Building Support-based Opponent Models in Persuasion Dialogues

Christos Hadjinikolis, Sanjay Modgil, and Elizabeth Black

King's College London
name.surname@kcl.ac.uk

Abstract. This paper deals with an approach to opponent-modelling in argumentation-based persuasion dialogues. It assumes that dialogue participants (agents) have models of their opponents' knowledge, which can be augmented based on previous dialogues. Specifically, previous dialogues indicate relationships of support, which refer to arguments as abstract entities, and to their logical constituents. The augmentation of an opponent model relies on these relationships. An argument external to an opponent model can augment that model with its logical constituents, if that argument shares support relationships with other arguments that can be constructed from that model. This likelihood varies according to support types. We therefore provide corresponding quantifications for each support type.

1 Introduction

Strategy development in agent dialogues is an area that has received ample research interest in the last years [2, 7, 12, 17]. Specifically, strategising in a dialogue concerns the selection of a particular locution among all available locutions, which by some measure is deemed optimal [5]. In competitive contexts, "optimal" is understood in terms of increasing a participant's self-interested utility. Since in such contexts the employed knowledge is usually distributed amongst the participants, agents are unaware of the locutions available to their opponents. Thus, they often assume models of their opponents' possible knowledge for simulating how a dialogue may evolve, and develop strategies accordingly.

A common assumption is that *opponent models (OMs)* can be constructed on the basis of a participant's accumulated dialogue experience [7, 12, 17]. Furthermore, in the context of argumentation-based persuasion dialogues, some researchers propose that these models are augmented with external content (arguments), assuming relationships between the latter and information already in the model [7]. These relationships can be based on the notion of *support*.

In [7], Hadjinikolis *et al.* rely on the $ASPIC^+$ framework for structured argumentation to define a dialogue system for persuasion. In this system agents are assumed to have models of their opponents' knowledge. This knowledge can be augmented based on a modeller's dialogue history using the notion of *reinstatement* support. For example, assume that two agents, Ag_1 (modeller) and Ag_2 (opponent) engage in a dialogue. Let A be an argument introduced by Ag_2

in a dialogue, countered by Ag_1 with argument B. If B is then countered by Ag_2 with a third argument C, we then assume a support relationship between A (the supported) and C (the supporter), in the sense that C reinstates A. Ag_1 can model such relationships as directed, weighted arcs, linking nodes that represent the associated arguments (e.g. $A \rightarrow C$), in what is referred to as a relationship graph ($\mathcal{R}G$). An arc weight represents the frequency with which a certain argument is followed by a supporter in dialogues in which the modeller has participated, e.g. how often does C follow after A. Relying on this graph Ag_1 can augment an existing OM of another agent (e.g. Ag_3), to include the logical constituents of supporters, assuming that the latter are related to arguments that can be instantiated from the current state of that OM. For example, including in an OM the constituents of argument C, given that A can already be instantiated by that OM. Arc weights can then be used for the calculation of a probability value assigned to these constituents, which represents the modeller's confidence that an opponent is indeed aware of them.

In addition to reinstatement, other kinds of relationships can be used to identify support between arguments. For example, let an argument A be attacked by two arguments B_1 and B_2. One may argue that B_1 and B_2 support each other since they share the same attack target. Furthermore, more expressive kinds of support can be identified between arguments if one inspects their structure. For example, one may assume that an argument X supports an argument Y if they share the same conclusion/claim. This could also be assumed if X's claim appears as a premise in Y or in the antecedent of a rule in Y.

The purpose of this paper is to extend the work in [7] in the following ways. Firstly, it extends the notion of a $\mathcal{R}G$ by including a new kind of support relationship, concerned with arguments which attack the same target, allowing for another modelling alternative. Secondly, by inspecting the structure of related arguments, a refined categorisation of different support types is proposed, according to which support relationships are distinguished between low-level logical relationships and high-level abstract relationships. The first are special instances of the latter. It is then argued that in addition to abstract relationships, logical ones suggest a stronger connection between related arguments, which can be interpreted as an increased likelihood of them being mutually known to a certain opponent. Finally, a more fine grained quantification of these likelihoods is proposed, which reflects the properties of the support relationships they concern.

The paper is organised as follows. Sections 2.1 & 2.2 respectively present the $ASPIC^+$ framework for structured argumentation [10], and the $ASPIC^+$-based dialogue framework for persuasion presented in [6]. Using a framework for structured argumentation is necessary for investigating both abstract as well as logical support relationships between arguments. $ASPIC^+$ is chosen as a general and expressive framework, which accommodates many existing logical approaches to argumentation [10], allowing us to claim an analogous generality for our research. Section 3 elaborates on the categorisation of different support relationships between arguments, and on how they are modelled as weighted directed arcs between nodes of arguments in a $\mathcal{R}G$. Section 4 shows how these weights are quantified in a way that reflects the relationships they concern. Finally, Section 5 discusses our work in relation to how the notion of support is generally used in the literature, while Section 6 summarises our contributions and presents future work.

2 Background

2.1 *ASPIC*$^+$

ASPIC$^+$ [10] instantiates Dung's [4] abstract approach by assuming an unspecified logical language $\mathcal{L}$, and by defining arguments as inference trees formed by applying strict or defeasible inference rules of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$, interpreted as 'if the *antecedents* $\varphi_1, \ldots, \varphi_n$ hold, then *without exception*, respectively *presumably*, the *consequent* φ holds'.

To define attacks, minimal assumptions on $\mathcal{L}$ are made; namely that certain wff (well formed formulæ) are a contrary or contradictory of certain other wff. Apart from this the framework is still abstract: it applies to any set of strict and defeasible inference rules, and to any logical language with a defined contrary relation. The basic notion of *ASPIC*$^+$ is an argumentation system.

