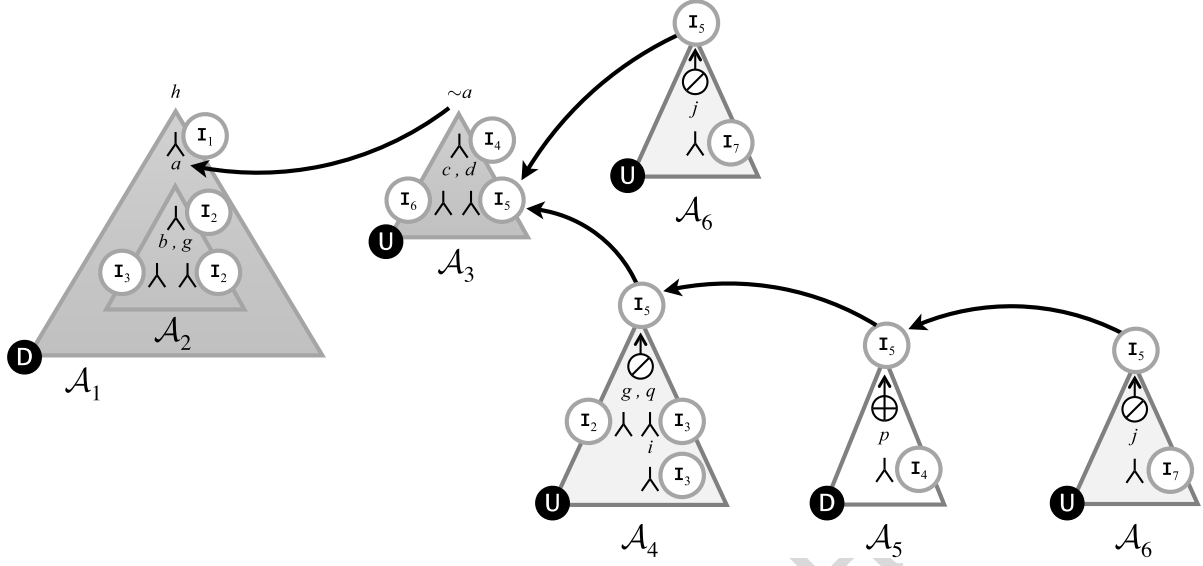
Definition 27 (Dialectical tree). Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, \mathbf{P}_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$ and \mathcal{A}_0 an argument built from $\mathbf{P}_{\mathbb{I}}$. A *dialectical tree for \mathcal{A}_0* , denoted $\mathcal{T}_{\mathcal{A}_0}$, is a tree such that its nodes are arguments built from $\mathbf{P}_{\mathbb{I}}$ and the edges denote defeats between those arguments, and the following conditions hold:

- (1) \mathcal{A}_0 is the root of $\mathcal{T}_{\mathcal{A}_0}$.
- (2) If $\Lambda = [\mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_n]$ is an acceptable argumentation line for \mathcal{A}_0 and there is no argument \mathcal{A}_m built from $\mathbf{P}_{\mathbb{I}}$ such that $\Lambda' = [\mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_n, \mathcal{A}_m]$ is an acceptable argumentation line for \mathcal{A}_0 , then \mathcal{A}_n is a leaf in $\mathcal{T}_{\mathcal{A}_0}$ and Λ is the branch of $\mathcal{T}_{\mathcal{A}_0}$ starting from the root \mathcal{A}_0 down to the leaf \mathcal{A}_n .

Once built, the dialectical tree is marked in order to determine the acceptance status of the root argument. The marking criterion will mark the nodes in the tree as *undefeated* (**U**) or as *defeated* (**D**). Briefly, leafs of the tree are **U**, while inner nodes are **D** if they are defeated by a **U** claim argument, a **U** backing argument or a **U** detracting argument originating a scd-defeat, or if they are defeated by **U** detracting arguments originating s-defeats that target an informant in each of its strength-determining sets of informants.

Definition 28 (Marking criterion). Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, \mathbf{P}_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$, \mathcal{A}_0 an argument built from $\mathbf{P}_{\mathbb{I}}$ and $\mathcal{T}_{\mathcal{A}_0}$ the dialectical tree for \mathcal{A}_0 . If N is a node in $\mathcal{T}_{\mathcal{A}_0}$, it will be marked as:

- **D** if N is the root of $\mathcal{T}_{\mathcal{A}_0}$ or it is an inner node corresponding to an argument \mathcal{A} , and either:
 - N has a child marked **U** corresponding to a c-defeater, an sd-defeater or an scd-defeater of \mathcal{A} ; or

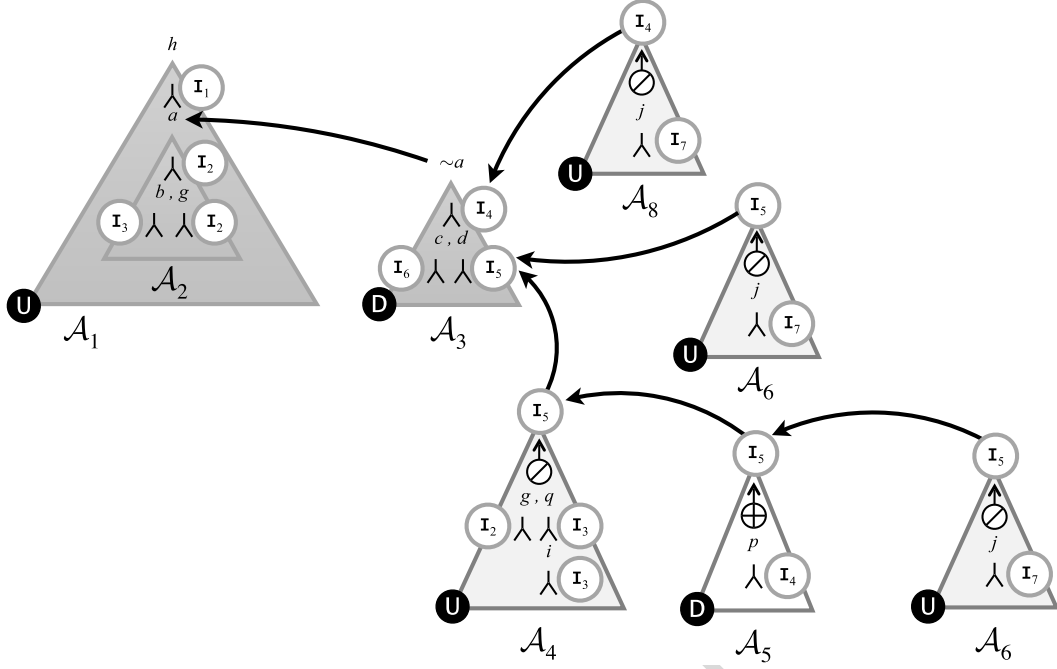
Fig. 5. Marked dialectical tree $\mathcal{T}_{(\mathcal{A}_1, h)}$ from Example 9.

- for each $\text{IS}_{\mathcal{A}} \in \text{StrDetInf}(\prec_{\mathcal{I}}, \mathcal{A}, \Lambda')$, N has a child marked **U** that corresponds to an s-defeater $(\mathcal{B}, \mathcal{I})$ of \mathcal{A} such that $\mathcal{I} \in \text{IS}_{\mathcal{A}}$, where Λ' is the argumentation line corresponding to the arguments in the branch starting from the root down to the node N .
- **U** otherwise.