Definition 1. *Let* $AS = (\mathcal{L}, ^-, \mathcal{R}, \leq)$ *be an* **argumentation system** *where:*

- $\mathcal{L}$ *is a logical language.*
- $^-$ *is a contrariness function from* $\mathcal{L}$ *to* $2^{\mathcal{L}}$*, such that:*
  - φ *is a* contrary *of* ψ *if* $\varphi \in \overline{\psi}$ *and* $\psi \notin \overline{\varphi}$*;*
  - φ *is a* contradictory *of* ψ *(denoted by '*$\varphi = -\psi$*'), if* $\varphi \in \overline{\psi}$ *and* $\psi \in \overline{\varphi}$*.*
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ *is a set of strict (*$\mathcal{R}_s$*) and defeasible (*$\mathcal{R}_d$*) inference rules such that* $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$*.*
- $\leq$ *is a pre-ordering on* $\mathcal{R}_d$*.*

Arguments are then constructed with respect to a knowledge base that is assumed to contain two kinds of formulæ.

Definition 2. *A* **knowledge base (KB)** *in an AS is a pair* $(\mathcal{K}, \leq')$ *where* $\mathcal{K} \subseteq \mathcal{L}$ *and* $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p$ *where these subsets of* $\mathcal{K}$ *are disjoint:* $\mathcal{K}_n$ *is the (necessary)* axioms *(which cannot be attacked); and* $\mathcal{K}_p$ *is the* ordinary premises *(on which attacks succeed contingent upon preferences), and where* $\leq'$ *is a pre-ordering on the ordinary premises* $\mathcal{K}_p$*.*

Arguments are now defined, where for any argument A, `Prem` returns all the formulas of $\mathcal{K}$ (*premises*) used to build A; `Conc` returns A's conclusion; `Sub` returns all of A's sub-arguments; and `Rules` returns all rules in A.

Definition 3. *An* **argument** A *on the basis of a knowledge base* $(\mathcal{K}, \leq')$ *in an argumentation system* $(\mathcal{L}, ^-, \mathcal{R}, \leq)$ *is:*

1. φ *if* $\varphi \in \mathcal{K}$ *with:* `Prem`$(A) = \{\varphi\}$*;* `Conc`$(A) = \varphi$*;* `Sub`$(A) = \{\varphi\}$*;* `Rules`$(A) = \emptyset$*.*

2. $A_1, \ldots A_n \rightarrow/\Rightarrow \psi$ *if* $A_1, \ldots, A_n$ *are arguments such that there exists a strict/defeasible rule* $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow/\Rightarrow \psi$ *in* $\mathcal{R}_s/\mathcal{R}_d$.
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$; $\texttt{Conc}(A) = \psi$;
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$;
   $\texttt{Rules}(A) = \texttt{Rules}(A_1) \cup \ldots \cup \texttt{Rules}(A_n) \cup$
   $\{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow/\Rightarrow \psi\}$

Three kinds of *attack* are defined for $ASPIC^+$ arguments. B can attack A by attacking a premise or conclusion of A, or an inference step in A. For the latter *undercutting* attacks, it is assumed that applications of inference rules can be expressed in the object language; the precise nature of this naming convention will be left implicit.

Definition 4. *A* **attacks** *B iff A* undercuts, rebuts *or* undermines *B, where:*

- *A* **undercuts** *argument B (on B') iff* $\texttt{Conc}(A) \in \overline{B'}$ *for some* $B' \in \texttt{Sub}(B)$ *of the form* $B''_1, \ldots, B''_n \Rightarrow \psi$.
- *A* **rebuts** *argument B (on B') iff* $\texttt{Conc}(A) \in \overline{\varphi}$ *for some* $B' \in \texttt{Sub}(B)$ *of the form* $B''_1, \ldots, B''_n \Rightarrow \varphi$. *In such a case A* contrary-rebuts *B iff* $\texttt{Conc}(A)$ *is a contrary of φ.*
- *A* **undermines** *B (on B') iff* $\texttt{Conc}(A) \in \overline{\varphi}$ *for some* $B' = \varphi$, $\varphi \in \texttt{Prem}(B) \setminus \mathcal{K}_n$. *In such a case A* contrary-undermines *B iff* $\texttt{Conc}(A)$ *is a contrary of φ.*

An undercut, contrary-rebut, or contrary-undermine attack is said to be preference-independent, *otherwise an attack is* preference-dependent.
Then, A **defeats** *B (denoted $A \rightarrow B$) iff A attacks B (denoted $A \rightharpoonup B$) on B', and either: $A \rightharpoonup B$ is* preference-independent, *or; $A \rightharpoonup B$ is* preference-dependent *and $A \not\prec B'$.*

Some kinds of attack succeed as *defeats* independently of preferences over arguments, whereas others succeed only if the attacked argument is not stronger than the attacking argument. The orderings on defeasible rules and non-axiom premises are assumed to be used in defining an ordering $\preceq$ on the constructed arguments. Unlike [10] a function p is explicitly defined in [6], that takes as input a KB in an AS (and so the defined arguments and orderings on rules and premises) and returns $\preceq$ (see [10] for ways in which such a function would define $\preceq$). Finally, $\prec$ is assumed to be the strict counerpart of $\preceq$. The combination of an argumentation system, a knowledge base and a function p, is called an *argumentation theory.*

Definition 5. *An* **argumentation theory** *is a triple $AT = (AS, KB, p)$ where AS is an argumentation system, KB is a knowledge base in AS and:*

$$p : AS \times KB \longrightarrow \preceq$$

such that $\preceq$ is an ordering on the set of all arguments that can be constructed from KB in AS.

2.2 The Dialogue Framework

In [6], Hadjinikolis *et al.* assume an environment of multiple agents $Ag_1, \ldots, Ag_\nu$, where each Ag_i can engage in persuasion dialogues in which its strategic selection of locutions may be based on what Ag_i believes its interlocutor (in the set $Ag_{j \neq i}$) believes. Each Ag_i maintains a model of its possible opponent agents that represents the logical information possible opponents may use to construct arguments and preferences, rather than just abstract arguments and their relations. All agents share the same contrary relation $^-$, the same language $\mathcal{L}$, and the same way of defining preferences over arguments based on the pre-orderings over non-axiom premises and defeasible rules (i.e., the same function p).

Definition 6. *Let $\{Ag_1, \ldots, Ag_\nu\}$ be a set of agents. For $i = 1 \ldots \nu$, the* **agent theory** *of Ag_i is a tuple:*

$$AgT_i =< S_{(i,1)}, \ldots, S_{(i,\nu)} >$$

where for $j = 1 \ldots \nu$, each sub-theory $S_{(i,j)}$ is what Ag_i believes is the argumentation theory $(AS_{(i,j)}, KB_{(i,j)}, p_{(i,j)})$ of Ag_j, and:

- *if $j = i$, $S_{(i,j)}$ is Ag_i's own argumentation theory.*
- *for any $i, j, k, m \in \{1 \ldots \nu\}$, it holds that:*

$$S_{(i,j)} = (AS_{(i,j)}, KB_{(i,j)}, p_{(i,j)}) \ \text{ and } \ S_{(k,m)} = (AS_{(k,m)}, KB_{(k,m)}, p_{(k,m)})$$

be any two distinct sub-theories, then:

$$p_{(i,j)} = p_{(k,m)}, \quad \mathcal{L}_{(i,j)} = \mathcal{L}_{(k,m)} \ \text{ and } \ ^-(i,j) = {}^-(k,m).$$

Essentially, the notion of an OM is captured by a sub-theory. For convenience, a simplified version of a sub-theory is assumed, of the form:

$$S_{(i,j)} = \{\mathcal{K}_{(i,j)}, \leq'_{(i,j)}, \mathcal{R}_{(i,j)}, \leq_{(i,j)}\}$$

which contains the discrete sets of logical elements assumed by the modeller (in this case Ag_i) to be known by each of its opponents ($Ag_{j \neq i}$), including the modeller's own sub-theory ($S_{(i,i)}$). Henceforth, we may omit subscripts identifying pre-orderings and rules specific to a given agent.

Dialogue participants are assumed to introduce arguments constructed in a common language $\mathcal{L}$, which attack those of their opponent, sharing an understanding of when one argument attacks another, based on the language dependent notion of conflict. Preferences may also be submitted in the dialogue against arguments, as a means of invalidating the success of an attack as defeat. *Commitment stores* are employed, to store the preferences and the logical constituents of the arguments introduced by each agent in a dialogue. These commitment stores are then used by the dialogue participants for directly updating the sub-theories (OMs) of their respective opponents, e.g. Ag_i can use the commitment store of its opponent Ag_j in a dialogue to *update* the contents of its sub-theory $S_{(i,j)}$.

Participants assume the roles of *proponent* (Pr) and *opponent* (Op), where the former submits an initial argument X, whose claim is the topic of the dialogue. The set of arguments $\mathcal{A}$ instantiated by the logical constituents submitted by both parties during the course of a dialogue, are assumed to be organised into a Dung framework, $AF = (\mathcal{A}, \mathcal{D})$, where $\mathcal{D}$ is the binary defeat relation on $\mathcal{A}$, i.e. $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$, defined on the basis of the attack relationships between the arguments, and the preferences introduced into the dialogue by both participants. Two sets of protocol rules are described: one for the grounded and one for the preferred semantics. These rules regulate turn-taking and the legal moves available to the participants in a dialogue, in a way that reflects their respective semantics. Conflicting preferences are resolved in favour of Op in the grounded case, and of Pr in the preferred[1]. Since the modelling of an opponent's preferences is not in the scope of this work, we only assume dialogue moves whose content is just arguments and leave the modelling of preferences to future work. Furthermore, since our interest is just to model opponent arguments in terms of how they appear in dialogues, rather than distinguishing between different dialogues with respect to different semantics (e.g. grounded, preferred), we define a general dialogue with minimal restrictions on the moves available to each participant at each point.

We define a dialogue $\mathscr{D}$ as a sequence of dialogue moves $< \mathcal{DM}_0, \ldots, \mathcal{DM}_n >$ of the form $\mathcal{DM} =< I, A >$, where $I \in \{Pr, Op\}$, $\overline{I} = Pr$ if $I = Op$ and vice-versa, and A is an argument in the set $\mathcal{A}_I$ instantiated from I's sub-theory ($S_{(I,I)}$) as well as from the commitment store of a participant's opponent. The content of $\mathcal{DM}_0$ is the initial argument for the topic of the dialogue. The legality of a dialogue move is regulated by explicit rules that, among others, account for the dialogical objective and a participant's role. For the purpose of this paper these are defined as follows:

Definition 7. *$\mathscr{D} =< \mathcal{DM}_0, \ldots, \mathcal{DM}_n >$ is a* **legal** *persuasion dialogue if:*

1. *$\mathcal{DM}_0 =< Pr, X >$ (the dialogue begins with Pr's move);*
2. *for $i = 0 \ldots n-1$, if $\mathcal{DM}_i =< I, A \in \mathcal{A}_I >$ then $\mathcal{DM}_{i+1} =< \overline{I}, B \in \mathcal{A}_{\overline{I}} >$ (Pr and Op take turns);*
3. *for $i = 1 \ldots n$, each $\mathcal{DM}_i$ is a reply to some $\mathcal{DM}_j$, $j < i$ (alternative replies are allowed), where $\mathcal{DM}_j =< I, A >$, $\mathcal{DM}_i =< \overline{I}, B >$ and B attacks A.*

Since we assume multi-reply protocols which allow participants to *backtrack* and reply to previous opponent moves, dialogues can be represented as trees rather than sequences of moves. An example is shown in Fig.1a, where Ag_1's moves $\mathcal{DM}_5$ & $\mathcal{DM}_7$ are used as alternative replies against Ag_2's move $\mathcal{DM}_1$.:

Definition 8. *Let $\mathscr{D} =< \mathcal{DM}_0, \ldots, \mathcal{DM}_n >$ be a dialogue and $\mathcal{M} = \{\mathcal{DM}_0, \ldots, \mathcal{DM}_n\}$ the set of moves in $\mathscr{D}$. Then $\mathcal{T} = \{\mathcal{M}, \mathcal{E}\}$ is a* **dialogue tree** *with root node $\mathcal{DM}_0$, and arcs $\mathcal{E} \subseteq \mathcal{M} \times \mathcal{M}$, such that for every two moves $\mathcal{DM}_i$ & $\mathcal{DM}_j$, $(\mathcal{DM}_i, \mathcal{DM}_j) \in \mathcal{E}$ means that $\mathcal{DM}_j$ is $\mathcal{DM}_i$'s* target *($\mathcal{DM}_i$ replies to $\mathcal{DM}_j$).*

[1] Note that if agents play logically perfectly they can be shown to win iff the argument they move is justified under the grounded respectively preferred semantics in the framework constructed during the dialogue [6].