Example 9. Consider the specification of agent \mathcal{I}_6 introduced in Example 3 and the arguments listed in Example 5. Figure 5 depicts the marked dialectical tree for argument (\mathcal{A}_1, h) .⁷ This tree has two acceptable argumentation lines, namely Λ_1 and Λ_3 from Example 8. The notation for arguments and defeats in the figure is the same as in Fig. 3. Also, each argument is marked in its left vertex as defeated or undefeated, with a circle containing a **D** or a **U**, respectively. Observe that the nodes corresponding to \mathcal{A}_6 are marked **U** (they are leaves). Then, the node corresponding to \mathcal{A}_5 is marked **D** because it has a child marked **U** corresponding to an scd-defeater. Consequently, since \mathcal{A}_5 is marked **D**, \mathcal{A}_4 is marked **U**. However, note that \mathcal{A}_3 is marked **U** even though it has two children corresponding to s-defeaters $(\langle \mathcal{A}_4, \mathcal{I}_5 \rangle)$ and $(\langle \mathcal{A}_6, \mathcal{I}_5 \rangle)$ that are marked **U**. This is because there is no s-defeater for \mathcal{A}_3 in $\mathcal{T}_{(\mathcal{A}_1, h)}$ targeting the informant \mathcal{I}_4 (recall that $\text{StrDetInf}(\prec_{\mathcal{I}}, \mathcal{A}_3, \Lambda'_1) = \{\{\mathcal{I}_4\}, \{\mathcal{I}_5\}\}$, with $\Lambda'_1 = [\langle \mathcal{A}_1, h \rangle, \langle \mathcal{A}_3, \sim a \rangle]$). Finally, since \mathcal{A}_3 is marked **U**, the root of $\mathcal{T}_{(\mathcal{A}_1, h)}$ is marked **D**.

A marked dialectical tree embodies a dialectical analysis considering every possible argument an agent can build for and against the root argument in the tree. Hence, if the root argument is marked as **U**, it means that the conclusion of that argument is warranted, and the agent can accept it as a belief. Moreover, the existence of *one* argument (\mathcal{A}, L) marked as **U** in its dialectical tree $\mathcal{T}_{(\mathcal{A}, L)}$ is sufficient for the agent to accept L as a belief. On the contrary, if every argument for L is marked as **D** in its own tree, then the literal L will not be warranted and thus, the agent will not accept it as a belief. The notions of *warranted argument* and *warranted literal* are formalized next:

⁷Note that, to save space, the tree is depicted horizontally.

Fig. 6. Marked dialectical tree $\mathcal{T}_{(\mathcal{A}_1, h)}$ from Example 11.

Definition 29 (Warranted claim argument – warranted literal). Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, P_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$ and $\langle \mathcal{A}, L \rangle$ a claim argument built from $P_{\mathbb{I}}$. We say that \mathbb{I} warrants the argument $\langle \mathcal{A}, L \rangle$ and the literal L iff $\langle \mathcal{A}, L \rangle$ is marked as **U** in the dialectical tree $\mathcal{T}_{(\mathcal{A}, L)}$.

Example 10. Consider agent \mathbb{I}_6 specified in Example 3 and the marked dialectical tree from Example 9. Agent \mathbb{I}_6 does not warrant the argument $\langle \mathcal{A}_1, h \rangle$ since, as depicted in Fig. 5, the root of $\mathcal{T}_{(\mathcal{A}_1, h)}$ is marked as **D**. Consequently, since there are no other claim arguments for “ h ”, agent \mathbb{I}_6 does not warrant the literal “ h ” (hence, it will not accept it as a belief).

Example 11. Suppose the detracting rule $(\mathbb{I}_4 \leftarrow \ominus j)$ is added to the IBDP $P_{\mathbb{I}_6}$ of agent \mathbb{I}_6 , leading to the new agent’s specification $(\mathbb{I}_6, (\Delta_{\mathbb{I}_6}, \Sigma_{\mathbb{I}_6} \cup \{(\mathbb{I}_4 \leftarrow \ominus j)\}), <_{co}^{\mathbb{I}_6}, <_{\mathbb{I}_6})$. In this scenario, the dialectical tree $\mathcal{T}_{(\mathcal{A}_1, h)}$ will be the one illustrated in Fig. 6. Differently from Example 9, $\langle \mathcal{A}_3, \sim a \rangle$ is now marked **D** because for each $\text{IS}_{\mathcal{A}_3} \in \text{StrDetInf}(<_{\mathbb{I}_6}, \langle \mathcal{A}_3, \sim a \rangle, \Lambda') = \{\{\mathbb{I}_4\}, \{\mathbb{I}_5\}\}$, $\langle \mathcal{A}_3, \sim a \rangle$ has a child marked **U** that corresponds to an s-defeater targeting an informant in $\text{IS}_{\mathcal{A}_3}$; specifically, the detracting arguments $\langle \mathcal{A}_4, \mathbb{I}_5 \rangle$ and $\langle \mathcal{A}_6, \mathbb{I}_5 \rangle$ are s-defeaters targeting \mathbb{I}_5 , whereas the detracting argument $\langle \mathcal{A}_8, \mathbb{I}_4 \rangle$ (with $\mathcal{A}_8 = \{(\mathbb{I}_4 \leftarrow \ominus j), (\mathbb{I}_7 : j \dashv)\}$) is an s-defeater targeting \mathbb{I}_4 . As a result, the root of $\mathcal{T}_{(\mathcal{A}_1, h)}$ is now marked **U** and thus, agent \mathbb{I}_6 warrants the argument $\langle \mathcal{A}_1, h \rangle$ and the literal “ h ”, accepting the latter as a belief.

As introduced above, the beliefs an agent has are determined by the set of literals it warrants. Given this, it is to be expected that the set of warranted beliefs from an agent is consistent. In the remainder of this section we will formally show that an agent does not warrant complementary literals, as expressed by Theorem 1. For this purpose, we will first show some intermediate results, given by Lemmas 1 and 2.

Lemma 1. *Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, \mathbf{P}_{\mathbb{I}}, <_{co}^{\mathbb{I}}, <_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$, and $\langle \mathcal{A}, L \rangle$ an argument built from $\mathbf{P}_{\mathbb{I}}$. If \mathbb{I} warrants $\langle \mathcal{A}, L \rangle$, then for every claim argument $\langle \mathcal{B}, H \rangle$ built from $\mathbf{P}_{\mathbb{I}}$ such that $\langle \mathcal{A}, L \rangle$ c-defeats $\langle \mathcal{B}, H \rangle$ it holds that the root node of $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$ has a child corresponding to argument $\langle \mathcal{A}, L \rangle$ that is marked **U**.*

Proof. Since $\langle \mathcal{A}, L \rangle$ c-defeats $\langle \mathcal{B}, H \rangle$, by Definitions 24, 26 and 27 it follows directly that there exists a child N of the root of $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$ corresponding to $\langle \mathcal{A}, L \rangle$. Now we have to prove that N is marked as **U** in $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$. Given that by hypothesis \mathbb{I} warrants $\langle \mathcal{A}, L \rangle$, by Definition 29 the root of the dialectical tree $\mathcal{T}_{\langle \mathcal{A}, L \rangle}$ is marked as **U**. By the characterization of dialectical trees, a part of $\mathcal{T}_{\langle \mathcal{A}, L \rangle}$ can appear as a sub-tree of $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$, rooted in the node N . Then we have the following cases:

- $\mathcal{T}_{\langle \mathcal{A}, L \rangle}$ (the complete tree) is a sub-tree of $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$, rooted in N . In this case, by Definition 28, the node N will be marked as **U** in $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$.
- A part of $\mathcal{T}_{\langle \mathcal{A}, L \rangle}$ (not the complete tree) is a sub-tree of $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$, rooted in N . In this case, there exists a node S in $\mathcal{T}_{\langle \mathcal{A}, L \rangle}$ that is not a node in the sub-tree of $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$ rooted in N . If the node S corresponds to an interfering argument in $\mathcal{T}_{\langle \mathcal{A}, L \rangle}$, then by Definition 28 it holds that not having S in $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$ does not change the marking of N , as it was marked as **U** in $\mathcal{T}_{\langle \mathcal{A}, L \rangle}$. Consequently, N will be marked as **U** in $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$. Let us now suppose that the node S corresponds to a supporting argument $\langle \mathcal{C}, X \rangle$ in $\mathcal{T}_{\langle \mathcal{A}, L \rangle}$. Therefore, there exists an acceptable argumentation line $\Lambda = [\langle \mathcal{A}, L \rangle, \dots, \langle \mathcal{D}, Y \rangle, \langle \mathcal{C}, X \rangle, \dots]$ corresponding to a branch of $\mathcal{T}_{\langle \mathcal{A}, L \rangle}$. Then, if the node S corresponding to $\langle \mathcal{C}, X \rangle$ does not appear in the sub-tree of $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$ rooted in N and the root of $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$ has a child corresponding to $\langle \mathcal{A}, L \rangle$, it should be the case that the argumentation line $\Lambda' = [\langle \mathcal{B}, H \rangle, \langle \mathcal{A}, L \rangle, \dots, \langle \mathcal{D}, Y \rangle, \langle \mathcal{C}, X \rangle]$ (i.e., the argumentation line starting with $\langle \mathcal{B}, H \rangle$ and continuing with the sub-sequence of Λ ending in $\langle \mathcal{C}, X \rangle$) is not acceptable because the inclusion of $\langle \mathcal{C}, X \rangle$ violates some condition of Definition 26. Hence, it must be the case that either:
 - The set of interfering arguments of Λ' (i.e., the set containing $\langle \mathcal{C}, X \rangle$ and every other argument in the interfering set of Λ' , such as $\langle \mathcal{A}, L \rangle$) is contradictory. However, such interfering arguments (including $\langle \mathcal{C}, X \rangle$ and $\langle \mathcal{A}, L \rangle$) appear in Λ as supporting arguments. Thus, this would imply that Λ is not an acceptable argumentation line because its set of supporting arguments would be contradictory as well. Contradiction.
 - $\langle \mathcal{C}, X \rangle$ is an argument appearing earlier in Λ' or is a disagreement sub-argument of an argument appearing earlier in Λ' . Without loss of generality, these two cases are covered by: $\langle \mathcal{C}, X \rangle$ is a sub-argument of an argument appearing earlier in Λ' . Then, we have to consider the following cases:
 - * $\langle \mathcal{C}, X \rangle$ is a sub-argument of an argument different from $\langle \mathcal{B}, H \rangle$ in Λ' . Since the only difference between Λ' and Λ prior to $\langle \mathcal{C}, X \rangle$ is the inclusion of $\langle \mathcal{B}, H \rangle$ as the first element in the sequence, $\langle \mathcal{C}, X \rangle$ would be a sub-argument of an argument appearing earlier in Λ , leading Λ to be a non-acceptable argumentation line. Contradiction.
 - * $\langle \mathcal{C}, X \rangle$ is a sub-argument of $\langle \mathcal{B}, H \rangle$. Since by hypothesis $\langle \mathcal{B}, H \rangle$ is a claim argument, by Definitions 9 and 12, $\langle \mathcal{C}, X \rangle$ is also a claim argument. Then, given that $\langle \mathcal{C}, X \rangle$ appears right after argument $\langle \mathcal{D}, Y \rangle$ in Λ and Λ' , we have that $\langle \mathcal{C}, X \rangle$ c-defeats $\langle \mathcal{D}, Y \rangle$; consequently, the set $\mathcal{C} \cup \mathcal{D}$ is contradictory. Moreover, since $\mathcal{C} \subseteq \mathcal{B}$ it holds that $\mathcal{B} \cup \mathcal{D}$ is also contradictory. Hence, $\Lambda''' = [\langle \mathcal{B}, H \rangle, \langle \mathcal{A}, L \rangle, \dots, \langle \mathcal{D}, Y \rangle]$ (i.e., the sub-sequence of Λ' ending in $\langle \mathcal{D}, Y \rangle$) would not be an acceptable argumentation line, which contradicts the assumption that Λ' was not acceptable because of the inclusion of $\langle \mathcal{C}, X \rangle$.

- $\langle \mathcal{C}, X \rangle$ leads to two consecutive blocking defeats in Λ' . Since $\langle \mathcal{C}, X \rangle$ is a supporting argument in Λ and, as previously shown, is different from $\langle \mathcal{A}, L \rangle$, it holds that $\langle \mathcal{C}, X \rangle$ is at least the third element in the sequence corresponding to Λ . Thus, the consecutive blocking defeats would also occur in Λ , meaning that Λ is not an acceptable argumentation line. Contradiction.
- $\langle \mathcal{C}, X \rangle$ is a detracting argument appearing right after argument $\langle \mathcal{D}, Y \rangle$ in Λ' , and there is no set of informants $\text{IS}_{\mathcal{D}} \in \text{StrDetInf}(\prec_{\mathbb{I}}, \langle \mathcal{D}, Y \rangle, \Lambda''')$ holding that $X \in \text{IS}_{\mathcal{D}}$, where $\Lambda''' = [\langle \mathcal{B}, H \rangle, \langle \mathcal{A}, L \rangle, \dots, \langle \mathcal{D}, Y \rangle]$ (i.e., the sub-sequence of Λ' ending in $\langle \mathcal{D}, Y \rangle$). Since $\langle \mathcal{C}, X \rangle$ is a supporting argument in Λ and, as previously shown, is different from $\langle \mathcal{A}, L \rangle$, it holds that $\langle \mathcal{C}, X \rangle$ is at least the third element in the sequence corresponding to Λ ; hence, $\langle \mathcal{D}, Y \rangle$ is at least the second element in Λ . Consequently, $\langle \mathcal{D}, Y \rangle$ is different from $\langle \mathcal{B}, H \rangle$ and the argument preceding $\langle \mathcal{D}, Y \rangle$ in Λ' (also, in Λ''') will appear preceding $\langle \mathcal{D}, Y \rangle$ in Λ as well. By Definition 25 the strength-determining sets of informants $\text{StrDetInf}(\prec_{\mathbb{I}}, \langle \mathcal{D}, Y \rangle, \Lambda''')$ will be obtained by comparing the informants of $\langle \mathcal{D}, Y \rangle$ with the informants of the argument preceding it in Λ''' (also, in Λ'), which is the same argument preceding it in Λ . Thus, it holds that $\text{StrDetInf}(\prec_{\mathbb{I}}, \langle \mathcal{D}, Y \rangle, \Lambda''') = \text{StrDetInf}(\prec_{\mathbb{I}}, \langle \mathcal{D}, Y \rangle, \Lambda'')$, where $\Lambda'' = [\langle \mathcal{A}, L \rangle, \dots, \langle \mathcal{D}, Y \rangle]$ (i.e., the sub-sequence of Λ ending in $\langle \mathcal{D}, Y \rangle$). As a result, if $\langle \mathcal{C}, X \rangle$ was targeting a non-determining informant of $\langle \mathcal{D}, Y \rangle$ in Λ' , it was doing so in Λ as well, meaning that Λ is not an acceptable argumentation line. Contradiction. \square