Every move in $\mathcal{M}$ that is not the target of another move is a *leaf-node*, while each distinct path from $\mathcal{DM}_0$ to a leaf node is a *dispute*. For a $\mathcal{T}$ with m leaf-nodes, $\Delta = \{d_1, \ldots, d_m\}$ is the set of all disputes in $\mathcal{T}$. Each new dispute results from a backtracking move by either of the participants. Note that for convenience we may represent a dialogue tree as $\mathcal{T} = \{d_1, \ldots, d_m\}$.

Provided a modeller's history of dialogues, Hadjinikolis *et al.* [7] model support relationships between arguments in these dialogues, in the form of a relationship graph ($\mathcal{R}G$). A $\mathcal{R}G$ is assumed to be incrementally constructed through a series of dialogues. It is composed of nodes which represent the set of all encountered opponent arguments (OAs) in a modeller's dialogue history, linked with directed, weighted arcs that represent support relationships between them.

Definition 9. *Let $\mathcal{H} = \{\mathscr{D}^1, \ldots, \mathscr{D}^k\}$ be an agent's* **history** *of dialogues and $\mathcal{A}^{\mathcal{H}}$ represent the set of arguments introduced by that agent's opponents in $\mathcal{H}$. Then a* **relationship graph ($\mathcal{R}G$)** *is a directed graph $\mathcal{R}G = \{\mathcal{A}^{\mathcal{H}}, R\}$, where $R \subseteq \mathcal{A}^{\mathcal{H}} \times \mathcal{A}^{\mathcal{H}}$ is a set of weighted arcs representing support relationships. For two arguments $A, B \in \mathcal{A}^{\mathcal{H}}$, we write r_{AB} to denote the arc $(A, B) \in R$, and denote the arc's weight as w_{AB} where $0 \leq w_{AB} \leq 1$.*

Note that arc weights are actually probability values. Thus, henceforth we may write $Pr(r_{AB})$ referring to r_{AB}'s weight w_{AB}, i.e. $Pr(r_{AB}) = w_{AB}$.

Finally, the augmentation process proposed in [7] consists of three steps. Let Ag_1 and Ag_2 be two agents about to engage in a dialogue. Let Ag_1 have a model of Ag_2's possible knowledge $S_{(1,2)}$ and let $\mathcal{R}G_1$ be Ag_1's relationship graph. First, instantiate a set $\mathcal{A}$ with all arguments that can be constructed from $S_{(1,2)}$. Second, identify a set $N_{\mathcal{A}}$ with arguments adjacent to $\mathcal{A}$ in $\mathcal{R}G_1$, where every $X \in N_{\mathcal{A}}$ is a supporter of some $Y \in \mathcal{A}$. Third, based on the arc weights on the support relationships between $\mathcal{A}$ and $N_{\mathcal{A}}$ compute and assign confidence values to the constituents of the arguments in $N_{\mathcal{A}}$, and augment $S_{(1,2)}$ with them.

3 Modelling Support Relationships

The modelling approach proposed in [7] assumes that if two arguments share a support relationship in a $\mathcal{R}G$, then if the supported in the relationship is assumed to be already known to a certain opponent, it is *likely* that the supporter is also known to that opponent. In contrast to [7] we assume more than just one type of support between arguments. This section presents four types of support, distinguishing them according to whether they are abstract or logical.

3.1 Abstract Support Relationships

Abstract support relationships are concerned with how *opponent arguments* appear in a dialogue structure. The first kind of support relationship we discuss is that of *reinstatement support (RS)*. Though not explicitly referred as such, this is the relationship on which the work in [7] relies on. The RS of an argument A (supported) by an argument B (supporter) is represented if B defends A as defined in [4]. An example of a RS identified in the dialogue tree of Fig.1a is the one between arguments A and C in the sense that C reinstates A by attacking A's attacker B. Other such relationships are those between A and G, A and I, and C and E, for which corresponding arcs are instantiated to construct the RS-$\mathcal{R}G$ of Fig.1b (where we assume Ag_2 is the modeller). The identification of these relationships as well as their representation in an $\mathcal{R}G$, is defined as follows.

Definition 10. *Let* $\mathcal{R}G=\{\mathcal{A}^{\mathcal{H}}, R\}$ *be a relationship graph and* $RS \subseteq R$ *be a subset of* R *representing all of the* **RS relationships**. *Let* A *and* B *be any two arguments respectively serving as the content of two opponent dialogue moves* $\mathcal{DM}_i \in d_k$ *and* $\mathcal{DM}_j \in d_l$ *in a dialogue tree* $\mathcal{T} = \{d_1, \ldots, d_m\}$. *Let* `level()` *be a function applied on a* $\mathcal{DM}$ *that returns the level of the move in* $\mathcal{T}$. *Then* $\exists r_{AB} \in RS$ *if:*

1. $k = l$ *(the two moves are in the same dispute);*
2. $i < j$ *(*$\mathcal{DM}_i$ *precedes* $\mathcal{DM}_j$ *in the dialogue);*
3. $\mathtt{level}(\mathcal{DM}_j) - \mathtt{level}(\mathcal{DM}_i) = 2$.