Lemma 2. *Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, \mathbf{P}_{\mathbb{I}}, \prec_{co}^{\mathbb{I}}, \prec_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$, and $\langle \mathcal{A}, L \rangle$ an argument built from $\mathbf{P}_{\mathbb{I}}$. If \mathbb{I} warrants $\langle \mathcal{A}, L \rangle$, then for every claim argument $\langle \mathcal{B}, H \rangle$ built from $\mathbf{P}_{\mathbb{I}}$ that is c-defeated by $\langle \mathcal{A}, L \rangle$ or that c-defeats $\langle \mathcal{A}, L \rangle$ it holds that \mathbb{I} does not warrant $\langle \mathcal{B}, H \rangle$.*

Proof. We have to consider two cases:

- $\langle \mathcal{A}, L \rangle$ c-defeats $\langle \mathcal{B}, H \rangle$. Since by hypothesis \mathbb{I} warrants $\langle \mathcal{A}, L \rangle$, by Lemma 1 the root of $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$ has a child node corresponding to $\langle \mathcal{A}, L \rangle$ that is marked **U**. Therefore, by Definition 28, the root of $\mathcal{T}_{\langle \mathcal{B}, H \rangle}$ will be marked as **D** and thus, by Definition 29, \mathbb{I} does not warrant $\langle \mathcal{B}, H \rangle$.
- $\langle \mathcal{B}, H \rangle$ c-defeats $\langle \mathcal{A}, L \rangle$. Let us suppose by contradiction that \mathbb{I} warrants $\langle \mathcal{B}, H \rangle$. If that were the case, by Lemma 1 the root of $\mathcal{T}_{\langle \mathcal{A}, L \rangle}$ will have a child node corresponding to $\langle \mathcal{B}, H \rangle$ that is marked **U**. Hence, by Definition 28, the root of $\mathcal{T}_{\langle \mathcal{A}, L \rangle}$ will be marked as **D**. Consequently, by Definition 29 $\langle \mathcal{A}, L \rangle$ would not be warranted by \mathbb{I} , which contradicts our hypothesis. \square

Theorem 1. *Let \mathbb{I} be a finite set of agent identifiers, $(\mathbb{I}, \mathbf{P}_{\mathbb{I}}, \prec_{co}^{\mathbb{I}}, \prec_{\mathbb{I}})$ the specification of an agent $\mathbb{I} \in \mathbb{I}$, and L, H two complementary literals. If \mathbb{I} warrants L , then \mathbb{I} does not warrant H .*

Proof. If \mathbb{I} warrants L , by Definition 29, there exists a claim argument $\langle \mathcal{A}, L \rangle$ built from $\mathbf{P}_{\mathbb{I}}$ that is warranted by \mathbb{I} . Let us now suppose by contradiction that \mathbb{I} warrants H . Then, by Definition 29, there would exist a claim argument $\langle \mathcal{B}, H \rangle$ that is warranted by \mathbb{I} . Hence, by Definition 13 $\langle \mathcal{A}, L \rangle$ and $\langle \mathcal{B}, H \rangle$ c-attack each other. Moreover, by Definition 20, it will be the case that $\langle \mathcal{A}, L \rangle$ is a proper c-defeater of $\langle \mathcal{B}, H \rangle$, $\langle \mathcal{B}, H \rangle$ is a proper c-defeater of $\langle \mathcal{A}, L \rangle$, or $\langle \mathcal{A}, L \rangle$ and $\langle \mathcal{B}, H \rangle$ are blocking c-defeaters of one another. Then, by Lemma 2 this would imply that \mathbb{I} does not warrant $\langle \mathcal{A}, L \rangle$, contradicting the hypothesis that L is warranted by \mathbb{I} . \square

6. Related work

The literature of argumentation offers a wide and extensive variety of approaches accounting for the notion of argument strength, some of which will be discussed in this section. For instance, approaches like [3,5] resolve conflicts by identifying the stronger arguments through a general preference relation. On the other hand, works like [13,28] define the strength of an argument through a formula yielding a numerical value. This formula is based on the intuition that the strength of an argument is inversely proportional to the strength of its attackers, with the aim of codifying the likelihood of an argument to be ultimately defeated. Similarly, [35] proposed a game-theoretic measure of argument strength, where the strength of each argument is calculated in a way such that if an argument is attacked then its strength falls, but if the attack is in turn attacked, then the strength of the original argument rises. Related to these, approaches to value-based argumentation [10] consider that the strength of an argument depends on the social values it advances, and determining whether the attack of one argument on another succeeds depends on the comparative strength of the values advanced by the arguments concerned. Also, in [41] an approach combining the ASPIC argumentation system and the fuzzy set theory is proposed, where argument strength is computed by using a t-norm aggregating the importance of its involved premises and rules. Similarly to us, the latter attaches information to the basic elements of their system to compute argument strength, which in turn is used to determine acceptable arguments. However, in that work it is not possible to challenge the strength of an argument (even when undercutting defeaters could be seen as similar to our detracting defeaters, an undercut challenges the applicability of a rule instead of the strength of the argument as detracting defeaters do).

We can also identify a group of approaches that, explicitly or implicitly, make use of a notion of credibility or trust to account for the arguments' strength. Given the close relationship between our proposal and these approaches, we devote the rest of this section to discuss the similarities and differences with some of them.

In [38] and [44] an argumentation formalism is proposed which, as part of its reasoning process, uses information about trust to measure the arguments' strength. This formalism is described as a set of graphs and, to determine an agent's beliefs, the authors propose a model that accounts for the trust in the information that is used for building the arguments. Like ours, their approach is presented in a multi-agent setting, where informant agents can have different levels of credibility and these credibilities are used to measure the arguments' strength. In contrast to our proposal, where each agent has its own credibility order (completely independent from the other agents'), they use a centralized notion of trust that is codified in a shared trust network. This global network holds information about how agents trust each other and can be used to obtain an agent-centric trust network that represents the viewpoint of a particular agent. Although from these graphs it is possible to determine a credibility order for each agent, these orderings are strongly dependent on the connections in the global network.

Similarly to our formalism, each piece of information in [38,44] is linked to an agent which establishes how credible that information is, and the strength of an argument is determined by the credibility of the information it is based on. Then, [38] and [44] use an argumentation inference mechanism to deal with a potentially contradictory belief base. In that context, arguments are built to support pieces of information that can be consistently inferred from the belief base, and the strength measures are used to decide between conflicting arguments. In addition to that, our approach allows to construct arguments for and against the consideration of informant agents. Therefore, differently from theirs, an agent in our approach will be able to reason and argue about the strength of its arguments, in addition to arguing about the domain information stored in its knowledge base.

Another significant difference between [38,44] and the work reported in this paper is that they use numerical values to establish the trust relation among agents, leading to a total order over the set of agent identifiers; in contrast, we use symbolic information in the form of a strict partial order over that set. Such contrast leads to different approaches for measuring the arguments' strength. On the one hand, they compute the strength of an argument using a formula based on numerical values assigned to the agents that provided the information used for building the argument. On the other hand, the strength of an argument in our approach is not absolute; instead, it is relative to the resolution of the conflicts the argument is involved in, and is established by the sets of strength-determining informants of the argument (according to a valid informant-based comparison criterion). Consequently, in their approach, the central component for determining the arguments' strength is the formula to be used, whereas in ours it is the adopted informant-based comparison criterion.

The work reported in [17] relates to our proposal in that they also make use of the credibility of informant agents as a source of argument strength. Similarly to us, they associate arguments with their information sources (and multiple sources can be associated with the same argument); however, they abstract away from the origin of such credibility. That is, their credibility function associates an informant with a set of arguments, without specifically attaching the informant to a particular piece of knowledge within the argument. In contrast, we associate each defeasible rule in an argument with its informant agent to conform a defeasible domain object. Then, in [18] the authors extend their work to combine credibilities of informants of facts, assumptions and conclusions in order to determine the arguments' strength. Like in our approach, defeats among arguments in [17,18] are defined by accounting for a notion of argument strength based on the credibility of their informants. However, a key aspect that differentiates our approach from theirs is the fact that their credibility function establishes a total order over the set of informants, whereas our credibility orders are assumed to be strict *partial* orders. Finally, as remarked above for other approaches, another difference between our work and [17,18] is the fact that we provide the means to reason about the arguments' strength by allowing to build arguments for and against their different informant agents; in contrast, the approach of [17,18] does not provide such a possibility.

In this paper we introduced a particular informant-based comparison criterion (see Definition 18) to measure and compare the arguments' strength. As discussed in Section 4, the single informant credibility criterion resembles the rules priorities criterion (see *e.g.*, [25]) which takes preferences among rules to then define preferences among arguments (which are in turn expressed as sets of rules). That is, the rules priorities criterion lifts preferences over rules to preferences over sets of rules. Similarly, an informant-based comparison criterion makes use of a credibility order (*i.e.*, a relationship over informant agents) to establish preferences over sets of informants; in other words, preferences over informant agents are lifted onto preferences over sets of informants. The literature offers a variety of approaches dealing with the issue of lifting preferences, such as [9,23,33,34] (see also [8] for an extensive overview on the issue of lifting an order relation on a set X to an order on the family of all non-empty subsets of X).

On the one hand, in [34] the authors propose natural extensions of the versions of five axioms introduced in [8] and also state that not every axiom is desirable in every situation. Regarding the axioms or conditions associated with a lifting principle (here, a valid informant-based comparison criterion), we can note the following difference. Whereas in our approach the credibility order (which is then lifted to define an informant-based comparison criterion) should be irreflexive, asymmetric and transitive, [34] assumes an order relation on a set X that has to be reflexive, antisymmetric, transitive and total. For instance, regarding the reflexivity vs. irreflexivity requirement, we can note that a valid informant-based comparison criterion is required to be irreflexive in order to align with the strict nature of the credibility

order it is based on. An exhaustive analysis of our approach in the context of their proposed axioms is left as future work, including the study of alternative or additional conditions to be imposed either on credibility orders or on valid informant-based comparison criteria.

In [9] the authors argue that, in the context of structured argumentation, the support provided by an argument for its conclusion is determined by the degree of support of its premisses, and by the degree of support provided by the inference rules applied in its construction. Then, among other things, the authors discuss several alternatives for lifting qualitative orderings on premisses and for lifting qualitative orderings on defeasible rules. It should be noted that, when accounting for orderings on different types of elements conforming the arguments, these orderings should be somehow combined when lifting preferences. Our proposal is such that domain knowledge in an IBDP is only expressed through the defeasible domain objects, which associate defeasible rules with their informant agents. In particular, our lifting criteria (the informant-based comparison criteria) only account for one dimension: the credibility order over the informant agents.

When considering the conditions imposed on valid informant-based comparison criteria and the conditions imposed by the different liftings discussed in [9], we can note the following: our conditions are meant to characterize a family of criteria whereas they describe specific kinds of liftings on qualitative orderings of elements. However, we can find a relationship between the liftings studied in [9] and the family of valid criteria characterized here, and we plan to explore this relationship as part of future work. For instance, consider the *max-max* criterion to lift orderings on premisses to orderings on sets of premisses X and Y : $Y \succ X$ iff for every $x \in X$ there is a $y \in Y$ such that $y \succ x$ (with \succ meaning *strictly preferred to*). Here, we can see that the *max-max* criterion satisfies the condition of being based on the original ordering (*i.e.*, at least one pair of elements of X and Y is related by \succ), as in the first clause of Definition 17. Then, it can be shown that the *max-max* criterion also satisfies the property of yielding irreflexive orderings when applied on transitive orderings over premisses. For this, consider a set X and suppose by contradiction that $X \succ X$. Hence, for every $x \in X$ there exists $x' \in X$ such that $x' \succ x$, and this must also be satisfied for such x' . So, because of the transitivity of the \succ relation over premisses, at some point a cycle will exist, contradicting the fact that it is a strict ordering. Consequently, the *max-max* lifting criterion satisfies irreflexivity (second clause of Definition 17). Finally, suppose the *max-max* criterion relates two sets of premisses X and Y as follows: $Y \succ X$ and $X \succ Y$. Here, if $Y \succ X$, by definition of *max-max* it holds that for every $x \in X$ there exists $y' \in Y$ such that $y' \succ x$. Then, if also $X \succ Y$, it must be the case that for every $y \in Y$ there is $x \in X$ such that $x \succ y$. So, because of the transitivity of the \succ relation over pairs of premisses, for every $y \in Y$ it holds that there exists $y' \in Y$ such that $y' \succ y$ (*i.e.*, $Y \succ Y$), contradicting the fact that the *max-max* criterion yields irreflexive orderings on sets of premisses. Consequently, the *max-max* criterion also satisfies the third clause of Definition 17 and thus, it could serve as a valid informant-based comparison criterion when considering a credibility order as the base \succ relation.

Another work that relates to our proposal is [32], where the authors use an argumentation mechanism based on trust as a layer of a belief revision process carried out by agents dealing with (potentially conflicting) opinions about their pairs. In their argumentation approach, trust is used for building a preference ordering amongst arguments, thus codifying their strength. For this, they first aggregate the information about different opinions for the same proposition. Then, using these aggregated propositions, they build arguments whose trustworthiness is assessed using a conjunctive fusion operator over the opinions forming the argument. This assessment considers the number of agents and information pieces that were needed for building the argument. Even though [32] does not account for the fact that arguments can be constructed to challenge an informant agent (hence, other arguments' strength), it could

be interesting to adapt their ideas in our proposal. For that purpose, we can think of two alternatives: either provide a comparison criterion that encodes the above mentioned strategies, or extend the notion of acceptable argumentation line (Definition 26) to consider preferences codifying those strategies.

A qualitative bipolar argumentative modeling of trust is proposed in [39]. Like in our proposal, their approach is qualitative and only a finite number of levels is assumed in the trust scale. On the other hand, in contrast with our proposal, they use a bipolar argumentative approach where trust and distrust can be independently assessed. There, an agent can evaluate its trust into an object X (that can be either a source or another agent) on the basis of two types of information: the observed behavior of X , and the reputation of X according to the other agents. Reputation information is then viewed as an input the agent uses for revising or updating its own trust evaluation, based on its perception. The approach proposed in [39] is such that two kinds of arguments in favor of trusting an agent (either establishing that a good point is reached or a bad point is avoided), as well as two kinds of arguments against trusting an agent (either indicating that a bad point is satisfied or that a good point is not reached) can be constructed. These four kinds of arguments are based on an inference rule and the trust evaluation of the agent, which is represented with an interval $[t-, t+]$ over a discrete scale, with the intended meaning that the trust is not larger than $t+$ nor smaller than $t-$. Clearly, our work relates to [39] in that we also allow to build arguments for or against an informant agent; however, our arguments do not encode reasons for trusting or distrusting such an agent, but they provide reasons for or against the consideration of information provided by it. Moreover, as another difference, even though [39] indicates some basic mechanisms leading to the revision of trust values, that paper is mainly focused on evaluating trust rather than on integrating the trust values as a measure of argument strength in an argumentation process.