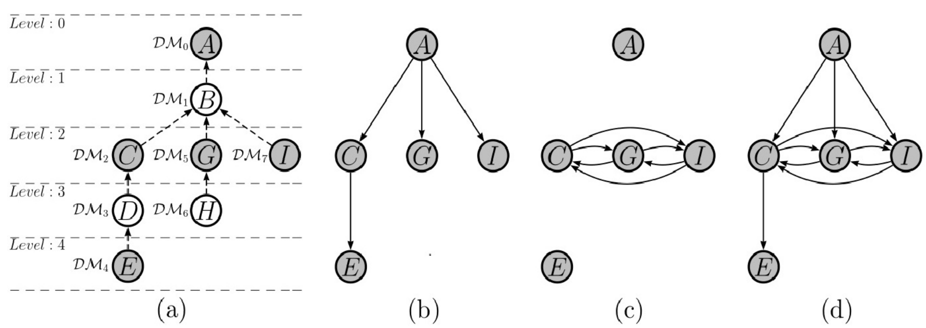

Fig. 1. a) A $\mathcal{T}$ between Ag_1(grey) & Ag_2, b) a RS-$\mathcal{R}G$, c) a CATS-$\mathcal{R}G$, d) the joint $\mathcal{R}G$.

In this paper we now introduce the additional notion of a *common attack target support (CATS)* relationship. Intuitively, arguments which attack the same target support each other in the sense that they serve the same objective; to invalidate that target. An example of a CATS identified in the dialogue tree of Fig.1a is the one between arguments C and G, in the sense that they both attack the same target (argument B). Notice that in contrast to RS relationships, each argument in a CATS relationship supports the other. Hence, arcs between these arguments are reciprocal. Referring to Fig.1a, CATS relationships exist between G and I, and C and I; hence the corresponding arcs in the CATS-$\mathcal{R}G$ of Fig.1c. In general, arguments are linked in a $\mathcal{R}G$ if they appear in distinct disputes in the same dialogue, in reply to the same modeller argument, i.e. attacking the same target in a dialogue. We formally express this as follows.

Definition 11. *Let* $\mathcal{R}G=\{\mathcal{A}^{\mathcal{H}}, R\}$ *be a relationships graph. Let* d' *be a sub-dispute of a dispute* $d \in \mathcal{T}$ *such that* $d' =< \mathcal{DM}_0, \ldots, \mathcal{DM}_k >$, *and* A *and* B *be two arguments respectively serving as the content of two opponent dialogue moves* $\mathcal{DM}_i$ *and* $\mathcal{DM}_j$ *in a dialogue tree* $\mathcal{T}$. *Then* $\exists r_{AB}, r_{BA} \in CATS \subseteq R$, *if both* $\mathcal{DM}_i$ *and* $\mathcal{DM}_j$ *extend* d' *in* $\mathcal{T}$, *and:*

– $\exists d'_1 =< \mathcal{DM}_0, \ldots, \mathcal{DM}_k, \mathcal{DM}_i >$, *where* d'_1 *is a sub-dispute of a* $d_1 \in \mathcal{T}$, *and;*
– $\exists d'_2 =< \mathcal{DM}_0, \ldots, \mathcal{DM}_k, \mathcal{DM}_j >$, *where* d'_2 *is a sub-dispute of a* $d_2 \in \mathcal{T}$.

3.2 Logical Support Relationships

We now turn to logical support relationships. As will be shown, they are in fact special instances of abstract support relationships and will therefore only affect the quantification of the weights assigned to the arcs they concern, i.e. no new arcs will be defined by identified logical supports. Note that we will motivate the need of using logical support relationships in Section 4, where we discuss the expectations implied by these relationships and how they positively affect the weighting of the arcs they concern.

In order to investigate logical relationships between arguments we turn to their structure. We exemplify by reference to an argumentation system $AS = (\mathcal{L}, ^-, \mathcal{R}, \leq)$, where: $\mathcal{L}$ is a language of propositional literals, composed from a set of propositional atoms $\{a, b, c, \ldots\}$ and the symbols $\neg$ and $\sim$ respectively denoting strong and weak negation (i.e., negation as failure). α is a strong literal if α is a propositional atom or of the form $\neg\beta$ where β is a propositional atom. α is a wff of $\mathcal{L}$ if α is a strong literal or of the form $\sim \beta$ where β is a strong literal; and for a wff α, α and $\neg\alpha$ are contradictories and α is a contrary of $\sim \alpha$. Assume then a dialogue between two agents (Ag_1 and Ag_2) with structured arguments, as it appears in Fig.2a.

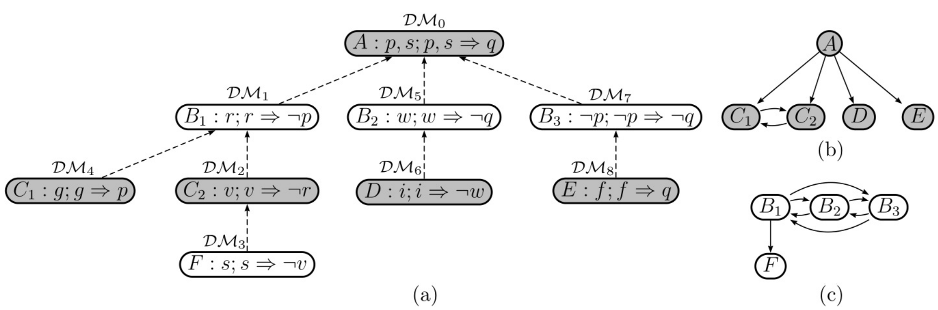

Fig. 2. a) A dialogue between Ag_1(grey) and Ag_2, b) the $\mathcal{R}G$s constructed by Ag_1 for Ag_2's arguments, and c) the $\mathcal{R}G$s constructed by Ag_2 for Ag_1's arguments.

The first logical relationship we identify is *backbone support (BBS)*. The BBS concerns relationships where the supporter's claim is a formula in the antecedent of a rule in another (supported) argument, in response to a challenge (attack) on that formula (as described in [9]). For instance, take the RS relationship r_{AC_1} (Fig.2b). Notice in Fig.2a that C_1's claim p is in the antecedent of rule $p, s \Rightarrow q$ in A, which means that C_1 serves as a BBS for A, in response to the B_1's attack on p. Note that strictly speaking, B_1 attacks the *premise* p in A_1, and in principle BBS support can of course respond to an attack on the conclusion of a defeasible rule (rather than a premise) that supplies the formula in the antecedent of the rule in the supported argument. However, we focus on the formula in the antecedent, to accommodate extensions of our dialogue framework that allow the use of enthymemes as modelled in $ASPIC^+$ [8]; e.g., when A_1 is simply moved as the rule $p, s \Rightarrow q$, and the supporting argument effectively backward extends A_1 [15] in response to the challenge. Finally, notice that though r_{AC_2} is a RS relationship, it cannot be characterised as BBS.