The authors in [29,30] adopt a symbolic approach to model credibility using two global relations: the trust relation and the distrust relation. These relations, together with the set of agents, constitute a trust system where a pair (a, b) in the trust (respectively, distrust) relation expresses that agent a trusts (respectively, distrusts) agent b . Their formalism aims at determining whether an agent trusts another taking into account the potential conflicts that may appear when the trust and distrust relations are jointly analyzed in the trust system. To that end, they follow an argumentation approach in which arguments represent a position for an agent to either trust or distrust a peer; those arguments are similar to our backing and detracting arguments, respectively. Additionally, when considering an advanced version of their system, each agent is also provided with a partial order defined amongst its peers, and they use this order to codify the efficacy in which an agent trusts its peers (aiming to model a degree of trustworthiness or reputation). Like us, they use these efficacy orders to provide strength to their arguments. However, the goal of the argumentation system in the two approaches differs. Their aim is to decide if an agent trusts another or not, given the trust and distrust relations. In contrast, we aim at establishing the beliefs an agent has, which correspond to the pieces of information (literals) it warrants. Notwithstanding this, similarly to ours, their proposal seeks to enrich argumentation systems with strength measures that account for (potentially conflicting) information about trust, and use these measures to decide which arguments and conclusions prevail.

The work reported in [4], similarly to ours, presents an agent-based argumentation approach for reasoning about beliefs and information received from other agents. There, beliefs are also used to represent how trustworthy the information sources are to a given agent. They identify six forms of trust that can appear as part of the formulas in the agent belief base. From their belief bases arguments are built, conflicts among them are identified and then resolved. In particular, besides the usual conflicts in structured argumentation, the authors identify several types of attack that arise from the semantics of the six forms

of trust considered. Therefore, like us, they allow to challenge and support the credibility or trustworthiness of an informant; nevertheless, there are some differences. In their approach each form of trust is binary: the agent either trusts or does not trust an informant. In contrast, in our approach agents are ordered using a strict partial order and thus, it is possible to establish whether an informant is more credible than another. There are also differences on how trust relates to argument strength. In their approach, trust forms are used for constructing arguments and do not directly affect argument strength. On the other hand, our approach uses the level of credibility as the measure to define the arguments' strength. They consider the notion of strength when they introduce graded beliefs, where a grade is attached to the beliefs operator. Using these grades they compute the strength of an argument as the weakest link. However, unlike our approach, they do not provide any mechanism to challenge the strength of an argument. Finally, it is worth to mention that these differences also apply to a comparison between our proposal and the work reported in [48].

7. Conclusion

In this paper we have presented an argumentative reasoning formalism where the credibility of informants plays a central role, as it allows to determine the arguments' strength. Our formalism was developed in a multi-agent setting where agents share domain knowledge. There, each agent may obtain information from other informant agents and also has an assessment of how credible these informants are. Agents are equipped with the argumentative machinery, allowing them to reason with the potentially conflicting information in their knowledge bases to finally determine their warranted beliefs. In our approach, defeasible rules (which represent domain knowledge) are associated with their informant agents. Also, we introduced two new kinds of rules (backing and detracting rules) in order to be able to argue about the contexts in which the domain knowledge provided by the informant agents should be used or not. In other words, these rules are used to express reasons for and against the consideration of informants, respectively. From all this knowledge, an agent will be able to construct arguments to support its inferences. In addition, each agent has a credibility order among its informant agents and a comparison criterion used to assess the strength of the conflicting arguments built from its knowledge base.

As shown before, our informant-based approach is such that the strength of an argument is determined by the credibility of its informant agents. To that end, the comparison criterion in an agent's specification is based on the agent's credibility order. In particular, we have shown that the strength of an argument in our approach is not absolute, but it is relative to the resolution of the conflicts the argument is involved in. Then, it could be the case that some informant providing information for building an argument is relevant for establishing the argument's strength in some cases, but not in others. In this context, the incorporation of backing and detracting rules allows agents to argue about the arguments' strength. Specifically, backing rules allow to express reasons for the consideration of informant agents, whereas detracting rules enable to express reasons against the consideration of information provided by them. Using these rules we defined new types of argument which, together with the classic arguments supporting conclusions, are considered by the argumentation machinery to establish the beliefs an agent has. Finally, we have formally shown that the warranting process employed by our argumentative approach is sound, preventing an agent from warranting contradicting conclusions.

It is worth noting that the defeasible domain objects within an informant-based Defeasible Logic Program (IBDP) establish a correspondence between defeasible rules and their informant agents. The

fact that, differently from standard DeLP programs, an IBDP does not include strict rules may appear as a limitation of our approach. However, recall that strict rules are provided in DeLP as a representational tool that gives the possibility of expressing the indefeasible nature of the relation between the body and head of such rules, making them indisputable. In contrast, in an IBDP, a domain object is expressed as a pair containing a rule and the informant of that rule. In this context, such rules are always defeasible since they come attached with the credibility of their informant agent. Therefore, our approach only accounts for defeasible rules (hence the name of the *defeasible domain objects*) since our main focus was on how the arguments' informants affect and determine the arguments' strength. Moreover, regarding the absence of strict rules, we would like to remark that a number of different applications of DeLP (e.g., see [1,2,16,24,26,27,47]) have been developed using defeasible logic programs without strict rules. Consequently, we consider that the current characterization of IBDPs without strict rules is not a real limitation for our approach's expressivity and applicability. Notwithstanding this consideration, and to maintain a general approach, it is possible to extend our approach to account for standard DeLP strict rules and facts, where these elements of the program should not come with an associated informant to reflect their strict and indisputable nature accurately; we plan to do this as part of future work.

Regarding the notion of argument strength accounted for in this work, we can highlight the fact that it is solely based on the credibility of the informant agents. This may lead one to think of this notion of argument strength as too narrow, or too specific. Nevertheless, as discussed in Section 4, an informant-based comparison criterion like the single informant comparison criterion does not stray too far from other existing criteria in the literature of structured argumentation, resembling the rules priorities criterion. At any rate, as part of future work, we will explore a generalization our notion of argument strength in order to combine different comparison criteria at different levels. For instance, we could adopt an approach similar to [15], where the rules priorities criterion is used first to resolve attacks into defeats and, in case of undecidedness, the generalized specificity criterion is considered later. More generally, we will explore the possibility of using the operators defined in [45], which allow to combine multiple argument comparison criteria.

As shown in the existing literature (see [12] for an overview), DeLP is among the four major approaches to structured argumentation. Extending a DeLP program into an IBDP, with the addition of informants of defeasible rules and informant rules allowing to argue for and against the consideration of information provided by the different informant agents, we believe we are contributing to expanding DeLP's applicability domain. Nevertheless, as part of our future work, we plan to study the possibility of further exploiting these ideas and apply them in the context of other major structured argumentation approaches such as ASPIC+ [36] or ABA [46]. In that regard, we consider that such an exploration could make for exciting advances in the area, and will most definitely bring on new challenges requiring a substantial transformation of those frameworks.

In Section 2 we argued that our approach is restricted to deal with credibility orders relating agents that are sources of information about the same topic. As future work, we intend to extend our approach to account for multi-topic credibility orders (*i.e.*, handle multiple credibility orders, one for each topic). In order to be able to deal with these, we also plan to extend the knowledge representation and reasoning capabilities of an IBDP by, among other things, expanding defeasible domain objects to state their topic explicitly. In particular, such an extension would allow us to better model scenarios like the one described in the example of the medical domain given after Definition 11. Then, for instance, we could represent the information provided by agent \mathcal{I}_X about medical treatments as belonging to one topic and, on the other hand, the information provided by \mathcal{I}_X stating that he is not a physician as belonging to another topic. In that way, detracting rules could also be extended to include the topic under which the

information provided by a given agent is challenged; consequently, for the example given in the paper, the detracting rule could target the defeasible domain objects provided by \mathcal{I}_X which correspond to medical advice but not the one referring to \mathcal{I}_X not being a physician.