The second kind of logical relationship is that of *common conclusion support (CCS)*. CCS bears on a participant's ability to support a certain claim in multiple ways. An example of a CCS can be identified in Fig.2a between arguments B_2 and B_3 since they share the same claim $\neg q$. Notice again that while $r_{B_1B_2}$, $r_{B_2B_1}$, $r_{B_1B_3}$, $r_{B_3B_1}$, and $r_{B_2B_3}$, $r_{B_3B_2}$ concern CATS relationships, only the last two additionally concern CCS relationships.

At this point we must clarify that it is not the case that just CATS can also be CCS relationships or that just RS relationships can be BBS relationships. Take for example the CATS relationship $r_{B_1B_3}$ (Fig.2c). Notice that B_1 serves as a BBS for argument B_3 (Fig.2a). Further notice that r_{AE} (Fig.2b) is a RS which is also a CCS (Fig.2a). These relationships are formally defined as follows.

Definition 12. *Let $\mathcal{RG} = \{\mathcal{A}^{\mathcal{H}}, R\}$ be a relationship graph where $R = RS \cup CATS$. Let A and B be two arguments in $\mathcal{A}^{\mathcal{H}}$ and $r_{AB} \in R$. Then $BBS \subseteq R$ is a subset of R representing all* **backbone support** *relationships, and $CCS \subseteq R$ is a subset of R representing all* **common conclusion support** *relationships, where if:*

- $\texttt{Conc}(A) = \phi$*, and* ϕ *is in the antecedent of a rule in* B*, then* $r_{AB} \in BBS$*;*
- $\texttt{Conc}(A) = \texttt{Conc}(B) = \phi$*, then* $r_{AB} \in CCS$*.*

Lastly, we stress once more that logical relationships are simply special instances of abstract relationships. As will be argued in the next section, they imply a stronger connection between the associated arguments, and consequently an increased likelihood that the modelled opponent is aware of the supporter.

4 Quantification of Support Relationships

All arcs in a $\mathcal{RG}$ assume assignment of numerical weight values. These values express likelihoods, in the form of probability values, that a supporter argument can be constructed by the modelled opponent, contingent on the latter being aware of the supported argument. Weight assignation depends on the abstract as well as on the logical support type of a relationship. Thus, we assume that arc weights (w) are produced from two distinct sub-weights; one abstract w^{α}, and one logical w^{λ}. We propose that logical weights should have a positive impact on the overall weight of an arc, given that we interpret abstract and logical weights as probability values concerned with the *same* random event; that of the supporter in a relationship being known to an opponent, contingent on knowledge of the supported. We therefore define the weight of an arc to be equal to the joint probability value of its two sub-weights. All weight values are independent and identically distributed (i.i.d.).

Definition 13. *Let $\mathcal{H}$ be a modeller's history of dialogues, A and B two arguments of a $\mathcal{R}G= \{\mathcal{A}^{\mathcal{H}}, R\}$ induced from $\mathcal{H}$, and $r_{AB} \in R$ with a weight w_{AB}, where w^{α}_{AB} and w^{λ}_{AB} are respectively the* **abstract and logical sub-weights** *of w_{AB}. Then:*

$$w_{AB} = w^{\alpha}_{AB} + w^{\lambda}_{AB} - w^{\alpha}_{AB} \cdot w^{\lambda}_{AB}$$

Different support types are quantified differently, and encode expectations regarding the awareness of the supporting arguments. For example, RS relationships imply that supporter opponent arguments are likely to follow after supported arguments in dialogues, as responses to challenges. The extent of that likelihood is defined by the frequency that this is shown to happen between two arguments, in a modeller's history of dialogues.

Assume, for example, that a modeller, Ag_1, monitors an opponent argument A, introduced by various opponents in a series of dialogues. Let A appear a total of 10 times in these dialogues, and is attacked by arguments introduced by Ag_1. Assume then that A is reinstated against those arguments by opponent arguments B, C and D, respectively 3, 4 and 1 times, while in two dialogues A is not reinstated by any argument. Then the abstract weights w^{α}_{AB}, w^{α}_{AC} and w^{α}_{AD} for the respective RS relationships r_{AB}, r_{AC} and r_{AD} will be $\frac{3}{10}$, $\frac{4}{10}$ and $\frac{1}{10}$. These weights represent how likely the supporting argument (e.g. B in r_{AB}) will be submitted by any given opponent, so as to reinstate the supported argument (e.g. A). This is the approach proposed in [7].

We produce an analogous ratio for the case of CATS, which represents how likely a modeller's argument will be attacked by a given pair of opponent arguments. As with RS relationships, CATS implies that supporting opponent arguments are likely to follow after supported arguments, as alternative attacks (replies). Again, this likelihood is defined by the frequency that this is shown to happen in a modeller's history of dialogues.

For example, suppose a modeller's argument A introduced in three dialogues with different agents, where Ag_1 monitors attacks on A. Assume that in the first dialogue A is attacked by opponent arguments B, C and D, in the second by B and C and in the third by B and D. One may then assume the following relationships: r_{BC}, r_{BD}, r_{CB}, r_{CD},r_{DC} and r_{DB}. The weight for each of these relationships will be the number of times the arguments in each relationship appear jointly, divided by the number of times that the supported argument in the relationship appeared in distinct disputes attacking the common target. For example, C follows B two out of the three times that B attacks A, thus $w_{BC} = \frac{2}{3}$, while B follows C every time that C attacks A, $w_{CB} = \frac{2}{2}$. Hence, $w_{BD} = \frac{1}{3}$, $r_{CD} = \frac{1}{2}$,$w_{DC} = \frac{1}{2}$ and $w_{DB} = \frac{2}{2}$. Definition 14 formally describes these quantifications.

Definition 14. *Let $\mathcal{H}$ be a modeller's history of dialogues and $\mathcal{R}G= \{\mathcal{A}^{\mathcal{H}}, R\}$ the relationship graph induced from $\mathcal{H}$, where $R = RS \cup CATS$ are respectively the sets of reinstatement and common attack target support. Let $r_{AB} \in R$ with a weight w_{AB}, where w^{α}_{AB} is the* **abstract sub-weight** *of w_{AB}. Also let:*

$$(a)\, \texttt{occurrences}_{RS}(\mathcal{H}, A, B) = C_{AB}, \quad (b)\, \texttt{instances}_{RS}(\mathcal{H}, A) = I_A$$

$$(c)\, \texttt{occurrences}_{CATS}(\mathcal{H}, A, B, C) = J_{AB}, \quad (d)\, \texttt{instances}_{CATS}(\mathcal{H}, A, C) = I_{AC}$$

be respectively: (a) a function that returns the number of times B follows after A in distinct disputes in $\mathcal{H}$; (b) a function that returns the number of times A appeared in distinct disputes in $\mathcal{H}$ though not as a leaf; (c) a function that returns the number of joint appearances of A and B against an argument C in the same dialogues in $\mathcal{H}$; (d) a function that returns the number of appearances of A against an argument C in all dialogue of $\mathcal{H}$. Then:

$$w^{\alpha}_{AB} = \begin{cases} \frac{C_{AB}}{I_A} & \textit{if } r_{AB} \in RS \\ \frac{J_{AB}}{I_{AC}} & \textit{if } r_{AB} \in CATS \end{cases}$$

Let us turn now to the quantification of logical relationships. As stated earlier, we assume that logical relationships imply a stronger connection between arguments that already share abstract relationships. Our intuitive expectation of this strengthening rests on two assumptions. The first is that generally, if one perceives argumentation as a way of characterising the reasoning one uses to arrive at certain beliefs about the world, it is then reasonable to expect that rational agents would ideally have explored all possible lines of reasoning with respect to a claim. Hence, if an agent makes use of a (sub)argument claiming p, then there is some likelihood that the agent will be aware of other arguments concluding p. Secondly, in real-world dialogues agents move incomplete arguments (enthymemes) in a dialogue (recall our discussion in Section 3.2), so that challenges on a formula in the antecedent of a rule, motivates submission of a supporting argument claiming that formula (and so effectively backward extending the incomplete argument).

Based on these two assumptions we *expect* that participants are likely to be aware of multiple ways of arguing for a claim, having been faced with responding to challenges on the claim, as well as having had to respond to challenges on a formula in order to argue why that formula is believed. These expectations are justified by the existence of logical support relationships between some arguments. Take for instance the case of the BBS relationship r_{AC_1} (Fig.2b). Here the introduction of $C_1 : g; g \Rightarrow p$ is caused by a challenge on $A : p; s; p, s \Rightarrow q$ that forces Ag_1 to reveal an alternative line of reasoning justifying p (in A p is already present as a premise), and hence an alternative line os reasoning justifying q (i.e. $g; g \Rightarrow p; s; p, s \Rightarrow q$). On the other hand, if the incomplete argument $A' : s; p, s \Rightarrow q$ had been allowed, then the supporting C_1 would have backward extended A' to yield $g; g \Rightarrow p; s; p, s \Rightarrow q$. Similarly, the CCS relationship $r_{B2,B3}$ (Fig.2c) suggests that Ag_2 has explored other alternatives for $\neg q$, which were revealed in the course of the dialogue, only when it became necessary.

Finally, there are many ways for quantifying the logical weights of an arc r_{AB}. One could focus on the supported argument in the relationship and using all the OMs available to a modeller, produce a ratio by counting the number of opponents that are aware of (can construct) logical supporters of A. Then, divide that number with the number of opponents that are aware of *any* supporters of A. This quantification approach focusses on the distinction between logical and abstract relationships. A drawback however is that all logical relationships where A is the supported argument will have the same weight. Other quantification perspectives with different objectives could focus on other aspects of these relationships, e.g. to further distinguish between BBS and CCS relationships. We therefore assume no absolute stance as to the exact way of quantifying logical support relationships and define a general logical weighting function as follows.

Definition 15. *Let AgT be an agent's agent theory containing all the sub-theories of its opponents, and $\mathcal{R}G = \{\mathcal{A}^{\mathcal{H}}, R\}$ a relationship graph where BBS, $CCS \subseteq R$ are respectively the sets of backbone and common conclusion support arcs in R. Let $r_{AB} \in \{CCS, BBS\}$ with a weight w_{AB}. Then:*

$$\texttt{weightL}(\mathcal{R}G, r_{AB}, AgT) \rightarrow w^{\lambda}_{AB}$$

is a function that returns w_{AB}'s **logical sub-weight** *w^{λ}_{AB}, where $0 \leq w^{\lambda}_{AB} \leq 1$.*

5 Related Work

The notion of support is multifaceted and is concerned with positive interactions between arguments [14]. Many types of support relationships have been identified in the literature so far with different applications [3, 11, 13]. The most common type is that of reinstatement which is implicit in Dung's framework [4], and is understood in the sense of counter-attack. However, as Amgoud *et al.* argue [1], support does not use the same method as attack and thus counter-attack cannot capture the notion of support completely.

In this respect, different perceptions of the notion have been formalised in the literature, giving rise to a class of acceptability semantics defined within *bipolar argumentation framework (BAF)*, in which interactions between arguments concern both attack as well as support relationships. For example, Boella *et al.* [3] distinguish between what they refer to as *deductive support*, according to which an argument A supports an argument B if the acceptance of A implies the acceptance of B, and *defeasible support* where the previous implication holds only by default and it can be attacked. Similarly, Nouioua [11] assumes a perception of support referred to as *necessary support*, according to which if an argument A necessarily supports B, then acceptance of A is required for the acceptance of B. Also, Oren and Norman [13] introduce the idea of *evidential support*, distinguishing "special" arguments which serve as prima-facie or indisputable sources of truth, and "standard" arguments whose claims are not sufficiently justified, and need to be supported by the former so as to be considered acceptable. Finally, accrual of arguments for the same claim [16] can also be interpreted as support, where the accruing arguments mutually support each other.