Finally, we would like to discuss other exciting prospects for future research. On the one hand, we plan to study how our approach could be extended to consider trust and distrust relations as presented in [29,30]; briefly, the idea would be to connect such relations with backing and detracting arguments. On the other hand, we will also study how an agent's credibility order should be updated when the warranted information is taken into account. For that purpose, backing and detracting arguments can also play a central role. For instance, suppose that the credibility order initially establishes that an informant \mathcal{I}_1 is less credible than another informant \mathcal{I}_2 . Then, if the argumentative machinery yields an undefeated backing argument for \mathcal{I}_1 and an undefeated detracting argument for \mathcal{I}_2 , there could be an indication that the credibility order has to be updated so that \mathcal{I}_1 becomes more credible than \mathcal{I}_2 . Furthermore, as another direction for future work, we intend to formally analyze the conditions characterizing our valid informant-based comparison criteria in the context of various lifting principles such as those addressed in [9], and also study (and redefine in the context of our approach, when required) different preference handling principles like the ones discussed in [22].

Acknowledgements

We would like to thank the reviewers for their helpful and insightful comments on the previous versions of this paper. This work has been supported by EU H2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 690974 for the project "MIREL: Mining and Reasoning with Legal texts", by CONICET (grant PIP 112-201701-00871) and by Universidad Nacional del Sur (grants PGI 24/N046 and 24/ZN32).

References

- [1] R.A. Agis, S. Gottifredi and A.J. García, Acquiring knowledge from expert agents in a structured argumentation setting, *Argument & Computation* **10**(2) (2019), 149–189. doi:[10.3233/AAC-190447](https://doi.org/10.3233/AAC-190447).
- [2] R.A. Agis, S. Gottifredi and A.J. García, An approach for distributed discussion and collaborative knowledge sharing: Theoretical and empirical analysis, *Expert Systems with Applications* **116** (2019), 377–395. doi:[10.1016/j.eswa.2018.09.016](https://doi.org/10.1016/j.eswa.2018.09.016).
- [3] L. Amgoud and C. Cayrol, A reasoning model based on the production of acceptable arguments, *Ann. Math. Artif. Intell.* **34**(1–3) (2002), 197–215. doi:[10.1023/A:1014490210693](https://doi.org/10.1023/A:1014490210693).
- [4] L. Amgoud and R. Demolombe, An argumentation-based approach for reasoning about trust in information sources, *Argument & Computation* **5**(2–3) (2014), 191–215. doi:[10.1080/19462166.2014.881417](https://doi.org/10.1080/19462166.2014.881417).
- [5] L. Amgoud and S. Vesic, A new approach for preference-based argumentation frameworks, *Ann. Math. Artif. Intell.* **63**(2) (2011), 149–183. doi:[10.1007/s10472-011-9271-9](https://doi.org/10.1007/s10472-011-9271-9).
- [6] L. Amgoud and S. Vesic, Rich preference-based argumentation frameworks, *Int. J. Approx. Reasoning* **55**(2) (2014), 585–606. doi:[10.1016/j.ijar.2013.10.010](https://doi.org/10.1016/j.ijar.2013.10.010).
- [7] K. Atkinson, P. Baroni, M. Giacomin, A. Hunter, H. Prakken, C. Reed, G.R. Simari, M. Thimm and S. Villata, Towards artificial argumentation, *AI Magazine* **38**(3) (2017), 25–36. doi:[10.1609/aimag.v38i3.2704](https://doi.org/10.1609/aimag.v38i3.2704).
- [8] S. Barberà, W. Bossert and P.K. Pattanaik, *Ranking Sets of Objects*, Springer, 2004, pp. 893–977.
- [9] M. Beirlaen, J. Heyninck, P. Pardo and C. Straßer, Argument strength in formal argumentation, *FLAP* **5**(3) (2018), 629–676.
- [10] T.J.M. Bench-Capon, Persuasion in practical argument using value-based argumentation frameworks, *J. Log. Comput.* **13**(3) (2003), 429–448. doi:[10.1093/logcom/13.3.429](https://doi.org/10.1093/logcom/13.3.429).
- [11] T.J.M. Bench-Capon and P.E. Dunne, Argumentation in artificial intelligence, *Artif. Intell.* **171**(10–15) (2007), 619–641. doi:[10.1016/j.artint.2007.05.001](https://doi.org/10.1016/j.artint.2007.05.001).

- [12] P. Besnard, A.J. García, A. Hunter, S. Modgil, H. Prakken, G.R. Simari and F. Toni, Introduction to structured argumentation, *Argument & Computation* **5**(1) (2014), 1–4. doi:[10.1080/19462166.2013.869764](https://doi.org/10.1080/19462166.2013.869764).
- [13] P. Besnard and A. Hunter, A logic-based theory of deductive arguments, *Artif. Intell.* **128**(1–2) (2001), 203–235. doi:[10.1016/S0004-3702\(01\)00071-6](https://doi.org/10.1016/S0004-3702(01)00071-6).
- [14] P. Besnard and A. Hunter, *Elements of Argumentation*, MIT Press, 2008. doi:[10.7551/mitpress/9780262026437.001.0001](https://doi.org/10.7551/mitpress/9780262026437.001.0001).
- [15] C.E. Briguez, M.C. Budán, C.A.D. Deagustini, A.G. Maguitman, M. Capobianco and G.R. Simari, Argument-based mixed recommenders and their application to movie suggestion, *Expert Syst. Appl.* **41**(14) (2014), 6467–6482. doi:[10.1016/j.eswa.2014.03.046](https://doi.org/10.1016/j.eswa.2014.03.046).
- [16] M. Capobianco, C.I. Chesñevar and G.R. Simari, Argumentation and the dynamics of warranted beliefs in changing environments, *Autonomous Agents and Multi-Agent Systems* **11**(2) (2005), 127–151. doi:[10.1007/s10458-005-1354-8](https://doi.org/10.1007/s10458-005-1354-8).
- [17] C.F. Chang, P. Harvey and A. Ghose, Source sensitive argumentation system, in: *ICEIS 2006 – Proceedings of the Eighth International Conference on Enterprise Information Systems: Databases and Information Systems Integration, Paphos, Cyprus, May 23–27, 2006*, 2006, pp. 39–46.
- [18] C.F. Chang, P. Harvey and A.K. Ghose, Combining credibility in a source sensitive argumentation system, in: *Advances in Artificial Intelligence, 4th Hellenic Conference on AI, SETN 2006, Heraklion, Crete, Greece, May 18–20, 2006, Proceedings*, 2006, pp. 478–481. doi:[10.1007/11752912_49](https://doi.org/10.1007/11752912_49).
- [19] A. Cohen, A.J. García and G.R. Simari, A structured argumentation system with backing and undercutting, *Eng. Appl. of AI* **49** (2016), 149–166. doi:[10.1016/j.engappai.2015.10.001](https://doi.org/10.1016/j.engappai.2015.10.001).
- [20] A. Cohen, S. Gottifredi, A.J. García and G.R. Simari, A survey of different approaches to support in argumentation systems, *Knowledge Eng. Review* **29**(5) (2014), 513–550. doi:[10.1017/S0269888913000325](https://doi.org/10.1017/S0269888913000325).
- [21] A. Cohen, S. Gottifredi, L.H. Tamargo and A.J. García, Extending Defeasible Logic Programming with informant-based argumentation, in: *Argumentation-Based Proofs of Endearment. Essays in Honor of Guillermo R. Simari on the Occasion of His 70th Birthday*, C.I. Chesñevar, M.A. Falappa, E. Fermé, A.J. García, A.G. Maguitman, D.C. Martínez, M.V. Martínez, R.O. Rodríguez and G.I. Simari, eds, College Publications, London, 2018, pp. 73–89.
- [22] K. Cyras and F. Toni, Properties of ABA+ for non-monotonic reasoning, in: *NMR 2016 – Proceedings of the 16th International Workshop on Non-Monotonic Reasoning, Cape Town, South Africa, 22–24 April 2016*, http://nmr2016.cs.uct.ac.za/proceedings_nmr2016_online.pdf.
- [23] S.K. Dyrkolbotn, T. Pedersen and J.M. Broersen, On elitist lifting and consistency in structured argumentation, *FLAP* **5**(3) (2018), 709–746.
- [24] A.J. García, N.D. Rotstein, M. Tucac and G.R. Simari, An argumentative reasoning service for deliberative agents, in: *Knowledge Science, Engineering and Management, Second International Conference, KSEM 2007, Melbourne, Australia, November 28–30, 2007, Proceedings*, 2007, pp. 128–139. doi:[10.1007/978-3-540-76719-0_16](https://doi.org/10.1007/978-3-540-76719-0_16).
- [25] A.J. García and G.R. Simari, Defeasible logic programming: An argumentative approach, *Theory and Practice of Logic Programming* **4**(1–2) (2004), 95–138. doi:[10.1017/S1471068403001674](https://doi.org/10.1017/S1471068403001674).
- [26] D.R. García, A.J. García and G.R. Simari, Planning and defeasible reasoning, in: *6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2007), Honolulu, Hawaii, USA, May 14–18, 2007*, 2007, p. 222.
- [27] D.R. García, A.J. García and G.R. Simari, Defeasible reasoning and partial order planning, in: *Foundations of Information and Knowledge Systems, 5th International Symposium, FoIKS 2008, Pisa, Italy, February 11–15, 2008, Proceedings*, 2008, pp. 311–328. doi:[10.1007/978-3-540-77684-0_21](https://doi.org/10.1007/978-3-540-77684-0_21).
- [28] S. Gottifredi, N.D. Rotstein, A.J. García and G.R. Simari, Using argument strength for building dialectical bonsai, *Ann. Math. Artif. Intell.* **69**(1) (2013), 103–129. doi:[10.1007/s10472-013-9338-x](https://doi.org/10.1007/s10472-013-9338-x).
- [29] W.T. Harwood, J.A. Clark and J.L. Jacob, Networks of trust and distrust: Towards logical reputation systems, in: *Logics in Security*, 2010.
- [30] W.T. Harwood, J.A. Clark and J.L. Jacob, A perspective on trust, security and autonomous systems, in: *New Security Paradigms Workshop*, 2010.
- [31] S. Kaci, L.W.N. van der Torre and S. Villata, Preference in abstract argumentation, in: *Computational Models of Argument – Proceedings of COMMA 2018, Warsaw, Poland, 12–14 September 2018*, 2018, pp. 405–412.
- [32] A. Koster, A.L.C. Bazzan and M. de Souza, Liar liar, pants on fire; or how to use subjective logic and argumentation to evaluate information from untrustworthy sources, *Artif. Intell. Rev.* **48**(2) (2017), 219–235. doi:[10.1007/s10462-016-9499-1](https://doi.org/10.1007/s10462-016-9499-1).
- [33] J. Maly, M. Truszczynski and S. Woltran, Preference orders on families of sets – when can impossibility results be avoided? in: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13–19, 2018, Stockholm, Sweden*, 2018, pp. 433–439.
- [34] J. Maly, M. Truszczynski and S. Woltran, Preference orders on families of sets – when can impossibility results be avoided?, *J. Artif. Intell. Res.* **66** (2019), 1147–1197. doi:[10.1613/jair.1.11879](https://doi.org/10.1613/jair.1.11879).
- [35] P. Matt and F. Toni, A game-theoretic measure of argument strength for abstract argumentation, in: *Logics in Artificial Intelligence, 11th European Conference, JELIA 2008, Dresden, Germany, September 28 – October 1, 2008. Proceedings*, 2008, pp. 285–297.

- [36] S. Modgil and H. Prakken, The *ASPIC*⁺ framework for structured argumentation: A tutorial, *Argument & Computation* **5**(1) (2014), 31–62. doi:[10.1080/19462166.2013.869766](https://doi.org/10.1080/19462166.2013.869766).
- [37] D. Nute, Defeasible reasoning: A philosophical analysis in PROLOG, in: *Aspects of Artificial Intelligence*, J.H. Fetzer, ed., Kluwer Academic Pub., 1988, pp. 251–288. doi:[10.1007/978-94-009-2699-8_9](https://doi.org/10.1007/978-94-009-2699-8_9).
- [38] S. Parsons, E. Sklar and P. McBurney, Using argumentation to reason with and about trust, in: *Argumentation in Multi-Agent Systems – 8th International Workshop, ArgMAS 2011, Taipei, Taiwan, May 3, 2011, Revised Selected Papers*, 2011, pp. 194–212.
- [39] H. Prade, A qualitative bipolar argumentative view of trust, in: *Scalable Uncertainty Management, First International Conference, SUM 2007, Washington, DC, USA, October 10–12, 2007, Proceedings*, 2007, pp. 268–276. doi:[10.1007/978-3-540-75410-7_20](https://doi.org/10.1007/978-3-540-75410-7_20).
- [40] I. Rahwan and G.R. Simari (eds), *Argumentation in Artificial Intelligence*, Springer, 2009. doi:[10.1007/978-0-387-98197-0](https://doi.org/10.1007/978-0-387-98197-0).
- [41] N. Tamani and M. Croitoru, A quantitative preference-based structured argumentation system for decision support, in: *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2014, Beijing, China, July 6–11, 2014*, IEEE, 2014, pp. 1408–1415.
- [42] L.H. Tamargo, A.J. García, M.A. Falappa and G.R. Simari, On the revision of informant credibility orders, *Artificial Intelligence* **212** (2014), 36–58. doi:[10.1016/j.artint.2014.03.006](https://doi.org/10.1016/j.artint.2014.03.006).
- [43] L.H. Tamargo, S. Gottifredi, A.J. García, M.A. Falappa and G.R. Simari, Deliberative DeLP agents with multiple informants, *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial* **15**(49) (2012), 13–30.
- [44] Y. Tang, K. Cai, P. McBurney, E. Sklar and S. Parsons, Using argumentation to reason about trust and belief, *J. Log. Comput.* **22**(5) (2012), 979–1018. doi:[10.1093/logcom/exr038](https://doi.org/10.1093/logcom/exr038).
- [45] J.C. Teze, S. Gottifredi, A.J. García and G.R. Simari, An approach to generalizing the handling of preferences in argumentation-based decision-making systems, *Knowl.-Based Syst.* **189** (2020).
- [46] F. Toni, A tutorial on assumption-based argumentation, *Argument & Computation* **5**(1) (2014), 89–117. doi:[10.1080/19462166.2013.869878](https://doi.org/10.1080/19462166.2013.869878).
- [47] M. Tucac, A.J. García and G.R. Simari, Using Defeasible Logic Programming with contextual queries for developing recommender servers, in: *The Uses of Computational Argumentation, Papers from the 2009 AAAI Fall Symposium, Arlington, Virginia, USA, November 5–7, 2009*, 2009.
- [48] S. Villata, G. Boella, D.M. Gabbay and L.W.N. van der Torre, A socio-cognitive model of trust using argumentation theory, *Int. J. Approx. Reason.* **54**(4) (2013), 541–559. doi:[10.1016/j.ijar.2012.09.001](https://doi.org/10.1016/j.ijar.2012.09.001).