Many of the above notions of support are motivated by logical relationships between the constituents of support-related arguments, such as those introduced in this paper. For example, in a similar sense to evidential support, BBS concerns relationships where the supporter is called to justify the antecedent of another argument (the supported), when the latter (which can be considered a standard argument) is challenged. In our case though, the supporter is not required to be a "special" argument. A case of deductive support can also be exemplified through BBS. Assume an argument $B : p; p \rightarrow q$ where p is an ordinary premise and $p \rightarrow q$ a strict rule which cannot be attacked. Assume then an argument $A : s; s \Rightarrow p$ which backbone-supports B. Since $p \rightarrow q$ cannot be attacked, acceptance of A should imply the acceptance of B. This is because any attack on B's premise p must also be by definition an attack on A's claim. Similarly, in the case of necessary support, one could say that A necessarily supports B. Also, CCS effectively models the accrual of mutually supporting arguments for the same claim.

Finally, we have considered extensions to our dialogical framework which allow for use of enthymemes, e.g., $A' : s; p, s \Rightarrow q$ supported by $C_1 : g; g \Rightarrow p;$. As argued by Modgil [9], logical instantiations of frameworks by given sets of formulae do not (for the purposes of argumentation-based inference where the claims of justified arguments identify the inferences from the instantiating formulae) warrant abstract representations of support relations in frameworks. Rather, support relations are useful in other settings, including dialogues. For example when incomplete arguments are moved (e.g., A'), and the missing elements are subsequently supported (e.g., with C_1). If one were to *start* with a set of formulae $\{g; g \Rightarrow p; s; p, s \Rightarrow q\}$, it would suffice to simply construct the argument $g; g \Rightarrow p; s; p, s \Rightarrow q$. In a dialogue, the latter is implicitly, and incrementally, constructed through moving A' and (in response to the challenge) the supporting $C1$. In these contexts, the use of such supporting relationships as well as relationships such as CATS, provide further value for opponent modelling purposes and for OM augmentation.

6 Conclusions & Future Work

This paper extended the work of Hadjinikolis *et al.* [7] in the following ways. Firstly, it extended the notion of a $\mathcal{R}G$ by introducing CATSs relationships between arguments, which is to the best of our knowledge a novel notion of support presented here, allowing for more modelling alternatives. Secondly, it proposed a distinction between abstract support relationships concerned with how arguments appear in the structure of dialogues, and logical relationships, concerned with relationships between the constituents of arguments already abstractly related. It then argued that (as in [9]) these relationships are redundant when considering logical instantiations for argumentation-based inference, but are needed in dialogical contexts, providing further value for opponent modelling purposes. Lastly, corresponding quantifications of the likelihoods implied by the presented support relationships were proposed.

Future research will focus on the development of a methodology towards evaluating our modelling approach and validating our assumptions on the increased likelihoods implied by logical support relationships between arguments.

References

1. L. Amgoud, C. Cayrol, M. C. Lagasquie-Schiex, and P. Livet. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23(10):1062–1093, 2008.

2. E. Black, A. Coles, and S. Bernardini. Automated planning of simple persuasion dialogues. In N. Bulling, L. van der Torre, S. Villata, W. Jamroga, and W. Vasconcelos, editors, *Computational Logic in Multi-Agent Systems*, volume 8624 of *Lecture Notes in Computer Science*, pages 87–104. Springer International Publishing, 2014.
3. G. Boella, D. M. Gabbay, L. W. N. van der Torre, and S. Villata. Support in abstract argumentation. In *Proc. of Computational Models of Argument COMMA'10*, pages 111–122, 2010.
4. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.
5. P. Dunne and P. McBurney. Concepts of optimal utterance in dialogue: Selection and complexity. In F. Dignum, editor, *Advances in Agent Communication*, volume 2922 of *Lecture Notes in Computer Science*, pages 310–328. Springer Berlin Heidelberg, 2004.
6. C. Hadjinikolis, S. Modgil, E. Black, P. McBurney, and M. Luck. Investigating strategic considerations in persuasion dialogue games. In *Proc. of the Starting AI Researchers' Symposium STAIRS'12*, pages 137–148. IOS Press, 2012.
7. C. Hadjinikolis, Y. Siantons, S. Modgil, E. Black, and P. McBurney. Opponent modelling in persuasion dialogues. In *Proc. of the International Joint Conference on Artificial Intelligence IJCAI'13*, pages 164–170. AAAI Press, 2013.
8. S. A. Hosseini, S. Modgil, and O. Rodrigues. Enthymeme construction in dialogues using shared knowledge. In *Proc. of Computational Models of Argument COMMA'14*, pages 325–332, 2014.
9. S. Modgil. Revisiting abstract argumentation frameworks. In *Proc. of the International Workshop on Theory and Applications of Formal Argumentation TAFA'13*, pages 1–15, 2013.
10. S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.
11. F. Nouioua. Afs with necessities: Further semantics and labelling characterization. In *Proc. of the International Conference on Scalable Uncertainty Management SUM'13*, pages 120–133, 2013.
12. N. Oren and T. Norman. Arguing Using Opponent Models. In *Argumentation in Multi-Agent Systems*, volume 6057 of *Lecture Notes in Computer Science*, pages 160–174. Springer Berlin Heidelberg, 2010.
13. N. Oren and T. J. Norman. Semantics for evidence-based argumentation. In *Proc. of Computational Models of Argument, COMMA'08*, pages 276–284, 2008.
14. S. Polberg and N. Oren. Revisiting support in abstract argumentation systems. In *Proc. of Computational Models of Argument COMMA'14*, pages 369–376, 2014.
15. H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Logic and Computation*, 15(6):1009–1040, 2005.
16. H. Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *Proc. of the International Conference on Artificial Intelligence and Law ICAIL'05*, pages 85–94, 2005.
17. R. Riveret, A. Rotolo, G. Sartor, H. Prakken, and B. Roth. Success chances in argument games: a probabilistic approach to legal disputes. In *Proc. of the Twentieth annual Conference on Legal Knowledge and Information Systems: JURIX*, pages 99–108, 2007.